Spectrum Sensing in Cognitive Radio: Bootstrap and Sequential Detection Approaches

Vom Fachbereich 18
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von
Fiky Y. Suratman, M.T.
geboren am 02.05.1976 in Jakarta (Indonesien)

Referent:	Prof. Dr.-Ing. Abdelhak M. Zoubir
Korreferent:	Prof. Dr.-Ing. Anja Klein
Tag der Einreichung:	19.11.2013
Tag der mündlichen Prüfung:	17.01.2014

D 17
Darmstadt, 2014

Bismillahirrahmanirrahim

Acknowledgments

I would like to express my sincere gratitude to Deutscher Akademischer Austausch Dienst(DAAD), the German Academic Exchange Service, for giving me financial support during my stay and study in Germany.

I am greatly indebted to many people who have helped and inspired me during my doctoral study. Foremost, I would like to deeply thank my supervisor, Prof. Dr.-Ing. Abdelhak Zoubir, for giving me the opportunity to work in Signal Processing Group (SPG) of TU-Darmstadt, for his continuous support, his patience and motivation, and his immense knowledge. His guidance helped me during the time of research and thesis writing. It has been a great pleasure to work under his supervision, technically and personally. I wish to thank Prof. Dr.-Ing. Anja Klein for co-supervising this thesis, and the instructive comments and suggestions that she has provided. I also wish to thank Prof. Dr.-Ing. Tran Quoc Khanh, Prof. Dr.-Ing. Rolf Jakoby, and Prof. Dr.-Ing. Klaus Hofmann who acted as chair and examiners in the PhD committee.

My thanks go to all my Indonesian friends at Indonesisches Kultur und Informationszentrum Darmstadt (IKID e.v.), whom I cannot mention one by one, and my former roommates at the SPG, Ahmed Mustofa and Waqas Sharif, for sharing each moment in Darmstadt and your friendships. I must also acknowledge Weaam Alkhaldi, his assistance at the SPG side before I came to Germany was very helpful. My appreciation goes to Michael Fauß for fruitful discussions and suggestions. I thank all my colleagues at the SPG of TU-Darmstadt. It has been great to know you all. Thank to Sara Al-Sayed, Mouhammad Alhumaidi, Toufik Boukaba, Nevine Demitri, Gökhan Gül, Jürgen Hahn, Lala Khadidja Hamaidi, Sahar Khawatmi, Michael Lang, Stefan Leier, Michael Leigsnering, Michael Muma, Tim Schäck, Adrian Sosic, Wassim Suleiman, Christian Weiss, and Feng Yin, as well as Renate Koschella and Hauke Fath. My thanks also goes to the former PhD students Christian Debes, Yacine Chakhchoukh, Raquel Fandos, Uli Hammes, Philipp Heidenreich, Zhihua Lu and Marco Moebus.

A special thanks to my family in Indonesia. Words cannot express how grateful I am to my father Bapak Suratman, my mother Ibu Euis Warjiah, and my sister Erna Fashanita. Your timeless sacrifice, prayers and encouragement are something that make me strong and even stronger throughout this life. I thank all big family members, in Ciwidey, Aceh, Garut, and wherever you are. Last but not least, I am most grateful to my wife Susan Agustia and my son Faiz G.S. Al-Farisi for their understanding, love, support and joy. I am finally going home to you.

Darmstadt, 07.02.2014

Kurzfassung

Diese Dissertation beschäftigt sich mit fortschrittlichen Techniken der Spektrumerkundung (spectrum sensing) für kognitive Funkgeräte (cognitive radio). Untersucht wird insbesondere das in der Spektrumerkundung häufig auftretende Problem geringer Stichprobengrößen, zu dessen Bewältigung lokale und verteilte Resampling-Methoden vorgeschlagen werden. Ein weiterer Schwerpunkt der Arbeit liegt auf dem Entwurf sequenzieller Testverfahren, wenn entscheidende Parameter nur teilweise oder gar nicht bekannt sind. Des Weiteren werden Verfahren zur kanalübergreifenden (multiband) Spektrumerkundung, zur Spektrumerkundung bei niedrigem Signal-zu-Rausch-Abstand (SNR) und zur Fusion harter, lokaler Entscheidungen in verteilten Szenarien entwickelt.

Die Annahme großer Stichprobengrößen erweist sich in der Spektrumerkundung dann als problematisch, wenn die Teststatistik in der Praxis aus nur wenigen Beobachtung gewonnen wird. Der Grund hierfür ist, dass die asymptotische Verteilung der Teststatistik unter der Nullhypothese bei kleinen Stichprobengrößen nicht der tatsächlichen Verteilung entspricht. Entsprechend weicht die unter der asymptotischen Verteilung berechnete Wahrscheinlichkeit für Fehlalarm und Fehldetektion deutlich von der beobachteten ab. Wir schlagen daher die Verwendung von Bootstrapping vor, wobei die Verteilung der Testsstatistik durch Resampling der beobachteten Daten geschätzt wird. Für die lokale Spektrumerkundung entwickeln wir den Bootstrap Null-Resampling Test, welcher bessere Ergebnisse als der Bootstrap Pivot Test und der asymptotische Test zeitigt. Für die verteilte Spektrumerkundung schlagen wir eine auf Resampling basierende Variante der Chain-Varshney Fusionsregel vor. Zur Schätzung der lokalen Detektionswahrscheinlichkeiten der Nutzer eines kognitiven Systems wird eine Kombination aus unabhängigem Resampling und Resampling eines gleitenden Blocks von Beobachtungen vorgeschlagen. Bei einer großen Anzahl von verteilten Nutzern kann parametrisches Bootstrapping am zentralen Knoten (fusion center) durchgeführt werden.

Der klassische Sequential Probability Ratio Test entscheidet zwischen zwei einfachen Hypothesen. Realistisch gesehen treten in der Spektrumerkundung jedoch zusammengesetzte Hypothesen auf, da viele Systemparameter nicht exakt bekannt sind. In dieser Dissertation wird der Anwendungsbereich des Sequential Probability Ratio Test auf zusammengesetzte Hypothesen ausgeweitet. Dazu werden die Entscheidungsschwellen mit Hilfe des parametrischen Bootstraps kontinuierlich angepasst. Im Gegensatz zu den meisten existierenden Verfahren, beruht der vorgeschlagene Test daher nicht auf der Annahme, dass asymptotische Verteilungen vorliegen und minimiert so

die Wahrscheinlichkeit von Fehlern, die durch ungenaue Parameterschätzungen entstehen. Dies gewährleistet eine verlässlichere Sprektrumerkundung bei gleichzeitiger Reduzierung der durchschnittlich benötigten Beobachten (average sample number, ASN). Des Weiteren schlagen wir eine Methode zur Senkung des Rechenaufwands vor-welcher notwendigerweise durch die Benutzung des Bootstraps entsteht-die für das Resampling eine Konvex-Kombination der letzten K Bootstrap Verteilungen verwendet. Das Verfahren erhöht den ASN-Wert nur unmaßgeblich und bietet zudem einen erhöhten Schutz gegen Entscheidungsfehler. Motiviert ist diese Arbeit durch die Tatsache, dass der Sequential Probability Ratio Test eine geringere Zeit zur Datenerfassung benötigt als ein äquivalenter Test mit einer festen Anzahl von Beobachtungen. Eine möglichst schnelle Entscheidungsfindung ist in der Spektrumerkundung wünschenswert, da sie den Datendurchsatz in kognitiven Systemen steigert.

Im letzten Abschnitt werden drei Themen behandelt. Zuerst wird die kanalübergreifende Spektrumerkundung untersucht und mehrere Testvarianten vorgestellt. Für Tests mit einer feste Anzahl von Beobachtungen wird die Verwendung des adaptiven Benjamini-Hochberg Verfahrens vorgeschlagen, da es eine bessere Balance zwischen den systemweiten Wahrscheinlichkeiten für Fehlalarme und Fehldetektionen bietet als das konventionelle Benjamini-Hochberg Verfahren. Für sequenzielle Tests schlagen wir eine auf den geordneten Stoppzeiten der einzelnen Tests beruhende Methode vor. Simulationsergebnisse zeigen, dass dieses Verfahren einen geringen ASN-Wert erzielt als das Bonferroni-Verfahren. Ein weiteres Problem in der Spektrumerkundung ist die Detektion eines Signals bei sehr geringem SNR. Für diesen Fall leiten wir einen lokal optimalen Detektor für modulierte Signale in Student-t verteiltem Rauschen her, dessen Zuverlässigkeit deutlich höher ist als die des Energiedetektors. Schließlich erweitern wir das Verfahren zur Fusion harter (1-Bit) Entscheidungen in der verteilten Spektrumerkundung um ein weiteres Bit, welches die Verlässlichkeit der Entscheidung beschreibt. Die Entscheidungsschwellen werden anschließend mit Hilfe geeignete Distanzmaße bestimmt. Durch die Verwendung zusätzlicher Informationen (2-Bit) wird so eine höhere Performance erreicht als bei Verfahren, welche nur harte Entscheidungen (1-Bit) zulassen.

Abstract

In this thesis, advanced techniques for spectrum sensing in cognitive radio are addressed. The problem of small sample size in spectrum sensing is considered, and resampling-based methods are developed for local and collaborative spectrum sensing. A method to deal with unknown parameters in sequential testing for spectrum sensing is proposed. Moreover, techniques are developed for multiband sensing, spectrum sensing in low signal to noise ratio, and two-bits hard decision combining for collaborative spectrum sensing.

The assumption of using large sample size in spectrum sensing often raises a problem when the devised test statistic is implemented with a small sample size. This is because, for small sample sizes, the asymptotic approximation for the distribution of the test statistic under the null hypothesis fails to model the true distribution. Therefore, the probability of false alarm or miss detection of the test statistic is poor. In this respect, we propose to use bootstrap methods, where the distribution of the test statistic is estimated by resampling the observed data. For local spectrum sensing, we propose the null-resampling bootstrap test which exhibits better performances than the pivot bootstrap test and the asymptotic test, as common approaches. For collaborative spectrum sensing, a resampling-based Chair-Varshney fusion rule is developed. At the cognitive radio user, a combination of independent resampling and moving-block resampling is proposed to estimate the local probability of detection. At the fusion center, the parametric bootstrap is applied when the number of cognitive radio users is large.

The sequential probability ratio test (SPRT) is designed to test a simple hypothesis against a simple alternative hypothesis. However, the more realistic scenario in spectrum sensing is to deal with composite hypotheses, where the parameters are not uniquely defined. In this thesis, we generalize the sequential probability ratio test to cope with composite hypotheses, wherein the thresholds are updated in an adaptive manner, using the parametric bootstrap. The resulting test avoids the asymptotic assumption made in earlier works. The proposed bootstrap based sequential probability ratio test minimizes decision errors due to errors induced by employing maximum likelihood estimators in the generalized sequential probability ratio test. Hence, the proposed method achieves the sensing objective. The average sample number (ASN) of the proposed method is better than that of the conventional method which uses the asymptotic assumption. Furthermore, we propose a mechanism to reduce the computational cost incurred by the bootstrap, using a convex combination of the latest K bootstrap distributions. The reduction in the computational cost does not impose a

significant increase on the ASN, while the protection against decision errors is even better. This work is motivated by the fact that the sequential probability ratio test produces a smaller sensing time than its counterpart of fixed sample size test. A smaller sensing time is preferable to improve the throughput of the cognitive radio network.

Moreover, multiband spectrum sensing is addressed, more precisely by using multiple testing procedures. In a context of a fixed sample size, an adaptive Benjamini-Hochberg procedure is suggested to be used, since it produces a better balance between the familywise error rate and the familywise miss detection, than the conventional Benjamini-Hochberg. For the sequential probability ratio test, we devise a method based on ordered stopping times. The results show that our method has smaller ASNs than the Bonferroni procedure. Another issue in spectrum sensing is to detect a signal when the signal to noise ratio is very low. In this case, we derive a locally optimum detector that is based on the assumption that the underlying noise is Student's t-distributed. The resulting scheme outperforms the energy detector in all scenarios. Last but not least, we extend the hard decision combining in collaborative spectrum sensing to include a quality information bit. In this case, the multiple thresholds are determined by a distance measure criterion. The hard decision combining with quality information performs better than the conventional hard decision combining.

Contents

Chapter 1
Introduction

Current spectrum management techniques are based on the command-and-control model. This can guarantee that the radio frequency spectrum will be exclusively allocated to a set of users (primary users) of a licensed network who won the radio frequency bidding and can use the spectra without any interference [1]. However, this approach can cause inefficient spectrum usage, because measurement campaigns have shown that most of the allocated spectra are under-utilized [2]. This fact has prompted the Federal Communications Commission (FCC) to propose opening licensed spectrum bands to unlicensed users. This has given birth to cognitive radio. The basic idea of cognitive radio is to allow unlicensed or cognitive radio (CR) users to use certain frequency bands that are not being used by primary users. In summary, the main functions of cognitive radio are [3]:

- *Spectrum sensing*: Detecting the unused spectrum and sharing the spectrum with other users without harmful interference.
- *Spectrum management*: Capturing the best available spectrum to meet user communication requirements.
- *Spectrum mobility*: Maintaining seamless communication requirements during transition to a better spectrum.
- *Spectrum sharing*: Providing a fair spectrum scheduling method among coexisting cognitive radio users.

For an overview of the first IEEE wireless standard based on cognitive radio (IEEE 802.22), in which CR users are allowed to use licensed analog and digital television bands, refer to [4,5]. Standardization activities in cognitive radio can be found in [6,7]. Cognitive radio is envisioned to be implemented, for example, in rural area networks, smart grid networks, public safety networks, cellular networks and wireless medical networks [8, 9].

Due to the complexity of cognitive radio, in this thesis we focus on the spectrum sensing function, as one of primary requirements for the successful implementation of cognitive radio. Note that what we mean here by cognitive radio, is the one that is featured with the interweave network paradigm [10].[1]

[1] In general, network paradigms in cognitive radio can be divided into underlay, overlay and interweave.

1.1 Motivation, research challenges and objectives

Spectrum sensing plays a critical role in cognitive radio networks. Several dedicated nodes, or possibly every CR user, need to be equipped with a spectrum sensor. It is to monitor continuously the spectrum activities of primary users in order to find a suitable spectrum band (spectrum holes) for cognitive access, and to avoid causing harmful interferences to the primary users. Since the primary users have higher priority of service than the CR users, spectrum sensing includes detection of possible collision when a primary user becomes active in the spectrum momentarily occupied by a CR user. Fig. 1.1 depicts a conceptual illustration of spectrum holes, in which different colors indicate different frequency bands.

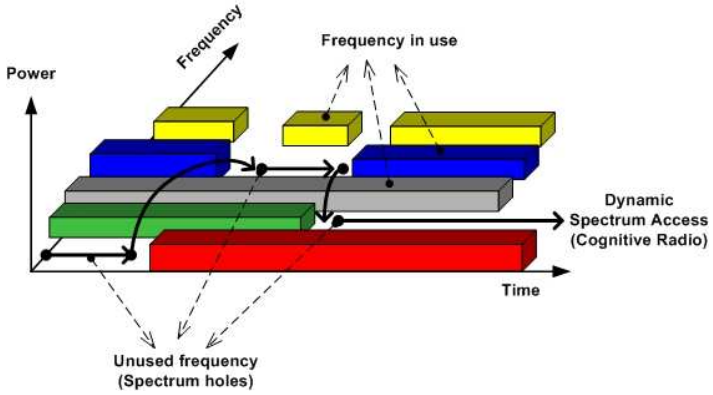


Figure 1.1: A conceptual illustration of spectrum holes over time and frequency. Within a certain geographical area and at a certain time, some frequency bands are not used by primary users.

Designing a detector for spectrum sensing in cognitive radio networks raises the following challenges:

- *Short sensing time*: Sensing time is defined as the time required by a detector (the required sample size) to declare that a specific band is unoccupied (null hypothesis) or occupied (alternative hypothesis). A short sensing time is a salient feature of spectrum sensing since it has a strong relation with the throughput that can be achieved by a cognitive radio network. However, this requirement mostly contradicts the sensing accuracy requirement, which in turn affects the overall throughput of the cognitive radio network and the resulting degree of interferences to the primary network.

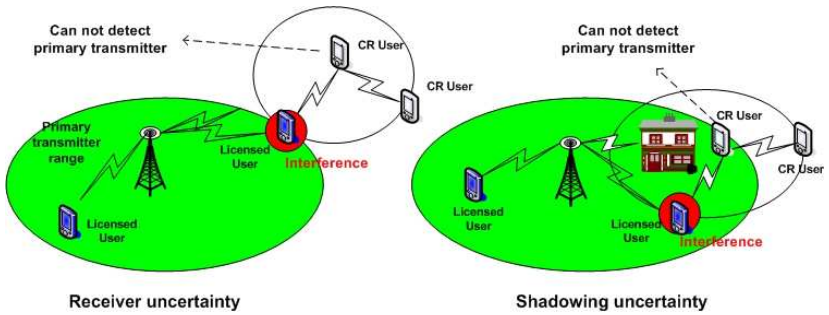


Figure 1.2: The hidden terminal problem occurs when a CR user experiences shadowing or lies outside coverage area of the transmitter of a primary user (licensed user).

- *Interference problem*: A cognitive radio network should be carefully designed so as to have a smaller degree of interference than the allowable level in a primary network. In spectrum sensing, this is measured by the probability of miss detection.

- *The hidden terminal problem*: It occurs when a CR user experiences shadowing or lies outside the coverage area of the primary user transmitter, as depicted in Fig. 1.2. The CR user in this condition could interfere with primary users. Thus, each CR user needs to deploy a detector that has good sensing accuracy at a low signal-to-noise ratio (SNR).

- *Noise uncertainty and outliers*: They occur in real world implementations and degrade the performance of a detector since the underlying noise deviates from the assumed model.

- *Various signal types*: The wideband feature of cognitive radio imposes a requirement of spectrum sensing to cope with various signal types. This includes different modulation schemes, data rates, transmission powers, etc., of primary users.

- *Limited resources*: These should be considered especially when one implements collaborative spectrum sensing algorithms. They include the number of collaborative users, the control channel bandwidth, the quantization level, etc.

To tackle these challenges altogether is not a trivial task. The general objective of this thesis is to develop and to implement efficient algorithms for spectrum sensing, which address the first four challenges. In particular, we aim for spectrum sensing with a short sensing time (a small sample size) to improve the agility of CR users.

1.2 State-of-the-Art

The essence of spectrum sensing is a binary hypothesis testing problem [11] for the presence or the absence of a primary user signal in a spectrum band. In general, there are two schemes for spectrum sensing, i.e., local spectrum sensing and collaborative (cooperative) spectrum sensing [1].

In *local spectrum sensing*, each CR user independently determines the presence or absence of a primary user, without any collaboration among them. This scheme, although simple in terms of computation and implementation, is sensitive to the hidden terminal problem. Several techniques that are based on classical detection have been proposed. They include energy detection, matched-filter based detection, cyclostationary based feature detection, covariance based detection, and eigenvalue based detection. In general, energy detection or radiometry [12,13] is preferred because it is of low complexity and does not require prior knowledge of the primary users' signal. However, it lacks robustness in noise power uncertainty, particularly in a low SNR [14]. Although a matched-filter detector [11, 15] is attractive for its optimality, it requires prior knowledge of primary users' signal, such as pilot symbols, modulation types, synchronization, etc., which in fact are not available to the CR users. Communication signals which have periodic patterns due to symbol rate, cyclic prefix, channel code, or chip rate, can be modeled as cyclostationary random process [16]. Therefore, a detector can be constructed by knowing some cyclic parameters of a signal and exploiting its cyclostationarity [17–20]. However, cyclostationary based feature detection has high computational complexity which becomes a limitation for agility of CR users. In addition, the cyclic parameters are not always available to the CR users in most practical scenarios. To circumvent the sensitivity of the energy detector to noise power uncertainty, covariance based signal detection is proposed [21, 22]. This method exploits the difference of statistical covariance matrices or autocorrelations of signal to noise. In practice, it is based on the sample covariance matrix, and prior information about statistics of the signal and the noise are not required. Another method to solve the noise uncertainty problem is eigenvalue based detection [23,24], i.e., based on the ratio of the maximum to the minimum eigenvalue of the covariance matrix of the received signal at the CR user, and also the ratio of the average eigenvalue to the minimum eigenvalue. Furthermore, local spectrum sensing with multiantenna receiver, to exploit spatial diversity and hence to improve detection performances, also has been widely studied in literature [24–31].

Collaborative spectrum sensing is developed to take advantage of spatial diversity of CR users that are geographically distributed. The main goal is to tackle the issue of

the hidden terminal problem due to shadowing or deep fading. It has been proven that collaborative spectrum sensing provides a diversity gain in a severe wireless environment [10]. In this scheme, a CR user is equipped with a detector and shares local sensing information with a fusion center. Accordingly, based on a fusion rule, the fusion center combines the shared local sensing information to make a global decision on the presence or absence of the primary user's transmission. The shared information can be soft or hard decisions made by CR users [32]. In soft-decision, each CR user computes the test statistic based on its local observations and forwards it directly to a common fusion center. In hard-decision, each CR user further processes the measured statistic to make an individual binary decision before finally forwarding it to the fusion center.

Various simpler techniques in soft decision combining (SDC) are employed in [12] and [33]. The performances of equal gain combining (EGC), selection combining (SC), and switch stay combining (SSC) are investigated for energy detector-based detection under Rayleigh fading and log-normal shadowing. In [34], by assuming Gaussian signal and noise, a simple fusion rule as a function of local SNRs of CR users is obtained. The authors show that this optimal soft decision combining scheme outperforms the EGC scheme. Meanwhile, to avoid a nonlinear operation produced by likelihood ratio based fusion, [35] uses a sub-optimum variant of SDC. Once the fusion center receives observations from CR users, the linear combining fusion rule is carried out, i.e., a soft decision from a particular user is weighted to represent its contribution to the global decision. By assuming very large sample sizes and using modified deflection coefficient as the objective function in optimization, [35] and [36] arrive at the optimal solutions for the weighting functions. For hard decision combining (HDC), some ad hoc fusion rules can be used [32], such as AND, OR, and Majority/Voting rules. The optimal fusion rules for Bayesian and Neyman-Pearson criteria have been derived in [37]. Implementations of the ad hoc and the optimal fusion rules in spectrum sensing are studied in [38] and the voting rule is explored in [39]. In addition, instead of assuming perfect receptions of transmitted bit decisions at the fusion center, multiuser diversity techniques are used in [40–42] to mitigate erroneous reporting channels between CR users and the fusion center. In general, SDC schemes have better performances than the corresponding HDCs. However, the HDC schemes require minimum bandwidth for the reporting channels which is attractive in collaborative spectrum sensing with limited resources. It is noteworthy that an ad hoc configuration is also possible. In this case, there is no dedicated fusion center and sensing information are shared among CR users. However, information dissemination to all CR users might lead to an unacceptable delay [10].

Most of the literature above are spectrum sensing schemes with a fixed sensing time,

i.e., the sample size is determined in advance to fulfill the requirement of the probability of errors. Sensing time is one of the most critical issues in spectrum sensing. It should be minimized to gain more transmission time for CR users and hence increase the throughput of cognitive radio network. However, sensing-throughput trade-off occurs in this case [43]. On the one hand, increasing the sensing time will increase the sensing accuracy, which in turn partially increases the throughput and reduces the degree of interference to the primary users. On the other hand, since the frame length is fixed, increasing the sensing time will decrease the transmission time, which in turn reduces the overall throughput of the cognitive radio network, and vice versa. To solve the issue, the optimality of the sequential detector [44,45] in reducing the sample size while maintaining probability of errors (false alarm and missed detection) is an attractive choice. Therefore, in the cognitive radio context, the degree of interference is fixed to an allowable value and CR users gain more time to transmit. This increases the throughput of the cognitive radio network. Note that in sequential detection, sample size is random and depends on the observations. In addition, sequential techniques also can be used to conserve energy since the sample size is minimized [46].

The earliest version of sequential detection is the so-called sequential probability ratio test (SPRT). It was pioneered by A. Wald in 1943 in a confidential report, which was published in 1945 [45]. The classical work of Bussgang and Middleton [47] applies the SPRT [44] to sequentially detect signals in noise. Subsequently, some analytical results and modified versions of the SPRT also have been introduced [48–52]. Some recent papers have studied the implementations of sequential techniques in spectrum sensing, e.g. [53–57]. Introducing the SPRT to local spectrum sensing that is based on cyclostationary feature of the received signals can be found in [53]. Sequential schemes in collaborative spectrum sensing are more attractive since they reduce the amount of data to be transmitted to the fusion center for identifying spectrum holes. In [54], current local log-likelihood ratios (LLRs) of CR users are transmitted to the fusion center where the SPRT is carried out. Should one of the two prespecified threshods at the fusion center be crossed, the test is terminated. Otherwise, the CR users take another sample and send their corresponding LLRs to the fusion center. In contrary, [55] considers a scenario where the local LLRs of CR users are sent in serial manner, i.e., one after another. They are combined sequentially at the fusion center to quickly detect the presence or the absence of primary user's signal. Sequential detection that aims for maximizing the average total achievable rate of a multi-channel cognitive radio systems under prescribed probability of missed detection is studied in [56]. The most recent work on collaborative spectrum sensing based on sequential detection can be found in [57]. Here, sequential tests are carried out at each CR user and the fusion center. First, each CR user computes the decision statistic based on local observation

and whenever it crosses predefined thresholds, the corresponding user sends quantized information (one or several bits) to the fusion center. The fusion center that asynchronously receives the bits from different CR users subsequently performs the SPRT to reach a global decision. However, most of the works mentioned above consider the case where the corresponding statistical parameters are known (simple hypotheses). Sequential detection for composite hypotheses is rarely addressed in literature.

1.3 Research context and contributions

The main difference to all the above literature, is that in this thesis we use resampling to develop the algorithms for spectrum sensing. It includes resampling in fixed sample size testing and also in sequential testing, using the non-parametric and parametric bootstrap. The reason of using bootstrap resampling is twofold. First, bootstrap based testing is able to work properly with a small sample size, when the devised asymptotic test fails. Second, bootstrap resampling is an appropriate choice to alleviate the problem of intractable distribution of a statistic. These bootstrap properties are explored to achieve the objective of developing spectrum sensing algorithms with a short sensing time. We summarize the contributions of the thesis as follow.

In fixed sample size testing, we consider the problem of detecting a correlated signal with a small sample size, when the underlying noise is uncorrelated.

- The bootstrap null resampling test: When the sample size is small, the asymptotic test often fails to work properly. The bootstrap pivot test, which is a common approach that is based on a pivotal statistic, also does not perform well in the scenario at hand. For this case, we propose to use the bootstrap null resampling test that is not based on a pivotal statistic. The bootstrap null resampling test outperforms the bootstrap pivot test.

- Resampling for Chair-Varshney fusion rule: In a changing wireless channel, between the transmitter of a primary user and the receiver of a cognitive radio user, it is difficult to apply the Chair-Varshney fusion rule due to the requirement of the fusion center to know each CR user performance indices (the probabilities of false alarm and miss detection). In this respect, we use the bootstrap null resampling for a local decision, and we estimate the local probability of detection in each measurement by estimating the distribution of the test statistic under the null hypothesis and the alternative hypothesis. This can be accomplished

by performing independent resampling to generate the distribution under the null hypothesis and moving-block bootstrap to generate the distribution under the alternative hypothesis. In addition, the distribution of the fusion rule under the null hypothesis, for global decision, is estimated by the parametric bootstrap when the number of CR users is large, instead of using all possible binary combinations. Thus, it significantly reduces the computational cost.

In sequential testing, we consider the problem of implementing sequential probability ratio tests for composite hypotheses.

- Bootstrap based sequential probability ratio tests for composite hypotheses: When we use the maximum likelihood estimator in the generalized sequential probability ratio test to estimate the unknown parameters, the two constant thresholds of the SPRT cannot be used directly. Instead, they should be updated according to the current sample size in order to avoid excessive decision errors. For this case, we develop a method to update the thresholds adaptively, using the parametric bootstrap. The resulting scheme has smaller average sample numbers (ASNs) than the scheme that is based on an asymptotic approach, while it still preserves the decision errors below the nominal values.
- Bootstrap based sequential probability ratio tests with reduced computational cost: We propose to use a convex combination of the latest distributions to reduce computational costs induced by the bootstrap. The reduction of computational costs does not significantly increase the ASN.

Apart from the bootstrap based approach, some works have been done during the doctoral program. They include:

- Multiple testing for multiband spectrum sensing: Multiple testing procedures can control decision errors on the system level (familywise). Therefore, they are suitable to be applied in multiband spectrum sensing where a group of spectrum bands are jointly detected. In a fixed sample size case, we propose to use an adaptive Benjamini-Hochberg procedure, since it is more powerful than any other multiple testing procedures. In a random sample size case, we devise a procedure that is based on the ordered stopping times of sequential probability ratio tests.
- Locally optimum detector for heavy-tailed noise: A locally optimum detector is derived where the underlying noise follows a Student's t-distributions. The approach is suitable for robust spectrum sensing where heavy-tailed noise occurs in a low SNR region.

- Hard decision combining with a quality information bit: Collaborative spectrum sensing using hard decision in combination with quality information is developed to improve the performance of the conventional HDC. An energy detector with a preset and two optimum thresholds is employed at each CR user. In this scheme, the fusion center performs the optimum Neyman-Pearson fusion rule.

1.4 Publications

The following publications have been produced during the period of doctoral candidacy.

Internationally refereed journal articles

- F.Y. Suratman and A.M. Zoubir, "Bootstrap based sequential probability ratio tests for spectrum sensing in cognitive radio," *IEEE Transactions on Signal Processing*, submitted.
- F.Y. Suratman and A.M. Zoubir, "Resampling based spectrum sensing for small sample sizes," *Digital Signal Processing*, to be submitted.

Internationally refereed conference papers

- F.Y. Suratman and A.M. Zoubir, Multiple Testing for Sequential Probability Ratio Tests with Application to Multiband Spectrum Sensing, in *The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, accepted.
- F.Y. Suratman and A.M. Zoubir, Bootstrap Based Sequential Probability Ratio Tests, in *The IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6352-6356, Vancouver, Canada, May 2013.
- F.Y. Suratman, A. Tetz and A.M. Zoubir , Collaborative Spectrum Sensing Using Sequential Detections: Soft Decision Vs. Hard Decision, in *IEEE International Conference on Information and Communication Technology*, pp. 1-6, Bandung, Indonesia, March 2013.
- F.Y. Suratman, and A.M. Zoubir, Effects of Model Errors on Multiple Antenna Based Spectrum Sensing Using Sequential Detection, in *The 1st International Conference on Communications, Signal Processing and Their Applications (ICCSPA)*, Sharjah, UAE, Feb. 2013.

- F.Y. Suratman, and A.M. Zoubir, Collaborative Spectrum Sensing in Cognitive Radio Using Hard Decision Combining with Quality Information, in *The IEEE Workshop on Statistical Signal Processing (SSP)*, pp. 377-380, Nice, France, Jun. 2011.
- V. Pohl, F.Y. Suratman, A. M. Zoubir and H. Boche, Spectrum Sensing for Cognitive Radio Architectures based on sub-Nyquist Sampling Schemes, in *The International ITG Workshop on Smart Antennas (WSA)*, pp. 1-8, Aachen, Germany, Feb. 2011.
- F.Y. Suratman, Y. Chakhchoukh, and A.M. Zoubir, Locally Optimum Detection in Heavy-Tailed Noise for Spectrum Sensing in Cognitive Radio, in *The 2nd International Workshop on Cognitive Information Processing (CIP)*, pp. 134-139, Elba, Italy, Jun. 2010.

1.5 Thesis overview

The thesis outline is as follows:

Chapter 2 begins with a general problem formulation for spectrum sensing in cognitive radio. The main part of the chapter is to introduce the basic theory for spectrum sensing. The introduction is generally divided into two parts, namely, detection with a fixed sample size and with a random sample size.

Chapter 3 presents resampling based techniques for spectrum sensing. After the asymptotic test is highlighted, it continues with the bootstrap pivot test and the bootstrap null resampling test as the proposed approach for local spectrum sensing. In collaborative spectrum sensing, resampling techniques for the Chair-Varshney fusion rule is presented. We end this chapter with multiband spectrum sensing using multiple testing procedures.

Chapter 4 extends the sequential probability ratio tests to composite hypothesis cases. The chapter introduces the thresholding problem in the generalized sequential probability ratio test. Then, it leads to the bootstrap approach, where we discuss a technique using the parametric bootstrap to solve the thresholding problem. Last, a method to reduce the computational cost of the bootstrap is discussed.

Chapter 5 discusses various approaches for spectrum sensing. First, a multiple testing procedure for sequential testing is introduced. The main application is multiband

spectrum sensing. Second, a robust locally optimum detector is presented. It is designed to detect a signal with low power in underlying heavy-tailed noise. Third, an extension of the HDC in collaborative spectrum sensing is discussed. The idea is to include a quality information bit in the transmitted data of a CR user to the fusion center.

Finally, conclusions and an outlook are presented in **Chapter 6**.

Chapter 2
Detection theory for spectrum sensing

This chapter is concerned with discussing the basic detection theory for spectrum sensing. It serves as a necessary background to understand the work presented in the following chapters. The general concept of binary hypothesis testing for spectrum sensing is presented in Section 2.1. In Section 2.2, different approaches to derive a detector for spectrum sensing with a fixed sample size are discussed. It includes the likelihood ratio test, the generalized likelihood ratio test, and the bootstrap test. In Section 2.3, the sequential probability ratio test is presented. The algorithm and the performance measure of the sequential probability ratio test are briefly explained.

2.1 The binary hypothesis testing problem for spectrum sensing

Spectrum sensing is a binary hypothesis testing problem, between a null hypothesis $\mathcal{H}_0$, to represent that the primary user's signal is absent (spectrum band unoccupied), against an alternative $\mathcal{H}_1$, to represent the opposite (spectrum band occupied). In general, the canonical spectrum sensing can be stated as follows. Let $\mathbf{x}_N = (\mathbf{x}[1]\ \mathbf{x}[2]\ \cdots\ \mathbf{x}[N])$ be a sequence of $N < \infty$ samples of a signal. It then decides whether the transmitted signal $\mathbf{r}_N = (\mathbf{r}[1]\ \mathbf{r}[2]\ \cdots\ \mathbf{r}[N])$ is present in $\mathbf{x}_N$. Formally, it can be written as

$$\mathcal{H}_0 : \mathbf{x}[n] = \mathbf{w}[n] \tag{2.1a}$$

$$\mathcal{H}_1 : \mathbf{x}[n] = \mathbf{r}[n] + \mathbf{w}[n], \quad n = 1, 2, \ldots, N, \tag{2.1b}$$

where $\mathbf{w}[n]$ denotes the additive noise component. Here, $\mathbf{x}[n]$ can be a vector or a scalar observation.

Now, if we assume that the data $\mathbf{x}_N$ comes from a specific distribution, the parametric model can be used. The sensor observes signals that follow the same family of distributions which can be described using a finite number of parameters, each under $\mathcal{H}_0$ and $\mathcal{H}_1$. For a theoretical description, the joint density functions under $\mathcal{H}_0$ and $\mathcal{H}_1$ are

$$\mathcal{H}_0 : f_{0,N}(\mathbf{x}_N; \boldsymbol{\theta}_0) \tag{2.2a}$$

$$\mathcal{H}_1 : f_{1,N}(\mathbf{x}_N; \boldsymbol{\theta}_1), \tag{2.2b}$$

where $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are the parameters, which could be scalars or vectors, under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. The respective parameter spaces are denoted as $\boldsymbol{\Theta}_0, \boldsymbol{\Theta}_1 \in \boldsymbol{\Theta}$.

Suppose that we consider $\boldsymbol{\theta}_i,\ i = 0, 1$, consisting of m parameters, i.e., $\boldsymbol{\theta}_i = \{\theta_1, \theta_2, \ldots, \theta_m\}$. If a hypothesis uniquely specifies each of the $\theta_k,\ k = 1, \ldots, m$, then it is called *simple*. A hypothesis is *composite* when one or more of the θ_i are not uniquely specified.

In some applications, we cannot rely on assumptions that the data are drawn from a given probability distribution. In this respect, the non-parametric version of the binary hypothesis testing problem is,

$$\mathcal{H}_0 : \boldsymbol{\psi} \in \boldsymbol{\Psi}_0 \tag{2.3a}$$

$$\mathcal{H}_1 : \boldsymbol{\psi} \in \boldsymbol{\Psi}_1, \tag{2.3b}$$

$$\tag{2.3c}$$

here, $\boldsymbol{\psi}$ represents specific statistics quantifying the presence or the absence of the desired signal, and $\boldsymbol{\Psi}_0$ and $\boldsymbol{\Psi}_1$ are the spaces under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Several types of the nonparametric detector can be found in [58].

In a cognitive radio context, $\mathcal{H}_0$ represents the condition that the observed spectrum band is unoccupied and hence CR users can establish a connection to exchange their data. Meanwhile, $\mathcal{H}_1$ is to represent that the band is occupied by incumbent primary users and CR users are not allowed to further establish a connection. The following general performance indices are usually used in hypothesis testing, each of which correspondens to some performance measure in cognitive radio.

- *Probability of false alarm (P_f)*: the probability of deciding for $\mathcal{H}_1$ when $\mathcal{H}_0$ is true. It indicates the probability of missing the opportunity to use an unoccupied band.

- *Probability of miss detection (P_m)*: the probability of deciding for $\mathcal{H}_0$ when $\mathcal{H}_1$ is true. It indicates the probability of causing interference to primary users.

- *Sample size (N)*: the minimum sample size to fulfill the requirement on P_m and P_f. It indicates the agility of CR users which in turn has an impact on the throughput of the cognitive radio network and the level of interference to the primary network.

Note that P_f, P_m and N are all interrelated for measuring the overall performance of a detector, particularly for spectrum sensing .

Considering the sample size, the design of the detector can be categorized into two main approaches. The first is to fix the sample size in advance based on some criterion. It will be referred to as *fixed sample size detector*. The second is to let the sample size be random, depending on the current observation, namely *sequential detector*.

2.2 Fixed sample size detector

There are several different approaches to choosing the form of a detector in hypothesis testing. Two of the more well-known criteria are the Bayes criterion and Neyman-Pearson (NP) criterion. In this thesis, we are mainly concerned with the NP criterion. It is not possible here to fully cover the concept of designing a detector based on the the NP criterion and a sub-optimal approach that follows afterwards. The details can be found, for example, in [11] and [59].

2.2.1 The Neyman-Pearson Criterion

Let $\mathbf{x}_N$ denote observation data of a fixed size N and let $T(\mathbf{x}_N)$ represent a test statistic as a function of the data. A generic fixed sample size detector can be written as follows

$$T(\mathbf{x}_N) \begin{cases} > \tau, & \text{accept } \mathcal{H}_1 \\ < \tau, & \text{accept } \mathcal{H}_0, \end{cases} \tag{2.4}$$

where τ is a threshold that separates two decision regions R_0 (accept $\mathcal{H}_0$) and R_1 (accept $\mathcal{H}_1$). The probabilities of false alarm P_f and miss detection P_m now can be defined formally as

$$\begin{aligned} P_f &= P_0(T(\mathbf{x}_N) > \tau) \\ P_m &= P_1(T(\mathbf{x}_N) < \tau), \end{aligned} \tag{2.5}$$

here $P_i(A)$ denotes the probability of event A under hypothesis $\mathcal{H}_i$. Note that it is not possible to reduce both error probabilities simultaneously. To have an optimal detector, a typical approach is to hold one error probability fixed while minimizing the other.

The so-called NP criterion states that a detector is optimal if the test statistic $T(\mathbf{x}_N)$ and the threshold τ are chosen such that the probability of miss detection P_m is minimized (the probability of detection $P_d = 1 - P_m$ is maximized) subject to the constraint that the probability of false alarm P_f is equal to a fixed nominal value α [11]. For simple hypotheses testing (2.2), i.e., $\boldsymbol{\Theta}_i = \{\boldsymbol{\theta}_i\},\ i = 0, 1$, the criterion is satisfied when the test statistic is the likelihood ratio (LR),

$$T(\mathbf{x}_N) = \frac{f_{1,N}(\mathbf{x}_N; \boldsymbol{\theta}_1)}{f_{0,N}(\mathbf{x}_N; \boldsymbol{\theta}_0)}, \tag{2.6}$$

and the threshold τ is found from

$$P_f = \int_{\{\mathbf{x}_N : T(\mathbf{x}_N) > \tau\}} f_0(\mathbf{x}_N; \boldsymbol{\theta}_0) d\mathbf{x}_N = \alpha. \tag{2.7}$$

The test statistic $T(\mathbf{x}_N)$ in (2.6) is termed the LR since it indicates for each value of $\mathbf{x}_N$ the likelihood of $\mathcal{H}_1$ versus the likelihood of $\mathcal{H}_0$. Commonly, the test statistic is expressed by taking the logarithm of the $T(\mathbf{x}_N)$, namely the log-likelihood ratio test (LLRT). The detector based on the NP criterion now becomes

$$Z_N = \ln \frac{f_{1,N}(\mathbf{x}_N; \boldsymbol{\theta}_1)}{f_{0,N}(\mathbf{x}_N; \boldsymbol{\theta}_0)} \begin{cases} > \ln \tau, & \text{accept } \mathcal{H}_1 \\ < \ln \tau, & \text{accept } \mathcal{H}_0. \end{cases} \tag{2.8}$$

Recall that the threshold τ is determined by the nominal value of the probability of false alarm α in (2.7).

As an example, we use the model of complex Gaussian signal in complex Gaussian noise. The competing hypotheses are

$$\begin{aligned} \mathcal{H}_0 :&\ \mathbf{x}[n] \sim \mathcal{CN}\left(0, \sigma_0^2\right), \\ \mathcal{H}_1 :&\ \mathbf{x}[n] \sim \mathcal{CN}\left(0, \sigma_1^2\right), \end{aligned} \tag{2.9}$$

where σ_0^2 and σ_1^2 are variances under hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Here, $\mathbf{x}[n]$ and $\mathbf{w}[n]$ are scalar. Let the observations be identically independent distribution (iid). Using (2.8), we obtain

$$Z_N = \sum_{n=1}^{N} \ln \frac{f_1(\mathbf{x}[n]; \sigma_1^2)}{f_0(\mathbf{x}[n]; \sigma_1^2)} = \left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right) \sum_{n=1}^{N} |\mathbf{x}[n]|^2 + N \ln \frac{\sigma_0^2}{\sigma_1^2}, \tag{2.10}$$

which is the energy detector.

2.2.2 The generalized likelihood ratio test

Previously, we had assumed complete knowledge of the distributions under $\mathcal{H}_0$ and $\mathcal{H}_1$ (simple hypotheses) to design an optimal detector. However, the approach is impractical in most situations where one or more parameters of the distributions are unknown.

For example in cognitive radio, a CR user might have no complete knowledge of the power or the frequency of the transmitted signal from a primary user. Similarly, the noise might be reasonably modeled as white Gaussian but with unknown variance. Therefore, it is important to design good detectors when the distributions under $\mathcal{H}_0$ and $\mathcal{H}_1$ are incompletely known. In this situation, we have to rely on sub-optimum methods.

One of the most well-known approaches for composite hypotheses testing is the generalized likelihood ratio test (GLRT). The GLRT replaces any unknown parameters by their maximum likelihood estimates (MLEs) [60]. In general, the GLRT has the form

$$\hat{T}(\mathbf{x}_N) = \frac{f_1(\mathbf{x}_N; \hat{\boldsymbol{\theta}}_1)}{f_0(\mathbf{x}_N; \hat{\boldsymbol{\theta}}_0)} \begin{cases} > \tau, & \text{accept } \mathcal{H}_1 \\ < \tau, & \text{accept } \mathcal{H}_0, \end{cases} \tag{2.11}$$

where $\hat{\boldsymbol{\theta}}_1$ is the MLE of $\boldsymbol{\theta}_1$ assuming $\mathcal{H}_1$ is true, and $\hat{\boldsymbol{\theta}}_0$ is the MLE of $\boldsymbol{\theta}_0$ assuming $\mathcal{H}_0$ is true. Formally,

$$\hat{\boldsymbol{\theta}}_i = \underset{\boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i}{\arg\max} \ln \left(f_i(\mathbf{x}_N; \boldsymbol{\theta}_i) \right), \ i = 0, 1. \tag{2.12}$$

The threshold τ, as in the simple hypotheses, is found from the nominal value of the probability of false alarm α.

2.2.3 p-value and the bootstrap test

Two major problems can be encountered in hypothesis testing. The first one occurs when the sample size is small and asymptotic methods do not apply. The second possible problem is that the distribution of the test statistic cannot be determined analytically. One can overcome both problems by using bootstrap techniques [61].

Suppose that T represent a test statistic of binary hypotheses. The basic idea of hypothesis testing is to compare the observed value of a test statistic t with the distribution that it would follow if the null hypothesis $\mathcal{H}_0$ was true. In general, we follow the convention that large values of T are evidence against $\mathcal{H}_0$. It is common to measure the level of evidence against $\mathcal{H}_0$ by using a *p-value*. Formally, it is defined as

$$p = P_0(T \geq t) = 1 - G_0(t), \tag{2.13}$$

where $G_0(T)$ is the cumulative distribution function of T under $\mathcal{H}_0$. For continues T, the p-value under $\mathcal{H}_0$ has a uniform distribution on $[0, 1]$. Therefore, the corresponding random variable P has distribution

$$P_0(P \leq p) = p. \tag{2.14}$$

The interpretation of the p-value is then equivalent to following a procedure which rejects $\mathcal{H}_0$ with error rate p [62].

Suppose that $G_0(T)$ is known. We would calculate the p-value and reject $\mathcal{H}_0$ whenever $p < \alpha$. This is equivalent to rejecting $\mathcal{H}_0$ whenever t exceeds the critical value $G_0^{1-\alpha}(T)$, which is the $1-\alpha$ quantile of $G_0(T)$. Unfortunately, in most practical scenarios $G_0(T)$ is unknown and hence we commonly rely on an asymptotic approximation, when it exists, for the distribution. However, this approach may or may not work well. As an alternative, we could use a bootstrap test which depends on resampling the available data to estimate the distributional information.

In a bootstrap test, we first generate B_s bootstrap data by resampling the observed data with replacements, which could be carried out parametrically or non-parametrically [63]. Each bootstrap data b, $b = 1, \cdots, B_s$, is then used to compute a bootstrap test statistic t^b, by the same procedure used to calculate t from the observed data. The bootstrap p-value $\hat{p}^*$, which is an estimate of the p-value in (2.13), can then be written as

$$\hat{p}^* = 1 - \hat{G}_0(t) = \frac{1 + \#\{t^b \geq t\}}{B_s + 1}, \tag{2.15}$$

where $\hat{G}_0(t)$ denotes the empirical distribution function (EDF) of t^b [62]. Note that in order to estimate the p-value using the bootstrap, resampling should be performed in a way that reflects the null hypothesis, even when the underlying data is far from the null [64]. This can be accomplished by some methods that we describe in Chapter 3.

2.3 The sequential probability ratio test

In fixed sample size detection, after N samples are recorded one of two possible decisions is taken: accept the null hypothesis $\mathcal{H}_0$ or accept the alternative hypothesis $\mathcal{H}_1$. In some cases the evidence based on the N samples might strongly support one of the two hypotheses. In other cases the evidence might be less convincing. Nevertheless, the decision has to be made. In sequential tests a third possible course of action is introduced to overcome the ambiguity, which is to take more observations. Therefore, a sequential test typically continues observing until the evidence strongly favors one of the two hypotheses. There are various stopping rules and decision rules that we can use in sequential tests [65]. Here, we are only concerned with the so-called *sequential probability ratio test* (SPRT) which was coined by A. Wald in 1945 [45].

2.3.1 The algorithm

Let $\mathbf{x}_N = (\mathbf{x}[1]\ \mathbf{x}[2]\ \cdots\ \mathbf{x}[N])$ be a sequence of i.i.d. observations of the signal recorded up to the sample N. Here, $\mathbf{x}[n]$ is assumed to admit the distribution described by the density function in (2.2) under each hypothesis. Suppose that the probability density functions of the observation $f_0(\mathbf{x}[n];\boldsymbol{\theta}_0)$ and $f_1(\mathbf{x}[n];\boldsymbol{\theta}_1)$ are exactly known. The log-likelihood ratio can then be written as

$$\begin{aligned}\bar{Z}_N &= \frac{f_{1,N}(\mathbf{x}_N;\boldsymbol{\theta}_1)}{f_{0,N}(\mathbf{x}_N;\boldsymbol{\theta}_0)} = \sum_{n=1}^{N} \ln \frac{f_1(\mathbf{x}[n];\boldsymbol{\theta}_1)}{f_0(\mathbf{x}[n];\boldsymbol{\theta}_0)} \\ &= \bar{Z}_{N-1} + \bar{z}_N, \quad N = 1, 2, \cdots \end{aligned} \tag{2.16}$$

where $\bar{Z}_0 = 0$ and

$$\bar{z}_N = \ln \frac{f_1(\mathbf{x}[N];\boldsymbol{\theta}_1)}{f_0(\mathbf{x}[N];\boldsymbol{\theta}_0)} \tag{2.17}$$

represents the increment at the time instance N. In Neyman-Pearson's approach [60] for testing a simple hypothesis $\mathcal{H}_0$ against a simple alternative $\mathcal{H}_1$, the sample size is pre-specified to $N = N_{fix}$ and a threshold τ is chosen based on the false alarm rate α for either accepting $\mathcal{H}_0$ ($\bar{Z}_N < \tau$) or accepting $\mathcal{H}_1$ ($\bar{Z}_N \geq \tau$). Meanwhile, the SPRT introduces a third option in addition to accepting $\mathcal{H}_0$ for small $\bar{Z}_N$ or accepting $\mathcal{H}_1$ for large $\bar{Z}_N$, namely it continues observing the data for intermediate values of $\bar{Z}_N$. More precisely, choose constants A and B where $-\infty < B < 0 < A < +\infty$ and increase $N \leftarrow N + 1$ to take more samples until the stopping time

$$N_s = \min_{N \geq 1} \{N :\ \bar{Z}_N \notin (B, A)\}. \tag{2.18}$$

It is obvious that the sample size is random, depending on the statistics of the observations. The test stops when either the upper threshold A or the lower threshold B is crossed for the first time. In summary, the SPRT can be defined as

$$\bar{Z}_N = \ln \frac{f_{1,N}(\mathbf{x}_N;\boldsymbol{\theta}_1)}{f_{0,N}(\mathbf{x}_N;\boldsymbol{\theta}_0)} \begin{cases} \geq A, & \text{accept } \mathcal{H}_1 \\ \leq B, & \text{accept } \mathcal{H}_0, \\ A < \bar{Z}_N < B, & N \leftarrow N + 1. \end{cases} \tag{2.19}$$

Using Wald's approximation [44], the two thresholds A and B can be found from

$$A \approx \ln \frac{1-\beta}{\alpha} \quad \text{and} \quad B \approx \ln \frac{\beta}{1-\alpha}, \tag{2.20}$$

where α and β are the nominal probabilities of false alarm and miss detection, respectively. By setting the thresholds according to (2.20), the actual probabilities of false

alarm P_f and of miss detection P_m of the SPRT might not be exactly equal to α and β, respectively. However, the two inequalities below are guaranteed [45]

$$P_f \leq \frac{\alpha}{1-\beta} \quad \text{and} \quad P_m \leq \frac{\beta}{1-\alpha}. \tag{2.21}$$

In real applications the values of α and β are usually small, such that $\alpha \approx \alpha/(1-\beta)$ and $\beta \approx \beta/(1-\alpha)$. Hence, from (2.21), the amount by which P_f may exceed α or P_m may exceed β is very small and can be neglected for all practical purposes [44]. The Algorithm 2.1 summarizes the implementation of the SPRT.

Algorithm 2.1 Implementation of the SPRT

Step 0) Initialize: $N = 0$, α and β to find the thresholds A and B from (2.20)

REPEAT

Step 1) Draw next sample $N \leftarrow N + 1$

Step 2) Calculate the increment $\bar{z}[N]$ from (2.17)

Step 3) Update $\bar{Z}_N$ by (2.16)

UNTIL $\bar{Z}_N \geq A$ or $\bar{Z}_N \leq B$

Step 4) If $\bar{Z}_N \geq A$, accept $\mathcal{H}_1$ and If $\bar{Z}_N \leq B$, accept $\mathcal{H}_0$

Some conditions should hold in order to have a finite stopping time in the SPRT. Suppose that the second moment of the increment (2.17) under each hypothesis is

$$E_i\left[(\bar{z})^2\right] \neq 0, \ i = 0, 1, \tag{2.22}$$

where $E_i[\cdot]$ denotes the expected value under $\mathcal{H}_i$.[1] With this condition, using the Lemma 1 in [66], it is proven that the SPRT with a finite stopping time N_s is a certain event, formally

$$P_i(N_s < \infty) = 1, \ i = 0, 1, \tag{2.23}$$

where $P_i(A)$ represents the probability of event A under $\mathcal{H}_i$.

2.3.2 Performance measures

Performances of the SPRT are described by the *operating characteristic* (OC) function and the *average sample number* (ASN) function. The OC function is related to the

[1] We drop the subscript N of the increment in (2.17) to signify that the random variable $\bar{z}_N$ is distributed independently of N.

decision errors and the ASN function signifies how fast the decision is made. Note that throughout the thesis we use the term sample size instead of sample number. However, in sequential testing, the term ASN is a more familiar term instead of average sample size (ASS).

Let $f(\mathbf{x}[n];\boldsymbol{\theta})$ represent the underlying distribution where the true parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The OC function $\mathcal{L}(\boldsymbol{\theta})$ is defined to be the probability of $\mathcal{H}_0$ acceptance under the distribution $f(\mathbf{x}[n];\boldsymbol{\theta})$. Suppose that the increment $\bar{z}$ under $f(\mathbf{x}_N;\boldsymbol{\theta})$ is very small compared to $A - B$. The OC function can be approximated by [45]

$$\mathcal{L}(\boldsymbol{\theta}) \approx \frac{e^{tA} - 1}{e^{tA} - e^{tB}}, \tag{2.24}$$

where $t \neq 0$ is a real number, which is the solution of the equation[2]

$$\int_{\mathbf{x}} \left[\frac{f_1(\mathbf{x};\boldsymbol{\theta}_1)}{f_0(\mathbf{x};\boldsymbol{\theta}_0)}\right]^t f(\mathbf{x};\boldsymbol{\theta}) d\mathbf{x} = 1. \tag{2.25}$$

The probability of miss detection P_m can be approximated by (2.24) when the underlying distribution $f(\mathbf{x}[n];\boldsymbol{\theta}) = f_1(\mathbf{x}[n];\boldsymbol{\theta}_1)$. In this case, $t = -1$ and hence

$$\mathcal{L}(\boldsymbol{\theta}_1) \approx \frac{e^{-A} - 1}{e^{-A} - e^{-B}}. \tag{2.26}$$

Meanwhile, the probability of false alarm P_f can be approximated when the underlying distribution $f(\mathbf{x}[n];\boldsymbol{\theta}) = f_0(\mathbf{x}[n];\boldsymbol{\theta}_0)$. In this case, $t = 1$ and hence

$$P_f = 1 - \mathcal{L}(\boldsymbol{\theta}_0) \approx \frac{1 - e^{B}}{e^{A} - e^{B}}. \tag{2.27}$$

The thresholds A and B are found from (2.20).

The ASN function is the expected value of random stopping time N_s of the SPRT when the underlying distribution is $f(\mathbf{x}[n];\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. By using the same argument that the increment $\bar{z}$ under $f(\mathbf{x}_N;\boldsymbol{\theta})$ is very small compared to $A - B$, the ASN function can be approximated by [45]

$$E_{\mathcal{H}}[N_s] \approx \frac{(1 - e^{tB})A + (e^{tA} - 1)B}{(e^{tA} - e^{tB})\, E_{\mathcal{H}}[\bar{z}]}. \tag{2.28}$$

where $E_{\mathcal{H}}[\cdot]$ denotes the expected value under $f(\mathbf{x}[n];\boldsymbol{\theta})$. The ASN under $\mathcal{H}_0$ and $\mathcal{H}_1$ can be found from (2.28) by substituting $t = 1$ and $t = -1$, respectively.

[2]We drop the time index n to indicate that the solution is time independent.

Chapter 3

Resampling based spectrum sensing for small sample sizes

Spectrum sensing is mostly devised based on the assumption that the sample size is large, by which the distribution of the test statistic can be approximated asymptotically. However, when we use a small sample size in the respective test statistic, the resulting approximation of the distribution could depart significantly from the true distribution. As a result, the performance of the spectrum sensing in terms of the probability of false alarm or miss detection is poor. In such situations, we propose to use the bootstrap. Unlike the asymptotic test, the bootstrap approach shows a proper balance between the probability of false alarm and miss detection.

The application of the bootstrap to various signal processing techniques can be found in [61, 67, 68] and references therein. Bootstrap-based detection methods have been successfully implemented in [69–71]. Literature on the implementation of bootstrap in spectrum sensing is scarce. In [72], the parametric bootstrap is used to generate the distribution of the likelihood ratio test under the null hypothesis. In [73], the authors suggest to use the bootstrap approach in multiantenna spectrum sensing that is based on eigenvalue distributions. In all the above literature, the detection methods are mainly devised from the appropriately constructed bootstrap confidence intervals [64], e.g., bootstrap pivoting. However, this might not work in some scenarios, including the one in this chapter. This is because, in a bootstrap based detection, there is a requirement to generate the bootstrap distribution of test statistics under a specific null hypothesis [74]. Thus, here we provide an alternative, where we use the null resampling method in order to have a better bootstrap approximation for the distribution under the null hypothesis. We also implement resampling methods in collaborative spectrum sensing. We apply the non-parametric bootstrap to estimate the probability of detection at each CR user and parametric bootstrap to estimate the distribution of the fusion rule under the null hypothesis, when the number of CR users is large. Furthermore, we study the implementation of the devised schemes to multiband spectrum sensing using multiple testing procedures [75].

The system model is explained in Section 3.1, including the asymptotic test for reference. In Section 3.2, we begin with the pivot bootstrap test and continue with the proposed null resampling test for local spectrum sensing. Resampling in collaborative spectrum sensing is explained in detail in Section 3.3. We proceed with multiband

spectrum sensing in Section 3.4. Simulation results are given in Section 3.5. Finally, we conclude in Section 3.6.

3.1 System Model

We consider the canonical spectrum sensing model (2.1), but now we assume that the received signal $\mathbf{r}[n],\ n = 1, \cdots, N$, is temporally correlated. Let us define the following vectors of M consecutive samples

$$\mathbf{x}_M = (\mathbf{x}[n]\ \mathbf{x}[n-1]\ \cdots\ \mathbf{x}[n-M+1])^T \tag{3.1}$$

$$\mathbf{r}_M = (\mathbf{r}[n]\ \mathbf{r}[n-1]\ \cdots\ \mathbf{r}[n-M+1])^T \tag{3.2}$$

$$\mathbf{w}_M = (\mathbf{w}[n]\ \mathbf{w}[n-1]\ \cdots\ \mathbf{w}[n-M+1])^T. \tag{3.3}$$

It is easy to verify that

$$\mathbf{R}_x = \mathbf{R}_r + \mathbf{R}_w \tag{3.4}$$

where

$$\mathbf{R}_x = E\left[\mathbf{x}_M \mathbf{x}_M^T\right],\ \ \mathbf{R}_r = E\left[\mathbf{r}_M \mathbf{r}_M^T\right],\ \ \mathbf{R}_w = E\left[\mathbf{w}_M \mathbf{w}_M^T\right] \tag{3.5}$$

are the covariance matrices of the corresponding signals and noise. Since we assume that the noise is uncorrelated, i.e., $\mathbf{R}_w = \sigma_w^2 \mathbf{I}_M$, $\mathbf{R}_x$ is a diagonal matrix when the signal $\mathbf{r}[n]$ is absent. Otherwise, some of the off-diagonal elements of $\mathbf{R}_x$ are non-zeros. Using this property, the authors in [22] propose to use the following statistic to detect the signal

$$T = \frac{\frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{M}|r_{ij}|}{\frac{1}{M}\sum_{i=1}^{M}|r_{ii}|}. \tag{3.6}$$

Here, r_{ij} denotes the element in the ith row and jth column of the matrix $\mathbf{R}_x$. Correspondingly, the binary hypotheses problem in this case can be rewritten as follows

$$\mathcal{H}_0 : T = 1 \tag{3.7a}$$

$$\mathcal{H}_1 : T > 1. \tag{3.7b}$$

In practice, the sample size is limited. Hence, the sample autocorrelation of the received signal is used, i.e.,

$$\hat{r}_{ij} = \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{x}[n]\mathbf{x}[n-|i-j|], \quad |i-j| = 0, 1, \cdots, M-1, \tag{3.8}$$

where N denotes the number of available samples and $M < N$ is the smoothing factor [22]. Therefore, the estimator of T is

$$\hat{T} = \frac{\frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{M}|\hat{r}_{ij}|}{\frac{1}{M}\sum_{i=1}^{M}|\hat{r}_{ii}|}. \tag{3.9}$$

It is noteworthy that the signal here is real. The extension to the complex signal is straightforward.

Using an asymptotic approximation, when $N \to \infty$, the authors in [22] derive the following threshold for the nominal probability of false alarm α,

$$\tau = \frac{1 + (M-1)\sqrt{\frac{2}{N\pi}}}{1 - \mathcal{Q}^{-1}(\alpha)\sqrt{\frac{2}{N}}}, \tag{3.10}$$

where $\mathcal{Q}^{-1}(\cdot)$ is the inverse of Q-function of the standard normal cumulative distribution function.[1] The advantage of using the test statistic (3.6) is that it does not need any prior information about the signal, channel and the noise. The performance of this approach is satisfying, as long as it receives a correlated signal and the underlying noise is uncorrelated. However, the performance degrades significantly when the sample size is not sufficiently large. This might be because the convergence rate to the approximating distribution is very slow. In this case, we propose to use the bootstrap to improve the performance when the sample size is small. In the sequel, we consider sample sizes $N \leq 100$.

3.2 The bootstrap test for local spectrum sensing

In general, resampling in bootstrap can be done in two ways, parametric or non-parametric [61]. In the parametric setting, an explicit expression for the sampling distribution of the data could be specified with a finite number of unknown parameters. Meanwhile, no particular forms are specified for the distributions in the non-parametric setting. The problem that we consider in this section conforms with the non-parametric setting. Therefore, we mainly focus on the non-parametric bootstrap in the sequel. The parametric case can be found, for example, in [62, 63].

[1] $\mathcal{Q}(x)$ is defined to be: $\mathcal{Q}(x) = \int_x^\infty 1/\sqrt{2\pi}e^{-t^2/2}dt$.

3.2.1 The pivot bootstrap test

The null hypothesis in (3.7) corresponds to a particular value of T, e.g., $T_0 = 1$. Thus, it is reasonable to use the equivalence between significance tests and confidence sets [62]. This means that if the value T_0 is outside a $1-\alpha$ confidence set for T, then T differs from T_0 with a p-value less than α. For the corresponding alternative hypothesis in (3.7), the choice based on this equivalence would be the pivot bootstrap test, which can be defined as

$$T_{piv} = \frac{\hat{T} - T_0}{\hat{S}_e} = \frac{\hat{T} - 1}{\hat{S}_e}, \tag{3.11}$$

where $\hat{T}$ denotes the estimator of T. Here, $\hat{S}_e$ is an estimator of the standard error of $\hat{T}$, which can be calculated, for example, by using bootstrap or jackknife [63]. When the resulting statistic is pivotal or approximate pivot, its sampling distribution should not depend upon which distribution F generated the data (i.e., not depend on the unknown value of T), even when F departs from the null distribution F_0. This pivotal property allows us to construct the bootstrap replication of T_{piv} using

$$T^b_{piv} = \frac{\hat{T}^b - \hat{T}}{\hat{S}^b_e}, \tag{3.12}$$

$\hat{T}^b$ and $\hat{S}^b_e$ denote the bootstrap statistics which can be obtained by the same procedure as $\hat{T}$ and $\hat{S}_e$, respectively. However, they use the bootstrap samples instead of the observed data.

In this approach, we use a moving block bootstrap for resampling to approximately maintain temporal dependencies that distinguish between signal present and absent. The principle of the moving block bootstrap is to draw k M-tuples from $\hat{F}$, which is an M-dimensional distribution formed by assigning probability mass $1/(N-M+1)$ to the overlapping blocks

$$\begin{aligned} BK_1 =& \{\mathbf{x}[1]\ \mathbf{x}[2]\ \cdots\ \mathbf{x}[M]\},\ BK_2 = \{\mathbf{x}[2]\ \mathbf{x}[3]\ \cdots\ \mathbf{x}[M+1]\}, \\ & \cdots,\ BK_{N-M+1} = \{\mathbf{x}[N-M+1]\ \mathbf{x}[N-M+2]\ \cdots\ \mathbf{x}[N]\}, \end{aligned}$$

and create the bootstrap data $\mathbf{x}^b_N = \left(\mathbf{x}^b[1]\ \mathbf{x}^b[2]\ \cdots\ \mathbf{x}^b[N]\right)$ by concatenating the k M-tuples, where $M < N$ and $k = \lceil N/M \rceil$. We refer to [76,77] for moving block bootstrap in details. The principle of the pivot bootstrap test using the moving block bootstrap is shown in Algorithm 3.1. We use a block size M in Algorithm 3.1 which is equal to the smoothing factor M in (3.8).

Algorithm 3.1 The pivot bootstrap test using the moving block bootstrap

Step 1) Draw N samples of the received signal $\mathbf{x}_N = (\mathbf{x}[1]\ \mathbf{x}[2]\ \ldots\ \mathbf{x}[N])$

Step 2) Apply the moving block bootstrap to the data $\mathbf{x}_N$,

$$\hat{F} \rightarrow \mathbf{x}_N^b = \left(\mathbf{x}^b[1]\ \mathbf{x}^b[2]\ \ldots\ \mathbf{x}^b[N]\right)$$

Step 3) Calculate T_{piv}^b using (3.12):

$\rightarrow \hat{T}^b$ is calculated from (3.9) by using the bootstrap data $\mathbf{x}_N^b$

$\rightarrow$ The moving block bootstrap is again applied to $\mathbf{x}_N^b$ to obtain $\mathbf{x}_N^{*b_s}$ and repeated B_{s1} times

$\rightarrow$ For each b_s, $b_s = 1, \ldots, B_{s1}$, calculate $\hat{T}^{b_s}$. $\hat{S}_e^b$ can be obtained by [63]

$$\hat{S}_e^b = \sqrt{\frac{\sum_{b_s=1}^{B_{s1}} \left[\hat{T}^{b_s} - \hat{T}^{(\cdot)}\right]^2}{B_{s1} - 1}}, \quad \hat{T}^{(\cdot)} = \sum_{b_s=1}^{B_{s1}} \frac{\hat{T}^{b_s}}{B_{s1}}.$$

Repeat **Step 2** and **Step 3** for $b = 1, 2, \ldots, B_s$.

Step 4) Calculate T_{piv} using (3.11) and $\hat{T}$ from (3.9), and then determine the p-value

$$\hat{p}_{piv}^* = \frac{1 + \#\{T_{piv}^b \geq T_{piv}\}}{B_s + 1}.$$

Step 5) If $\hat{p}_{piv}^* < \alpha$, accept $\mathcal{H}_1$, otherwise accept $\mathcal{H}_0$.

3.2.2 The null-resampling bootstrap test

The fact that we need to calculate the estimate of the standard error $\hat{S}_e$ in Algorithm 3.1, increases the computational cost of the pivot bootstrap test by B_{s1}. In addition, we show later in Subsection 3.2.3 that the pivoting test applied to the model in Section 3.1 does not perform well. This motivates the use of another approach, namely null-resampling bootstrap test. The test has a smaller computational cost and has a better performance than the pivoting test in the scenario at hand.

The construction of the bootstrap hypothesis test can be different to the construction of the confidence interval due to the requirement of generating the bootstrap distribution of test statistics under a specific null hypothesis. In Algorithm 3.1, we perform resampling from the EDF $\hat{F}$ according to the moving block bootstrap, since it is suggested to be used in the confidence interval estimation for dependent data [61]. Now, however, we must resample data from a distribution, say $\hat{F}_0$, which constitutes the relevant null hypothesis $\mathcal{H}_0$, namely null resampling distribution [62]. According to (3.7), the relevant

$\hat{F}_0$ here is the distribution that yields (approximately) $T = 1$. This can be obtained by resampling the observed data $\mathbf{x}_N$ independently since it then removes temporal dependencies among samples. Therefore, $\hat{F}_0$ is formed by assigning equal probabilities $1/N$ to each sample value $\mathbf{x}[n],\ n = 1, \ldots, N$, and independently resampling the data with replacements to produce bootstrap data $\mathbf{x}_N^b = \left(\mathbf{x}^b[1]\ \mathbf{x}^b[2]\ \ldots\ \mathbf{x}^b[N]\right),\ b = 1, \ldots, B_s$. In this way, we only need to replicate $\hat{T}$ by using the bootstrap data to estimate $G_0(T)$ in (2.13), and there is no need to estimate $\hat{S}_e$. Thus, it is simpler and it has lower computational cost than the pivot bootstrap test. The implementation of the null-resampling bootstrap test is shown in Algorithm 3.2.

Algorithm 3.2 The null-resampling bootstrap test

Step 1) Draw N samples of the received signal $\mathbf{x}_N = (\mathbf{x}[1]\ \mathbf{x}[2]\ \ldots\ \mathbf{x}[N])$ and calculate $\hat{T}$ using (3.9)

Step 2) Apply the null-resampling to the observed data

$$\hat{F}_0 \rightarrow \mathbf{x}_N^b = \left(\mathbf{x}^b[1]\ \mathbf{x}^b[2]\ \ldots\ \mathbf{x}^b[N]\right)$$

Step 3) Calculate $\hat{T}^b$ in (3.9) using the bootstrap data $\mathbf{x}_N^b$

Repeat **Step 2** and **Step 3** for $b = 1, 2, \ldots, B_s$.

Step 4) Calculate the p-value

$$\hat{p}_{null}^* = \frac{1 + \#\{\hat{T}^b \geq \hat{T}\}}{B_s + 1}.$$

Step 5) If $\hat{p}_{null}^* < \alpha$, accept $\mathcal{H}_1$, otherwise accept $\mathcal{H}_0$.

3.2.3 Approximate distributions of the test statistics

In this section, we analyze the resulting approximated distributions in the asymptotic test, the pivot bootstrap test and the null-resampling bootstrap test. The aim is to show how close the approximate distribution of the respective test is to the true distribution under $\mathcal{H}_0$, which is then used to explain the decision errors that might be committed by the test statistic. We specifically point out the reason why the asymptotic test and the pivot bootstrap test fail to perform well. To corroborate the analysis in this section, experimental results will be shown in Section 3.5.

First, let us rewrite the binary hypotheses testing (3.7) into an equivalent form

$$\mathcal{H}_0 : T_i = \frac{1}{T} = 1 \tag{3.13a}$$

$$\mathcal{H}_1 : 0 < T_i = \frac{1}{T} < 1, \tag{3.13b}$$

where T is from (3.6). Suppose that we have an observed value of the test statistic $\tilde{t}$, the p-value in this case is

$$p = P_0(T_i \leq \tilde{t}) = G_0(\tilde{t}). \tag{3.14}$$

Note that the inequality is in reverse compared to (2.13), due to (3.13).

According to [22], the estimator of T_i under $\mathcal{H}_0$, denoted as $\hat{T}_i$, is approximately Gaussian distributed when N is finite and large, i.e.

$$\hat{T}_i \dot{\sim} \mathcal{N}\left(\mu_{\hat{T}_i}, \sigma^2_{\hat{T}_i}\right) \tag{3.15}$$

where the mean and the variance are

$$\mu_{\hat{T}_i} = \frac{1}{1 + (M-1)\sqrt{\frac{2}{\pi N}}}, \quad \sigma^2_{\hat{T}_i} = \left(\frac{\sqrt{\frac{2}{N}}}{1 + (M-1)\sqrt{\frac{2}{\pi N}}}\right)^2. \tag{3.16}$$

Therefore, the two parameters only depend on smoothing factor M and sample size N.

The pivot bootstrap test.
The equivalent test statistic of (3.11) is

$$T_{i,piv} = \frac{\hat{T}_i - T_0}{\hat{S}_{i,e}} = \frac{\hat{T}_i - 1}{\hat{S}_{i,e}}. \tag{3.17}$$

and the bootstrap replication of $T_{i,piv}$ is

$$T^b_{i,piv} = \frac{\hat{T}_i^{\,b} - \hat{T}_i}{\hat{S}^b_{i,e}}. \tag{3.18}$$

Note that $\hat{S}_{i,e} = \sigma_{\hat{T}_i}$. The distributions of $T_{i,piv}$ and $T^b_{i,piv}$ under $\mathcal{H}_0$ are

$$T_{i,piv} \dot{\sim} \mathcal{N}\left(\mu_{T_{i,piv}}, 1\right), \quad \mu_{T_{i,piv}} = -\frac{1 - \mu_{\hat{T}_i}}{\sigma_{\hat{T}_i}} = -\frac{M-1}{\sqrt{\pi}}, \tag{3.19}$$

and

$$T^b_{i,piv} \dot{\sim} \mathcal{N}(0, 1). \tag{3.20}$$

Accordingly, we conclude that, first, the variances are stabilized, and so that the variance stabilization such as in [78] is not an issue.[2] Second, the distribution of $T_{i,piv}$ is the same as the distribution of $T^b_{i,piv}$, but shifted $\frac{M-1}{\sqrt{\pi}}$ to the left. If we repeat the test with size α many times, the test tends to have small values of $T_{i,piv}$, and hence has a significant number of the resulting p-values that are smaller than α. In this respect, the rejection of $\mathcal{H}_0$ is very high. This also means that the distribution of the p-values is not uniform, but they are mostly concentrated on small values. Using (3.19) and (3.20), the relation between the actual value of the probability of false alarm P_f and the nominal value α could be approximated by

$$P_f \approx 1 - \left[\mathcal{Q}\left(\mathcal{Q}^{-1}(1-\alpha) + \frac{M-1}{\sqrt{\pi}}\right)\right]. \tag{3.21}$$

For example, when $M = 5$, we will have the curve as depicted in Fig. 3.1. It shows that P_f is considerably larger than α. It is noteworthy that even the relation between P_f and α seems to depend only on M according to (3.21), the sample size N still affects the relation. This can be confirmed by (3.16). If we were to have $P_f \approx \alpha$ using this method, then it needs a very large N. When N is very large, $N \to \infty$, then the mean value $\mu_{\hat{T}_i} \to 1$ and $T_{i,piv} \dot{\sim} \mathcal{N}(0,1)$.[3] Unfortunately, this would take $N \gg 100$, since P_f is still much larger than α when $N = 100$, as shown by simulation in the sequel. In this situation, the bootstrap approach is not an appropriate option. However, this does not mean that the bootstrap approach performs poorly for very large N. But, if we consider the trade-off between performance and computational cost, the asymptotic test might offer advantages over the bootstrap approach.

The asymptotic test.

For the asymptotic test, the true distribution of $\hat{T}_i$ under $\mathcal{H}_0$, which is unknown, is approximated by the Gaussian distribution (3.15). Then, the threshold can be determined for the preset value of the probability of false alarm α.[4] The performance of the asymptotic test can be explained by comparing its density function with the true density function of the test statistic $\hat{T}_i$, as shown in Fig. 3.2. Since we do not know the true density function, it is generated by Monte Carlo runs. The figure shows that the true values of the test statistic have smaller spreads than the assumed asymptotic Gaussian distribution, particularly for small N. In terms of the p-value, if the test statistic with size α is evaluated many times under $\mathcal{H}_0$, when α is preset to

[2] According to [22], when N is large, the distribution of $\hat{T}_i$ under $\mathcal{H}_1$ is also approximately Gaussian with a certain mean and variance, depending on the SNR and the degree of correlation. However, the standardized statistics $T_{i,piv}$ and $T^b_{i,piv}$, both approximately have variances equal to 1.

[3] In this case, even though the variance $\hat{T}_i$ is close to 0, the variance of $T_{i,piv}$ is still close to 1.

[4] Suppose that the threshold for the preset α, found using the distribution (3.15), is denoted as τ_i. The threshold in (3.10) is basically $\tau = 1/\tau_i$. Thus, the performance of the asymptotic test using the statistic (3.13) is equal to the performance of the asymptotic test using the statistic (3.6).

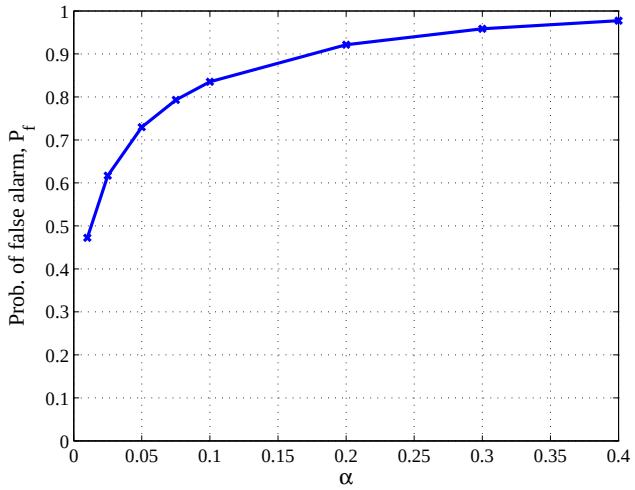


Figure 3.1: The curve between the actual value of the probability of false alarm P_f and the nominal value α, using Equation (3.21) with $M = 5$.

be sufficiently small, it would have a significant number of resulting p-values that are larger than α. In this case, the acceptance rate of $\mathcal{H}_0$ under $\mathcal{H}_0$ would be considerably high. For a reasonable value of α that is used in spectrum sensing, i.e., $0.01 \leq \alpha \leq 0.2$, and for $N \leq 100$, we have confirmed that P_f is always significantly smaller than α. Even for a large sample size $N = 1000$ as shown in the figure, the lower tail of the asymptotic distribution still differs significantly from the true distribution. This might be advantageous at some point since we have smaller values than the target values, but, it would lead to a significantly reduced power of test (large probability of miss detection), which is inadmissible in spectrum sensing. Note that when α is preset to be sufficiently large, the result would be different and we might have a significant number of p-values that are smaller than α.[5] However this case is not relevant in practice.

The null-resampling bootstrap test.

It is always a good idea to look at the bootstrap data graphically in order to see how close the bootstrap statistics are to the true statistics, by which we can estimate how good the performance of the bootstrap approach would be. In Fig. 3.2, it shows that the bootstrap approximated density functions are close to the true density functions of the test statistic $\hat{T}_i$, in particular for $N = 50$ and 100. In this case, we will have the distribution of the p-value to be approximately uniform, and so the actual probability of false alarm will be close to the nominal value, $P_f \approx \alpha$. For $N = 20$, the lower

[5]It is noteworthy that this dependency on α does not happen in the bootstrap pivot test. This is because the true distribution (the distribution of $T_{i,piv}$) under $\mathcal{H}_0$, differs in mean from the bootstrap approximate distribution (the distribution of $T^b_{i,piv}$). Meanwhile, in the asymptotic test with a small N, the difference is dominated by their variances.

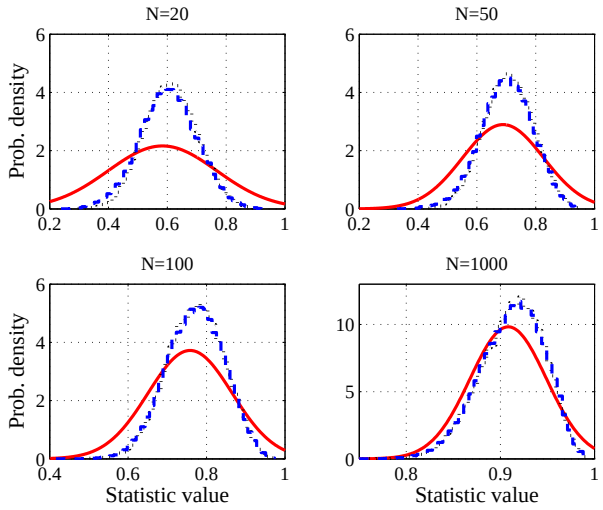


Figure 3.2: Density functions of the test statistic under $\mathcal{H}_0$, i.e., the true density (dotted line), the approximated density functions using the null-resampling bootstrap (dashed line) and the asymptotic normal density from (3.15) (solid line). The true densities were generated by 5×10^4 Monte Carlo runs of $\hat{T}_i$, and the bootstrap approximated densities were generated by 5×10^4 bootstrap replications $\hat{T}_i^b$. The smoothing factor was set to $M = 5$, with different values of sample size N.

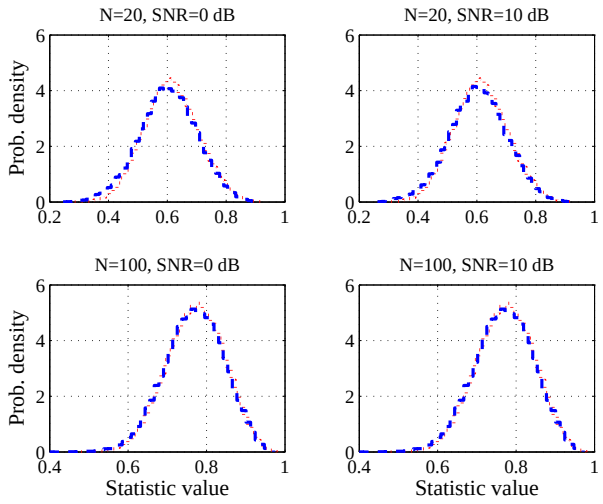


Figure 3.3: The true density functions of the test statistic under $\mathcal{H}_0$ (dotted line) and the approximated density functions using the null-resampling bootstrap (dashed line). The bootstrap approximated densities were generated when the observations are under $\mathcal{H}_1$ with the SNRs $= 0$ dB and 10 dB. The true densities were generated by 5×10^4 Monte Carlo runs of $\hat{T}_i$, and the bootstrap approximated densities were generated by 5×10^4 bootstrap replications $\hat{T}_i^b$. The smoothing factor was set to $M = 5$, with sample sizes $N = 20$ and 100.

tail of the bootstrap density is noticeably different from the true density. This makes $P_f < \alpha$, but the difference is still acceptable and not as large as in the asymptotic test. To corroborate the result, in Fig. 3.3 we compare the true density functions under $\mathcal{H}_0$ with the approximated density functions using the null-resampling bootstrap when the observations are under $\mathcal{H}_1$, with SNR = 0 dB and SNR = 10 dB. Here, we assume that the correlated observations under $\mathcal{H}_1$ are due to the channel time dispersion (see Section 3.5 for details). In general, the figure shows that the bootstrap approximated density functions are in close agreement with the true density functions, irrespective to the SNR of the observations. This can be explained as follows. Suppose that a CR user receives a correlated signal from a primary user with power σ_r^2 and that noise power at the receiver is σ_w^2. When independent resampling is applied to the combined signal, since we apply the null-resampling bootstrap test, the bootstrap data can be seen as data from uncorrelated noise with power $\sigma_r^2 + \sigma_w^2$. For this case, the covariance matrix of the bootstrap data will be approximately a diagonal matrix. Then, referring to (3.6), the resulting bootstrap replications of the test statistic will be $\hat{T}_i^b \approx 1$, regardless of the signal power. Therefore, if we replicate the bootstrap data B_s times, the empirical distribution of $\hat{T}^b$ can be used to approximate the true distribution of $\hat{T}_i$ under $\mathcal{H}_0$, even if the observations come from $\mathcal{H}_1$ with various powers.

In this part, we mainly focus on the probability of false alarm. However, the closeness of P_f to α also affects the probability of miss detection P_m. When we have a test that has $P_f \ll \alpha$, the power of the test will be small (P_m will be large), and vice-versa. Hence, since we have a small sample size N, the aim is to make P_f as close as possible to α, which then results in acceptable P_m for the specified N. This only can be accomplished by the null-resampling bootstrap test, as we explained above.

3.3 Resampling based collaborative spectrum sensing

It is well known that collaborative spectrum sensing provides a solution for the hidden terminal problem, because it is unlikely that all channels between the primary transmitter and the CR users would undergo deep faded simultaneously when the distances between CR users are sufficiently large. In addition, collaborative spectrum sensing improves the detection performance and reduces the detection time [10]. Here, we consider the collaborative spectrum sensing with resampling for small sample sizes. From this point on, we will use the abbreviation CSS to represent collaborative spectrum sensing. Note that we only consider the bootstrap null-resampling test as the local detector in sequel.

Consider the case where L CR users observe a spectrum band, each with an observation length N. The signal model for each user can then be written as

$$\mathcal{H}_0 : \mathbf{x}_l[n] = \mathbf{w}_l[n], \tag{3.22a}$$

$$\mathcal{H}_1 : \mathbf{x}_l[n] = \sum_{j=1}^{J-1} \mathbf{h}_l[j]\mathbf{s}[n-j] + \mathbf{w}_l[n], \tag{3.22b}$$

where the channel coefficients $\mathbf{h}_l[j],\ l = 1, \ldots, L,\ j = 1, \ldots, J$, are independent. Each CR user implements the null-resampling bootstrap test as the local detector, which has been described in Subsection 3.2.2. The implementation of an adhoc fusion rule such as OR, AND, and MAJORITY rule [32], is straightforward and there is no difference to non-resampling based detections. Each CR user decides in favor of $\mathcal{H}_0$ or $\mathcal{H}_1$ and forwards the decision bit to the fusion center (FC) where the respective fusion rule is executed.

Now, suppose that the Chair-Varshney fusion rule is used at the FC. The test statistic at the FC reads [32]

$$T_{fc} = \sum_{l=1}^{L} \left[u_l \ln \frac{P_d^l}{P_f^l} + (1 - u_l) \ln \frac{1 - P_d^l}{1 - P_f^l} \right], \tag{3.23}$$

where $u_l \in \{0, 1\}$ is the decision bit of the user l. Here, $u_l = 0$ and $u_l = 1$ denote the CR user to favor $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. P_d^l and P_f^l represent its probability of detection and the probability of false alarm, respectively. Therefore, the local performance measures, P_d^l and P_f^l, of each user should be known at the FC. Note that to determine the threshold in the local detector using the Neyman-Pearson approach, the probability of false alarm is set to a nominal value α. This prior information, i.e., $P_f^l = \alpha$, can be used at the FC.[6] However, $P_d^l,\ l = 1, \ldots, L$ are not easily obtained since each of which depends on the received signal at the CR user. More precisely, it depends on the channel condition which includes the distance of the CR user to the primary transmitter. In this case, $G_1^l(T)$, the distribution of the local test statistic for the CR user l under $\mathcal{H}_1$, varies. Accordingly, estimating $G_1^l(T)$ as a way to estimate P_d^l is necessary to be conducted at the CR user. Here, we provide such method to be able to implement the Chair-Varshney fusion rule in a situation where the channels between the primary transmitter and the CR users change.

[6]The requirement to have $P_f^l = \alpha$ could be closely approximated by the null-resampling bootstrap test, as we have explained in Subsection 3.2.3.

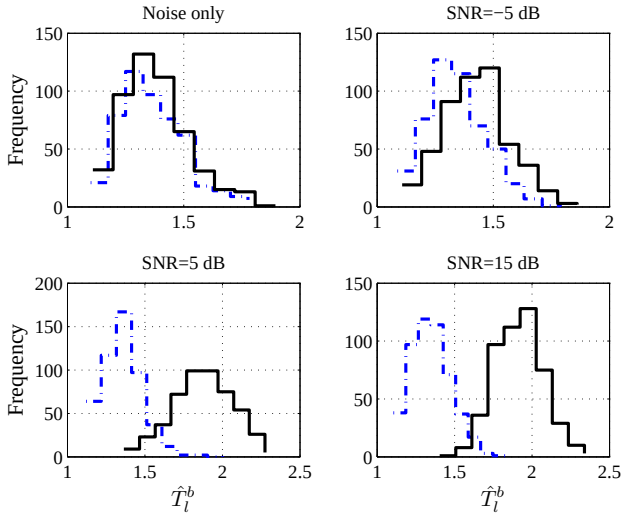


Figure 3.4: Histograms of 500 bootstrap replications of the test statistics $\hat{T}_l^b$ when the true hypothesis is $\mathcal{H}_0$ (noise only) and $\mathcal{H}_1$ (for different values of SNR). Dashed-dotted lines represent the estimates for the distribution under $\mathcal{H}_0$ and solid lines represent the estimates for the distribution under $\mathcal{H}_1$. The simulation setup to generate the figure is explained in Section 3.5.

3.3.1 Procedures at a CR user

Consider that the CR user l observes the data $\mathbf{x}_{l,N} = (\mathbf{x}_l[1]\ \mathbf{x}_l[2]\ \ldots\ \mathbf{x}_l[N])$ and hence the observed test statistic $\hat{T}_l$ is obtained. The resampling-based CSS is derived based on the following observations. Suppose that $\hat{T}_l$ is small. This can happen either under F_0^l, the distribution of the data under $\mathcal{H}_0$ for the CR user l, or F_1^l, the distribution under $\mathcal{H}_1$. As explained in Subsection 3.2.2, the null-resampling can be implemented to estimate $G_0^l(T_l)$, the distribution of the test statistic under $\mathcal{H}_0$ for the CR user l, no matter from which distribution the data comes from. Meanwhile, assuming the data originates from F_1^l, the moving block bootstrap can be implemented to estimate $G_1^l(T_l)$ since it maintains temporal dependencies of the data. The fact that the data is originally from F_0^l would not necessarily matter since this case is similar to the data originating from F_1^l with a very small SNR. Therefore, the EDF $\hat{G}_0^l(\hat{T}_l)$ produced by the null-resampling and the EDF $\hat{G}_1^l(\hat{T}_l)$ produced by the moving block bootstrap would be sufficiently separated. This would be more pronounced when the data is originally from F_1^l. The larger the SNR, the larger the separation between the two EDFs, as illustrated by the histograms in Fig. 3.4.

Suppose that the two EDFs $\hat{G}_0^l(\hat{T}_l)$ and $\hat{G}_1^l(\hat{T}_l)$ have been obtained and the probability of false alarm for the CR user l is selected to be $P_f^l = \alpha^l,\ l = 1, 2, \ldots, L$. The CR user l produces a local decision bit u_l based on the p-value as shown in Algorithm 3.2.

Now, to estimate the probability of detection of the local detector, first, the threshold is calculated using

$$\tau_l = \hat{G}_0^{(1-\alpha)l}(\hat{T}_l), \tag{3.24}$$

which is the $1-\alpha$ quantile of $\hat{G}_0^l(\hat{T}_l)$. Practically, it is the (B_s+1-k)th largest value of $\hat{T}_l^{0b},\ b=1,\ldots,B_s$, where $k=\lfloor(B_s+1)\alpha\rfloor$ and $\hat{T}_l^{0b}$ denotes the replication of $\hat{T}$ using the null-resampling at the CR user l. Recall that B_s is the number of bootstrap replications. Then, the probability of detection P_d^l is estimated by

$$\hat{P}_d^l = \frac{1+\#\{\hat{T}_{l1}^{1b} \geq \tau_l\}}{B_s+1}, \tag{3.25}$$

where $\hat{T}_l^{1b},\ b=1,\ldots,B_s$, are the replications of $\hat{T}$ using the moving block bootstrap at the CR user l. Afterwards, the decision bit u_l and the estimate of the probability of detection $\hat{P}_d^l$ are transmitted to the FC. Note that the FC can use the prior information $P_f^l=\alpha^l$ and no need to be transmitted when it is assumed to be constant. It is possible to use a random threshold at the CR user, e.g., by randomly selecting α^l from a pre-specified set S_α. However, this will increase the bandwidth requirement of the reporting channel since the selected α^l should also be transmitted to the FC, along with u_l and $\hat{P}_d^l$. The algorithm of the local detector for the resampling-based CSS is shown in Algorithm 3.3.

3.3.2 Fusion rule at the FC

Let us assume that the FC already knows u_l, α_l, and $\hat{P}_d^l,\ l=1,\ldots,L$, of all CR users that joint the CSS. Let $\mathbf{u}=(u_1\ u_2\ \ldots\ u_L)$ represent a vector of decision bits from L CR users. Here, we assume that the decision bits are independent. The test statistic (3.23) at the FC should now read

$$T_{fc}(\mathbf{u}) = \sum_{l=1}^{L}\left[u_l \ln\frac{\hat{P}_d^l}{\alpha^l} + (1-u_l)\ln\frac{1-\hat{P}_d^l}{1-\alpha^l}\right]. \tag{3.26}$$

Note that we need to have $G_0(T_{fc}(\mathbf{u}))$, the distribution of $T_{fc}(\mathbf{u})$ under $\mathcal{H}_0$, to make a global decision at the FC. However, since now $\hat{P}_d^l,\ l=1,\ldots,L$, are assumed to change from one measurement to another, we have to generate the distribution function for each measurement. In this case, either all possible binary combinations can be considered or the parametric bootstrap can be used.

For L CR users in the CSS, there are 2^L possible values of $\mathbf{u}$, i.e., $\mathbf{u}^k = \left(u_1^k\ u_2^k\ \ldots\ u_L^k\right),\ k=1,2,\ldots,2^L$. Consequently, based on (3.26), $T_{fc}(\mathbf{u})$ could also

Algorithm 3.3 The decision and the estimation of the probability of detection at the CR user

Step 1) Draw N samples of the received signal $\mathbf{x}_{l,N} = (\mathbf{x}_l[1]\ \mathbf{x}_l[2]\ \ldots\ \mathbf{x}_l[N])$ and calculate $\hat{T}$ using (3.9)

Step 2) Apply the null-resampling to the observed data

$$\hat{F}_0 \rightarrow \mathbf{x}_{l,N}^{0b} = \left(\mathbf{x}_l^{0b}[1]\ \mathbf{x}_l^{0b}[2]\ \ldots\ \mathbf{x}_l^{0b}[N]\right)$$

Step 3) Apply the moving block bootstrap to the observed data

$$\hat{F}_1 \rightarrow \mathbf{x}_{l,N}^{1b} = \left(\mathbf{x}_l^{1b}[1]\ \mathbf{x}_l^{1b}[2]\ \ldots\ \mathbf{x}_l^{1b}[N]\right)$$

Step 4) Calculate $\hat{T}_l^{0b}$ using the bootstrap data $\mathbf{x}_N^{0b}$ and $\hat{T}_l^{1b}$ using the bootstrap data $\mathbf{x}_N^{1b}$

Repeat **Step 2** to **Step 4** for $b = 1, 2, \ldots, B_s$.

Step 5) Determine the decision bit by calculating the p-value

$$\hat{p}^* = \frac{1 + \#\{\hat{T}_l^{0b} \geq \hat{T}\}}{B_s + 1},$$

and if $\hat{p}^* < \alpha$, accept $\mathcal{H}_1$ ($u_l = 1$), otherwise accept $\mathcal{H}_0$ ($u_l = 0$).

Step 6) Estimate the local probability of detection by first sorting $\hat{T}_l^{0b},\ b = 1, \ldots, B_s$

$$\hat{T}_{l,(1)}^{0b} \leq \hat{T}_{l,(2)}^{0b} \leq \ldots \leq \hat{T}_{l,(B_s)}^{0b}$$

and the threshold $\tau_l = \hat{T}_{l,(B_s+1-k)}^{0b}$, where $k = \lfloor (B_s+1)\alpha \rfloor$. The estimate of the local probability of detection is then found from (3.25).

Step 7) Transmit u_l and $\hat{P}_d^l$ to the FC.

have a value drawn from a finite set $S_{T_{fc}} = \{T_{fc}(\mathbf{u}^1)\ T_{fc}(\mathbf{u}^2)\ \ldots\ T_{fc}(\mathbf{u}^{2^L})\}$, each of which has a probability of appearance under $\mathcal{H}_0$ equal to

$$\begin{aligned} P_0(T_{fc}(\mathbf{u}^k)) &= \prod_{l=1}^{L} \left[u_l^k P_0(u_l^k = 1) + (1 - u_l^k) P_0(u_l^k = 0)\right] \\ &= \prod_{l=1}^{L} \left[u_l^k \hat{P}_f^l + (1 - u_l^k)(1 - \hat{P}_f^l)\right], \quad u_l^k \in \{0, 1\}. \end{aligned} \tag{3.27}$$

Thus, the distribution function of $T_{fc}(\mathbf{u})$ under $\mathcal{H}_0$ can be written as

$$G_0(T_{fc}(\mathbf{u})) = \sum_{k=1}^{2^L} g_0(T_{fc}(\mathbf{u}^k)) \mathbf{1}_{\{T_{fc}(\mathbf{u}^k) \leq T_{fc}(\mathbf{u})\}}, \tag{3.28}$$

where $\mathbf{1}_{\{.\}}$ is the indicator function and $g_0(\cdot)$ denotes the probability mass function of

Algorithm 3.4 The global decision at the fusion center

Step 1) Receive the decision bit u_l and the local probability of detection $\hat{P}_d^l$ from L CR users.

Step 2) Calculate the observed test statistic $\hat{T}_{fc}(\mathbf{u})$ using (3.26).

If $2^L \leq B_s$, continue with **Step 3a** to **Step 5a**, otherwise, continue with **Step 3b** to **Step 5b**.

Step 3a) Calculate $T_{k,FC}(\mathbf{u}^k)$ and the corresponding probability $P_0(T_{k,FC}(\mathbf{u}^k))$ from (3.26) and (3.27), for all possible values of $\mathbf{u}^k,\ k = 1, \ldots, 2^L$.

Step 4a) Obtain the distribution function $G_0(T_{fc}(\mathbf{u}))$ in (3.28).

Step 5a) Calculate the p-value $\hat{p}$ using (3.30), and if $\hat{p} < \alpha$, accept $\mathcal{H}_1$ as a global decision, otherwise accept $\mathcal{H}_0$.

Step 3b) Generate parametrically the bootstrap data using the parameters in 3.31 for the vector of decision bit

$$\hat{F}_{0,\{\alpha^l, l=1,\ldots,L\}} \rightarrow \mathbf{u}^b = \left(u_1^b\ u_2^b\ \cdots\ u_L^b\right),$$

Step 4b) Calculate $T_{fc}(\mathbf{u}^b)$ from (3.26).

Repeat **Step 3b** and **Step 4b** for $b = 1, 2, \cdots, B_s$.

Step 5b) Calculate the p-value $\hat{p}^*$ using (3.32), and if $\hat{p}^* < \alpha$, accept $\mathcal{H}_1$ as a global decision, otherwise accept $\mathcal{H}_0$.

$T_{fc}(\mathbf{u})$ under $\mathcal{H}_0$, i.e.,

$$g_0(T_{fc}(\mathbf{u})) = \sum_{k=1}^{2^L} P_0(T_{fc}(\mathbf{u}^k))\delta(T_{fc}(\mathbf{u}) - T_{fc}(\mathbf{u}^k)), \tag{3.29}$$

here $\delta(\cdot)$ is the Kronecker delta function. For the observed test statistic $\hat{T}_{fc}(\mathbf{u})$, the p-value can then be calculated by

$$\hat{p} = 1 - G_0(\hat{T}_{fc}(\mathbf{u})) = \sum_{k=1}^{2^L} P_0(T_{k,FC})\mathbf{1}_{\{T_{fc}(\mathbf{u}^k) > \hat{T}_{fc}(\mathbf{u})\}}. \tag{3.30}$$

If $\hat{p} < \alpha$, then accept $\mathcal{H}_1$ as a global decision, otherwise accept $\mathcal{H}_0$. Here, α denotes a nominal value of the global probability of false alarm.

The complexity of the algorithm to yield a global decision using all possible binary combinations depends on the number of CR users L. For example, when $L = 10$, it

requires $2^{10} = 1024$ combinations and exponentially increases as L increases. In this case, to reduce the complexity when L is large, we can use the parametric bootstrap to generate $\hat{G}_0(T_{fc}(\mathbf{u}^b))$, the EDF under $\mathcal{H}_0$, instead of using all possible combinations. This is possible because the decision bit under $\mathcal{H}_0$ from each CR user can be simulated as Bernoulli trials with probabilities

$$P_0(u_l = 0) = 1 - \alpha^l, \quad P_0(u_l = 1) = \alpha^l, \quad l = 1, \ldots, L. \tag{3.31}$$

Therefore, the bootstrap data for the vector of decision bits $\mathbf{u}^b = \left(u_1^b \; u_2^b \; \ldots \; u_L^b\right)$, $b = 1, \ldots, B_s$, can be generated parametrically. Accordingly, the bootstrap test statistic $T_{fc}(\mathbf{u}^b),\ b = 1, \ldots, B_s$, can be calculated from (3.26) and $\hat{G}_0(T_{fc}(\mathbf{u}^b))$ is obtained. For the observed test statistic $\hat{T}_{fc}(\mathbf{u})$, the p-value is therefore

$$\hat{p}^* = 1 - \hat{G}_0^(\hat{T}_{fc}(\mathbf{u})) = \frac{1 + \#\{T_{fc}(\mathbf{u}^b) \geq \hat{T}_{fc}(\mathbf{u})\}}{B_s + 1}, \tag{3.32}$$

and then the global decision is made. In brief, when the number of CR users L results in a number of combinations $2^L \leq B_s$, then use the all binary combinations to produce $G_0(T_{fc}(\mathbf{u}))$, otherwise switch to the parametric bootstrap to have $\hat{G}_0^(T_{fc}(\mathbf{u}^b))$. The implementation at the FC for the resampling-based CSS is shown in Algorithm 3.4.

3.4 Multiple hypothesis testing for multiband spectrum sensing

Multiple hypothesis testing refers to the testing of more than one hypothesis at a time [79]. It is intended to solve the multiplicity effect by making the individual tests more conservative to arrive at rejecting $\mathcal{H}_k$ hypothesis. The procedure to conduct multiple hypothesis tests is often called *multiple comparison procedure* (MCP) or *multiple test procedure* (MTP) [80]. We prefer to use MTP in sequel. We begin with the definition of some measures in MTP that will be repeatedly referred to, in this section and also in Section 5.1. Afterwards, we detail the implementation of MTPs in spectrum sensing.

3.4.1 Performance measures in MTP

Consider a situation where we have $K > 1$ pairs of null and alternative hypotheses to be tested jointly, i.e. $\mathcal{H}_{0,k}$ vs $\mathcal{H}_{1,k},\ k = 1, \ldots, K$. Let us assume that K_0 out of K are true null hypotheses. Numbers of correct and false decisions are represented in Table

Table 3.1: Number of correct and false decisions for testing K null hypotheses

	Declared $\mathcal{H}_0$	**Declared $\mathcal{H}_1$**	**Total**
True $\mathcal{H}_0$	U	V	K_0
True $\mathcal{H}_1$	T	S	$K - K_0$
	$K - R$	R	K

3.1. Note that U, V, S, T and R are random variables, where R is observable and the others are unobservable.

Suppose that now we are more interested in a measure of errors for the family instead of the individual tests. A common measure of errors in classical MTP is called *familywise error rate* (FWE). It is defined as the probability of committing any type I error or false alarm in families of comparisons, formally

$$\text{FWE} = P_0(V \geq 1). \tag{3.33}$$

There are two kinds of FWE. The FWEC, is calculated under the complete null hypotheses (all $H_{0,k}$ are true), and the FWEP, is calculated under a partial null hypotheses (some subsets of nulls, say $\mathcal{H}_{0,j1}, \ldots, \mathcal{H}_{0,jk}$, are true). An MTP is said to control the FWE in the weak sense if FWEC $\leq \alpha$, and it controls it in the strong sense if FWEP $\leq \alpha$ regardless of which subsets of null hypotheses is true [81]. A simple example for the MTP that is devised to control the FWE is the Bonferroni procedure, in which each hypothesis k, $k = 1, \ldots, K$, is tested at level α/K, and it guarantees to have FWE $\leq \alpha$. Another procedure is proposed by Holm [82], which is to improve the power of the Benferroni procedure. However, FWE-based procedures are still conservative and generally lead to a significant reduction in the power of test. To resolve this issue, Benjamini and Hochberg propose to use another measure [83], namely *false discovery rate* (FDR). It is defined as the expected value of the proportion of the rejected null hypotheses which are erroneously rejected, formally

$$\text{FDR} = E\left[\frac{V}{V+S}\right] = E\left[\frac{V}{R}\right]. \tag{3.34}$$

To evaluate the performance, the following additional measures are used. We define *familywise miss detection* (FWM). It refers to the probability of committing any type II error or miss detection in families of comparisons, formally

$$\text{FWM} = P_1(T \geq 1). \tag{3.35}$$

Following [84], the *false alarm ratio* (FAR) is defined as the expected value of the ratio of the number of false alarms and the number of true null hypotheses, i.e.,

$$\text{FAR} = E\left[\frac{V}{K_0}\right]. \tag{3.36}$$

and the *missing ratio* (MR) is defined as the expected value of the ratio of the number of miss detections and the number of true alternative hypotheses, i.e.,

$$\mathrm{MR} = E\left[\frac{T}{K - K_0}\right]. \tag{3.37}$$

The implementation of MTP to source enumeration can be found in [85] and to the identification of optimal sensor position in [86]. It also has been implemented in distributed detections and wireless sensor networks, such as in [87–91]. In the following subsection, we elaborate on the implementation of MTP in spectrum sensing.

3.4.2 The implementation of MTP in multiband spectrum sensing

We assume that the primary network operates over a wide frequency bandwidth which is divided into K nonoverlapping subbands, such as in multicarrier-based system. Whenever possible, a primary user can be assigned to use K_p subbands simultaneously, where $1 \leq K_p \leq K$. Now, the binary hypothesis testing problem for the subband k, is

$$\mathcal{H}_{k,0} : \mathbf{x}_k[n] = \mathbf{w}_k[n], \tag{3.38a}$$

$$\mathcal{H}_{k,1} : \mathbf{x}_k[n] = \mathbf{r}_k[n] + \mathbf{w}_k[n], \quad k = 1, \ldots, K, \ n = 1, \ldots, N. \tag{3.38b}$$

Suppose that within a particular time interval K_0 subbands might not be used by the primary users are available for cognitive access. Let us assume that the CR network supports some CR users to use several unoccupied subbands simultaneously. The number of subbands K_c assigned to a specific CR user is, say, based on priority. Here, $1 \leq K_c \leq K_0$ and $0 \leq K_0 \leq K$. Therefore, we need an overall view of the performance, e.g. false alarm and miss detection for each possible value of K_c. In particular, it is convenient if we have control over the type I error (false alarm) for all possible values of K_c to increase the throughput, while we can minimize the type II error (miss detection) for the subbands that are occupied, to suppress the interferences to the primary users. In this case, MTP should be implemented in multiband spectrum sensing to provide control over the decision errors at the system level, not only per subband.

3.4.2.1 Controlling FDR, an appropriate choice for multiband spectrum sensing

Testing each subband individually at level $\alpha_k = \alpha, \ k = 1, \ldots, K$, will produce multiplicity effect. Assuming independent test statistics and K_0 unoccupied subbands, the

resulting probability of false alarm due to the multiplicity can be written as

$$\text{FWE} = P_0(V > 1) = 1 - (1 - \alpha)^{K_0}. \tag{3.39}$$

Therefore, the larger K_0, the larger the FWE. A simple way to have control over the false alarm for each value of K_0 is to control the FWE in the strong sense. However, it then reduces the power of test significantly. More precisely, it increases the degree of interferences to the primary users, as a consequence of restrictively controlling the false alarm. In particular, primary users that use $K_p > 1$, receive larger aggregate interferences for the larger K_p.[7] In these situations, full protection resulting from controlling FWE is too restrictive. On the other hand, multiplicity control by testing individually is too permissive. Thus, handling appropriately the tradeoff between the two is pivotal.

Compared to the FWE controlling procedure, the FDR has the two following properties [83]. First, if all null hypotheses are true, $K_0 = K$, the FDR is equivalent to the FWE. Therefore, controlling the FDR in this case also means controlling the FWE in the weak sense. Second, if $K_0 < K$, the FDR is smaller than or equal to the FWE. Thus, any procedures devised to control the FWE also controls the FDR, but not in reverse. These indicate that the FDR controlling procedure is more advantageous than the FWE controlling procedure to be implemented in spectrum sensing, since the former has a better tradeoff than the latter. More precisely, the FDR controlling procedure provides relaxed restriction on the FWE and thus gains more power than the FWE controlling procedure.

The FDR controlling procedure has been used in [84] and [93] for spectrum sensing to jointly detect the K subbands and evaluate the performance in the system level. However, we want to emphasize here that the use of FAR and MR as the system measures could be misleading. This is because when we test each hypothesis k individually at nominal value $\alpha_k = \alpha,\ k = 1, \ldots, K$, then it is automatically guaranteed that FAR $\leq \alpha$ [94] (see Fig. 3.11 for the simulation results).[8] In addition, for a target FAR $\leq \alpha$, jointly testing with MTP results in a higher MR than that of testing individually.

3.4.2.2 The procedures

We propose to use the adaptive Benjamini-Hochberg procedure [94], hereafter A-BHP, for multiband spectrum sensing. It is based on the FDR which is the extension of the

[7] Aggregate interference is also used in [92]. It is one of the constraints for the opportunistic rate optimization problem in multiband spectrum sensing.

[8] FAR in [94] is referred to as the effective per comparison error rate.

Algorithm 3.5 The Benjamini-Hochberg procedure

Step 1) Calculate the p-values of $p_1, p_2, \ldots, p_K$ for the subbands $k = 1, 2, \ldots, K$, using Algorithm 3.2 or 3.4.

Step 2) Rank the p-values in ascending order, $p_{1^*} \leq p_{2^*} \leq p_{3^*} \leq \cdots \leq p_{K^*}$ for the corresponding hypotheses $\mathcal{H}_{0,1^*}, \mathcal{H}_{0,2^*}, \mathcal{H}_{0,3^*}, \ldots, \mathcal{H}_{0,K^*}$.

Step 3) Calculate $k^*_{max} = \max\{1 \leq k \leq K : \ p_{k^*} \leq \dfrac{k\alpha^*}{K}\}$.

Step 4) If such k^*_{max} exists, reject the null hypotheses $\mathcal{H}_{0,1^*}, \mathcal{H}_{0,2^*}, \ldots, \mathcal{H}_{0,k^*_{max}}$ (the corresponding subbands are declared occupied). Otherwise, reject nothing (all subbands are unoccupied).

original work, i.e., the Benjamini-Hochberg procedure (BHP) [83], to gain more power. For convenience, we first elaborate on the BHP briefly and then the A-BHP will follow.

Suppose that the the binary hypothesis (3.38) for the subband k, $k = 1, \ldots, K$, is tested either using Algorithm 3.2, for the local spectrum sensing, or using Algorithm 3.4, for the collaborative spectrum sensing. Let $p_1, p_2, \ldots, p_K$ denote the resulting p-values corresponding to the respective null hypotheses $\mathcal{H}_{0,1}, \mathcal{H}_{0,2}, \ldots, \mathcal{H}_{0,K}$.[9] For independent test statistics and for any configuration of alternative hypotheses, the BHP shown in Algorithm 3.5 controls the FDR at α^*. As mentioned previously, in terms of power the BHP is better than MTPs based on the FWE controlling procedure, such as simple Bonferroni or Holm's sequentially rejective method. However, when not all null hypotheses are true (not all subbands unoccupied), $K_0 < K$, the BHP controls the FDR at a level which is too low. To solve this issue, the A-BHP estimates the current configuration of the tested hypotheses, and then the joint test adapts accordingly. Using this procedures, the power of the test is further improved, so the interferences to the primary users is reduced.

The A-BHP basically combines two steps. The first is to estimate K_0, and the second is to apply the BHP according to the estimate. In order to estimate K_0, the A-BHP relies on the following observation. Suppose that the test statistics are independent and all null hypotheses are true, $K_0 = K$. Since the p-values p_k, $k = 1, \ldots, K$, are uniformly distributed over $[0, 1]$, the ordered set of p-values, denoted as $p_{1^*} \leq p_{2^*} \leq p_{3^*} \leq \cdots \leq p_{K^*}$, behave as an ordered snapshot from the uniform distribution. Hence, the expected value

$$E\left[p_{k^*}\right] \approx \frac{k}{K+1}, \quad k = 1, \ldots, K. \tag{3.40}$$

[9] The p-values are from Step 4 for the local spectrum sensing, or from Step 5a or 5b for the collaborative spectrum sensing.

Algorithm 3.6 The adaptive Benjamini-Hochberg procedure (A-BHP)

Step 1) Calculate the p-values of $p_1, p_2, \ldots, p_K$ for the subbands $k = 1, 2, \ldots, K$, using Algorithm 3.2 or 3.4.

Step 2) Rank the p-values in ascending order, $p_{1^*} \leq p_{2^*} \leq p_{3^*} \leq \cdots \leq p_{K^*}$ for the corresponding hypotheses $\mathcal{H}_{0,1^*}, \mathcal{H}_{0,2^*}, \mathcal{H}_{0,3^*}, \ldots, \mathcal{H}_{0,K^*}$.

Step 3) Compare the ordered p-values using the condition $p_{k*} \geq \dfrac{k\alpha^*}{K}$, $\forall k = 1, \ldots, K$. If all satisfy it, reject no hypothesis (all subbands declared unoccupied) and stop. Otherwise, continue to the next step

Step 4) Start with $k = 1$, calculate the slope using (3.41) sequentially, until the first $\hat{S}_{k+1} < \hat{S}_k$. $\hat{K}_0$ can be found from (3.42).

Step 5) From the largest p-value p_{K^*} to the smallest value, compare the condition $p_{k^*} > \dfrac{k\alpha^*}{\hat{K}_0}$, until the first $k = k_f$ violates the condition.

Step 6) Reject all hypotheses $\mathcal{H}_{0,1^*}, \mathcal{H}_{0,2^*}, \ldots, \mathcal{H}_{0,k_f^*}$ (the corresponding subbands are declared occupied), and do not reject the rest (the corresponding subbands are declared unoccupied).

Briefly, the plot of p_{k^*} versus k should show linearity along a line with slope $S = 1/(K+1)$ which connects the origin and the point $(K+1, 1)$. However, when $K_0 < K$, the p-values corresponding to alternative hypotheses tend to be smaller and concentrate on the left side of the plot. Therefore, using the line with slope $\hat{S}$ that fits a suitable number of the largest p-values (that supposedly contributed from K_0 null hypotheses only), the estimate of K_0 can be calculated. To determine the suitable number of the largest p-values, the lowest slope estimator

$$\hat{S}_k = (1 - p_{k^*})/(K + 1 - k) \tag{3.41}$$

is calculated sequentially starting from $k = 1$ towards larger k as long as $\hat{S}_{k+1} \geq \hat{S}_k$, and otherwise stop. The estimated number of true null hypotheses is thus

$$\hat{K}_0 = \min\left[(1/\hat{S}_{k+1}), K\right]. \tag{3.42}$$

Afterwards, the BHP procedure is applied using the estimate value $\hat{K}_0$. The complete A-BHP is shown in Algorithm 3.6. For rigorous arguments and proofs, refer to [83, 94].

3.5 Simulation results

Numerical experiments were conducted to evaluate the performances of the proposed methods in local, collaborative, and multiband spectrum sensing. For the local spectrum sensing, we consider the sample sizes $N = 20, 50, 100$, and for collaborative spectrum sensing we fix $N = 100$. For all simulations, the smoothing factor is fixed to $M = 5$. The number of bootstrap replications is set to $B_s = 499$ for all cases. Note that we have confirmed that to use larger B_s for bootstrapping does not improve the performance significantly. For the pivot bootstrap test in local spectrum sensing, the number of second bootstrap replications to estimate $\hat{S}_e$ is set to $B_{s1} = 25$.[10] To generate results, we use 5×10^3 Monte Carlo runs.

Here, we assume that the noise $\mathbf{w}[n]$ follows an i.i.d. Gaussian distribution $\mathcal{N}(0, \sigma_w^2)$, where we set the noise power to $\sigma_w^2 = 1$.[11] However, we do not lose generality, since the test based on (3.6) under $\mathcal{H}_0$ does not depend on the noise power. The signal transmitted by a primary user $\mathbf{s}[n]$ also follows an i.i.d. Gaussian distribution. We consider the channel to be a multipath fading channel, and hence the signal arrives at a CR user as

$$\mathbf{r}[n] = \sum_{j=1}^{J-1} \mathbf{h}[j]\mathbf{s}[n-j], \tag{3.43}$$

where $\mathbf{h}[j],\ j = 1, \ldots, J-1$, denotes the discrete-time channel impulse response between the primary transmitter and a CR user's receiver and J is the number of resolvable paths. The coefficients $\mathbf{h}[j],\ j = 1, \ldots, J-1$, are assumed to be constant during the period of an observation and we set $J = 10$. However, they indeed change between measurements and each follows an i.i.d. Gaussian distribution with average power path gain (in dB relative to the first path) as follow

$$[0 \quad -0.5 \quad -1 \quad -3 \quad -7 \quad -10 \quad -10 \quad -10 \quad -20 \quad -20]. \tag{3.44}$$

In this case, the received signal $\mathbf{r}[n]$ is temporally correlated due to the channel impulse response. It is noteworthy that the mechanism that produces correlations on the received signal is irrelevant to the performance of the tests. The received signal could be temporally correlated due to various reasons, such as oversampling, channel time dispersion (as we simulate here), or correlations in the original signal itself.

[10] To estimate the standard error of an estimator, the larger the number of replications the better. However, $B_{s1} = 25$ still gives a good performance [63].

[11] Note that from the beginning we do not assume the signal and noise to follow a specific distribution. Therefore, the tests can still be used, even for cases which depart from our assumptions here.

3.5.1 Local spectrum sensing

Here, the performances of the null-resampling bootstrap test, as the proposed method, is compared with the bootstrap pivot test and the asymptotic test [22]. In general, the objectives are to verify how close the actual probability of false alarm of a test is to the nominal value, and to evaluate the power of the tests (probability of detection) for various SNR values. At the same time, we strengthen the arguments that we mention in Section 3.2.3.

Fig. 3.5 and 3.6 show that the bootstrap null-resampling test has a better performance than the bootstrap pivot test and the asymptotic test, in terms of maximizing the probability of detection, subject to a nominal value of the probability of false alarm. The actual probability of false alarm of the bootstrap pivot test is too large for the respective nominal value and it fails to meet the objective $P_f \leq \alpha$. For example, for a nominal $\alpha = 0.05$ and 0.1, it yields $P_f \approx 0.6$ and 0.7, respectively. The false alarm decreases as the sample size increases, but it is still far from the nominal value. A common consequence of having a larger probability of false alarm is to have a better probability of detection. This can be seen in Fig. 3.6, where the probability of detection of the bootstrap pivot test is better for all SNRs. However, the condition is not desired for the spectrum sensing. This is because the pivot bootstrap test results in a better protection against interference to the primary network, but misses out on many opportunities to use unoccupied bands, which in turn heavily reduces the throughput of the CR network. Note that the curves for the pivot bootstrap test in Fig. 3.5 are similar to the approximation curve from (3.21) that is depicted in Fig. 3.1.

Meanwhile, Fig. 3.5 shows that the asymptotic test is the opposite of the bootstrap pivot test. The actual probability of false alarm is far below the nominal value, especially for small sample sizes. It gets closer to the nominal value as the sample size increases. However, even for $N = 100$, the gap is still considerably large, as large as 0.1 (10%) for $\alpha > 0.1$. This indicates that the underlying distribution of the test statistic, under $\mathcal{H}_0$, slowly converges to the assumed asymptotic distribution as the sample size N increases. In contrast, the null-resampling bootstrap test could maintain the actual probability of false alarm close to the nominal value, especially for $N = 50$ and $N = 100$. Consequently, the resulting probability of detection of the null-resampling bootstrap test is larger than the asymptotic test. This is shown in Fig. 3.6. For $N = 100$, the gap is approximately $7\% - 15\%$ over the SNR range and more pronounced for smaller sample sizes. In this case, the asymptotic test is practically too conservative in protecting the probability of false alarm. It is noteworthy that the simulation results supports the arguments in Subsection 3.2.3.

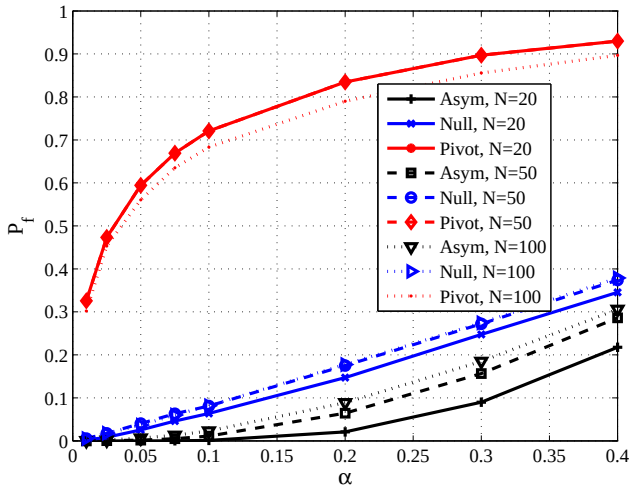
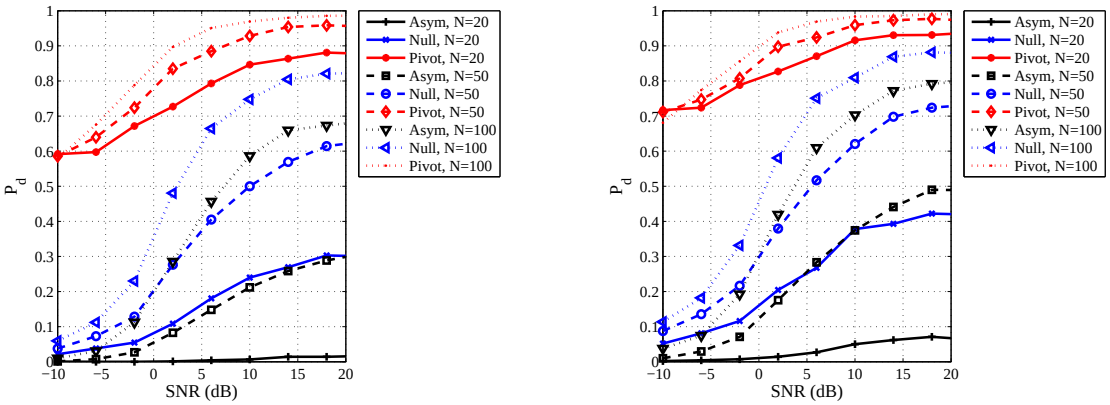


Figure 3.5: Actual probability of false alarm P_f vs. nominal value α for the asymptotic approach (Asym), the null-resampling bootstrap test (Null), and the bootstrap pivot test (Pivot) with different sample sizes N.



(a) Nominal prob. of false alarm $\alpha = 0.05$ (b) Nominal prob. of false alarm $\alpha = 0.1$

Figure 3.6: Probability of detection P_d vs. SNR (dB) for the asymptotic approach (Asym), the null-resampling bootstrap test (Null), and the bootstrap pivot test (Pivot) with different sample sizes N.

In summary, the simulation results have shown that the null-resampling test gives a better estimate of the true distribution G_0 of the test statistic under $\mathcal{H}_0$ than the bootstrap pivot test and the asymptotic test. The fact that the probability of detection is relatively small for small SNRs is merely a consequence of using small sample sizes. Fortunately, the performance can be improved by a collaborative spectrum sensing scheme which is to be elaborated in the next section.

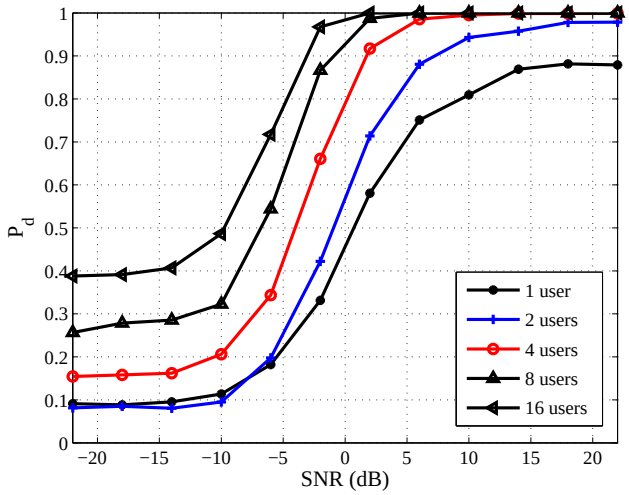
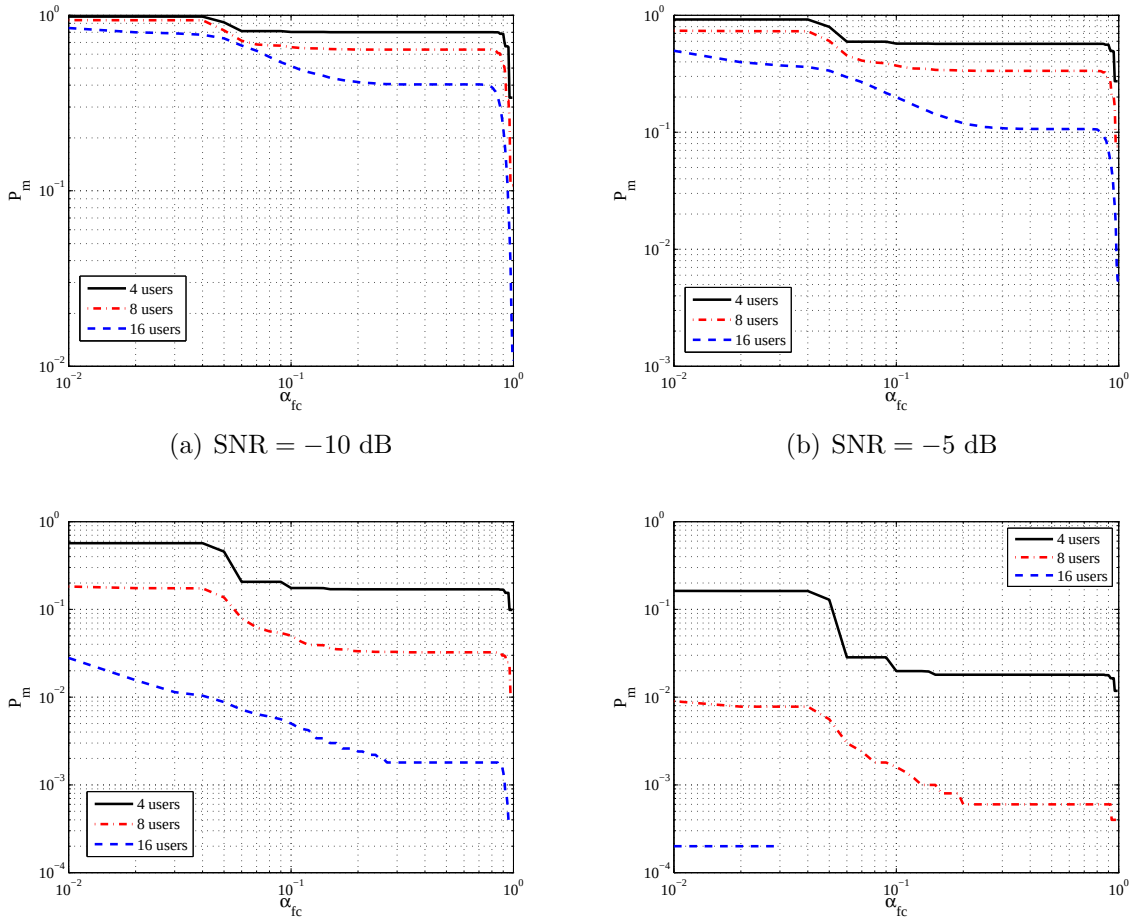


Figure 3.7: Probability of detection P_d vs. SNR (dB) for various number of CR users.

3.5.2 Collaborative spectrum sensing

In this part, performances of collaborative spectrum sensing are shown for numbers of CR users $L = 2, 4, 8$ and 16. The objectives are to quantify the improvement on the probability of detection (or miss detection) due to collaboratively detecting a primary user's signal, instead of locally detecting, and to evaluate the Chair-Varshney fusion rule using the bootstrap approach in comparison to the Chair-Varshney fusion rule using all binary combinations. Note that the received signal component at each CR user follows (3.43). The average path gain of all CR users follow (3.44), and the channel coefficients are independent between CR users.

Fig. 3.7 depicts the probability of detection as a function of SNR for various number of CR users. It shows that the SNR gain of the collaborative spectrum sensing is significant. For example, for target probability of detection $P_d \geq 0.8$, the local spectrum sensing can only achieve this when the SNR $\geq$ 10 dB. Meanwhile, the target can be achieved at the SNR ≥ -5 dB when 16 CR users detect collaboratively. The SNR gain is approximately 15 dB in this case. To corroborate, Fig. 3.8 shows the complementary receiver operating characteristic (ROC) of collaborative spectrum sensing with $4, 8$ and 16 CR users, for different values of SNR. In the figure, the probability of miss detection P_m for some curves remains constant for different nominal values of the global probability of false alarm α_{fc}, particularly for fewer CR users. This is due to discrete observations at the FC. Note that, to solve this issue, we cannot apply randomized test, as suggested in [95]. This is because the distributions of the test statistics change from one measurement to another, as we aim to anticipate the varying channels (varying



(a) SNR = −10 dB

(b) SNR = −5 dB

(c) SNR = 0 dB

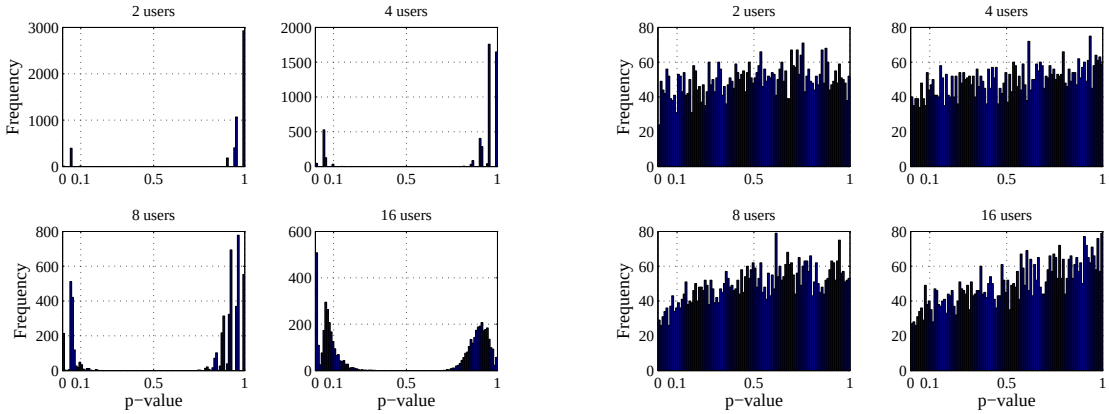
(d) SNR = 5 dB

Figure 3.8: Complementary ROC, probability of miss detection P_m vs. nominal values of the global probability of false alarm α_{fc}, for different numbers of CR users and SNRs.

average SNR of the received signal) between a primary transmitter and a CR user's receiver.

The discrete observations also have an implication on the actual value of the global probability of false alarm P_f, as follows. To generate the curves in Fig. 3.7 and Fig. 3.8, each CR user was set to have $\alpha^l = 0.05, l = 1, \ldots, L$ and the nominal value of the global probability of false alarm was set to $\alpha_{fc} = 0.1$. The actual value P_f deviates from the nominal value. They are 0.08, 0.15, 0.27 and 0.38 for $2, 4, 8$ and 16 CR users, respectively. This is inevitable because the distribution of the test statistic at the FC under $\mathcal{H}_0$ is discrete. Therefore, the distribution of the p-value is not uniform [62]. In our case, it is concentrated on low and high values as shown in the histograms of Fig 3.9.a. It can be extracted from the figure and from (2.14) that

$$P_f = P_0(P \leq 0.1) > 0.1, \tag{3.45}$$



(a) Discrete (b) Continue

Figure 3.9: Distributions of p-value under $\mathcal{H}_0$ for the resampling Chair-Varshney fusion rule (discrete) and the equal combining fusion rule (continue), for $2, 4, 8$ and 16 CR users.

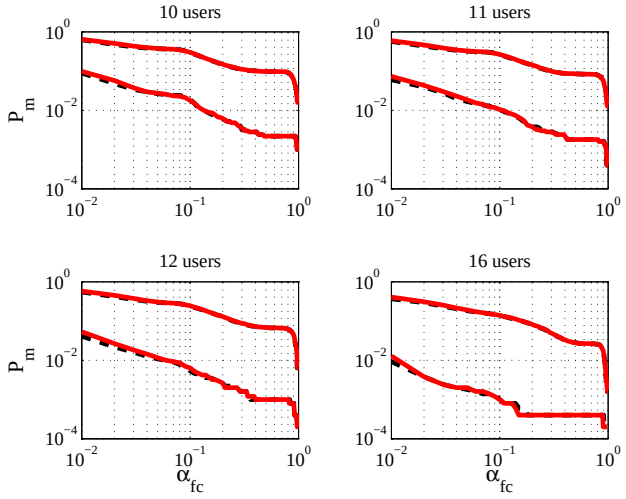


Figure 3.10: Complementary ROC, probability of miss detection P_m vs. nominal values of the global probability of false alarm α_{fc}, when the distributions of Chair-Varshney fusion rule under $\mathcal{H}_0$ are generated by the all combinations (solid line) and the bootstrap approach (dashed line), for $10, 11, 12$ and 16 CR users. The upper and the lower curves are for SNR$= -5$ dB and SNR$= 0$ dB, respectively.

for number of CR users $L > 2$. In contrast, when the observations at the FC is continuous the distribution of the p-value is approximately uniform as depicted in Fig.3.9.b. To generate the p-value distributions in Fig.3.9.b, we use equal gain combining fusion

rule at the FC. Formally,

$$T_{egc} = \sum_{l=1}^{L} \hat{T}_l, \tag{3.46}$$

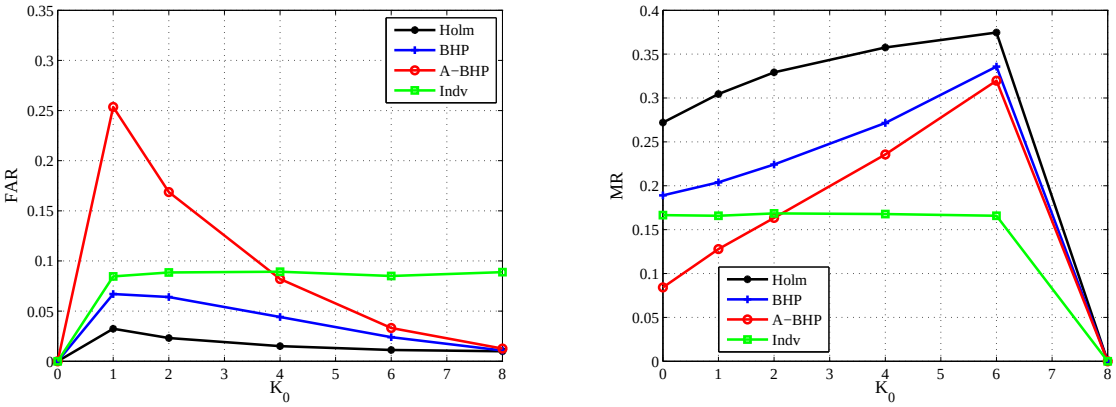
where $\hat{T}_l$ is the test statistic (3.9) from the CR user l. To get a p-value, $\hat{G}_0(T_{egc})$, the EDF of T_{egc}, is generated using the bootstrap replications of $\hat{T}_l,\ l = 1, \ldots, L$. In this case, each CR user applies null-resampling to the observed data and then calculates the bootstrap replications $\hat{T}_l^b,\ b = 1, \ldots, B_s$. The results are transmitted to the FC to be combined and yield bootstrap replications $T_{egc}^b,\ b = 1, \ldots, B_s$. Therefore, the p-value can be calculated as usual. The resulting actual probability of false alarm in this setup is thus $P_f = P_0(P \leq 0.1) < 0.1$, as can be extracted from the figure. More precisely, the probabilities of false alarm are approximately $0.08, 0.08, 0.07$ and 0.07 for $2, 4, 8$ and 16 CR users, respectively. In terms of false alarm rate, this setup is attractive. However, it consumes a much higher bandwidth for the reporting channels than the resampling Chair-Varshney fusion rule. This is because each bootstrap replication should be transmitted to the FC from each CR user.[12] Note that the simulation setups for the two schemes to generate the p-value distributions are the same. Hence, we believe that the proposed resampling Chair-Varshney fusion rule is more appropriate to be used in collaborative spectrum sensing.

Fig. 3.10 shows the complementary ROC curves for the Chair-Varshney fusion rule where the distributions under $\mathcal{H}_0$ are generated based on the all binary combinations and the bootstrap approach. The two curves in each subfigure are almost identical. Therefore, the bootstrap approach significantly reduces the computational cost without causing significant degradations in performance. For example, when the number of CR users is 16, there will be 65536 combinations, while in the bootstrap approach we only use $B_s = 499$ combinations. This corroborates the convincing results of the proposed resampling based collaborative spectrum sensing.

3.5.3 Multiband spectrum sensing

Performances of multiband spectrum sensing using Adaptive Benjamini-Hochberg procedure (A-BHP) are shown in this part. We compare the A-HBP with the conventional

[12] This scheme could also be conducted by allowing CR user to forward their sampled data to the FC without any processing. Then, the FC will calculate the bootstrap replications on behalf of the CR users. This might save the bandwidth if the sample size N taken by each user is smaller than the number of bootstrap replications B_s, like in our case. However, the bandwidth is still much higher compared to the resampling Chair-Varshney fusion rule.



(a) False alarm ratio (FAR)

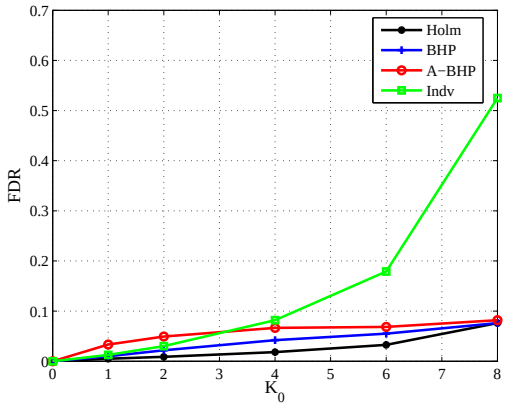
(b) Missing ratio (MR)

Figure 3.11: FAR and MR vs. K_0 for the local spectrum sensing using Holm's approach (Holm), Benjamini-Hochberg (BHP), adaptive Benjamini-Hochberg (A-BHP), and the individual testing procedure (Indv).

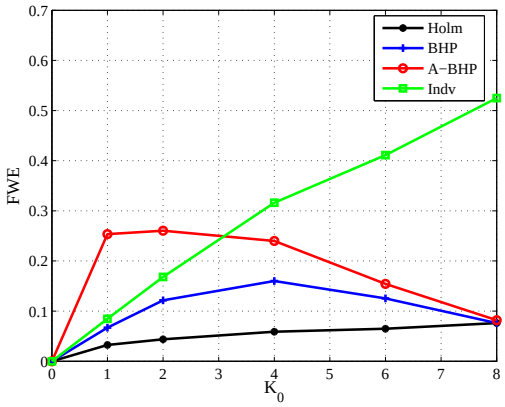
Benjamini-Hochberg procedure (BHP) and Holm procedure (HP). Local spectrum sensing is used in Fig. 3.11 and 3.12. In this case, we set the total number of subbands to $K = 8$. The number of unoccupied subbands varies from $K_0 = 0$ to 8. Note that for each K_0, the unoccupied subbands were selected randomly from K subbands. The performance improvement by collaborative spectrum sensing is shown in Fig.3.13. In all simulations, the signals in occupied subbands received by each CR user are assumed to have the same SNRs. In addition, we use the nominal value FDR = 0.1.

As explained previously, when we use FAR and MR as the system measures, such as in [84] and [93], the results could be misleading. With these measures, the individual testing (without MTP) yields better results than the one with an MTP. This is because the FAR and MR only indicate performances per subband, not familywise. In general, the individual testing has lower MRs than the schemes with an MTP, while it still preserves FAR ≤ 0.1. This condition is shown in Fig. 3.11. However, the results, and hence the conclusions, are different if we use familywise measures, such as FDR, FWE and FWM.

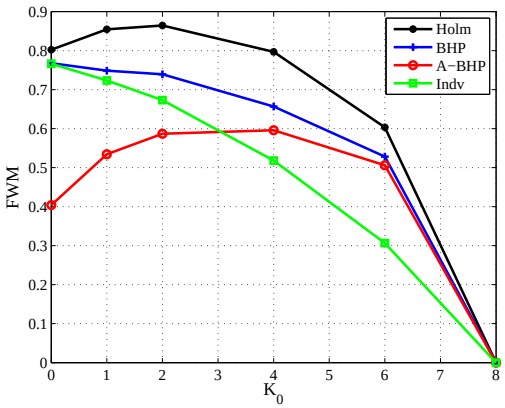
In multiband spectrum sensing, even though we use FDR as the design criteria for BHP and A-BHP, the use of FWE and FWM as system measures is more meaningful. Fig. 3.12.a. shows that the three MTPs fulfill the requirement to have the actual FDR ≤ 0.1. It is obvious from Fig 3.12.b. that the HP controls the FWE in the strong sense, while the BHP and A-BHP control the FWE in the weak sense. In this respect, for the HP, regardless of how many available subbands might be opportunistically used by a CR user, the probability of missing an opportunity to use the respective subbands



(a) False discovery rate (FDR)



(b) Familywise error rate (FWE)



(c) Familywise miss detection rate (FWM)

Figure 3.12: Familywise measures FDR, FWE and FWM vs. K_0 for the local spectrum sensing using Holm's approach (Holm), Benjamini-Hochberg (BHP), adaptive Benjamini-Hochberg (A-BHP), and the individual testing procedure (Indv).

is no larger than $\alpha = 0.1$. This condition does not hold for the BHP and A-BHP. However, strongly controlling the FWE has a consequence to increase the FWM as shown in Fig. 3.12.c., where the HP performs the worst. Meanwhile, the procedure to test each subband individually performs the best in terms of the FWM, but the worst in terms of the FWE. The case is different with the A-BHP, which has a better tradeoff between the FWE and the FWM, than the other procedures. When the number of unoccupied subbands K_0 increases, the A-BHP pays more attention to the unoccupied subbands. As a result, its FWE is better than individually testing and slightly different from the BHP and the HP. Meanwhile, when the number of occupied subbands increases (K_0 decreases), the A-BHP relaxes the FWE to gain power, that decreases the FWM. More precisely, A-BHP slightly reduces the throughput of the CR network compared to BHP and HP, when the majority of subbands are unoccupied. However, it provides better

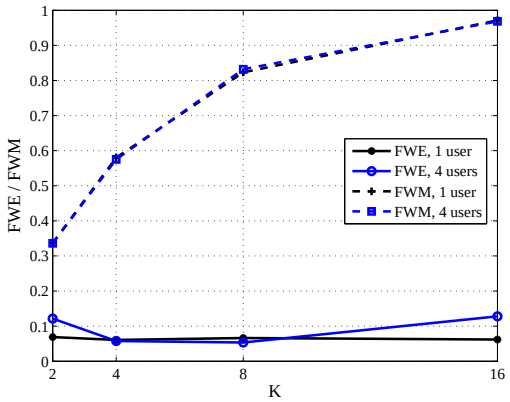
protection to the primary network against receive interferences, when the majority of subbands are occupied. Based on this fact, the A-BHP is more attractive to be implemented in multiband spectrum sensing than the other MTPs, including the individual testing.

Fig. 3.13 shows the performances of local spectrum sensing and collaborative spectrum sensing using 4 CR users. Each CR user was set to have a nominal value of false alarm $\alpha^l = 0.1$. Here, we assume that the probability of each subband being occupied by a primary user is equal, i.e., $P(\mathcal{H}_{k,1}) = P_1,\ k = 1, \ldots, K$. We consider equal SNRs for occupied subbands, precisely, SNR $= -10$ dB and SNR $= 10$ dB, and the subbands are considered to be busy ($P_1 = 0.8$), mildly busy ($P_1 = 0.5$) and sparse ($P_1 = 0.2$). The number of subbands varies from $K = 2$ to $K = 16$. The advantage of using collaborative spectrum sensing in multiband cases is not perceivable at low SNR. However, it is more pronounced when the SNR increases. When the SNR $= 10$ dB, the collaborative spectrum sensing for $K = 16$ suppresses the interferences to the primary network by approximately 70% (FWM decreases from ≈ 0.8 to 0.1), 55% and 60% for mildly occupied, sparse and busy subbands, respectively. In terms of the FWE, the collaborative spectrum sensing has performances comparable to the local spectrum sensing when the occupancies are sparse and busy. The collaborative spectrum sensing will miss up to 15% more opportunities to use unoccupied subbands compared to the local spectrum sensing, when the subbands are mildly occupied. However, overall, the gain that we obtain by using collaborative spectrum sensing for multiband cases outweighs the losses.

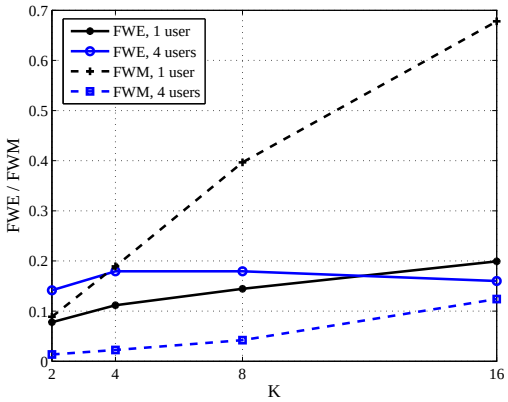
In fixed sample size cases, as we consider in this chapter, we only have control over the type I error, not both with the type II error. Thus, we focus on improving the test power to suppress the interferences. It would be convenient, if we have control over the two types of error. Thus, we can set the probability of missing unoccupied subbands, and we can limit the interferences to the primary network at the same time. This case will be considered in Chapter 5.1, where we develop a multiple testing procedure for sequential testing.

3.6 Conclusion

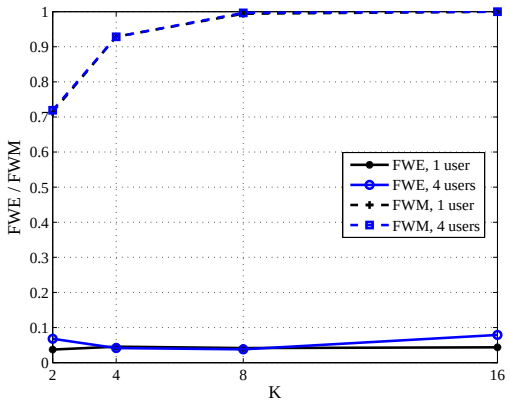
In this chapter, we address the problem of spectrum sensing with a small sample size. It includes local spectrum sensing and collaborative spectrum sensing. In addition, we apply the two schemes to multiband spectrum sensing, where controlling the familywise decision errors is more important than controlling single band decision errors. More



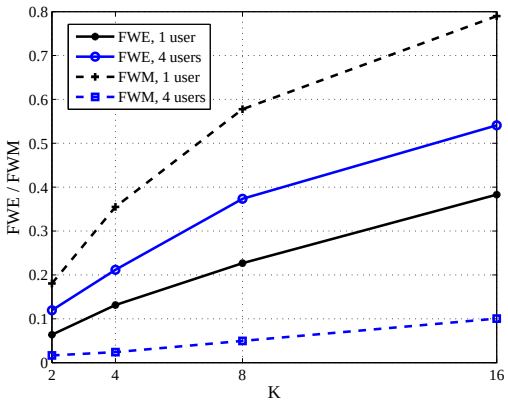
(a) Sparse, SNR = −10 dB



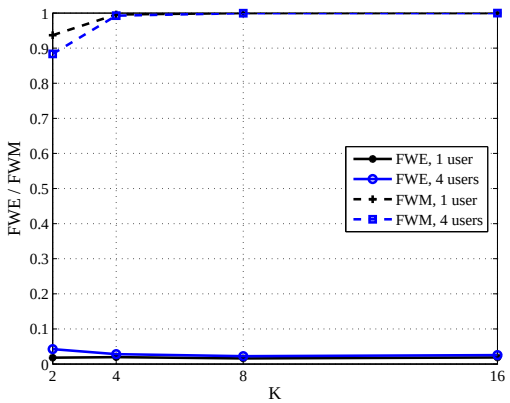
(b) Sparse, SNR = 10 dB



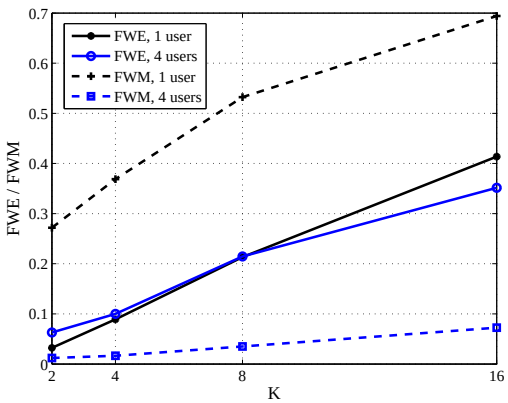
(c) Mild, SNR = −10 dB



(d) Mild, SNR = 10 dB



(e) Busy, SNR = −10 dB



(f) Busy, SNR = 10 dB

Figure 3.13: FWE and FWM vs. K for the local spectrum sensing and the collaborative spectrum sensing using 4 CR users. The subbands are busy ($P_1 = 0.8$), mildly busy ($P_1 = 0.5$) and sparse ($P_1 = 0.2$).

precisely, we considered correlated primary user's signal in uncorrelated noise. In local spectrum sensing, we propose to use the null-resampling bootstrap test. The simulations results show that the null-resampling bootstrap test has better performances than the pivot bootstrap test and the asymptotic test. The actual probabilities of false alarm of the null-resampling bootstrap test are close to the nominal values. The actual probabilities of false alarm of the asymptotic test are too conservative and hence result in a smaller power than the null-resampling bootstrap test. This means that the probability of causing interference to the primary network is considerably high. In contrary, the pivot bootstrap test has too high false alarm rates. This indicates that the CR network misses to many opportunities to use unoccupied spectrum bands. In this case, the null-resampling bootstrap test provides a better balance between the probabilities of false alarm and miss detection. For collaborative spectrum sensing, we propose to use the Chair-Varshney fusion rule. In each CR user, the distribution of the test statistic under the null hypothesis is estimated by independent bootstrap resampling, and the distribution of the test statistic under alternative hypothesis is estimated by moving-block bootstrap resampling. At the fusion center, the distribution of the fusion rule under null hypothesis is estimated by the parametric bootstrap when the number of CR users is large. The simulation results show that the Chair-Varshney fusion rule using the parametric bootstrap is comparable to the one that uses all binary combinations of CR users. At the end, we apply the local and the collaborative sensing to multi-band cases. We highlight the advantages of using the adaptive Benjamini-Hochberg for jointly testing multiple subbands, instead of the original Benjamini-Hochberg and Holm's procedure. The simulation results confirm the advantages.

Chapter 4

Bootstrap based sequential tests for spectrum sensing

In Section 2.3, we present the SPRT for testing a simple hypothesis against a simple alternative. However, the simple hypothesis approach is not suitable in most practical scenarios, such as spectrum sensing. Unknown parameters might exist which enforce the use of non-exact distributions. Therefore, extending the SPRT to composite hypotheses is necessary, as presented in this chapter.

Some studies on sequential detection for composite hypotheses have focused on asymptotic optimality [96,97]. In this setup, the probability of false alarm and the probability of miss detection are assumed to converge to zero or, equivalently, the constant thresholds are assumed to go to infinity. This work, to some extend, is based on the framework used in [54]. However, the authors in [54] mainly focus on an asymptotic bound on the error term of the log-likelihood ratio, such that the threshold margin for the upper and the lower threshold, which is a function of the sample size, can be replaced by some positive constant. Eventually, constant thresholds are used in this earlier work. The present work differs from the previous work in several aspects. Here, the SPRT is extended to cope with composite hypotheses by combining the generalized SPRT and the parametric bootstrap to obtain its thresholds. The thresholds are not assumed to be constant but are updated for each new sample. Therefore, the proposed procedure has smaller sample sizes than asymptotic methods. In addition, our method avoids exhaustive simulations to determine the thresholds under worst case assumptions, such as in [54]. A method which significantly reduces the computational costs caused by the bootstrap is also proposed. The resulting test has performance comparable to the test which is without complexity reductions.

The sensing objective is explained in Section 4.1. The problem formulation and a detailed discussion of the generalized SPRT are given in 4.2. Section 4.3 begins with a brief review of the conventional method to do thresholding for the generalized SPRT. Subsequently, our proposed method using the parametric bootstrap is explained in detail and a scheme for reducing the computational cost induced by the bootstrap is presented. An example and simulation results are given in Section 4.4. Finally, we conclude in Section 4.5.

4.1 Preliminaries

As mentioned in Section 2.1, the spectrum sensing accuracy in cognitive radio is measured by the probability of detection errors, i.e. probabilities of false alarm and miss detection. To distinguish between the nominal and the actual values, we use α and β to denote the nominal values of the probabilities of false alarm and miss detection, respectively, whereas P_f and P_m are the corresponding actual values.

Sensing time is one of the most critical issue in cognitive radio. In general, the time frame of a CR user is divided into two periods of time, sensing time and transmission time. The sensing time should be minimized in order to gain more transmission time and hence an increased throughput of cognitive radio network. However, sensing-throughput trade-off occurs in this case [43]. On the one hand, reducing the sensing time increases the transmission time which in turn increases the overall throughput of the cognitive radio network. On the other hand, reducing the sensing time decreases the sensing accuracy, which in turn partially decreases the throughput (P_f increases) and increases the degree of interference to the primary network (P_m increases).[1] Note that increasing the sensing time is the opposite. Therefore, the sensing objective is to guarantee a certain degree of interference and a certain percentage of missed transmission opportunities, i.e.

$$P_f \leq \alpha \quad \text{and} \quad P_m \leq \beta, \tag{4.1}$$

while using as few samples as possible. In this case, the degree of interference to the primary network is fixed to an allowable value and CR users gain more transmission time to increase the throughput of the cognitive radio network.

4.2 The generalized sequential probability ratio test

To begin, suppose that a CR user receives a signal which is distributed according to (2.2) under each hypothesis. Suppose that the sequence of observations $\mathbf{x}_N$ is indepen-

[1] Note that it can as well be the other way around: the faster the sensing time is, the shorter it takes the CR users to release the frequency band after being re-occupied by the primary users. Hence, the effect that a change in the sensing time has on the degree of interference needs to be investigated case by case.

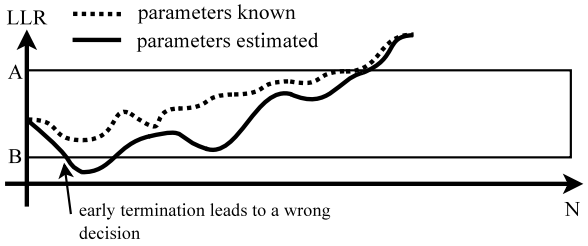


Figure 4.1: An early termination that leads to a wrong decision (miss detection) due to the difference between the log-likelihood ratio (LLR) when the parameters are estimated and when the parameters are known. The difference happens due to the estimation error.

dent and identically distributed (i.i.d.). The joint density functions of $\mathbf{x}_N$ are

$$\mathcal{H}_0 : f_{0,N}(\mathbf{x}_N; \boldsymbol{\theta}_0) = \prod_{n=1}^{N} f_0(\mathbf{x}[n]; \boldsymbol{\theta}_0), \ \boldsymbol{\theta}_0 \in \boldsymbol{\Theta}_0 \tag{4.2a}$$

$$\mathcal{H}_1 : f_{1,N}(\mathbf{x}_N; \boldsymbol{\theta}_1) = \prod_{n=1}^{N} f_1(\mathbf{x}[n]; \boldsymbol{\theta}_1), \ \boldsymbol{\theta}_1 \in \boldsymbol{\Theta}_1, . \tag{4.2b}$$

Here, the parameters $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are unknown under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Therefore, it is necessary to replace the unknown parameters by their estimates. Similar to the case of a fixed sample size detector in Section 2.2.2, the generalized log-likelihood ratio (GLLR) can then be written as

$$\hat{Z}_N = \ln \left(\frac{f_{1,N}(\mathbf{x}_N; \hat{\boldsymbol{\theta}}_1^{(N)})}{f_{0,N}(\mathbf{x}_N; \hat{\boldsymbol{\theta}}_0^{(N)})} \right) \tag{4.3}$$

where

$$\hat{\boldsymbol{\theta}}_i^{(N)} = \underset{\boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i}{\arg\max} \ln \left(f_{i,N}(\mathbf{x}_N; \boldsymbol{\theta}_i) \right) \tag{4.4}$$

is the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}_i$ assuming that hypothesis $\mathcal{H}_i, i = 0, 1$, is true. As a consequence of using the estimates of the unknown parameters, the two thresholds of the SPRT (2.20) cannot be used directly in this setup. A threshold modification is required to compensate the estimation errors introduced by the MLEs. Otherwise, the actual probability of false alarm P_f or miss detection P_m increases (depending on the underlying distribution), and hence the sensing objective (4.1) is violated. The situation is illustrated in Fig. 4.1, where the two thresholds of the simple hypotheses SPRT are employed unaltered and the test commits a miss detection due to an estimation error on the log-likelihood ratio.

Theoretically, the estimation error is statistically large for small N and vice versa. To reflect this condition, an appropriate method is needed to set the thresholds adaptively

based on N. In this way, the thresholds should be large to tolerate large estimation errors when N small and get tighter as N increases. Let A_N and B_N denote the new time dependent upper and lower thresholds. The generalized SPRT for composite hypotheses can now be defined as

$$\hat{Z}_N \begin{cases} \geq A_N, & \text{accept } \mathcal{H}_1 \\ \leq B_N, & \text{accept } \mathcal{H}_0, \\ A_N < \hat{Z}_N < B_N, & N \leftarrow N+1. \end{cases} \tag{4.5}$$

The authors in [54] suggest that A_N and B_N are functions of the sample size N with the following equations

$$A_N = A + \Delta_N^0, \quad B_N = B - \Delta_N^1, \tag{4.6}$$

where A and B are determined from (2.20), and $\Delta_N^i \geq 0, i = 0, 1$, is the modified value for the threshold at the stage N under $\mathcal{H}_i$. Note that Δ_N^0 and Δ_N^1 are used to prevent terminations that lead to wrong decisions due to estimation errors under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Thus, the sensing objective (4.1) is finally preserved.

Before we proceed, we briefly elaborate the convergence behavior of the MLEs $\hat{\boldsymbol{\theta}}_0^{(N)}$ and $\hat{\boldsymbol{\theta}}_1^{(N)}$ under both hypotheses. Recall that by assuming some regularity conditions on the family of distributions, the MLE is asymptotically efficient and consistent [98]. When $\hat{\boldsymbol{\theta}}_i^{(N)}$ is evaluated under $\mathcal{H}_j$ characterized by the distribution $f_j(\mathbf{x}_N; \boldsymbol{\theta}_j)$, as $N \to \infty$, it converges in probability to [54]

$$\bar{\boldsymbol{\theta}}_i = \underset{\check{\boldsymbol{\theta}}_i \in \boldsymbol{\Theta}_i}{\arg\min}\ KL(f_j(\mathbf{x}[n]; \boldsymbol{\theta}_j) || f_i(\mathbf{x}[n]; \check{\boldsymbol{\theta}}_i)),\ i, j = 0, 1, \tag{4.7}$$

where $KL(f_j||f_i)$ is Kullback-Leibler (KL) distance between distributions f_j and f_i. Note that when $\hat{\boldsymbol{\theta}}_i^{(N)}$ is evaluated under $\mathcal{H}_i$, i.e. $i = j$, it converges in probability to a true value $\boldsymbol{\theta}_i$. In this case the KL distance is equal to zero since $\bar{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_j (= \boldsymbol{\theta}_i)$. To summarize, under each hypothesis,

$$\mathcal{H}_0 : \hat{\boldsymbol{\theta}}_0^{(N)} \xrightarrow{\text{P}} \boldsymbol{\theta}_0 \quad \hat{\boldsymbol{\theta}}_1^{(N)} \xrightarrow{\text{P}} \bar{\boldsymbol{\theta}}_1, \quad \boldsymbol{\theta}_0 \in \boldsymbol{\Theta}_0, \bar{\boldsymbol{\theta}}_1 \in \boldsymbol{\Theta}_1 \tag{4.8a}$$

$$\mathcal{H}_1 : \hat{\boldsymbol{\theta}}_0^{(N)} \xrightarrow{\text{P}} \bar{\boldsymbol{\theta}}_0 \quad \hat{\boldsymbol{\theta}}_1^{(N)} \xrightarrow{\text{P}} \boldsymbol{\theta}_1, \quad \bar{\boldsymbol{\theta}}_0 \in \boldsymbol{\Theta}_0, \boldsymbol{\theta}_1 \in \boldsymbol{\Theta}_1, \tag{4.8b}$$

here $\xrightarrow{\text{P}}$ denotes convergence in probability as $N \to \infty$ (see Appendix 4.6.1 for the details). Throughout the paper, we make the following assumptions:

AS1) The parameter spaces $\boldsymbol{\Theta}_0$ and $\boldsymbol{\Theta}_1$ are assumed to be known and completely specified and the solution of (4.7) is unique for each value of $\boldsymbol{\theta}_j \in \boldsymbol{\Theta}_j,\ j = 0, 1$. Consequently, $\bar{\boldsymbol{\theta}}_i$ under $\mathcal{H}_j$ where $i \neq j$ can be computed offline (see Section 4.4 for example).

AS2) The expected value of

$$\ln\left(\frac{f_1(\mathbf{x}[n];\bar{\boldsymbol{\theta}}_1)}{f_0(\mathbf{x}[n];\boldsymbol{\theta}_0)}\right) \tag{4.9}$$

under $\mathcal{H}_0$ is negative and finite, and the expected value of

$$\ln\left(\frac{f_1(\mathbf{x}[n];\boldsymbol{\theta}_1)}{f_0(\mathbf{x}[n];\bar{\boldsymbol{\theta}}_0)}\right) \tag{4.10}$$

under $\mathcal{H}_1$ is positive and finite.

Now, by using (4.8) the log-likelihood ratio in (4.3) under each hypothesis can be re-written as follows,

$$\begin{aligned}\mathcal{H}_0 : \hat{Z}_N &= \ln\left(\frac{f_{1,N}(\mathbf{x}_N;\bar{\boldsymbol{\theta}}_1)}{f_{0,N}(\mathbf{x}_N;\boldsymbol{\theta}_0)}\right) + \ln\left(\frac{f_{1,N}(\mathbf{x}_N;\hat{\boldsymbol{\theta}}_1^{(N)})}{f_{1,N}(\mathbf{x}_N;\bar{\boldsymbol{\theta}}_1)}\frac{f_{0,N}(\mathbf{x}_N;\boldsymbol{\theta}_0)}{f_{0,N}(\mathbf{x}_N;\hat{\boldsymbol{\theta}}_0^{(N)})}\right)\\ &= \bar{Z}_N^0 + \Delta Z_N^0 \qquad (4.11a)\\ \mathcal{H}_1 : \hat{Z}_N &= \ln\left(\frac{f_{1,N}(\mathbf{x}_N;\boldsymbol{\theta}_1)}{f_{0,N}(\mathbf{x}_N;\bar{\boldsymbol{\theta}}_0)}\right) + \ln\left(\frac{f_{1,N}(\mathbf{x}_N;\hat{\boldsymbol{\theta}}_1^{(N)})}{f_{1,N}(\mathbf{x}_N;\boldsymbol{\theta}_1)}\frac{f_{0,N}(\mathbf{x}_N;\bar{\boldsymbol{\theta}}_0)}{f_{0,N}(\mathbf{x}_N;\hat{\boldsymbol{\theta}}_0^{(N)})}\right)\\ &= \bar{Z}_N^1 + \Delta Z_N^1, \qquad (4.11b)\end{aligned}$$

where ΔZ_N^i is the error term under $\mathcal{H}_i, i = 0, 1$, due to the estimation errors. Virtually, $\hat{Z}_N$ can be considered as an estimate of the log-likelihood ratio $\bar{Z}_N^i$ under hypothesis $\mathcal{H}_i, i = 0, 1$. By inspecting equation (4.11) and using the convergence properties of the MLEs in (4.8), $\hat{Z}_N$ approaches $\bar{Z}_N^i$ as N increases. Note that $\bar{Z}_N^i$ is the SPRT in (2.19) under hypothesis $\mathcal{H}_i$ between $\boldsymbol{\theta} = \boldsymbol{\theta}_i$ and $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}_j,\ j \neq i$. In this respect, the generalized SPRT has thresholds A_N and B_N that should approach the thresholds A and B of the SPRT. In particular, equations (4.6) are adequate for the thresholds since Δ_N^i coincides with ΔZ_N^i as N increases. Therefore, the relation between Δ_N^i and ΔZ_N^i in the generalized SPRT needs to be formalized (see Section 4.3).

Note that equation (4.11) enables us to relate the actual probabilities of false alarm and miss detection of the generalized SPRT to those of the SPRT. Assume that the generalized SPRT terminates at stopping time $N = N_s$, the probability of false alarm P_f and miss detection P_m can then be represented by

$$\begin{aligned}P_f &= P_0(\hat{Z}_{N_s} \geq A_{N_s})\\ &= P_0(\hat{Z}_{N_s} \geq A_{N_s}, \bar{Z}_{N_s}^0 \geq A) + P_0(\hat{Z}_{N_s} \geq A_{N_s}, \bar{Z}_{N_s}^0 < A)\\ &\leq P_0(\bar{Z}_{N_s}^0 \geq A) + P_0(\hat{Z}_{N_s} \geq A_{N_s}, \bar{Z}_{N_s}^0 < A), \qquad (4.12a)\\ P_m &= P_1(\hat{Z}_{N_s} \leq B_{N_s})\\ &= P_1(\hat{Z}_{N_s} \leq B_{N_s}, \bar{Z}_{N_s}^1 \leq B) + P_1(\hat{Z}_{N_s} \leq B_{N_s}, \bar{Z}_{N_s}^1 > B)\\ &\leq P_1(\bar{Z}_{N_s}^1 \leq B) + P_1(\hat{Z}_{N_s} \leq B_{N_s}, \bar{Z}_{N_s}^1 > B). \qquad (4.12b)\end{aligned}$$

By inspection of the right hand side of the inequalities (4.12a) and (4.12b) we see that the first terms are essentially the probabilities of false alarm and miss detection of the SPRT. Correspondingly, they are upper bounded by the nominal values α and β, as explained in Section 2.3. Furthermore, suppose that we can choose A_N and B_N such that the second terms are also limited by some small values,

$$P_0(\hat{Z}_{N_s} \geq A_{N_s}, \bar{Z}^0_{N_s} < A) \leq \varepsilon_0 \tag{4.13a}$$

$$P_1(\hat{Z}_{N_s} \leq B_{N_s}, \bar{Z}^1_{N_s} > B) \leq \varepsilon_1, \tag{4.13b}$$

where $0 < \varepsilon_0, \varepsilon_1 < 1$. The inequalities in (4.12) now become

$$P_f \leq \ \alpha + \varepsilon_0 \ \text{ and } \ P_m \leq \beta + \varepsilon_1. \tag{4.14}$$

This implies that to achieve the sensing objective (4.1) for the generalized SPRT, we have to choose A_N and B_N so as the probability of the event in (4.13) under each hypothesis is close to zero.

4.3 Thresholding for the generalized sequential probability ratio test

In Section 4.2, we discuss the generalized SPRT and in particular emphasize the need to modify the thresholds in order to achieve the sensing objective. In this section, we propose methods to accomplish this goal using the bootstrap. Note that we specifically aim for moderate values of the decision errors, which are typical for spectrum sensing in cognitive radio. In particular, we consider probabilities of false alarm and miss detection not less than 1%.[2] We highlight this aim since we conjecture that the generalized SPRT has asymptotic optimum properties without the need to modify the thresholds [97],[3] more precisely when $\min\{e^A, e^{-B}\} \to \infty$ as $\max\{\alpha, \beta\} \to 0$. However, this condition does not apply for moderate values of α and β. Hence, the threshold modifications are indispensable in this case. This will also be shown by our simulation results in the next section.

To modify the thresholds, we use equation (4.6) as the basis. Next, we formalize the relation between the modified value of the threshold Δ^i_N and the error term in the

[2]Spectrum sensing demands a detector which can work properly under very low signal to noise ratio (SNR). For example, it requires $P_f, P_m = 0.1$ at SNR ≈ -20 dB in the IEEE 802.22 standard [99]. For $P_f, P_m \ll 10^{-1}$, the sample size will be very large.

[3]We rely on this reference to make the statement since the basic ideas are similar, although the concerns are not the same. To the best of our knowledge, a formal proof for the asymptotic properties of the generalized SPRT as we investigate here has not yet been done in literature.

generalized SPRT ΔZ_N^i. Let the left hand side of equations (4.13) be expressed as functions of N instead of N_s. The two inequalities below are confirmed,

$$\begin{aligned} P_0(\hat{Z}_N \geq A_N, \bar{Z}_N^0 < A) &\leq P_0(|\hat{Z}_N - \bar{Z}_N^0| \geq A_N - A) \\ &= P_0(|\Delta Z_N^0| \geq \Delta_N^0) \qquad (4.15a) \\ P_1(\hat{Z}_N \leq B_N, \bar{Z}_N^1 > B) &\leq P_1(|\hat{Z}_N - \bar{Z}_N^1| \geq B - B_N) \\ &= P_1(|\Delta Z_N^1| \geq \Delta_N^1). \qquad (4.15b) \end{aligned}$$

Theoretically, the probabilities on the right hand side cannot be set equal to zero to achieve the sensing objective. However, we can choose a sufficiently small tolerance value that determines Δ_N^i, once the distribution of the random variable $|\Delta Z_N^i|$ under hypothesis $\mathcal{H}_i$ is specified for each stage N. Unfortunately, to derive a closed-form expression for the distribution of $|\Delta Z_N^i|$ appears to be non-trivial or even intractable for most cases, particularly when N is small. This is because $|\Delta Z_N^i|$ is the error term of the log-likelihood ratio which depends on the true parameters and their sequentially estimated values. To point out the difference between our approach to solving this issue and the existing method, we start with a review of the latter. In brief, the existing method relies on asymptotic assumptions, while the proposed methods use bootstrap techniques to deal with analytically intractable expressions.

4.3.1 The existing method

The authors in [54] prove that when the modified values of the thresholds in (4.6) are set according to

$$\Delta_N^0 = \sqrt{\frac{E_0[(\Delta Z_N^0)^2]}{\epsilon}}, \quad \Delta_N^1 = \sqrt{\frac{E_1[(\Delta Z_N^1)^2]}{\epsilon}} \qquad (4.16)$$

for any $0 < \epsilon < 1$, where $E_i[(\cdot)^2]$ denotes the second moment under $\mathcal{H}_i$, it then follows that

$$P_f \leq \alpha + \epsilon \text{ and } P_m \leq \beta + \epsilon. \qquad (4.17)$$

To circumvent the difficulties in calculating the second moment of ΔZ_N^i the authors rely on asymptotic approaches. More precisely, it is shown that the second moment of ΔZ_N^i under $\mathcal{H}_i$, $i = 0, 1$, is upper bounded by a constant for $N \to \infty$. Thus, as N increases, it is legitimate to replace the modified values that originally depend on N by positive constants, i.e. $\Delta_N^0 = C_A$ and $\Delta_N^1 = C_B$, to form the thresholds

$$A_N = A + C_A, \quad B_N = B - C_B. \qquad (4.18)$$

Recall that A and B are the SPRT thresholds found from (2.20). In summary, C_A and C_B in (4.18) can be chosen such that the sensing objective of (4.1) is fulfilled with a smallest possible ASN. However, a systematic way to select proper values for C_A and C_B is unknown and hence it can only rely on exhaustive Monte-Carlo simulations. In addition, C_A and C_B may vary with the parameters being considered in the model, e.g. the worst case condition for signal to noise ratio (SNR). Different values of α and β result in different C_A and C_B. This condition is exacerbated for cooperative spectrum sensing, since it demands flexibility in terms of frequent parameters changes among the cooperating CR users. Therefore, an alternative method for determining the thresholds needs to be developed.

4.3.2 The bootstrap based sequential probability ratio test (B-SPRT)

In principle, the bootstrap replaces any unknown parameters with estimates and re-uses data to provide a way to approximate distributional information, using the observed samples as a basis. Therefore, the application of the bootstrap to the case considered here is well suited. We use the parametric bootstrap since we have a well-defined probability model for data in (4.2). The only difference from the nonparametric bootstrap is that the samples are drawn from a parametric estimate of the population rather than the nonparametric estimate [63].

Now, suppose that the probability is preset by a sufficiently small tolerance value ϵ_i under hypothesis $\mathcal{H}_i, i = 0, 1$, namely

$$P_0(|\Delta Z_N^0| \geq \Delta_N^0) = \epsilon_0, \quad P_1(|\Delta Z_N^1| \geq \Delta_N^1) = \epsilon_1. \tag{4.19}$$

Since the distributions are unknown, we estimate them using ideal bootstrap[4] distributions

$$P_0^*(|\Delta Z_N^{0b}| \geq \Delta_N^{0*}) = \epsilon_0, \quad P_1^*(|\Delta Z_N^{1b}| \geq \Delta_N^{1*}) = \epsilon_1, \tag{4.20}$$

where Δ_N^{i*} is an ideal bootstrap estimate of Δ_N^i. In the implementation, we do not calculate the conditional probabilities (4.20), instead we generate B_s bootstrap replications of $|\Delta Z_N^{ib}|, i = 0, 1$ to form empirical distributions. Therefore, $\hat{\Delta}_N^i$ is obtained as a bootstrap estimate of Δ_N^i by which the thresholds A_N and B_N in (4.6) are completely determined. The detailed explanations are as follows.

[4]Here, we use the terms ideal bootstrap estimate Δ_N^{i*} and bootstrap estimate $\hat{\Delta}_N^i$. The ideal bootstrap estimate is the limit of the bootstrap estimate as the number of bootstrap replications B_s goes to infinity [63], i.e. in our case $\Delta_N^{i*} = \lim_{B_s \to \infty} \hat{\Delta}_N^i$.

We initially observe the signal up to the sample $N = N_I$. The actual hypothesis test only starts after this period.[5] This is to avoid very large estimation errors due to very small sample sizes. In addition, initial samples are also important to produce a reliable historical data set in our proposed method for reducing the computational cost of the bootstrap (See Section 4.3.3). Even though, the sequential test begins at $N = N_I + 1$, N_I is indeed included in the overall sample size of the sequential test. To avoid confusion between the data obtained by the sampling process and the data generated by the bootstrap mechanism, we use the term 'data' for the former and 'bootstrap replication' or 'bootstrap data' for the latter.

Suppose that we are at the stage $N-1$ of the sequential test. The following steps for calculating the thresholds need to be performed using recorded data $\mathbf{x}_{N-1} = (\mathbf{x}[1]\ \mathbf{x}[2]\ \cdots\ \mathbf{x}[N-1])$. Let $F_{\boldsymbol{\theta}}$ be the distribution of the observations, parameterized by $\boldsymbol{\theta}$. B_s bootstrap replications, each of size N, are generated from parametric estimates of the population $\hat{F}_{\hat{\boldsymbol{\theta}}_i^{(N-1)}}$ by assuming $\mathcal{H}_i$ is true, i.e. for $b = 1, \cdots, B_s$,

$$\mathcal{H}_0 : \hat{F}_{\hat{\boldsymbol{\theta}}_0^{(N-1)}} \rightarrow \mathbf{x}_N^{0b} = (\mathbf{x}[1]^{0b} \cdots \mathbf{x}[N]^{0b}) \tag{4.21a}$$

$$\mathcal{H}_1 : \hat{F}_{\hat{\boldsymbol{\theta}}_1^{(N-1)}} \rightarrow \mathbf{x}_N^{1b} = (\mathbf{x}[1]^{1b} \cdots \mathbf{x}[N]^{1b}), \tag{4.21b}$$

where $\hat{F}_{\hat{\boldsymbol{\theta}}_i^{(N-1)}}$ is parameterized by $\hat{\boldsymbol{\theta}}_i^{(N-1)}, i = 0, 1$, which are the MLEs of $\boldsymbol{\theta}_i$. Note that we generate the bootstrap replications for $\mathcal{H}_0$ and $\mathcal{H}_1$ to get the upper threshold A_N and the lower threshold B_N, respectively. Furthermore, the replications of the log-likelihood ratios $\hat{Z}_N$ and $\bar{Z}_N$ of (4.11) are computed for all bootstrap data $b = 1, \cdots, B_s$. Let the log-likelihood ratios be given by $\hat{Z}_N^{ib}$ and $\bar{Z}_N^{ib}$, respectively, under hypothesis $\mathcal{H}_i$. Under each assumed hypothesis we obtain

$$\mathcal{H}_0 : \hat{Z}_N^{0b} = \ln\left(\frac{f_{1,N}(\mathbf{x}_N^{0b}; \hat{\boldsymbol{\theta}}_1^{(N)0b})}{f_{0,N}(\mathbf{x}_N^{0b}; \hat{\boldsymbol{\theta}}_0^{(N)0b})}\right), \quad \bar{Z}_N^{0b} = \ln\left(\frac{f_{1,N}(\mathbf{x}_N^{0b}; \bar{\boldsymbol{\theta}}_1)}{f_{0,N}(\mathbf{x}_N^{0b}; \hat{\boldsymbol{\theta}}_0^{(N-1)})}\right), \tag{4.22a}$$

$$\mathcal{H}_1 : \hat{Z}_N^{1b} = \ln\left(\frac{f_{1,N}(\mathbf{x}_N^{1b}; \hat{\boldsymbol{\theta}}_1^{(N)1b})}{f_{0,N}(\mathbf{x}_N^{1b}; \hat{\boldsymbol{\theta}}_0^{(N)1b})}\right), \quad \bar{Z}_N^{1b} = \ln\left(\frac{f_{1,N}(\mathbf{x}_N^{1b}; \hat{\boldsymbol{\theta}}_1^{(N-1)})}{f_{0,N}(\mathbf{x}_N^{1b}; \bar{\boldsymbol{\theta}}_0)}\right), \tag{4.22b}$$

for $b = 1, 2, \cdots, B_s$, where $\hat{\boldsymbol{\theta}}_0^{(N)ib}$ and $\hat{\boldsymbol{\theta}}_1^{(N)ib}$ are the MLEs based on the bootstrap data $\mathbf{x}_N^{ib}, i = 0, 1$, from (4.21). Up to this point, we have

$$\mathcal{H}_0 : |\Delta Z_N^{0b}| = |\hat{Z}_N^{0b} - \bar{Z}_N^{0b}| \tag{4.23a}$$

$$\mathcal{H}_1 : |\Delta Z_N^{1b}| = |\hat{Z}_N^{1b} - \bar{Z}_N^{1b}|, \tag{4.23b}$$

for $b = 1, 2, \cdots, B_s$. Here, we replace the unknown parameters as follows

[5] In all our simulations, we choose $N_I = 20$.

- The SPRT $\bar{Z}_N^{0b}$ is given by $\bar{Z}_N^0$ in (4.11) with the unknown parameter $\boldsymbol{\theta}_0$ replaced by its estimate $\hat{\boldsymbol{\theta}}_0^{(N-1)}$.
- $\bar{Z}_N^{1b}$ is given by $\bar{Z}_N^1$ with the unknown parameter $\boldsymbol{\theta}_1$ replaced by its estimate $\hat{\boldsymbol{\theta}}_1^{(N-1)}$.
- $\bar{\boldsymbol{\theta}}_0$ and $\bar{\boldsymbol{\theta}}_1$ are calculated offline using (4.7) as mentioned in Section 4.2.

The resulting values in (4.23) are then ranked in increasing order to obtain

$$\mathcal{H}_0 : |\Delta Z_N^{01^*}| \leq |\Delta Z_N^{02^*}| \leq \cdots \leq |\Delta Z_N^{0B_s^*}| \tag{4.24a}$$

$$\mathcal{H}_1 : |\Delta Z_N^{11^*}| \leq |\Delta Z_N^{12^*}| \leq \cdots \leq |\Delta Z_N^{1B_s^*}|, \tag{4.24b}$$

from which the distributions in (4.19) are estimated by

$$\frac{\#(|\Delta Z_N^{0b*}| \geq \hat{\Delta}_N^0 | \mathcal{H}_0)}{B_s} = \epsilon_0, \quad \frac{\#(|\Delta Z_N^{1b*}| \geq \hat{\Delta}_N^1 | \mathcal{H}_1)}{B_s} = \epsilon_1 \tag{4.25}$$

where $\#(\cdot|\mathcal{H}_i)$ indicates cardinality under hypothesis $\mathcal{H}_i$. Therefore, $\hat{\Delta}_N^0$ and $\hat{\Delta}_N^1$ are the estimates of Δ_N^0 and Δ_N^1 in (4.19), respectively, which are chosen such that (4.25) fulfilled. At this point, the modified thresholds A_N and B_N from (4.6) can be determined, i.e.

$$A_N = A + \hat{\Delta}_N^0, \quad B_N = B - \hat{\Delta}_N^1. \tag{4.26}$$

In summary, the thresholds for the N-th stage are calculated at the stage $(N-1)$ based on the data observed thus far. Hence, upon receiving the Nth sample, the log-likelihood ratio $\hat{Z}_N$ is computed and directly compared to the calculated thresholds (A_N, B_N). The algorithm of the bootstrap based sequential probability ratio test (B-SPRT) is summarized in Algorithm 4.1.

The way this method uses the bootstrap distribution is similar to the problem of finding confidence intervals of an estimator, more precisely the percentile interval method [63]. Conditioned on a particular sequence of data $\mathbf{x}_N = (\mathbf{x}[1]\ \mathbf{x}[2]\ \ldots\ \mathbf{x}[N])$, $|\Delta Z_N| = |\hat{Z}_N - \bar{Z}_N|$ is the true value which is unknown and needs to be estimated at the stage N. The bootstrap data is generated based on the observed samples and the bootstrap statistics $|\Delta Z_N^{ib}|, i = 0, 1$, are computed accordingly to produce the bootstrap empirical distributions. Then, the endpoints of the confidence interval (the lower and the upper limits) are calculated from the distributions for a pre-specified confidence level. In our case, the lower limit is equal to zero so we only need to determine the upper limit which is $\hat{\Delta}_N^0$ or $\hat{\Delta}_N^1$, depending on the assumed hypothesis. Suppose that the tolerance values ϵ_0 and ϵ_1 in (4.25) have been given. We can state that: By using the proposed method,

Algorithm 4.1 Bootstrap based SPRT

Step 1) Initialize α, β, A and B from (2.20), B_s, ϵ_0, ϵ_1

Step 2) Draw $N = N_I$ samples and obtain initial MLEs $\hat{\boldsymbol{\theta}}_1^{(N)}$ and $\hat{\boldsymbol{\theta}}_0^{(N)}$ from (4.4)

REPEAT

Step 3) Update the thresholds using the bootstrap:

$\rightarrow$ Generate B_s bootstrap replications from (4.21)

$\rightarrow$ For $i = 0, 1, b = 1, \cdots, B_s$, compute $|\Delta Z_{N+1}^{ib}|$ from (4.23)

$\rightarrow$ Rank in increasing order as (4.24)

$\rightarrow$ Calculate $\hat{\Delta}_{N+1}^{i}, i = 0, 1$ from (4.25) and update the thresholds by (4.26)

Step 4) Draw next sample $N \leftarrow N + 1$ and obtain the MLEs $\hat{\boldsymbol{\theta}}_1^{(N)}$ and $\hat{\boldsymbol{\theta}}_0^{(N)}$

Step 5) Calculate the test statistic $\hat{Z}_N$ from (4.5)

UNTIL $\hat{Z}_N \geq A_N$ or $\hat{Z}_N \leq B_N$

Step 6) If $\hat{Z}_N \geq A_N$, accept $\mathcal{H}_1$ and if $\hat{Z}_N \leq B_N$, accept $\mathcal{H}_0$

should the data $\mathbf{x}_N$ originate from $\mathcal{H}_0$, then $|\Delta Z_N|$ will be in the interval $[0, \hat{\Delta}_N^0)$ with $(1-\epsilon_0)\times 100\%$ confidence, otherwise, should it originate from $\mathcal{H}_1$, then $|\Delta Z_N|$ will be in the interval $[0, \hat{\Delta}_N^1)$ with $(1-\epsilon_1)\times 100\%$ confidence. Note that the percentile method to determine the endpoint of the confidence interval via the bootstrap is *first-order accurate* [63].[6] The accuracy can be improved to *second-order accurate* by choosing other methods such as bias-corrected and accelerated (BC_a) [100] and bootstrap-t interval [63] methods. However, these methods have higher computational demands than the percentile method since they generally require iterated bootstrap re-sampling. In addition, our simulations using the percentile method, which is much simpler than the other methods, show satisfactory results.

4.3.3 B-SPRT with reduced computational cost

The B-SPRT algorithm explained in Subsection 4.3.2 is computationally expensive since B_s bootstrap replications for each hypothesis need to be generated at each stage N. Inspired by the work of [101], we propose a B-SPRT with reduced computational

[6] Let η be a real-valued parameter of interest for which we want to have a confidence interval. Consider a single endpoint $\hat{\eta}[\alpha]$ with one-side coverage α, i.e. $P(\eta \leq \hat{\eta}[\alpha]) \approx \alpha$. An approximate confidence point $\hat{\eta}[\alpha]$ is first-order accurate if $P(\eta \leq \hat{\eta}[\alpha]) = \alpha + O(1/\sqrt{N})$ and second-order accurate if $P(\eta \leq \hat{\eta}[\alpha]) = \alpha + O(1/N)$. This means that first-order accurate and the second-order accurate methods have errors in matching $P(\eta \leq \hat{\eta}[\alpha]) = \alpha$ that go to zero with rate $1/\sqrt{N}$ and $1/N$, respectively.

cost in this subsection. For simplicity, we refer to the B-SPRT in Subsection 4.3.2 as B-SPRT-1 and use B-SPRT-2 for the one with reduced computational cost.

In B-SPRT-1, the bootstrap distributions of $|\Delta Z_N|$ at stage N is constructed using B_s bootstrap replications at stage $(N-1)$, based on the observed data up to this stage. For simplicity, let F_N^{i*} represent the ideal cumulative bootstrap distribution at stage N under hypothesis $\mathcal{H}_i, i = 0, 1$. In B-SPRT-2, we propose to use a convex combination of the latest $K > 1$ bootstrap distributions to reduce computational cost. Suppose that $\check{F}_N^{i*}$ denotes the convex combination of the distributions at stage N under hypothesis $\mathcal{H}_i, i = 0, 1$. It can be written as

$$\check{F}_N^{i*} = \sum_{k=0}^{K-1} w_k F_{N-k}^{i*}, \quad \sum_{k=0}^{K-1} w_k = 1, \tag{4.27}$$

where $w_k \geq 0$ is the weight for the distribution at the stage $N-k$. The ideal bootstrap estimates Δ_N^{i*} at stage N can then be found from

$$\check{F}_N^{0*}(\Delta_N^{0*}) = 1 - \epsilon_0, \quad \check{F}_N^{1*}(\Delta_N^{1*}) = 1 - \epsilon_1. \tag{4.28}$$

Therefore, we only need to generate a fraction of the bootstrap replications needed by B-SPRT-1 for each stage N.

Suppose that we use equal weights $w_k = 1/K, k = 0, \cdots, K-1$, in the convex combination. We only need $B_k = B_s/K$ new bootstrap replications instead of B_s in each stage to get an empirical distribution function. In this case, the two set replications in (4.23) become

$$\mathcal{H}_0 : |\Delta Z_{N-K+1}^{01}|, \cdots, |\Delta Z_{N-K+1}^{0B_k}|, \cdots, |\Delta Z_N^{01}|, \cdots, |\Delta Z_N^{0B_k}|, \tag{4.29a}$$

$$\mathcal{H}_1 : |\Delta Z_{N-K+1}^{11}|, \cdots, |\Delta Z_{N-K+1}^{1B_k}|, \cdots, |\Delta Z_N^{11}|, \cdots, |\Delta Z_N^{1B_k}|. \tag{4.29b}$$

Each time a data sample is recorded, B_k bootstrap replications are generated and replace the oldest data, i.e. $(|\Delta Z_{N-K+1}^{i1}|, \ldots, |\Delta Z_{N-K+1}^{iB_k}|), i = 0, 1$, so the overall number of bootstrap replications at each stage is still equal to B_s. It is clear that the computational cost of the B-SPRT-2 is lower than that of the B-SPRT-1 by a factor $1/K$. The rest of the algorithm is similar to the BSPRT-1 in Algorithm 4.1. Note that, however, at the initial stage $N = N_I$, we need to generate B_s bootstrap replications to construct historical data sets which are not yet available. Another possible scheme to reduce the computational cost is to update the thresholds only at $N = N_I + lK,\ K > 1, l = 1, 2, \ldots$. The idea is indeed similar, namely to reduce the computational cost by lowering the refresh rate of the bootstrap data.

We end this subsection with a theorem regarding the stopping time of B-SPRT (B-SPRT-1 and B-SPRT-2).

Theorem 1: Suppose that the assumption AS2 holds. The stopping time N_s^B of the B-SPRT is finite under hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, i.e.

$$P_i(N_s^B < \infty) = 1, \quad i = 0, 1 \tag{4.30}$$

Proof: See Appendix 4.6.2.

4.3.4 Some remarks on the B-SPRT

Until today, it is not clear how to optimally apply the bootstrap in sequential testing. This subsection serves to highlight the differences between the B-SPRT and a similar method suggested in [101]. This is also to argue that if we implement the bootstrapping as in the B-SPRT, the computational cost is reduced significantly. At the end, we highlight the tolerance values ϵ_0 and ϵ_1 for the B-SPRT.

In contrast to our method, the approach given in [101] bootstraps the statistic $\hat{Z}_N$ itself. Suppose that we are at the stage $N-1$. The thresholds A_N and B_N are then found from

$$\mathcal{H}_0 : P_0^* \left(\max_{N_I+1 \le N \le N_{T_r}} \hat{Z}_N^{0b} \ge A_N \right) \le \alpha \tag{4.31a}$$

$$\mathcal{H}_1 : P_1^* \left(\max_{N_I+1 \le N \le N_{T_r}} \hat{Z}_N^{1b} \le B_N \right) \le \beta, \tag{4.31b}$$

for $b = 1, \ldots, B_s$, where $\hat{Z}_N^{ib}$, $i = 0, 1$, are the log-likelihood ratios from (4.22). Here, N_{T_r} is the truncated sample size which is chosen to be sufficiently large ($N_{T_r} \gg N_I$) so as to guarantee the probability that the stopping time N_s is higher than N_{T_r} is sufficiently small under each hypothesis, formally $P_i(N_s > N_{T_r}) = \epsilon$, $i = 0, 1$.[7] Note that N_{T_r} should be determined under the worst case scenario. Therefore, this method demands a very high computational cost since B_s bootstrap data, each with size N_{T_r}, need to be generated at each stage, i.e. for $b = 1, \ldots, B_s$,

$$\mathcal{H}_0 : \hat{F}_{\hat{\boldsymbol{\theta}}_0^{(N-1)}} \to \mathbf{x}_{N_{T_r}}^{0b} = (\mathbf{x}[1]^{0b} \cdots \mathbf{x}[N_{T_r}]^{0b}) \tag{4.32a}$$

$$\mathcal{H}_1 : \hat{F}_{\hat{\boldsymbol{\theta}}_1^{(N-1)}} \to \mathbf{x}_{N_{T_r}}^{1b} = (\mathbf{x}[1]^{1b} \cdots \mathbf{x}[N_{T_r}]^{1b}). \tag{4.32b}$$

To reduce the computational costs of this method, one might consider to use bootstrap data of size N at each stage, similar to the B-SPRT. Unfortunately, this is not possible

[7] Conventional, non-generalized SPRTs are often truncated to avoid very large sample sizes in case of parameter mismatches (see the detail, for example in [48, 49]).

for the following reason. To begin, it is worth to emphasize that the expression (4.31) uses the nominal values of the probabilities of false alarm α and miss detection β, since the log-likelihood ratio $\hat{Z}_N$ is the statistic to be replicated by the bootstrap. A sequential test with the nominal values α and β cannot be interpreted as a sequence of tests conducted at each stage N, each with associated nominal values $\alpha_N = \alpha$ and $\beta_N = \beta,\ N = 1, 2, \ldots$. Instead, the nominal values α and β are associated with the sequential test performed until the random stopping time N_s is reached. The calculations of $\hat{Z}_N^{ib}$ should replicate the original test, i.e. be performed for $N_I + 1 \leq N \leq N_{T_r}$ for each b, which guarantees that the test has stopped before N_{T_r}-th sample with high probability. Only in this way can the distribution of the maximum values in (4.31) be estimated accurately. Generating bootstrap data of size N at each stage, however, is insufficient and leads to violations of the sensing objective.

To circumvent these difficulties, we first introduce the SPRT $\bar{Z}_N^i$ whose performance is guarantee to fulfill the sensing objective (4.1). Secondly, we estimate the deviation of the generalized SPRT $\hat{Z}_N$ from the $\bar{Z}_N^i$, represented by the statistic $|\Delta Z_N^i|$. Now, we rely on the statistic $|\Delta Z_N^i|$ instead of $\hat{Z}_N$ to update the thresholds. Note that the distribution of the statistic $|\Delta Z_N^i|$ with associated tolerance values ϵ_i to update the threshold, depends only on the log-likelihood ratios computed at stage N. We only require bootstrap data of size N according to (4.21). The approach indeed imposes additional cost since we have to compute $\bar{Z}_N^i,\ i = 0, 1$. However, the complexity reduction that we gain by reducing the size of the bootstrap data is higher than the additional cost of computing $\bar{Z}_N^i$.

So far, we have shown that the estimation errors are reflected in the differences between the log-likelihood ratios of the generalized SPRT $\hat{Z}_N$ and the SPRT $\bar{Z}_N^i$. The main idea of the B-SPRT is to avoid premature terminations of the generalized SPRT due to the estimation errors. But, it is supposed to terminate for the correct reason, i.e. enough data have been collected to make a decision as reliable as that of an SPRT unaffected by estimation errors. This is accomplished by estimating two modified values for the thresholds $\Delta_N^i, i = 0, 1$, such that the probability that the absolute difference between the log-likelihood ratios of the generalized SPRT and the SPRT at each stage N is higher than the modified value Δ_N^i is not more than the tolerance value ϵ_i under assumed hypothesis $\mathcal{H}_i$. Briefly, the procedure prohibits termination when the estimation errors are still considerably large and delays it to a time when enough data have been observed to guarantee that the test makes the same decision as an SPRT with known parameters. Accordingly, the increases of the actual probabilities of false alarm P_f and miss detection P_m due to the estimation errors can be minimized and the sensing objective is approximately preserved. We want to emphasize here that the pre-specified tolerance values ϵ_0 and ϵ_1 in (4.25) do not correspond to the increases

of P_f and P_m, which are due to the estimation errors. However, we make sure that these increases are small by choosing sufficiently small tolerance values. An analytic expression for the relation between the two is difficult to derive, though, even in an asymptotic sense.

Note that the selection of the tolerance values ϵ_0 and ϵ_1 also affects the number of replications B_s required for the bootstrap according to (4.25). The smaller the tolerance values, the larger the number of bootstrap replications that we need to estimate the tail of the distributions. In this case, we believe that choosing tolerance values $\epsilon_0 = \epsilon_1 = 0.01$ and $B_s = 500$ is sufficient.

4.4 Example

To evaluate the performance of the proposed algorithms, the signal $\mathbf{r}[n]$ and the noise $\mathbf{w}[n]$ in (2.1) are assumed to follow a zero mean circularly symmetric complex Gaussian distribution. Therefore, the composite hypotheses that need to be tested are

$$\mathcal{H}_0 : \ \mathbf{x}[n] \sim \mathcal{CN}\left(0, \sigma_0^2\right), \quad 0 < \sigma_0^2 \leq \sigma_{0_{\max}}^2 \tag{4.33a}$$

$$\mathcal{H}_1 : \ \mathbf{x}[n] \sim \mathcal{CN}\left(0, \sigma_1^2\right), \quad \sigma_{1_{\min}}^2 \leq \sigma_1^2 < \infty, \tag{4.33b}$$

where $\sigma_{1_{min}}^2 > \sigma_{0_{max}}^2$, σ_0^2 and σ_1^2 are the unknown parameters. Here, $\mathbf{x}[n]$ and $\mathbf{w}[n]$ are scalar. The MLEs for σ_0^2 and σ_1^2 are given by

$$\hat{\sigma}_0^{2(N)} = \min\left\{\frac{1}{N}\sum_{n=1}^{N}|\mathbf{x}[n]|^2, \sigma_{0_{\max}}^2\right\}, \tag{4.34a}$$

$$\hat{\sigma}_1^{2(N)} = \max\left\{\frac{1}{N}\sum_{n=1}^{N}|\mathbf{x}[n]|^2, \sigma_{1_{\min}}^2\right\}. \tag{4.34b}$$

According to (4.8), the two estimators under each hypothesis $\mathcal{H}_i$

$$\mathcal{H}_0 : \hat{\sigma}_0^{2(N)} \xrightarrow{\mathrm{P}} \sigma_0^2, \quad \hat{\sigma}_1^{2(N)} \xrightarrow{\mathrm{P}} \bar{\theta}_1 = \sigma_{1_{\min}}^2 \tag{4.35a}$$

$$\mathcal{H}_1 : \hat{\sigma}_0^{2(N)} \xrightarrow{\mathrm{P}} \bar{\theta}_0 = \sigma_{0_{\max}}^2, \quad \hat{\sigma}_1^{2(N)} \xrightarrow{\mathrm{P}} \sigma_1^2, \tag{4.35b}$$

as $N \to \infty$. Here, $\bar{\theta}_1 = \sigma_{1_{min}}^2$ and $\bar{\theta}_0 = \sigma_{0_{max}}^2$ are found from (4.7). The log-likelihood ratio from (4.3) becomes

$$\begin{aligned}\hat{Z}_N &= \sum_{n=1}^{N} \ln\left(\frac{f_1(\mathbf{x}[n]; \hat{\sigma}_1^{2(N)})}{f_0(\mathbf{x}[n]; \hat{\sigma}_1^{2(N)})}\right) \\ &= \left(\frac{1}{\hat{\sigma}_0^{2(N)}} - \frac{1}{\hat{\sigma}_1^{2(N)}}\right)\sum_{n=1}^{N}|\mathbf{x}[n]|^2 + N\ln\frac{\hat{\sigma}_0^{2(N)}}{\hat{\sigma}_1^{2(N)}}.\end{aligned} \tag{4.36}$$

In this case, the log-likelihood ratios to generate bootstrap replications according to (4.3) are

$$\mathcal{H}_0 : \hat{Z}_N^{0b} = \sum_{n=1}^{N} \ln\left(\frac{f_1(\mathbf{x}[n]^{0b}; \hat{\sigma}_1^{2(N)0b})}{f_0(\mathbf{x}[n]^{0b}; \hat{\sigma}_0^{2(N)0b})}\right),$$

$$\bar{Z}_N^{0b} = \sum_{n=1}^{N} \ln\left(\frac{f_1(\mathbf{x}[n]^{0b}; \sigma_{1_{\min}}^2)}{f_0(\mathbf{x}[n]^{0b}; \hat{\sigma}_0^{2(N-1)})}\right), \tag{4.37a}$$

$$\mathcal{H}_1 : \hat{Z}_N^{1b} = \sum_{n=1}^{N} \ln\left(\frac{f_1(\mathbf{x}[n]^{1b}; \hat{\sigma}_1^{2(N)1b})}{f_0(\mathbf{x}[n]^{1b}; \hat{\sigma}_0^{2(N)1b})}\right),$$

$$\bar{Z}_N^{1b} = \sum_{n=1}^{N} \ln\left(\frac{f_1(\mathbf{x}[n]^{1b}; \hat{\sigma}_1^{2(N-1)})}{f_0(\mathbf{x}[n]^{1b}; \sigma_{0_{\max}}^2)}\right), \tag{4.37b}$$

where $\hat{\sigma}_1^{2(N)ib}$ and $\hat{\sigma}_0^{2(N)ib}$ are the MLEs evaluated using the bootstrap data $\mathbf{x}_N^{ib} = (\mathbf{x}[1]^{ib} \dots \mathbf{x}[N]^{ib})$, $i = 0, 1,\ b = 1, \dots, B_s$. For all simulations, we choose $\sigma_{0_{\max}}^2 = 1$ and $\sigma_{1_{\min}}^2 = 1.1$. Thus, in terms of SNR the worst case scenario is $10 \log_{10}\left(\frac{\sigma_{1_{\min}}^2 - \sigma_{0_{\max}}^2}{\sigma_{0_{\max}}^2}\right) = -10\ dB$. It happens when σ^2 lies at the boundary of the parameter space, i.e. either at $\sigma_1^2 = 1.1$ under $\mathcal{H}_1$ or at $\sigma_0^2 = 1$ under $\mathcal{H}_0$.

We pre-specify the initial sample sizes to be $N_I = 20$ for the B-SPRT, so the sequential test begins at $N = 21$. The tolerance values are set to $\epsilon_0 = \epsilon_1 = 0.01$ and the number of bootstrap replications is $B_s = 500$. Hence, $\hat{\Delta}_N^i$ is the 496th largest of the ordered $|\Delta Z_N^{ib^*}|, b = 1, \dots, B_s$. For the bootstrap method with reduced computational cost (BSPRT-2), the number of bootstrap replications is initially set to $B_s = 500$ at N_I to provide a historical data set in the first place and continue with B_k bootstrap replications for $N > N_I$. We use equal weights $w_k = 1/K$ and evaluate the performance for $K = 4$ ($B_k = 500/4 = 125$), 20 ($B_k = 25$), and 100 ($B_k = 5$). We evaluate the performance under nominal probabilities of false alarm and miss detection $\alpha = \beta$ ranging from 0.01 to 0.1. The simulation results were generated using 5×10^3 Monte-Carlo runs. The performance measures that we use are the actual probability of false alarm P_f, the actual probability of miss detection P_m and the ASN.

To get tighter stopping conditions for B-SPRT-1 and B-SPRT-2, an adjustment for $|\Delta Z_N^{ib}|, i = 0, 1$, during the bootstrapping is necessary. The idea is to not update the thresholds, i.e. to set $|\Delta Z_N^{ib}| = 0$, whenever the corresponding $\bar{Z}_N^{ib}$ for every sequence $\mathbf{x}_N^{ib} = (\mathbf{x}[1]^{ib} \dots \mathbf{x}[N]^{ib}), b = 1, \dots, B_s$ has terminated, except for a sequence $\mathbf{x}_N^{0b} = (\mathbf{x}[1]^{0b} \dots \mathbf{x}[N]^{0b})$ which produces $\bar{Z}_N^{0b} \leq B$ and $\hat{Z}_N^{0b} > B$ under $\mathcal{H}_0$ and for a sequence $\mathbf{x}_N^{1b} = (\mathbf{x}[1]^{1b} \dots \mathbf{x}[N]^{1b})$ which produces $\bar{Z}_N^{1b} \geq A$ and $\hat{Z}_N^{1b} < A$ under $\mathcal{H}_1$. This is because we want to make the probability of having the B-SPRT-1 or the B-SPRT-2 to terminate with the same decision as the corresponding SPRT as high as possible.

Table 4.1: Constant values C_A and C_B of the MT-SPRT and the approximated minimum value $N_{\min}$ of the FSND

$\alpha = \beta$	C_A	C_B	$N_{\mathbf{min}}$
0.01	1.1	1.3	2387
0.025	1.0	1.3	1694
0.05	1.0	1.2	1193
0.075	0.8	1.2	914
0.1	0.8	1.2	724

4.4.1 Benchmarks

We compare the proposed method to the conventional one (Section 4.3.1) which we will refer to as MT-SPRT in the sequel. The constant values C_A and C_B for the thresholds of the MT-SPRT in (4.18) are depicted in Table 4.1 for each $\alpha = \beta$. These constants were found by exhaustive simulations under the worst case scenario ($SNR = -10\ dB$). The tests were conducted for 2×10^3 Monte-Carlo runs under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$. The thresholds A_N and B_N were chosen as constants C_A and C_B, respectively, were determined via grid search over values ≥ 0.1 with step size 0.1. As optimum values for C_A and C_B, we picked those which fulfill the sensing objective while minimizing $\max\{\mathrm{ASN}_0, \mathrm{ASN}_1\}$ as suggested in [54]. Here, ASN_i denotes the average sample number under hypothesis $\mathcal{H}_i$.

We also compare the B-SPRT-1 and B-SPRT-2 to the corresponding best fixed sample size detector (FSND). For the FSND, the log-likelihood ratio in (4.36) is used as the test statistic. Based on Neyman-Pearson's approach, a threshold τ which separates the two decision regions is selected according to the nominal value of the probability of false alarm α. For this detector type, the minimum required sample size $N_{\min}$ is determined in advance and cannot be changed over the unknown σ^2 values in the parameter spaces. Hence, it should be calculated in order to fulfill the sensing objective under the worst case scenario, i.e. $\sigma_0^2 = \sigma_{0_{\max}}^2$ and $\sigma_1^2 = \sigma_{1_{\min}}^2$. The minimum sample size can be obtained by

$$N_{\min} \approx \left(\frac{\mathcal{Q}^{-1}(\beta)\left(\frac{\sigma_{1_{\min}}^2}{\sigma_{0_{\max}}^2} - 1\right) + \mathcal{Q}^{-1}(\alpha)\left(1 - \frac{\sigma_{0_{\max}}^2}{\sigma_{1_{\min}}^2}\right)}{\frac{\sigma_{1_{\min}}^2}{\sigma_{0_{\max}}^2} + \frac{\sigma_{0_{\max}}^2}{\sigma_{1_{\min}}^2} - 2} \right), \tag{4.38}$$

where $\mathcal{Q}^{-1}(\cdot)$ is the inverse Q-function of the standard normal cumulative distribution function. Table 4.1 shows $N_{\min}$ for each $\alpha = \beta$.

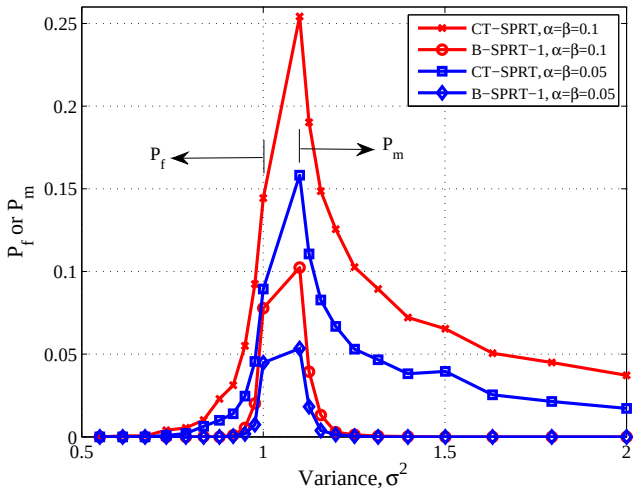


Figure 4.2: Performance improvement of B-SPRT-1 in terms of P_f and P_m compared to the generalized SPRT using constant thresholds A and B for $\alpha = \beta = 0.1$ and 0.05.

4.4.2 Simulation results

First, we concentrate on the proposed method B-SPRT-1. The B-SPRT-2 will be discussed afterwards. To quantify the improvement of the B-SPRT-1 in terms of P_f and P_m, we compare it to the generalized SPRT whose two thresholds are held constant (CT-SPRT), i.e. $A_N = A$ and $B_N = A$. This is also to show the effectiveness of the B-SPRT-1 in avoiding decision errors due to the estimation errors. The result is depicted in Fig. 4.2, which shows the improvement of the B-SPRT-1 over the CT-SPRT for all permitted σ^2 values under each hypothesis. The improvement is more pronounced when σ^2 is close to the boundaries of the parameter spaces (low SNR region). For nominal values $\alpha = \beta = 0.1$, when $\sigma^2 = 1$ the true error probabilities are reduced from $P_f \approx 0.15$ for the CT-SPRT to ≈ 0.08 for the B-SPRT-1 and when $\sigma^2 = 1.1$, P_m reduces from ≈ 0.25 to ≈ 0.1. Likewise, for $\alpha = \beta = 0.05$, the error probabilities go down from $P_f \approx 0.09$ and $P_m \approx 0.16$ for CT-SPRT to ≈ 0.04 and ≈ 0.05 for the B-SPRT-1. P_f and P_m of both detectors decrease as σ^2 departs from the boundaries. In summary, choosing the thresholds as functions of the sample size using the parametric bootstrap leads to convincing results in sequential detection for composite hypotheses, meaning that the sensing objectives are preserved.

Fig. 4.3 shows the actual error probabilities P_f and P_m as functions of the nominal probabilities ($\alpha = \beta$) under $\sigma^2 = 1$ ($\mathcal{H}_0$) and $\sigma^2 = 1.1$ ($\mathcal{H}_1$) for the CT-SPRT, the B-SPRT-1 and the MT-SPRT. For all nominal values $\alpha = \beta$, MT-SPRT and B-SPRT-1 perform comparably well in preserving the sensing objective. However, satisfactory

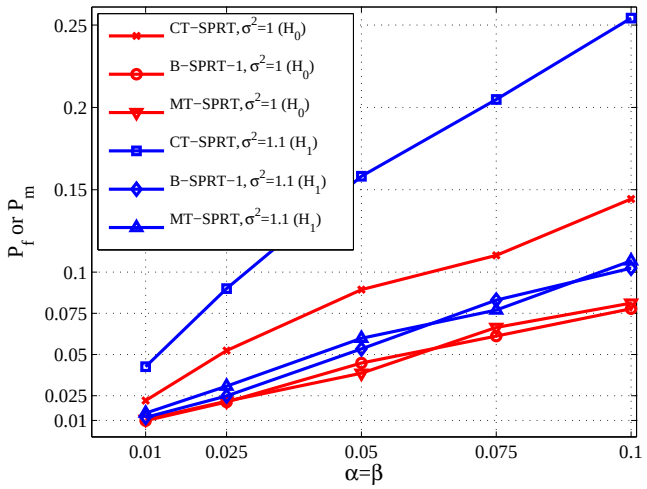


Figure 4.3: Performance comparisons in terms of P_f and P_m for CT-SPRT, B-SPRT-1 and MT-SPRT, as functions of nominal values $\alpha = \beta$, evaluated at $\sigma^2 = 1$ (under $\mathcal{H}_0$) and $\sigma^2 = 1.1$ (under $\mathcal{H}_1$).

results in terms of P_f and P_m are not the only concern of sequential tests. Fig. 4.4 shows a comparison of MT-SPRT and B-SPRT-1 in terms of the ASN for low SNR (top sub-figure) and high SNR (bottom sub-figure). From the top sub-figure we can see that the proposed method B-SPRT-1 outperforms the conventional MT-SPRT over all nominal values of $\alpha = \beta$ under each hypothesis. The improvement ranges from $\approx$ 130 to 200 samples under $\mathcal{H}_0$ and $\approx$ 70 to 150 samples under $\mathcal{H}_1$. On the other hand, the MT-SPRT is better than the B-SPRT-1 for high SNR as shown in Fig. 4.4 (bottom sub-figure). Here, the margin ranges from $\approx$ 10 to 14 samples under $\mathcal{H}_0$ and $\mathcal{H}_1$. A possible explanation for this behavior is the fact that the B-SPRT-1 starts only after some initial samples $N_I = 20$ have been taken. This eliminates the possibility that it terminates at stopping times $N_s \leq 20$. At high SNR, the probability that a test terminates before $N_I = 20$ is much higher than at low SNR, since the average increment of the log-likelihood ratio increases with the SNR. Meanwhile, the MT-SPRT is not affected by this problem since it starts the sequential test with the first sample.

We have already stressed the fact that the average of the estimation errors is larger for a smaller number of samples N and vice versa. Hence, the smaller the sample size, the larger the second order moment of the error term $|\Delta Z_N^i|$. This relation is reflected in the empirical distribution of $|\Delta Z_N^{ib}|$, which is obtained from B_s bootstrap replications. It has a smaller second order moment if all bootstrap replications are generated from the current $N - 1$ samples, than if a convex combination of the latest K bootstrap distributions is used. The larger K (smaller B_k) is, the more old data are incorporated in the distribution, i.e., the refresh rate of the data is proportional to $1/K$. On average,

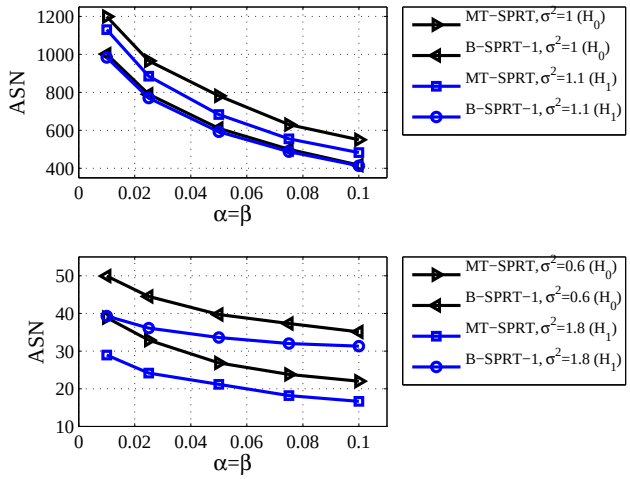


Figure 4.4: The ASNs of B-SPRT-1 and MT-SPRT when σ^2 lies at the boundaries of parameter spaces (top sub-figure) and far from the boundaries (bottom sub-figure).

this yields a larger estimate for the modified threshold value $\hat{\Delta}_N^i, i = 0, 1$. Thus, the thresholds A_N and B_N of B-SPRT-2 with larger $K > 1$ take longer to converge than the B-SPRT-1. This is displayed in Fig. 4.5, in which the thresholds have been averaged over 500 Monte-Carlo runs. In all sub-figures, the differences are hardly noticeable, except for the B-SPRT-2 with $B_k = 5$. However, when σ^2 departs from the boundaries of parameter spaces (higher SNR), their thresholds are close to each other and their convergence rates are of course faster.

The conditions on the thresholds have implications on the performance of the B-SPRT-1 and B-SPRT-2 as shown in Fig. 4.6 and 4.7. In Fig. 4.6, the B-SPRT-1 and B-SPRT-2 approximately achieve the sensing objective. In general, it shows that the B-SPRT-2 which has looser thresholds than B-SPRT-1 has better protection against decision errors, particularly for large K ($B_k = 5$). However, this also implies that the ASN of the B-SPRT-2 is larger than that of the B-SPRT-1 as depicted in Fig. 4.7. The largest margin between the B-SPRT-1 and the B-SPRT-2 can be observed for $B_k = 5$ and $\alpha = \beta = 0.1$, namely ≈ 40 samples at $\sigma^2 = 1$ ($\mathcal{H}_0$) and ≈ 45 samples at $\sigma^2 = 1.1$ ($\mathcal{H}_1$). For σ^2 further away from the boundaries, i.e. at $\sigma^2 = 0.6$ ($\mathcal{H}_0$) and $\sigma^2 = 1.8$ ($\mathcal{H}_1$), the ASNs are approximately the same. This is because at high SNR the average increment of the log-likelihood ratio is much larger than the threshold differences between the tests. These results indicate that the performance of the B-SPRT-2, which offers a significant reduction in computational costs, is still comparable to that of the B-SPRT-1. Note that for the same reason as discussed before, the B-SPRT-2 has smaller ASNs than the MT-SPRT for low SNR even for the largest K (in this case $B_k = 5$), as can be extracted from Fig. 4.4 and 4.7. Again, the opposite is true for high SNR.

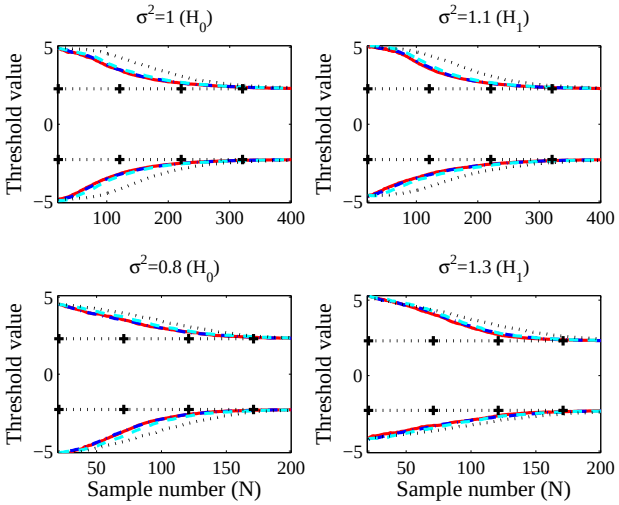


Figure 4.5: The thresholds for B-SPRT-1 (solid line) and B-SPRT-2 with $K = 4/B_k = 125$ (dashdot), $K = 20/B_k = 125$ (dashed), $K = 100/B_k = 5$ (dotted) as functions of sample size N. The constant thresholds A and B for the SPRT are also plotted (dotted with + sign). The top sub-figures are for σ^2 at the boundaries and the bottom sub-figures for σ^2 away from the boundaries. All the thresholds are for nominal error probabilities $\alpha = \beta = 0.1$.

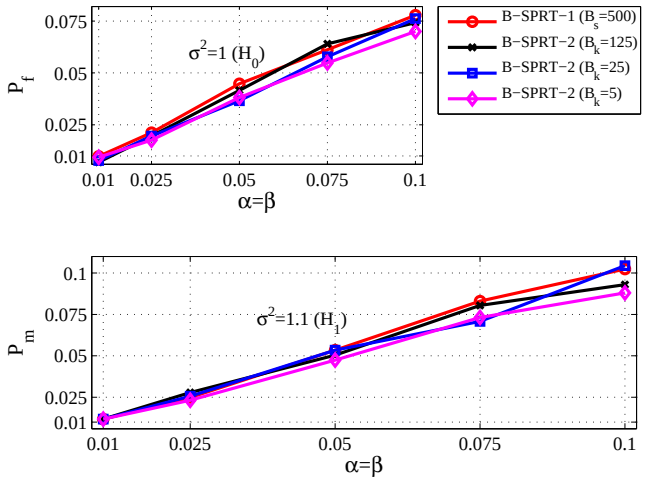


Figure 4.6: Performance comparisons in terms of P_f and P_m for B-SPRT-1 and B-SPRT-2 with various values of B_k at $\sigma^2 = 1$ ($\mathcal{H}_0$) and $\sigma^2 = 1.1$ ($\mathcal{H}_1$).

Fig. 4.8 and 4.9 show the relative efficiencies of the B-SPRT-1 and B-SPRT-2 with $B_k = 5$ for various values of σ^2 under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Here, the relative efficiency is defined as the ratio between the ASN of the B-SPRT and the minimum sample size ($N_{\min}$) required by the FSND to fulfill the same sensing objective, i.e. $P_f \leq \alpha$ and $P_m \leq \beta$. Table 4.1 shows the values of $N_{\min}$ for different nominal probabilities

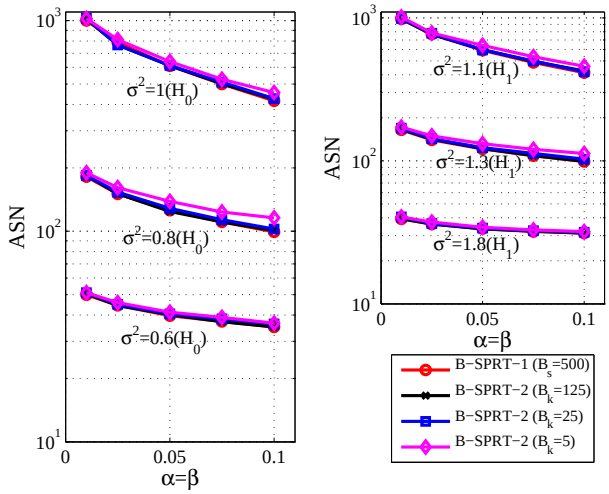


Figure 4.7: Performance comparisons on the ASN for B-SPRT-1 and B-SPRT-2 with various values of B_k and σ^2. The left sub-figure is under $\mathcal{H}_0$ and the right under $\mathcal{H}_1$.

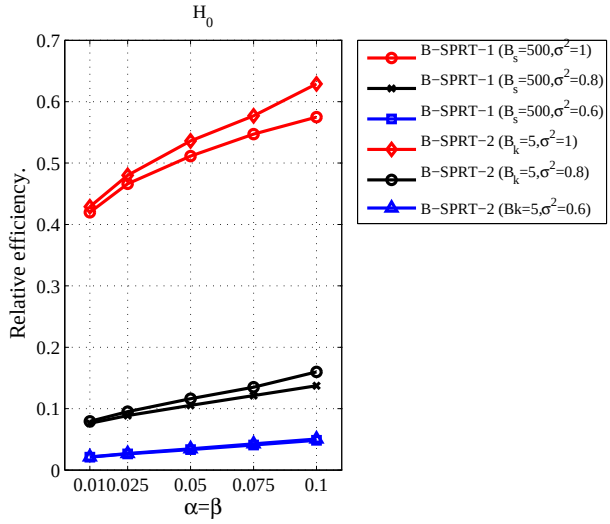


Figure 4.8: Relative efficiencies of B-SPRT-1 and B-SPRT-2 with $B_k = 5$ under $\mathcal{H}_0$ for various values of σ^2.

$\alpha = \beta$. In general, Fig. 4.8 and 4.9 show the superiority of the B-SPRT over the fixed sample size detectors for all cases considered.

The increase in computational cost due to the bootstrap in sequential detection is inevitable. However, we provide a solution to reduce the cost significantly. In addition, in an era of exponentially declining computational costs, bootstrap-based methods, such as the one presented, are becoming a bargain and more attractive. Based on the results presented in the section, the proposed method is a promising technique. It is

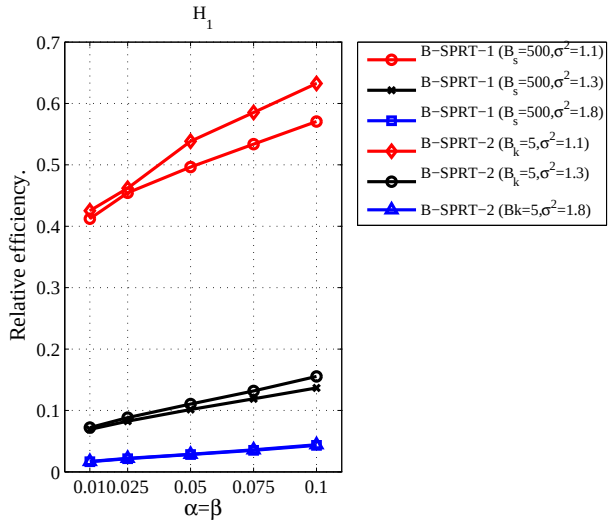


Figure 4.9: Relative efficiencies of B-SPRT-1 and B-SPRT-2 with $B_k = 5$ under $\mathcal{H}_1$ for various values of σ^2.

suitable for situations where time is critical, such as in spectrum sensing for cognitive radio, and offers a viable solution to the problem of sensing-throughput trade-off [43].

4.5 Conclusion

In this chapter, we present the implementation of the parametric bootstrap in sequential detection for composite hypothesis testing. The bootstrap is used to update the thresholds in adaptive manner according to the current observations. It improves the probabilities of false alarm and miss detection of the generalized sequential probability ratio test by minimizing decision errors due to erroneous estimators. Simulation results have shown that the proposed method has smaller ASNs than the conventional, meanwhile their actual probabilities of false alarm and miss detection are comparable. In addition, we propose a mechanism to reduce computational costs incurred by the bootstrap. The results show that the reduction in computational cost does not significantly affect the performance of the boostrap based sequential probability ratio test, in terms of the ASN. The margin is even unnoticeable for high SNR. The superiority of the proposed method to the fixed sample size test is also shown. Hence, it is an attractive candidate to be implemented in spectrum sensing for cognitive radio where the delay constraint is a critical issue.

4.6 Appendix

4.6.1 Clarification of expressions (4.7) and (4.8)

Let the received signal $\mathbf{x}[n]$ under hypothesis $\mathcal{H}_0$ admit the distribution $f_0(\mathbf{x}[n];\boldsymbol{\theta}_0)$. Assume the minimum Kullback-Leibler distance between $f_0(\mathbf{x}[n];\boldsymbol{\theta}_0)$ and $f_1(\mathbf{x}[n];\check{\boldsymbol{\theta}}_1)$, $\check{\boldsymbol{\theta}}_1 \in \boldsymbol{\Theta}_1$, is achieved when $\check{\boldsymbol{\theta}}_1 = \bar{\boldsymbol{\theta}}_1$. The expression (4.4) for the maximum likelihood estimators of $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ under $\mathcal{H}_0$ can be rewritten as

$$\hat{\boldsymbol{\theta}}_0^{(N)} = \underset{\check{\boldsymbol{\theta}}_0 \in \boldsymbol{\Theta}_0}{\arg\max} \frac{1}{N} \sum_{n=1}^{N} \ln\left(\frac{f_0(\mathbf{x}[n];\check{\boldsymbol{\theta}}_0)}{f_0(\mathbf{x}[n];\boldsymbol{\theta}_0)}\right) \tag{4.39a}$$

$$\hat{\boldsymbol{\theta}}_1^{(N)} = \underset{\check{\boldsymbol{\theta}}_1 \in \boldsymbol{\Theta}_1}{\arg\max} \frac{1}{N} \sum_{n=1}^{N} \ln\left(\frac{f_1(\mathbf{x}[n];\check{\boldsymbol{\theta}}_1)}{f_1(\mathbf{x}[n];\bar{\boldsymbol{\theta}}_1)}\right). \tag{4.39b}$$

Then, by the weak law of large numbers [98], as $N \to \infty$

$$\frac{1}{N} \sum_{n=1}^{N} \ln\left(\frac{f_0(\mathbf{x}[n];\check{\boldsymbol{\theta}}_0)}{f_0(\mathbf{x}[n];\boldsymbol{\theta}_0)}\right) \xrightarrow{\mathrm{P}} -KL\left(f_0(\mathbf{x}[n];\boldsymbol{\theta}_0)||f_0(\mathbf{x}[n];\check{\boldsymbol{\theta}}_0)\right) \tag{4.40}$$

for the term in equation (4.39a), and

$$\begin{aligned}\frac{1}{N} \sum_{n=1}^{N} \ln\left(\frac{f_1(\mathbf{x}[n];\check{\boldsymbol{\theta}}_1)}{f_1(\mathbf{x}[n];\bar{\boldsymbol{\theta}}_1)}\right) \xrightarrow{\mathrm{P}} \big\{&KL\left(f_0(\mathbf{x}[n];\boldsymbol{\theta}_0)||f_1(\mathbf{x}[n];\bar{\boldsymbol{\theta}}_1)\right) \\ &-KL\left(f_0(\mathbf{x}[n];\boldsymbol{\theta}_0)||f_1(\mathbf{x}[n];\check{\boldsymbol{\theta}}_1)\right)\big\}\end{aligned} \tag{4.41}$$

for the term in equation (4.39b). Therefore, we have

$$\hat{\boldsymbol{\theta}}_0^{(N)} \xrightarrow{\mathrm{P}} \underset{\check{\boldsymbol{\theta}}_0 \in \boldsymbol{\Theta}_0}{\arg\min}\ KL\left(f_0(\mathbf{x}[n];\boldsymbol{\theta}_0)||f_0(\mathbf{x}[n];\check{\boldsymbol{\theta}}_0)\right) = \boldsymbol{\theta}_0, \tag{4.42a}$$

$$\hat{\boldsymbol{\theta}}_1^{(N)} \xrightarrow{\mathrm{P}} \underset{\check{\boldsymbol{\theta}}_1 \in \boldsymbol{\Theta}_1}{\arg\min}\ KL\left(f_0(\mathbf{x}[n];\boldsymbol{\theta}_0)||f_1(\mathbf{x}[n];\check{\boldsymbol{\theta}}_1)\right) = \bar{\boldsymbol{\theta}}_1, \tag{4.42b}$$

from combining (4.39a) with (4.40) and (4.39b) with (4.41), respectively. By similar arguments, the MLEs under hypthesis $\mathcal{H}_1$, corresponding to the distribution $f_1(\mathbf{x}[n];\boldsymbol{\theta}_1)$, it can be shown that

$$\hat{\boldsymbol{\theta}}_0^{(N)} \xrightarrow{\mathrm{P}} \underset{\check{\boldsymbol{\theta}}_0 \in \boldsymbol{\Theta}_0}{\arg\min}\ KL\left(f_1(\mathbf{x}[n];\boldsymbol{\theta}_1)||f_0(\mathbf{x}[n];\check{\boldsymbol{\theta}}_0)\right) = \bar{\boldsymbol{\theta}}_0, \tag{4.43a}$$

$$\hat{\boldsymbol{\theta}}_1^{(N)} \xrightarrow{\mathrm{P}} \underset{\check{\boldsymbol{\theta}}_1 \in \boldsymbol{\Theta}_1}{\arg\min}\ KL\left(f_1(\mathbf{x}[n];\boldsymbol{\theta}_1)||f_1(\mathbf{x}[n];\check{\boldsymbol{\theta}}_1)\right) = \boldsymbol{\theta}_1. \tag{4.43b}$$

To put (4.42a) to (4.43b) in a compact form: When the $\hat{\boldsymbol{\theta}}_i^{(N)}$ is evaluated under $\mathcal{H}_j$ with $\mathbf{x}[n]$ distributed according to $f_j(\mathbf{x}[n];\boldsymbol{\theta}_j)$, then it converges in probability to

$$\bar{\boldsymbol{\theta}}_i = \underset{\check{\boldsymbol{\theta}}_i \in \boldsymbol{\Theta}_i}{\arg\min}\ KL(f_j(\mathbf{x}[n];\boldsymbol{\theta}_j)||f_i(\mathbf{x}[n];\check{\boldsymbol{\theta}}_i)). \tag{4.44}$$

for $i, j = 0, 1$. This proves (4.7) and (4.8).

4.6.2 Proof of Theorem 1

For $N \to \infty$, the estimated parameters converge according to (4.8). Therefore, the sequence of the log-likelihood ratio $\hat{Z}_N$ in (4.11) under each hypothesis, as $N \to \infty$ can be approximated by

$$\mathcal{H}_0 : \hat{Z}_N \approx \hat{Z}_{N-1} + \bar{z}_N^0, \quad \mathcal{H}_1 : \hat{Z}_N \approx \hat{Z}_{N-1} + \bar{z}_N^1 \tag{4.45}$$

where $\bar{z}_N^i$ denotes the increment under hypothesis $\mathcal{H}_i$, i.e.,

$$\mathcal{H}_0 : \bar{z}_N^0 = \ln\left(\frac{f_1(\mathbf{x}[N];\bar{\boldsymbol{\theta}}_1)}{f_0(\mathbf{x}[N];\boldsymbol{\theta}_0)}\right), \tag{4.46a}$$

$$\mathcal{H}_1 : \bar{z}_N^1 = \ln\left(\frac{f_1(\mathbf{x}[N];\boldsymbol{\theta}_1)}{f_0(\mathbf{x}[N];\bar{\boldsymbol{\theta}}_0)}\right). \tag{4.46b}$$

Let us drop the index N to indicate that the random variable $\bar{z}_N^i$ is distributed independently of N. The expected values of the increment $\bar{z}^0$ and $\bar{z}^1$ is given by

$$\mathcal{H}_0 : E_0\left[\bar{z}^0\right] = -KL(f_0(\mathbf{x};\boldsymbol{\theta}_0)||f_1(\mathbf{x};\bar{\boldsymbol{\theta}}_1)), \tag{4.47a}$$

$$\mathcal{H}_1 : E_1\left[\bar{z}^1\right] = KL(f_1(\mathbf{x};\boldsymbol{\theta}_1)||f_0(\mathbf{x};\bar{\boldsymbol{\theta}}_0)), \tag{4.47b}$$

are finite based on the assumption AS2. Up to this point, the requirements for a finite stopping time for constant threholds stated by Lemma 1 in [66], namely

$$\mathcal{H}_0 : E_0\left[\left(\bar{z}^0\right)^2\right] \neq 0, \quad \mathcal{H}_1 : E_1\left[\left(\bar{z}^1\right)^2\right] \neq 0 \tag{4.48}$$

are fulfilled, i.e., implied by (4.47). However, this fact does not immediately guarantee that the test will terminate with a finite stopping time N_s since the thresholds now are functions of the sample size N.

Recall that the B-SPRT observes data up to N_I and the sequential test is performed afterwards. First, let us consider B-SPRT-2 with $K = \infty$ ($B_k = 0$) and refer to this test as B-SPRT-CT. Choosing $B_k = 0$ means observing the data up to N_I and then constructing the two thresholds using the parametric bootstrap once and for all.

The sequential test is then performed using the resulting thresholds for all N with no update. As a result, the B-SPRT-CT has looser stopping conditions than the B-SPRT-1 and the B-SPRT-2 with $1 < K < \infty$. Note that the B-SPRT-1 is B-SPRT-2 with $K = 1$. Hence, when the stopping time of the B-SPRT-CT is guaranteed to be finite, the stopping times of the B-SPRT-1 and B-SPRT-2 are guaranteed to be finite as well. In this respect, the only task left is to show that the constant thresholds constructed by B-SPRT-CT are bounded for large N_I.

Applying Markov's inequality to the random variable $|\Delta Z_N^i|, i = 0, 1$, we obtain

$$\mathcal{H}_0 : P_0\left(|\Delta Z_N^0| \geq c_0\right) \leq \frac{\sqrt{E_0\left[|\Delta Z_N^0|^2\right]}}{c_0} \tag{4.49a}$$

$$\mathcal{H}_1 : P_1\left(|\Delta Z_N^1| \geq c_1\right) \leq \frac{\sqrt{E_1\left[|\Delta Z_N^1|^2\right]}}{c_1}. \tag{4.49b}$$

where $c_0, c_1 > 0$. For the left hand side of the inequalities, we put the pre-specified values ϵ_0 and ϵ_1 according to (4.19). At stage N_I, we have

$$\mathcal{H}_0 : c_0 \leq \frac{\sqrt{E_0\left[|\Delta Z_{N_I}^0|^2\right]}}{\epsilon_0}, \tag{4.50a}$$

$$\mathcal{H}_1 : c_1 \leq \frac{\sqrt{E_1\left[|\Delta Z_{N_I}^1|^2\right]}}{\epsilon_1}. \tag{4.50b}$$

Therefore, to have bounded thresholds for the B-SPRT-CT as $N_I \to \infty$, i.e., bounded c_0 and c_1, the second moments $E_i\left[|\Delta Z_{N_I}^0|^2\right],\ i = 0, 1$, have to be bounded. This has already been shown by the authors of [54] so that we have

$$E_i\left[|\Delta Z_{N_I}^i|^2\right] = O(1), \quad i = 0, 1. \tag{4.51}$$

The fact that the above arguments are given by referring to the population distributions instead of the bootstrap distributions, does not invalidate the arguments if we refer to the bootstrap domain. It is only for the sake of convenience.

4.6.3 The derivation of (4.38)

Suppose that we have a simple hypothesis against a simple alternative described by the model in (2.9). Under $\mathcal{H}_0$, the test statistic (2.10) can be rewritten as

$$Z_N = \sum_{n=1}^{N} \left[\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right) \mathbf{y}[n] + \log \frac{\sigma_0^2}{\sigma_1^2}\right], \tag{4.52}$$

where $\mathbf{y}[n] = \left(\frac{|\mathbf{x}[n]|}{\sigma_0^2}\right)^2$ is exponentially distributed with the expected value $\mu = 1$ and variance $\sigma_e^2 = 1$. Let us rewrite (4.52) further to be

$$Z_N = \sqrt{N}\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right)\left\{\sqrt{N}\left[\sum_{n=1}^{N}\frac{1}{N}\mathbf{y}[n] - 1\right] + \sqrt{N}\right\} + \sum_{n=1}^{N}\log\frac{\sigma_0^2}{\sigma_1^2}. \tag{4.53}$$

Using the central limit theorem, as $N \to \infty$, the term

$$\sqrt{N}\left[\sum_{n=1}^{N}\frac{1}{N}\mathbf{y}[n] - 1\right] \xrightarrow{\mathrm{D}} \mathbf{y} \sim \mathcal{N}(0,1), \tag{4.54}$$

where $\xrightarrow{\mathrm{D}}$ denotes convergence in distribution. Therefore, when $N \to \infty$

$$Z_N \dot{\sim} \mathcal{N}\left(N\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right) + \sum_{n=1}^{N}\log\frac{\sigma_0^2}{\sigma_1^2}, N\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right)^2\right), \tag{4.55}$$

where $\dot{\sim}$ denotes approximately distributed. The probability of false alarm can now be determined from

$$\begin{aligned} P_f &= P_0(Z_N > \tau') \\ &\approx \mathcal{Q}\left(\frac{\tau' - N\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right) - N\log\frac{\sigma_0^2}{\sigma_1^2}}{\sqrt{N\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right)^2}}\right), \end{aligned} \tag{4.56}$$

where $\mathcal{Q}(\cdot)$ is the Q-function of the standard normal distribution. The threshold τ' is then

$$\tau' \approx \mathcal{Q}^{-1}(P_f)\sqrt{N\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right)^2} + N\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right) + N\log\frac{\sigma_0^2}{\sigma_1^2}. \tag{4.57}$$

Using similar arguments, under $\mathcal{H}_1$

$$Z_N \dot{\sim} \mathcal{N}\left(N\left(\frac{\sigma_1^2}{\sigma_0^2} - 1\right) + \sum_{n=1}^{N}\log\frac{\sigma_0^2}{\sigma_1^2}, N\left(\frac{\sigma_1^2}{\sigma_0^2} - 1\right)^2\right). \tag{4.58}$$

Therefore, the probability of miss detection is

$$\begin{aligned} P_m &= P_1(\bar{Z}_N < \tau') \\ &\approx \mathcal{Q}\left(\frac{N\left(\frac{\sigma_1^2}{\sigma_0^2} - 1\right) + N\log\frac{\sigma_0^2}{\sigma_1^2} - \tau'}{\sqrt{N\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right)^2}}\right). \end{aligned} \tag{4.59}$$

Substituting (4.57) into (4.59) and after some arrangements, the minimum sample size $N_{\min}$ required to fulfill $P_f = \alpha$ and $P_m = \beta$ is

$$N_{\min} \approx \left(\frac{\mathcal{Q}^{-1}(\beta)\left(\frac{\sigma_1^2}{\sigma_0^2} - 1\right) + \mathcal{Q}^{-1}(\alpha)\left(1 - \frac{\sigma_0^2}{\sigma_1^2}\right)}{\frac{\sigma_1^2}{\sigma_0^2} + \frac{\sigma_0^2}{\sigma_1^2} - 2} \right). \tag{4.60}$$

Chapter 5

Miscellaneous works

In this chapter, three other selected works that have been done during the PhD project are presented. All the works do not use bootstrap as an approach for detection. In Section 5.1, we propose a sequential multiple testing procedure. This work is motivated by an application in multiband spectrum sensing for cognitive radio, in which a primary user or a cognitive radio user can use several bands at a time. The proposed procedure simultaneously controls the false alarm and miss detection rate not only for a single band, but also for the system (familywise). The common method to individually tests the hypotheses fails to achieve this. In Section 5.2, we provide a scheme to improve spectrum sensing performances at low SNR using locally optimum detection. It is derived based on the assumption that the underlying noise is Student's t-distributed. The resulting scheme outperforms energy detection in all scenarios. In addition, it is more robust to outliers than the conventional scheme which is derived based on Gaussian noise assumption. Next, we describe collaborative spectrum sensing using hard decision in combination with quality information in Section 5.3. This scheme necessitates to determine three thresholds at the local detector. One threshold is determined by a pre-specified local probability of false alarm, and the other two thresholds are found using distance measure criteria. The results show that the proposed scheme outperforms conventional hard decision combining. We also evaluate the scheme under imperfect reporting channels, modeled as binary symmetric channels with varying binary error probabilities.

5.1 Multiple testing for sequential probability ratio tests

Multiband spectrum sensing is less well studied than single band (narrow-band) spectrum sensing [46]. To mention a few references, multiband spectrum sensing can be found in [35, 92, 102]. In Section 3.4, we discuss the implementation of MTPs for multiband spectrum sensing. The motivation behind the use of MTPs in multiband spectrum sensing is the capability of MTPs to provide control over decision errors not only per single band but also for the overall system. The results show a better tradeoff between type I errors and type II errors at the system level, compared to individual

testing procedures. In Section 3.4 and references therein, MTPs are used in the context of a fixed sample size. In this part, we propose an MTP for sequential detection, more precisely for sequential probability ratio tests (SPRT), where the sample size is random. Again we make use of the additional degree of freedom in sequential testing to control not only the FWE, but also the FWM. To the best of our knowledge, the study of MTP for the SPRT is scarce. Note that with an MTP for the SPRT we do not mean sequential multihypothesis testing, such as in [103–105] and the like. At least, three points mark the differences between the implementation of an MTP in fixed sample size cases (FSN) and in the SPRT we propose here. First, unlike FSN, the p-value [62] as the level of evidence is difficult to get in the SPRT since the number of samples N changes. However, we can use random stopping times (random sample sizes) instead. The smallest stopping time can be considered to be equivalent to the smallest p-value. Simply said, an MTP in FSN works based on the ordered p-values, while in the SPRT it works based on the ordered stopping times. Second, ordering p-values in FSN is done after all tests finished, while ordering stopping times in the SPRT is done while the SPRTs are still in progress (successively one test after another, depending on the one that finished earlier). Third, the main objective in MTP for FSN is to maximize the power, while in the SPRT context, the aim is to minimize the ASN.

5.1.1 Formulation and method

The model that we use here is similar to the model in Section 3.4.2. We assume that the primary network operates over a wide frequency bandwidth which is divided into K nonoverlapping subbands, and a primary user can be assigned to use a number of subbands K_p simultaneously, where $1 \leq K_p \leq K$. The binary hypothesis testing problem for spectrum sensing of the subband k is

$$\begin{aligned} &\mathcal{H}_{k,0} : f_{k,0}(\mathbf{x}_k[n]; \boldsymbol{\theta}_{k,0}) \\ &\mathcal{H}_{k,1} : f_{k,1}(\mathbf{x}_k[n]; \boldsymbol{\theta}_{k,1}), \quad k = 1, \ldots, K. \end{aligned} \tag{5.1}$$

where $\mathbf{x}_k[n]$ denotes a scalar or a vector observation, and $f_{k,i}(\cdot)$ is the density function of the subband k under hypothesis $\mathcal{H}_i,\ i = 0, 1$. $\boldsymbol{\theta}_{k,0}$ and $\boldsymbol{\theta}_{k,1}$ are the parameters for the subband k which could be scalars or vectors, under the respective hypotheses. Here, we assume that the observations are identically independent distributed (i.i.d.) within subbands and also independent accross subbands. Suppose that K_0 out of K are available for cognitive access within particular time intervals. Moreover, we consider the case where the CR network also supports some CR users to use several unoccupied subbands simultaneously. The number of subbands $K_c,\ 1 \leq K_c \leq K_0$, assigned to a specific CR user is, say, based on priority and currently active CR users.

To accommodate the use of multiple subbands by the primary and the CR users, we need an overall view on the performance of spectrum sensing, e.g., false alarm and miss detection not only per subband, but for $1 \leq K_c \leq K_0$ and $1 \leq K_p \leq (K - K_0)$, respectively. Note that K_0 could span from 0 (all subbands occupied) to K (all subbands unoccupied).

For the sake of clarity, we give an illustration. Suppose that each subband is individually tested with nominal probabilities of false alarm $\alpha_k = \alpha$ and miss detection $\beta_k = \beta,\ k = 1, \ldots, K$. Let us assume that there are $K_0 = 5$ unoccupied subbands and two active CR users, that is one with high priority (can use $1 \leq K_c \leq 4$ subbands) and the other with low priority ($K_c = 1$). The low priority user could have a subband to use, while the high priority user might defer to use all four unoccupied subbands simultaneously, and thus uses lower K_c, due to the higher false alarm rate caused by the multiplicity effect (the actual probability of false alarm for the four subbands $= 1-(1-\alpha)^4$). With the same argument, any primary user that uses a higher number of subbands K_p simultaneously will experience a higher aggregate interference level due to a higher probability of miss detection. In this case, an MTP should be implemented in multiband spectrum sensing to jointly detect the subbands and to provide control over the decision errors at the system level.

Some performance measures for multiband spectrum sensing have been mentioned in Section 3.4.1. However, since we do sequential detection, other important measures are the average sample number over all subbands that are under $\mathcal{H}_0$, denoted as ASN_0, and over all subbands that are under $\mathcal{H}_1$, denoted as ASN_1. They are defined formally as

$$\mathrm{ASN}_0 = E_0\left[\frac{\sum_{i=1}^{K_0} N_{s,i}}{K_0}\right],\ \mathrm{ASN}_1 = E_1\left[\frac{\sum_{l=1}^{K-K_0} N_{s,l}}{K-K_0}\right], \tag{5.2}$$

where $N_{s,i}$ denotes the stopping time of the subband S_i. Note that a false discovery rate controlling procedure is now commonly used in the fixed sample size case to get more powerful tests, as we explain in Section 3.4.2.1. However, the use of the FDR in MTP for the SPRT is unclear due to the capability of the SPRT to simultaneously control the probabilities of false alarm and miss detection.

Let $\mathbf{x}_{k,N} = (\mathbf{x}_k[1]\ \mathbf{x}_k[2]\ \cdots\ \mathbf{x}_k[N])$ be a sequence of i.i.d. observations of a signal recorded up to the sample N at subband S_k. Here, $\mathbf{x}_k[n]$ is assumed to admit the distribution described by the density function in (5.1) under each hypothesis. The sequential probability ratio test (SPRT), according to (2.18), is

$$\bar{Z}_{k,N} = \begin{cases} \geq A, & \text{accept } \mathcal{H}_{k,1} \\ \leq B, & \text{accept } \mathcal{H}_{k,0}, \\ A < \bar{Z}_{k,N} < B, & N \leftarrow N+1. \end{cases} \tag{5.3}$$

where $Z_{k,N} = \sum_{n=1}^{N} \log \dfrac{f_{k,1}(\mathbf{x}_k[n]; \boldsymbol{\theta}_{k,1})}{f_{k,0}(\mathbf{x}_k[n]; \boldsymbol{\theta}_{k,0})}$, $k = 1, 2, \ldots, K$. For the implementation, we use the thresholds [45]

$$A \approx \log \frac{1}{\alpha'}, \quad B \approx \log \beta', \tag{5.4}$$

to have actual probabilities of false alarm $P_{f,k} \leq \alpha'$ and miss detection $P_{m,k} \leq \beta'$ in the subband S_k, where α' and β' denote the respective nominal values of the probabilities of false alarm and miss detection per subband basis. Note that we use the same thresholds for all subbands. We further assume that the SPRTs start simultaneously in all K subbands whenever the sensing period starts, and the SPRT running in each subband fulfills the condition of having a finite random stopping time $N_{s,k},\ k = 1, \ldots, K$ [66]. The objective is to have

$$\text{FWE} \leq \alpha, \quad \text{FWM} \leq \beta, \tag{5.5}$$

with ASN_0 and ASN_1 as small as possible, by jointly testing all K subbands.

Simple Bonferroni procedure (SBF) for the SPRT. For K hypotheses (subbands), the simplest way of conducting an MTP is to follow the simple Bonferroni procedure [81], like in the fixed sample size case. More precisely, we test each subband individually at the level $\alpha_k = \alpha/K$ and $\beta_k = \beta/K,\ \forall k \in \{1, \ldots, K\}$ and set the thresholds accordingly using (5.4). Thus, it will guarantee to fulfill the sensing objective (5.5). However, this approach is too conservative in protecting decision errors and hence it results in large ASN_0 and ASN_1 for the SPRT. To solve this issue, we propose a stepwise procedure that will be explained in the sequel.

Stepwise procedure (SWP) for the SPRT. Inspired by the work of Holm [82] that uses the ordered p-values for the fixed sample size case, the SWP is based on the ordered stopping times $N_{s,(1)} \leq N_{s,(2)} \leq \cdots \leq N_{s,(K)}$ corresponding to subbands $S_{(1)}, S_{(2)}, \ldots, S_{(K)}$. The procedure in the SPRT is more involved since we have two thresholds to consider and we perform ordering while the sample size N is increasing. In principle, we start the SPRT in each subband with the largest value of the threshold A and the lowest value of the threshold B. The largest A and the lowest B depend on the nominal values of the FWE and FWM and the number of subbands K. Whenever one or more SPRTs stop and favor $\mathcal{H}_1$ (the subbands are declared occupied), update A with a smaller value to conduct the SPRT in the other subbands, and simultaneously, whenever one or more SPRTs stop and favor $\mathcal{H}_0$ (the subbands are declared unoccupied), update B with a larger value to conduct the SPRT in the other subbands. In addition, after any SPRT stops we proceed with a procedure to find the other SPRTs

that have crossed the respective smaller value of A or larger value of B in the past. If any, declare the respective subbands as occupied or unoccupied and repeat the procedure with the next smaller A or larger B. Otherwise, the SPRTs proceed for the rest of subbands. The whole process continues until all subbands have been declared occupied or unoccupied.

Suppose that the nominal values of the FWE and FWM are α and β, respectively. The detail of the SWP is as follows:

Step 1). Initialize the sample size $N = 0$, variables $I_1 = 0$ and $I_0 = 0$, and a set of terminated subbands $\mathbf{S}_T = \emptyset$. The two sets of thresholds

$$\begin{aligned}\mathbf{\Lambda}_A =&\{\log(1/\alpha), \log(2/\alpha), \ldots, \log((K - I_1)/\alpha)\}\\ \mathbf{\Lambda}_B =&\{\log(\beta), \log(\beta/2), \ldots, \log(\beta/(K - I_0))\},\end{aligned} \tag{5.6}$$

correspond to the FWE and FWM of Benferroni's method for the number of subbands $1, 2, \ldots, K$. The size of the sets $\mathbf{\Lambda}_A$ and $\mathbf{\Lambda}_B$ will be shrinking while in progress, which depends on the variables I_1 and I_0. I_1 indicates the number of subbands which have been declared occupied and I_0 indicates the number of subbands which have been declared unoccupied.

Step 2). For all subbands $S_k \notin \mathbf{S}_T$, take a sample $N \leftarrow N + 1$, calculate and then compare $Z_{k,N}$ according to (5.3) where $A = \max\{\lambda_A : \lambda_A \in \mathbf{\Lambda}_A\}$ and $B = \min\{\lambda_B : \lambda_B \in \mathbf{\Lambda}_B\}$. If one of the two thresholds is crossed in any subband, continue with **Step 3**, otherwise repeat **Step 2**. Note that, at each stage N, $Z_{k,N}$ is also inspected by the processor in the subband k for whether it has crossed the upper thresholds $\{\log(1/\alpha), \ldots, \log((K - I_1 - 1)/\alpha)\}$ or the lower thresholds $\{\log(\beta), \ldots, \log(\beta/(K - I_0 - 1))\}$. The results are stored, say, in a memory $\mathbf{u}_k = \left[u_{k,1}, u_{k,2}, \ldots, u_{k,(K-1)}\right]$, where $u_{k,l}$ is set to 1 if $Z_{k,N} \geq \log(l/\alpha)$ or set to 0 if $Z_{k,N} \leq \log(\beta/l)$, otherwise keep $u_{k,l}$ empty.

Step 3). If the SPRTs in a subset of subbands $\mathbf{S}_1 = \{S_{(1)}, \ldots, S_{(L_1)}\}$ stop, due to the respective $Z_{(k),N} \geq A$, the subbands $S_{(1)}, \ldots, S_{(L_1)}$ are declared occupied, and the variable $I_1 \leftarrow I_1 + L_1$ is updated. Simultaneously, if the SPRTs in a subset of subbands $\mathbf{S}_0 = \{S_{1^*}, \ldots, S_{L_0^*}\}$, where $\mathbf{S}_0 \cap \mathbf{S}_1 = \emptyset$, stop, due to the respective $Z_{k^*,N} \leq B$, the subbands $S_{1^*}, \ldots, S_{L_0^*}$ are declared unoccupied, and the variable $I_0 \leftarrow I_1 + L_0$ is updated. The set of terminated subbands should also be updated, i.e., $\mathbf{S}_T = \{S_T, S_1, S_0\}$.[1] If not all subbands have been declared occupied or unoccupied, the decisions for the rest

[1]Note that the possibility to have the total number of terminations $L_1 + L_0 > 1$ is small when the SNRs (more precisely, the increments) are small, since under this condition, the variance of the stopping time at each subband is large and thus the SPRTs most likely will not stop at the same time (either $\mathbf{S}_0$ or $\mathbf{S}_1$ is most probably an empty set). However, when the SNRs are high, the possibility is larger, since the variance of stopping time in each subband is small.

of the subbands depend on the results of the following procedure, which inspects the memory $\mathbf{u}_k$, $\forall S_k \notin \mathbf{S}_T$. Initially, set index variables $i_0 = 1$ and $i_1 = 1$,

(a) If none of $S_k \notin \mathbf{S}_T$ has $u_{k,(K-I_1-i_1)} = 1$ or $u_{k,(K-I_1-i_1)} = 0$, update the sets $\mathbf{\Lambda}_A$ and $\mathbf{\Lambda}_B$ in (5.6) and repeat **Step 2**.

(b) Otherwise, if $u_{k,(K-I_1-i_1)} = 1$ in a subset of subbands $\mathbf{S}'_1 = \{S_{1'}, \ldots, S_{l'_1}\}$, the subbands $S_{1'}, \ldots, S_{l'_1}$ are declared occupied, update the variable $i_1 \leftarrow i_1 + l_1$. Similarly, if $u_{k,(K-I_1-i_0)} = 0$ in a subset of subbands $\mathbf{S}'_0 = \{S'_{1^*}, \ldots, S'_{l_{0^*}}\}$ where $\mathbf{S}'_0 \cap \mathbf{S}'_1 = \emptyset$, the subbands $S'_{1^*}, \ldots, S'_{l_0^*}$ are declared unoccupied, and then update the variable $i_0 \leftarrow i_0 + l_0$. The set of terminated subbands should also be updated, i.e., $\mathbf{S}_T = \{\mathbf{S}_T, \mathbf{S}'_1, \mathbf{S}'_0\}$. As long as $\mathbf{S}'_0 \cup \mathbf{S}'_1 \neq \emptyset$, continue from the beginning of (b). Otherwise, if not all subbands have been declared occupied or unoccupied, update the variables $I_1 \leftarrow I_1 + i_1 - 1$ and $I_0 \leftarrow I_0 + i_0 - 1$ and accordingly the sets $\mathbf{\Lambda}_A$ and $\mathbf{\Lambda}_B$ in (5.6), then repeat **Step 2**.

Note that we only update the upper threshold A (not A and B) whenever one or more SPRTs stop and favor $\mathcal{H}_1$. The opposite holds, when favoring $\mathcal{H}_0$. This can be explained as follows. Suppose that an SPRT at the subband S_k stops and favors $\mathcal{H}_{k,1}$ with the thresholds $A = \log(K/\alpha)$ and $B = \log(\beta/K)$. In this case, when $\mathcal{H}_{k,0}$ has been rejected, using the nominal value of the probability of false alarm $\alpha' = \alpha/K$ ($A = \log(K/\alpha)$), we should believe that $\mathcal{H}_{k,0}$ is false ($\mathcal{H}_{k,1}$ is true). Therefore, there are only $K-1$ null hypotheses which might be still true, implying the critical value now to be $\alpha' = \alpha/(K-1)$ (update the upper threshold to $A = \log((K-1)/\alpha)$). However, for the lower threshold B, we should believe that there are still K alternative hypotheses which might be true (including the one that has been declared true), since we have no evidence that any of $\mathcal{H}_{k,1}$ has been rejected. This implies the nominal value for the probability of miss detection is still $\beta' = \beta/K$ (the lower threshold is maintained at $B = \log(\beta/K)$). The same argument applies when an SPRT at the subband S_k stops and favors $\mathcal{H}_{k,0}$. If we were to update both thresholds each time an SPRT stops, regardless of which hypothesis is rejected, then the objective (5.5) will not be achieved.

5.1.2 Simulation results

As an example, we assume that a CR user receives complex Gaussian signals in each subband, i.e

$$\begin{aligned} &\mathcal{H}_{k,0} : \mathbf{x}_k[n] \sim \mathcal{CN}\left(0, \sigma^2_{k,0}\right), \\ &\mathcal{H}_{k,1} : \mathbf{x}_k[n] \sim \mathcal{CN}\left(0, \sigma^2_{k,1}\right), \quad k = 1, 2, \ldots, K. \end{aligned} \tag{5.7}$$

The log-likelihood ratio $Z_{k,N}$ can then be calculated from (5.3) to perform spectrum sensing in K subbands. For all simulations we assume that the noise power $\sigma^2_{k,0} = 1,\ k = 1, \ldots, K$, where $\sigma^2_{k,1}$ depends on the SNR in each subband, which is defined as $\mathrm{SNR}_k = 10 \log_{10}((\sigma^2_{k,1} - \sigma^2_{k,0})/\sigma^2_{k,0})$. The nominal values of the FWE and FWM are set to $\alpha = \beta = 0.1$. All the results are generated using 10^4 Monte Carlo runs.

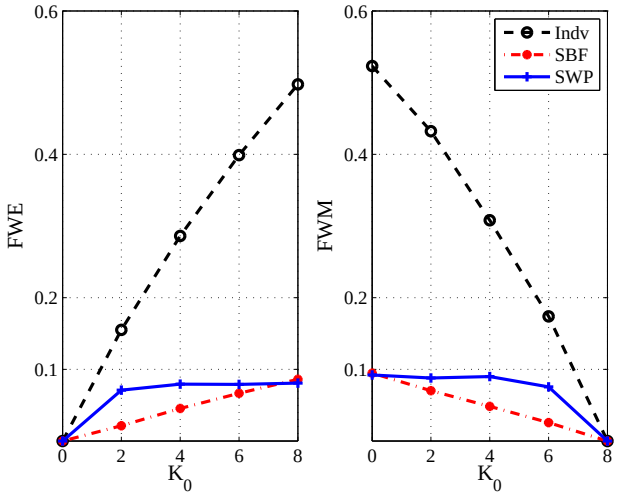


Figure 5.1: FWE and FWM vs. K_0 for individually testing without MTP (Indv), the SBF, and the proposed SWP.

Fig. 5.1 shows the FWE and FWM for the SBF and the SWP. The number of subbands that are jointly tested is $K = 8$ and the number of unoccupied subbands K_0 varies from 0 to 8. Note that K_0 unoccupied subbands were randomly selected from the K subbands in each Monte Carlo run. Here, the occupied subbands have the same SNR$= -10$ dB. The results when all subbands are individually tested without MTP at the nominal values $\alpha' = \beta' = 0.1$, are also shown. In general, Fig. 5.1 indicates that the SBF and SWP control the FWE and the FWM in the strong sense. More precisely, regardless of how many available subbands might be opportunistically used by a CR user, the probability of missing an opportunity to use the respective subbands is no larger than $\alpha = 0.1$, and regardless of how many subbands are used by a primary user, the probability of the respective primary user to receive interferences is no larger than $\beta = 0.1$. However, full protection resulting from the SBF is too restrictive, particularly for the FWE when K_0 is large and for the FWM when K_0 is small. Meanwhile, lack of multiplicity control by testing individually is too permissive, hence reducing the overall throughput of the CR network (due to higher FWE for higher numbers of unoccupied subbands) and increasing the interferences to the primary users (due to higher FWM for higher number of occupied subbands), especially for the primary users that use several subbands at a time. The proposed procedure SWP handles the problem appropriately.

Table 5.1: ASN_0 and ASN_1 for the SBF and the SWP as a function of K_0.

		K_0				
		0	2	4	6	8
ASN_0	SBF	-	980	977	977	978
	SWP	-	928	910	861	779
ASN_1	SBF	934	935	934	932	-
	SWP	743	820	864	888	-

It pulls the FWE and the FWM closer to the nominal values while still preserving the objective (5.5). As a result, the SWP has smaller ASNs than the SBF, as shown in Table 5.1. By using the SWP, the gain that we obtain on ASN_1 is larger when the subbands are busier (mostly occupied), and the gain is larger for ASN_0 when the subbands are sparser (mostly unoccupied).

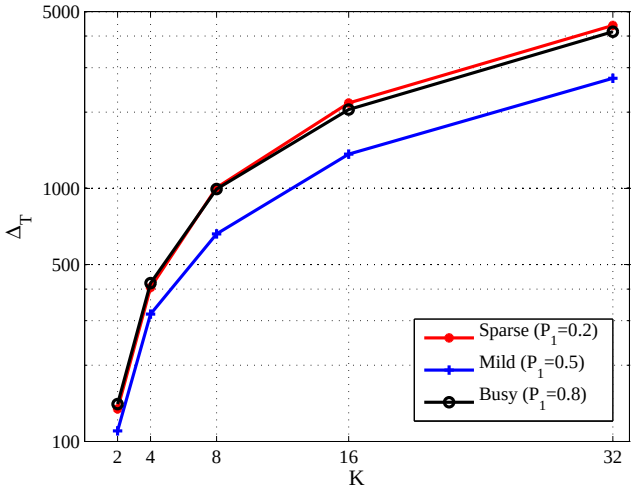


Figure 5.2: The gap Δ_T between the total ASN of the SBF and the SWP vs. K, when the subbands occupancies are busy, mild and sparse.

According to (5.2), ASN_0 and ASN_1 represent the average sample numbers per subband basis. The total average sample number can be defined as

$$\begin{aligned}\text{ASN}_T =& E\left[\sum_{i=1}^{K} N_{s,i}\right] = E\left[\sum_{i=1}^{K_0} N_{s,i} + \sum_{l=1}^{K-K_0} N_{s,l}\right] \\ \approx& K\left\{(1-P_1)\text{ASN}_0 + P_1\text{ASN}_1\right\}, \end{aligned} \tag{5.8}$$

where we have assumed that the probability of each subband being occupied by a primary user is equal, i.e., $P(\mathcal{H}_{k,1}) = P_1,\ k = 1, \ldots, K$. In Fig. 5.2, we plot the gap

$\Delta_T = \mathrm{ASN}_T(\mathrm{SBF}) - \mathrm{ASN}_T(\mathrm{SWP})$, between the total ASN of the SBF and the SWP against the number of subbands K, in which the occupied subbands have the same SNR$= -10$ dB. We evaluate the performance when the channel occupancy is considered to be busy ($P_1 = 0.8$), mildly busy ($P_1 = 0.5$) and sparse ($P_1 = 0.2$). It can be remarked that the gap is larger when either K subbands are busy or sparse. The reasoning is as follows. When the subbands are busy, the alternative hypotheses dominate the MTP. In this case, the probability that the SWP updates the upper threshold A down to the lowest value in each realization is high, and hence the probability that SPRTs stop with smaller sample sizes is also high. The same case applies when the subbands are sparse, namely the null hypotheses dominate the MTP. In this case, the probability that the SWP updates the lower threshold B up to the largest value in each realization is high, and hence the probability of SPRTs to stop with smaller sample sizes is also high. However, this does not apply when the subbands are mildly busy since, then the null hypotheses and the alternative hypotheses are competing. In this case, all SPRTs will mostly have stopped immediately after reaching the $[K/2]$ smallest of the thresholds A and the $[K/2]$ largest of the thresholds B. Therefore, the ASN of the SWP of mildly occupied subbands is higher than that of busy and sparse subbands. This explains the smaller gap when the subbands are mildly occupied.

5.2 Locally optimum detection in heavy-tailed noise for spectrum sensing

Generally, energy detection or radiometry is preferred for its low complexity. However, it lacks robustness in noise power uncertainty, especially in a low SNR regime [14,106]. In contrast, locally optimum detection (LOD) is designed to behave optimality in a low SNR regime [60,107]. Hence, implementing LOD in spectrum sensing is well motivated. In this work, we propose a robust version of the LOD derived in [108] in terms of dealing with heavy-tailed noise and outliers. We shall refer to the proposed LOD as Robust Locally Optimum Detection (RLOD). Note that the LOD in [108] is derived based on a Gaussian noise assumption. Meanwhile, the RLOD is derived based on the assumption that the noise is Student's t-distributed. This is motivated by the fact that the latter distribution is very suitable for modeling heavy-tailed noise and the presence of large outliers that can cause noise uncertainty in a low SNR regime.

5.2.1 Formulation and method

We assume several frequency bands, each of which can be occupied by a primary user. Every potential CR user can use a block of N samples to perform spectrum sensing in a certain band or several bands independently. The channel is assumed to be flat over each frequency band. The received signal in each band can be written as:

$$\mathbf{x}[n] = \theta e^{j\phi}\mathbf{s}[n] + \mathbf{w}[n], \quad n = 0, 1, \ldots, N-1 \tag{5.9}$$

where ϕ is a channel phase shift, uniformly distributed over $(-\pi, \pi]$, and considered constant during each sensing period. Here, $\mathbf{x}[n]$ is a scalar observation. The signal constellation is represented by $\mathbf{s}[n]$ which is drawn with equal probabilities from the M-ary symbols $(s_0, s_1, s_2, ..., s_{M-1})$, whereas $\mathbf{w}[n]$ denotes the noise. The problem of detecting weak (low SNR) primary signals in spectrum sensing, based on a received signal $\mathbf{x}[n]$, can be formulated as

$$\mathcal{H}_0 : \ \theta = 0 \tag{5.10a}$$

$$\mathcal{H}_1 : \ \theta > 0. \tag{5.10b}$$

Let $\mathbf{x}_N = (\mathbf{x}[1]\ \mathbf{x}[2]\ \cdots\ \mathbf{x}[N])$ be a finite sequence of N i.i.d. observations of a signal. We can write the joint pdf of the observations as

$$f_N(\mathbf{x}_N; \theta) = \prod_{n=1}^{N} f(\mathbf{x}[n]; \theta). \tag{5.11}$$

Since $\mathbf{w}[n]$ is assumed to follow a Student's t-distribution with ν degrees of freedom and the M-ary symbols $s_m,\ m = 0, ..., M-1$ are transmitted with equal probability, $f(\mathbf{x}[n]; \theta)$ is given by

$$f(\mathbf{x}[n]; \theta) = \frac{1}{M} \sum_{m=0}^{M-1} \frac{1}{2\pi\sigma} \left(1 + \frac{|\mathbf{x}[n] - \theta e^{j\phi} s_m|^2}{\nu\sigma^2}\right)^{-(\frac{\nu+2}{2})}. \tag{5.12}$$

Based on [107], the test statistic of a locally optimum detector for the null hypothesis $\mathcal{H}_0$ versus the alternative hypothesis $\mathcal{H}_1$ is obtained as the ratio

$$T(\mathbf{x}_N) = \frac{f_N^{(i)}(\mathbf{x}_N; \theta)|_{\theta=0}}{f(\mathbf{x}_N; 0)}, \tag{5.13}$$

where $f^{(i)}(\mathbf{x}_N; \theta)$ is the first non zero i-th derivative of the joint pdf of the observation vector $\mathbf{x}_N$ under $\mathcal{H}_1$ and $f(\mathbf{x}_N; 0)$ is the joint pdf of the observation vector $\mathbf{x}_N$ under $\mathcal{H}_0$. The first derivative $f^{(1)}(\mathbf{x}[n]; \theta)$ is equal to zero under the assumption of symmetric

signal constellations, such as BPSK, QPSK, QAM and etc. Therefore, we compute the second derivative of (5.11) and stop, since $f^{(2)}(\mathbf{x}[n];\theta)$ is different from zero at $\theta = 0$. Finally, after substituting the results and assuming BPSK, the test statistic (5.13) becomes

$$T(\mathbf{x}_N) = \sum_n \left\{ \frac{(\nu+4)}{2\nu} \frac{\Re[(\mathbf{x}^*[n])^2 e^{j2\phi}] + |\mathbf{x}[n]|^2}{\sigma^2\Big(1 + \frac{|\mathbf{x}[n]|^2}{\nu\sigma^2}\Big)^2} - \frac{1}{1 + \frac{|\mathbf{x}[n]|^2}{\nu\sigma^2}} \right\}, \tag{5.14}$$

where $\Re[z]$ and z^* are the real part and complex conjugate of a complex number z, respectively. This is an important expression which depends on the choice of ν and thus takes different forms for different distributions. When $\nu = 1$, we obtain the test statistic for Cauchy distributed noise. When $\nu \to \infty$, $T(\mathbf{x}_N)$ reduces to the test statistic derived for the Gaussian case in [108]:

$$T1(\mathbf{x}_N) = \Re\left[e^{j2\phi} \left(\frac{1}{N} \sum_{n=0}^{N-1} (\mathbf{x}^*[n])^2 \right) \right] + \frac{1}{N} \sum_{n=0}^{N-1} |\mathbf{x}[n]|^2.$$

By further assuming that the channel phase shift ϕ can be estimated using the phase of pseudo-variance multiplication of the received signal $\mathbf{x}[n]$, and that $\mathbf{s}[n]$ has higher signal constellations size than BPSK, the test statistic above reduces to the energy detector [108]:

$$T2(\mathbf{x}_N) = \frac{1}{N} \sum_{n=0}^{N-1} |\mathbf{x}[n]|^2.$$

Since it is very difficult to find a closed form expression for the distribution of the test statistic under $\mathcal{H}_0$ and $\mathcal{H}_1$, we will determine the probability of false alarm P_f and miss detection P_m via simulations. In the following section, we present simulation results for the performances of our proposed RLOD, the classical LOD based on the Gaussian noise assumption and the energy detector. They are evaluated for Gaussian distributed noise, Student's t-distributed noise, and contaminated Gaussian noise.

5.2.2 Simulation results

For all simulations, we assumed that a primary user transmits BPSK signals. The signals experience a large attenuation and a phase shift which is distributed uniformly over $(-\pi, \pi]$. Thus, the CR user has to detect a signal in a low SNR regime. The pdf of the test statistic under $\mathcal{H}_0$ and $\mathcal{H}_1$ was obtained via 10^4 Monte Carlo runs for each condition. Most of the simulations have been done under the assumption that allows

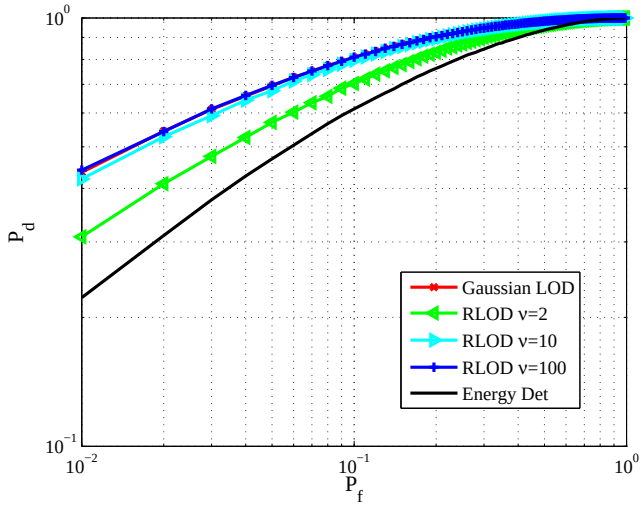


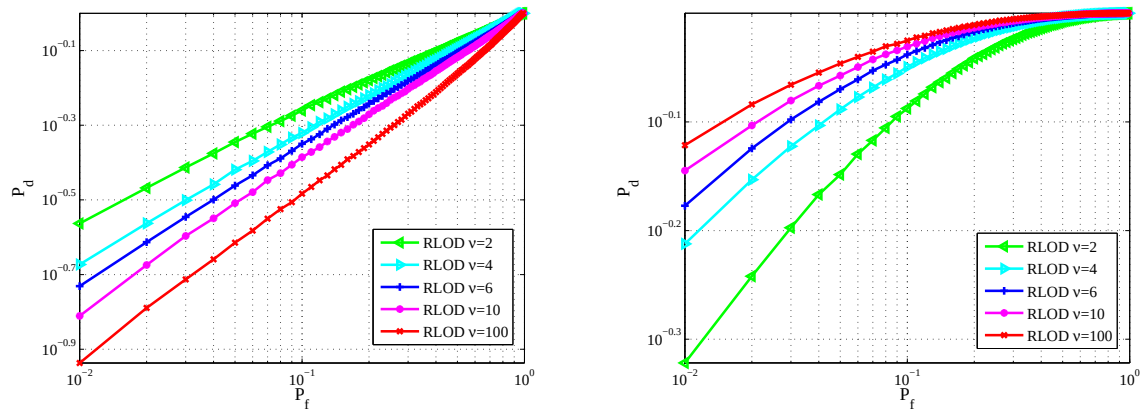
Figure 5.3: Receiver operating characteristics for Gaussian (conventional) LOD, energy detector, and RLOD with various degree of freedom ν, under Gaussian noise.

for a minimum signal strength $\theta_{\min}$, which produces an SNR of -10 dB. For the case of heavy-tailed noise or outliers, the actual SNR will be lower than the assumed SNR. For the first four figures, we have used sample size $N = 1000$.

Fig. 5.3 shows the receiver operating characteristic (ROC) for Gaussian noise. It can be seen that the best performance is obtained for $\nu = 100$ and that the proposed detector performs equally well as the detection under the Gaussian assumption. This is to be expected, since the classical LOD is a special case of the proposed RLOD. The performance decreases with decreasing ν. The worst results are obtained with the energy detector. In fact, the RLOD outperforms the energy detector for any degrees of freedom ν in the Gaussian case.

Fig. 5.4.*a* shows the results obtained with noise following a Student's t-distribution with a small degree of freedom ($\nu = 2$). It is obvious that the best performance is obtained for the proposed test with $\nu = 2$ (no mismatch between the assumption and the truth). The performance of the RLOD decreases with increasing ν. The results when the generated noise follows Student's t-distribution with a large degree of freedom $\nu = 100$ is depicted in Fig. 5.4.*b*. The performance of the RLOD decreases with decreasing ν.

Suppose now that we have a signal with outliers which is modeled by a Gaussian mixture model [107], where the outliers follow a Gaussian with a larger variance. In this case, the noise is generated from $(1 - \epsilon)\mathcal{N}(0, \sigma^2) + \epsilon\mathcal{N}(0, \kappa\sigma^2)$. The results for $\kappa = 10$ and contamination $\epsilon = 10\%$ are depicted in Fig. 5.5. It can be noticed that



(a) Degree of freedom $\nu = 2$ (b) Degrees of freedom $\nu = 100$

Figure 5.4: Receiver operating characteristic for RLOD under Student's t-distributed noise with $\nu = 2$ and $\nu = 100$ degrees of freedom.

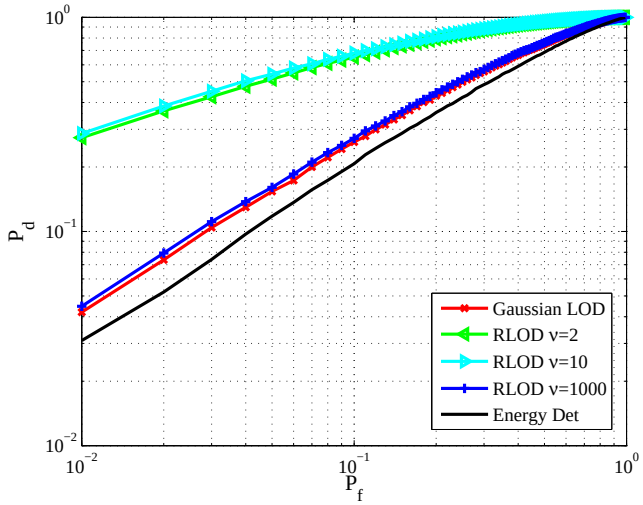


Figure 5.5: Receiver operating characteristics for Gaussian (conventional) LOD, energy detector, and RLOD with various degrees of freedom ν, under Gaussian mixture model with $\kappa = 10$ and $\epsilon = 10\%$.

the test statistic derived from the Student's t-distribution with 2 and 10 degrees of freedom give a clear increase in robustness, compared to the conventional test.

In cognitive radio, the aim is to keep the quality of service (QoS) of the primary network to a certain level in terms of the degree of interference. From a CR user's perspective, maintaining the degree of interference in a primary network based on a QoS requirement can be interpreted as maintaining the probability of miss detection P_m to a nominal value. Hence, to evaluate the performance we fix P_m at a certain

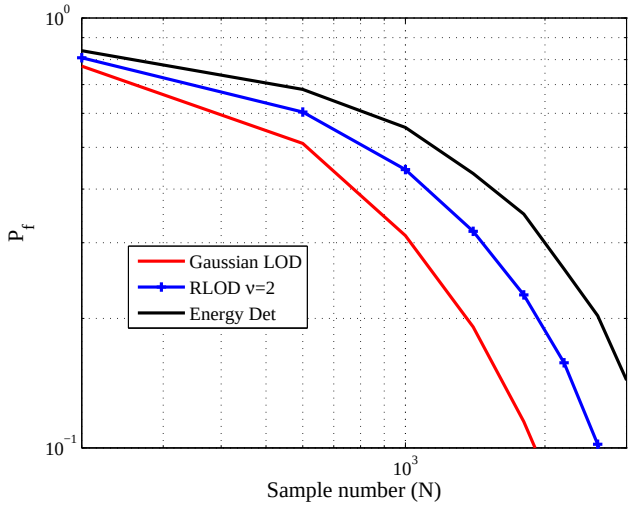


Figure 5.6: Probability of false alarm P_f for Gaussian (conventional) LOD, energy detector, and RLOD with $\nu = 2$ degree of freedom, as a function of sample size N without noise contamination.

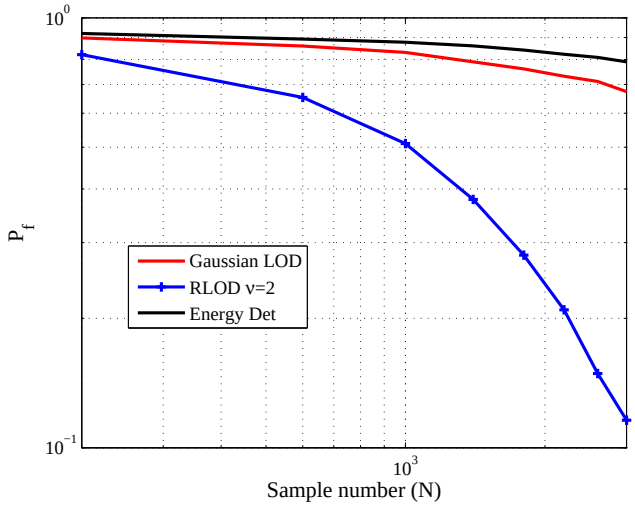


Figure 5.7: Probability of false alarm P_f for Gaussian (conventional) LOD, energy detector, and RLOD with $\nu = 2$ degree of freedom, as a function of sample size N under Gaussian mixture model with $\kappa = 10$ and $\epsilon = 10\%$.

level and compute P_f. Figs. 5.6 and 5.7 depict the curves of the probability of false alarm P_f with respect to the sensing time (sample size N). Fig. 5.6 illustrates the results when the signal does not contain outliers ($\epsilon = 0$). Fig. 5.7 is obtained with 10% contaminated data generated from the Gaussian mixture with $\kappa = 10$. In both figures, the probability of miss detection is fixed to $P_m = 0.05$ and $\nu = 2$ for the RLOD test.

Fig. 5.6 shows that without noise contamination the performance of the RLOD is better

than the performance of the energy detector and only slightly worse than that of the optimum Gaussian LOD. When the signal contains outliers, the RLOD has the best behavior which is shown in Fig. 5.7. For a value of $N = 1000$, $P_f \approx 0.5$ for the RLOD, while in the absence of outliers it is ≈ 0.45. This shows that the RLOD as decreased in performance in the presence of 10 % outliers. However, this decrease is limited when compared to the Gaussian LOD test or the energy detector which are not robust. For the energy detector, P_f increases from ≈ 0.55 to 0.9 in the presence of outliers and from ≈ 0.32 to 0.83 for the Gaussain LOD.

5.3 Collaborative spectrum sensing using HDC with quality information bit

Collaborative spectrum sensing has the main goal to tackle the issue of the hidden terminal problem. To this end, it takes advantage of spatial diversity in order to increase the global probability of detection, especially in a harsh wireless environment, by allowing collaboration between CR users. In a parallel scheme, each CR user is equipped with a detector and forwards local sensing information to a fusion center (FC) which finally makes a global decision about the existence of a primary user's signal according to a fusion rule, such as the soft decision combining (SDC) or the hard decision combining (HDC) rule [32].

Note that the HDC requires a minimum channel bandwidth between each CR user and the FC, but has poorer performance than the SDC, as shown in [34]. In this work we extend the HDC in the cognitive radio context by means of quality information, so as to improve the performance in terms of the probability of miss detection. We detail how quality information bits can be derived and incorporated into the fusion rule.

5.3.1 Formulation and method

Let L denote the number of CR users that sense a certain frequency band collaboratively and forward their decision results to an FC. Each CR user employs energy detection with N samples for the local decision. The n-th scalar sample of the l-th CR user, $1 \leq n \leq N$, $1 \leq l \leq L$, can be expressed by $\mathbf{x}_l(n)$ and the detection problem formulated as follows

$$\mathcal{H}_0 \quad : \quad \mathbf{x}_l[n] = \mathbf{w}_l[n] \tag{5.15a}$$

$$\mathcal{H}_1 \quad : \quad \mathbf{x}_l[n] = \mathbf{h}_l \mathbf{s}[n] + \mathbf{w}_l[n], \tag{5.15b}$$

where $\mathbf{s}[n]$ is the primary user's signal which is assumed to be deterministic and unknown, $\mathbf{w}_l[n]$ is white Gaussian noise and $\mathbf{h}_l$ is the scalar channel gain between the primary user and the l-th CR user. The received signal energy Y_l at the l-th CR user follows a central chi-square distribution under $\mathcal{H}_0$ and a non-central chi-square distribution under $\mathcal{H}_1$ [12], each with N degrees of freedom and non-centrality parameter $N\gamma$, where γ is the SNR at the CR user. The SNR is defined as $\gamma = \dfrac{P}{N_0 W}$ with P being the power of the primary user's signal, N_0 and W are the one-sided noise power spectral density and bandpass filter bandwidth, respectively. Hence, the probability of false alarm with local decision threshold τ_l, can be written as follows [12, 33]

$$P_f = P(Y_l > \tau_l | \mathcal{H}_0) = \frac{\Gamma(N/2, \tau_l/2)}{\Gamma(N/2)} \triangleq G_{N/2}(\tau_l), \tag{5.16}$$

where Γ denotes the upper incomplete gamma function [109]. When h_l varies due to fading or shadowing, the probability of detection P_d should be averaged over fading statistics. In this case, it can be written as [12],

$$P_d = \int_{\gamma} Q_{N/2}(\sqrt{N}x, \sqrt{G^{-1}_{N/2}(P_f)}) f_\gamma(x) dx \triangleq f_l(P_f). \tag{5.17}$$

Here, $f_\gamma(x)$ is the pdf of the SNR under fading and $Q_{N/2}$ is the generalized Marcum Q-function. Meanwhile, P_f remains the same in the presence or absence of fading, since it is independent of γ. For a closed-form expression of P_d under different fading statistics, refer to [12]. Basically, (5.17) is the receiver operating characteristics (ROC) of the energy detector for fading channels.

Consider the case where each CR user not only transmits a decision bit, but along with it a quality information bit, indicating the degree of confidence that the CR user has in its decision [95]. Fig. 5.8 illustrates how the decision bit u_l and the quality bit c_l of the l-th CR user are defined. We assume that $u_l = 0$ and 1 indicate no primary user's signal detected and primary user's signal detected, respectively. When the measured energy Y_l falls into the region within a strip (τ_{lL},τ_{lU}), defined as the "no confidence" region, then the bit $c_l = 0$ is transmitted along with the decision bit. Otherwise, when Y_l falls in the confidence regions ($Y_l < \tau_{lL}$ or $Y_l > \tau_{lU}$), $c_l = 1$ is transmitted.

As Fig. 5.8 shows, we are supposed to determine three thresholds, the middle threshold τ_l, the lower threshold τ_{lL}, and the upper threshold τ_{lU}. The middle threshold τ_l is pre-specified by the local probability of false alarm of the l-th CR user P_f^l using the inverse function of (5.16). The method to obtain the other two optimum thresholds is explained in the following. Let $\mathbf{D} = (d_1, d_1, ..., d_L)$ denote the observation vector at the combining node which consists of local decisions $d_l = (u_l, c_l)$, and let P_{ij}^{kl} be the

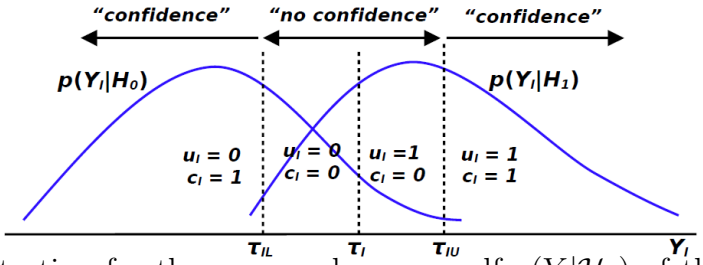


Figure 5.8: Illustration for the measured energy pdf $p(Y_l|\mathcal{H}_k)$ of the CR user under hypothesis $\mathcal{H}_0$ and $\mathcal{H}_1$, and the definition of the decision and quality information bits for each region.

joint probability distribution of each region when $u_l = i$ and $c_l = j$ under hypothesis $\mathcal{H}_k$ for the l-th CR user,

$$P_{ij}^{kl} = P(u_l = i, c_l = j|\mathcal{H}_k) \quad i, j, k = 0, 1. \tag{5.18}$$

The optimum decision rule based on the Neyman-Pearson criterion at the combining node is given by [32]:

$$T(\mathbf{D}) = \log \frac{P(\mathbf{D}|\mathcal{H}_1)}{P(\mathbf{D}|\mathcal{H}_0)} \begin{cases} > \tau_0, & u_0 = 1 \\ < \tau_0, & u_0 = 0, \\ = \tau_0, & \text{randomization}, \end{cases} \tag{5.19}$$

where u_0 is the global decision and τ_0 is the decision threshold determined by the preset global probability of false alarm α. It is necessary to include a randomization step, where $u_0 = 1$ with probability ω and $u_0 = 0$ with probability $1 - \omega$, since the pdf of $T(\mathbf{D})$ is discrete under $\mathcal{H}_0$ and $\mathcal{H}_1$. Assuming independence between CR users and using the notation in (5.18), the test statistic can be written as

$$\begin{aligned} T(\mathbf{D}) &= \log \frac{P(\mathbf{D}|\mathcal{H}_1)}{P(\mathbf{D}|\mathcal{H}_0)} = \sum_{l=1}^{n_{01}} \log \frac{P_{01}^{1l}}{P_{01}^{0l}} + \sum_{l=1}^{n_{00}} \log \frac{P_{00}^{1l}}{P_{00}^{0l}} \\ &+ \sum_{l=1}^{n_{10}} \log \frac{P_{10}^{1l}}{P_{10}^{0l}} + \sum_{l=1}^{n_{11}} \log \frac{P_{11}^{1l}}{P_{11}^{0l}} = \sum_{l=1}^{L} w_l \end{aligned} \tag{5.20}$$

where $w_l = \log \frac{P_{ij}^{1l}}{P_{ij}^{0l}}$, n_{ij} is the number of CR users which decide $d_l = (u_l = i, c_l = j)$, and $n_{01} + n_{00} + n_{10} + n_{11} = L$. Therefore, the test statistic is a weighted sum of local decisions, which are functions of the partition $\{P_{ij}^{kl}\}_{l=1}^{N}$ determined by the threshold τ_{lL},τ_l and τ_{lU} at each CR user.

In order to find optimum sub-partitioning $\{P_{ij}^{kl}\}_{l=1}^{L}$, which in turn is used to find the optimum threshold, we follow the steps in [110], in which J-Divergence is employed as the

objective function for optimization. J-Divergence is a member of the Ali-Silvey class of distance measures, which define the distance between two probability distributions [111]. In our case, it can be written as

$$\begin{aligned} J &= E_{\mathcal{H}_1}[T(\mathbf{D})] - E_{\mathcal{H}_0}[T(\mathbf{D})] \\ &= \sum_{l=1}^{N}(E_{\mathcal{H}_1}[w_l] - E_{\mathcal{H}_0}[w_l]) = \sum_{l=1}^{L} J_l, \end{aligned} \tag{5.21}$$

where $E_{\mathcal{H}_k}[\cdot]$ denotes expectation under hypotheses $\mathcal{H}_k$. It is obvious from (5.21) that sub-partitioning each local decision region of a CR user is independent of those of others. From (5.20) and (5.21), the J-Divergence contributed by the lth CR user J_l can be expressed as

$$\begin{aligned} J_l &= (P_{01}^{1l} - P_{01}^{0l}) \log \frac{P_{01}^{1l}}{P_{01}^{0l}} + (P_{00}^{1l} - P_{00}^{0l}) \log \frac{P_{00}^{1l}}{P_{00}^{0l}} \\ &+ (P_{10}^{1l} - P_{10}^{0l}) \log \frac{P_{10}^{1l}}{P_{10}^{0l}} + (P_{11}^{1l} - P_{11}^{0l}) \log \frac{P_{11}^{1l}}{P_{11}^{0l}}. \end{aligned} \tag{5.22}$$

Using several relations that can be derived from Fig. 5.8 and the definition of (5.18), differentiating J_l with respect to P_{11}^{0l} and P_{00}^{0l}, and setting both resulting equations to zero, leads to the following equations

$$f_l'(P_{11}^{0l}) = \frac{\dfrac{P_d^l - f_l(P_{11}^{0l})}{P_f^l - P_{11}^{0l}} - \dfrac{f_l(P_{11}^{0l})}{P_{11}^{0l}} + \log\left\{ (\dfrac{P_d^l - f_l(P_{11}^{0l})}{P_f^l - P_{11}^{0l}})(\dfrac{P_{11}^{0l}}{f_l(P_{11}^{0l})}) \right\}}{-\dfrac{P_f^l - P_{11}^{0l}}{P_d^l - f_l(P_{11}^{0l})} + \dfrac{P_{11}^{0l}}{f_l(P_{11}^{0l})} + \log\left\{ (\dfrac{P_d^l - f_l(P_{11}^{0l})}{P_f^l - P_{11}^{0l}})(\dfrac{P_{11}^{0l}}{f_l(P_{11}^{0l})}) \right\}} \tag{5.23}$$

$$f_l'(P_f') = \frac{\dfrac{1 - f_l(P_f')}{1 - P_f^l - P_{00}^{0l}} - \dfrac{f_l(P_f') - P_d^l}{P_{00}^{0l}} + \log\left\{ (\dfrac{1 - f_l(P_f')}{1 - P_f^l - P_{00}^{0l}})(\dfrac{P_{00}^{0l}}{f_l(P_f') - P_d^l}) \right\}}{-\dfrac{P_f^l - P_{11}^{0l}}{P_d^l - f_l(P_f')} + \dfrac{P_{11}^{0l}}{f_l(P_f')} + \log\left\{ (\dfrac{P_d^l - f_l(P_f')}{P_f^l - P_{11}^{0l}})(\dfrac{P_{11}^{0l}}{f_l(P_f')}) \right\}} \tag{5.24}$$

where $P_f' = P_f^l + P_{00}^{0l}$, P_d^l and P_f^l are the probabilities of detection and false alarm of the l-th CR user, and $f_l(\cdot)$ is the ROC given in (5.17). Since P_f^l is pre-specified and P_d^l can be obtained from (5.17), we can have an optimum sub-partitioning of P_{11}^{0l} by solving (5.23) numerically. Therefore, we can obtain τ_{lU} from (5.16). Obtaining τ_{lL} can be accomplished by first solving (5.24) numerically in order to get P_{00}^{0l}.

Referring to (5.23) and (5.24), $f_l'(\cdot)$ needs to be determined, which is basically the derivative with respect to P_f of the ROC in (5.17). By using the chain rule and inverse

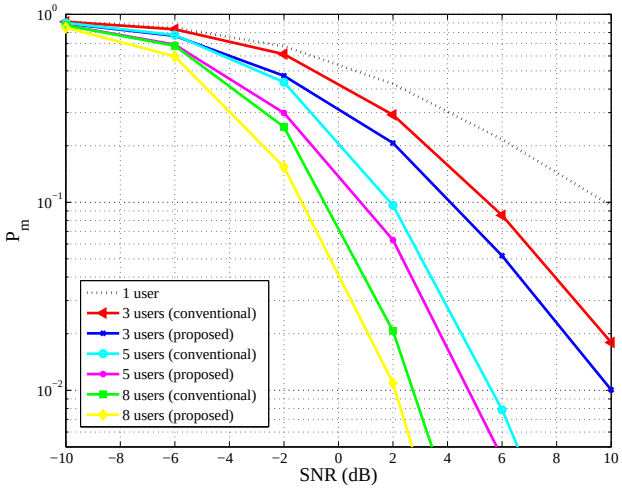


Figure 5.9: Probabilities of miss detection P_m of HDC 1-bit (conventional) and HDC 2-bits (proposed) as a function of SNRs for various number of users.

function rule on the derivative, $f_l'(\cdot)$ under fading can be expressed by

$$f_l'(P_f) = dP_d/dP_f = \int_\gamma \frac{2^{N-\frac{1}{2}}}{\left\{\sqrt{NxG_{N/2}^{-1}(P_f)}\right\}^{N/2-1}} \times \Gamma(N/2)e^{(-\frac{N}{2}x)} I_{\frac{N}{2}-1}\left(\sqrt{2NxG_{N/2}^{-1}(P_f)}\right) f_\gamma(x)dx \tag{5.25}$$

where $I_{\frac{N}{2}-1}(\cdot)$ is the $(\frac{N}{2}-1)$-th order modified Bessel function of the first kind, $\Gamma(\cdot)$ is the Gamma function and $G_{N/2}^{-1}(P_f)$ is the inverse of the incomplete Gamma function of (5.16) [109].

5.3.2 Simulation results

We use the probability of miss detection P_m as a performance measure, since it is more relevant than the probability of detection P_d in a cognitive radio context, because it indicates the degree of interferences to primary users. In Fig. 5.9 and 5.10, the reporting channels are assumed to be perfect. Note that the HDC 1-bit is the optimal scheme according to [32].

Fig. 5.9 and 5.10 show that the HDC scheme with quality information bits (HDC 2-bit) outperforms the HDC 1-bit scheme. For a targeted $P_m = 10\%$, with a number of users varying from 2 to 8, the SNR gain that could be achieved by incorporating quality information bits ranges from 1 dB to 2 dB. Meanwhile, Fig. 5.10 shows that for

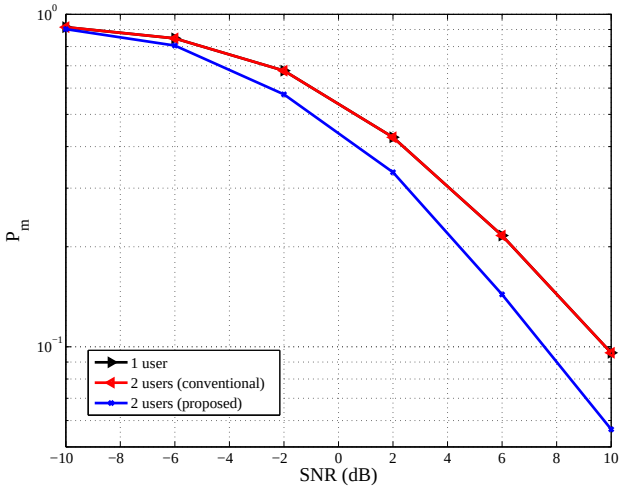


Figure 5.10: Performance of 2 collaborating users using HDC 1-bit and 2-bits in terms of probability of miss detection P_m as a function of SNRs, compared to 1 user.

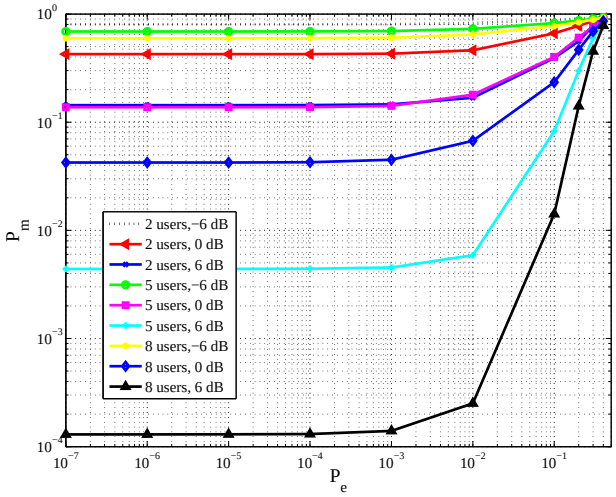


Figure 5.11: Performance of the HDC 2-bits under imperfect reporting channel. The probability of miss detection P_m is plotted against bit error probability P_e for various users and SNRs.

2 collaborating users, spatial diversity cannot be exploited by the HDC 1-bit scheme, indicated by the overlap of both curves. This has been proven in [95], which states that in a configuration of L similar detectors, a better performance than a single user, can only be achieved for $L > 2$. This is no longer true for the HDC 2-bit.

The assumption that the reporting channel between each collaborating user and the FC is perfect may not be valid in practice. It could be subjected to propagation

effects, collisions and interference, which cause the transmitted information from each collaborating user to be erroneous at the FC. These errors surely will cause performance loss. In order to quantify the effect of imperfect reporting channels on the performance of our scheme, we assume the channels to be binary symmetric channels, with equal bit error probabilities (BEP). Using the Hamming distance $h_{l,j}$ [112], between 2-bit sequences (d_l, d_j), the joint distribution at the FC $P^{kl}_{ij,fc}$ for hypothesis $\mathcal{H}_k$ can be written as

$$P^{kl}_{ij,fc} = \sum_{j=1}^{4} P_e^{h_{l,j}} (1 - P_e)^{2-h_{l,j}} P^{kl}_{ij} \tag{5.26}$$

where P_e is the BEP and P^{kl}_{ij} is from (5.18).

Fig. 5.11 depicts the performance of our scheme under imperfect reporting channels for a different number of users and an average SNR. We assume the BEP to be equal for all channels and vary it from $P_e = 0$ (not showing in the figure), which is equal to the perfect channel, to 0.4. The performances do not change significantly for $P_e \leq 10^{-3}$ compared to the performance in the perfect channel case. They depart noticeably starting from $P_e = 10^{-2}$, and they get worse when P_e increases above 0.4. This suggests that in the design of the fusion rule at the FC, the imperfect channel effect should be taken into account.

5.4 Conclusion

The multiple testing procedure has been implemented in sequential probability ratio tests. The results show that the proposed method fulfills the objective to have control over the decision errors at the system level which is required in a scenario where the primary and the cognitive radio networks provide their users accesses to several bands at a time. The total ASN of the proposed method is significantly smaller than that of the simple Bonferroni procedure. Therefore, the proposed method is a promising technique to increase the overall throughput of cognitive radio networks without causing harmful interferences to the primary networks.

Extensive simulations lead to the conclusion that the RLOD that we propose, outperforms the energy detector in a low SNR region in all cases, even in the Gaussian noise case. Our simulation results also show that the proposed detector is more robust to outliers than the energy detector or the Gaussian LOD. This makes it a better candidate for being implemented in spectrum sensing for cognitive radio.

A new scheme that incorporates quality information into the HDC 1-bit scheme has been presented. In summary, for all considered numbers of users and SNR ranges, the performance in terms of the probability of miss detection improves compared to the conventional HDC 1-bit scheme. We also have considered the performance when facing imperfect channels. Under binary symmetric channels, it is suggested to incorporate the channel effect into the design of the fusion rule when $P_e > 10^{-3}$. As a disadvantage, the scheme requires the fusion center to know the local detector performance indices to implement the optimum likelihood ratio based fusion rule.

Chapter 6

Conclusions and outlook

In this thesis, advanced techniques for spectrum sensing with short sensing times have been developed. These techniques comprise bootstrap based local and collaborative test with a fixed sample size, and bootstrap based sequential probability ratio tests. Moreover, we developed methods for multiband spectrum sensing, locally optimum detection and two bits collaborative spectrum sensing. In this part, we draw conclusions in Section 6.1, and outline directions for future works in Section 6.2.

6.1 Conclusions

In the beginning, we considered the case where the primary user signal is temporally correlated and the noise is uncorrelated. We highlighted the drawback of using the asymptotic test when the sample size is small. Moreover, we showed that the common approach of using the bootstrap pivoting test also failed to achieve satisfactory performance. Thus, we developed an alternative method that is based on a null resampling approach. Unlike the bootstrap pivoting test, simulation results show that the actual probability of false alarm of the bootstrap null resampling test is close to the nominal value. This means that the proposed method yields better estimates of the true distribution of the test statistic under the null hypothesis than the bootstrap pivoting test. The former has a smaller computational cost. Furthermore, we extended the resampling based approach to collaborative spectrum sensing using the Chair-Varshney fusion rule, where the local probability of detection is estimated in each measurement. When the number of CR users is large, the parametric bootstrap is used to estimate the distribution of the fusion rule under the null hypothesis. As a result, the computational cost is significantly reduced. The simulation results demonstrate that the performance of collaborative spectrum sensing using the parametric bootstrap does not significantly differ from the one that uses all possible binary combinations.

We tackled further the issue of implementing the sequential probability ratio test under composite hypotheses. When the constant thresholds for the sequential probability ratio test are used, the generalized sequential probability ratio test for composite hypotheses produces large decision errors, particularly for low SNR. This is because many terminations occur due to maximum likelihood estimator at earlier stages, when the

number of samples is still small. In order to alleviate the problem, we developed techniques that can update the thresholds according to the sample size. This is accomplished by estimating the distribution of the error terms in the likelihood ratios, using the parametric bootstrap. Simulations have been conducted and the results indicate that the decision errors are reduced, and thus the sensing objective is preserved. Furthermore, the proposed method has smaller ASNs than the conventional method, whose thresholds are found by Monte Carlo simulations. In addition, we highlighted the advantage in terms of computational cost of using the proposed approach, instead of using a method that has been suggested in the literature. In general, additional computational costs due to the bootstrap are inevitable. However, we provided a solution by combining the K latest bootstrap distributions, by which the computational cost is reduced by a factor as large as a hundred. Simulation results show that the ASN of the bootstrap based sequential probability ratio test with reduced computational cost is still comparable to the one without a computational cost reduction scheme.

In the last part, we considered various problems in spectrum sensing. First, we used multiple testing procedures for multiband spectrum sensing. For fixed sample size cases, we suggested to use an adaptive Benjamini-Hochberg procedure, since it provides a better compromise between the familywise error rate and the familywise miss detection than the original Benjamini-Hochberg procedure. In the context of sequential testing, we developed a stepwise procedure that based on the ordered stopping times of the sequential probability ratio tests. Simulation results show that the devised procedure maintains the actual values of decision errors close to the nominal values, regardless of the number of subbands occupied by a primary user, and regardless of the number of subbands that might be opportunistically used by a CR user. The ASN of the stepwise procedure is smaller than that of the Bonferroni procedure. Second, the problem of detecting low SNR signal in the presence of outliers has been addressed. For this case, we developed a robust locally optimum detector, which was derived based on the assumption that the noise is Student's t-distributed. We carried out simulations and the results show that the robust locally optimum detector outperforms the energy detector, and performs even better than the locally optimum detector that is based on a Gaussian noise assumption, when outliers are present. Third, we sought to improve the performance of hard decision combining in collaborative spectrum sensing. In the derived scheme, each CR user transmits not only a decision bit, but also a quality information bit which indicates a degree of confidence in the decision that the CR user has made. Each CR user has three thresholds, one of which is pre-specified and the other two are determined by means of the J-Divergence. The performance improves significantly, as shown by simulations.

6.2 Outlook

We identify at least three possible works that can be conducted in the future. They include:

- Resampling based spectrum sensing for correlated signals: The problem of detecting correlated signals in uncorrelated noise has been treated in this thesis. However, interference from a source that has a degree of temporal correlations might appear at the receiver of a spectrum sensor. Consequently, the assumption that we used in this thesis does not hold anymore. More precisely, we also have correlated observations under the null hypothesis. Should we know the degree of correlation of the respective interferer, the challenge would be to construct a null resampling method under this condition. However, the case is more challenging if the degree of correlation of the respective interferer is unknown. In addition, for multiband spectrum sensing, we have assumed that the observations between subbands are independent, under the null and the alternative hypotheses. However, the approach must be different if we assume that the observations between subbands are dependent. As a next stage, investigating resampling-based multiple testing procedure for dependent observations seems relevant.

- Bootstrap based sequential probability ratio tests: In this thesis, the modifications of the two thresholds are treated independently. The upper threshold is modified to control the probability of false alarm using the parametric bootstrap assuming the data is generated under the null hypothesis. In fact, the modification also affects the required sample size to correctly decide, whether or not the data originate from the alternative hypothesis. The same considerations apply to the lower threshold. Therefore, joint optimization to modify the two thresholds by bootstrapping could be an appropriate choice in the future work.

- Multiple testing procedure for multiband spectrum sensing using sequential testing: In sequential testing, we developed a stepwise procedure that is based on a heuristic approach. An analytical approach such as using optimization theory is possible. From our perspective, some questions are still open: What is the optimum way to conduct MTP in the SPRT? What is the role of FDR and some other measures in MTP for sequential testing? Is it possible to use these measures to further reduce the sample size and, if so, how? Answering these questions for the analytical approach will be a challange.

List of Acronyms

A-BHP	Adaptive Benjamini-Hochberg procedure
AS	Assumption
ASN	Average sample number
BHP	Benjamini-Hochberg procedure
B-SPRT	Bootstrap based sequential probability ratio test
BPSK	Binary Phase Shift Keying
CR	Cognitive radio
CSS	Collaborative spectrum sensing
EDF	Empirical distribution function
EGC	Equal gain combining
FAR	False alarm ratio
FC	Fusion center
FCC	Federal communications commission
FDR	False discovery rate
FSN	Fixed sample number
FWE	Familywise error rate
FWM	Familywise miss detection
GLRT	Generalized likelihood ratio test
HDC	Hard decision combining
HP	Holm procedure
i.i.d.	independent and identically distributed
KL	Kullback-Leibler
LLR	Likelihood ratio
LLRT	Log-likelihood ratio test

LOD	Locally optimum detector
MCP	Multiple comparison procedure
MLE	Maximum likelihood estimator
MR	Missing ratio
MTP	Multiple testing procedure
NP	Neyman-Pearson
OC	Operating characteristic
pdf	Probability density function
QAM	Quadrature amplitude modulation
QoS	Quality of service
QPSK	Quadrature phase shift keying
RLOD	Robust locally optimum detector
SBF	Simple Bonferroni procedure
SC	Selection combining
SDC	Soft decision combining
SNR	Signal-to-noise ratio
SPRT	Sequential probability ratio test
SSC	Switch stay combining
SWP	Stepwise procedure

List of Symbols

$\mathbf{1}_{\{\cdot\}}$	the indicator function
A	Constant upper threshold in the SPRT
A_N	Adaptive upper threshold in the generalized SPRT
B	Constant lower threshold in the SPRT
B_N	Adaptive lower threshold in the generalized SPRT
B_s	Number of bootstrap replications
C_A	Constant modified value for the threshold A
C_B	Constant modified value for the threshold B
$\mathcal{CN}(0, \sigma^2)$	Zero mean circularly symmetric complex Gaussian distribution
$E_i[\cdot]$	Expected value under hypothesis $\mathcal{H}_i$
$f_{i,N}$	pdf under hypothesis $\mathcal{H}_i$ for the observation with N sample number
$F_{\boldsymbol{\theta}}$	Distribution of the observations parameterized by $\boldsymbol{\theta}$
$\hat{F}_i$	Empirical distribution of an observation under $\mathcal{H}_i$
$g_0(\cdot)$	pmf under $\mathcal{H}_0$
$G_i(T)$	True distribution function of T under $\mathcal{H}_i$
$\hat{G}_i(T)$	Empirical distribution function of T under $\mathcal{H}_i$
$\mathbf{h}$	Channel coefficient
$\mathcal{H}_0$	Null hypothesis
$\mathcal{H}_1$	Alternative hypothesis
$\mathcal{H}_{k,i}$	Hypothesis $\mathcal{H}_i$ for the subband k
$\mathbf{I}_M$	$M \times M$ identity matrix
K	Total number of subbands
K_0	Number of unoccupied subbands
$\hat{K}_0$	Estimated number of unoccupied subbands
$KL(f_j \| \| f_i)$	Kullback-Leibler distance between f_j and f_i
L	Number of users
$\mathcal{L}(\boldsymbol{\theta})$	The OC function
M	Smoothing factor
N	Sample number or stage in sequential testing
N_I	Initial number of samples
N_s	Stopping time
p	p-value
$\hat{p}$	Estimates of a p-value

$P_i(A)$	Probability of event A under hypothesis $\mathcal{H}_i$
P_d	Probability of detection
P_f	Actual probability of false alarm
P_m	Actual probability of miss detection
$\mathcal{Q}(\cdot)$	Q-function
r	Autocorrelation
$\hat{r}$	Sample autocorrelation
$\mathbf{r}$	Received signal component
$\mathbf{r}_N$	Received signal component with N sample number
R_i	Decision region to accept $\mathcal{H}_i$
$\mathbf{R}$	Covariance matrix
S_k	Subband k
$\mathbf{S}_T$	A set of terminated subbands
$\hat{S}_e$	Estimates of the standard error of a test statistic
$\hat{S}_k$	Lowest slope estimator
T	Test statistic
T_l^b	Bootstrap replication of a test statistic for the user l
$\mathbf{u}$	Vector of decision bits
$\mathbf{x}$	A scalar or vector observation
$\mathbf{x}_N$	An observation with N sample number
$\mathbf{w}$	A noise component n
$\bar{z}_N$	increment at time instance N
Z_N	Log-likelihood ratio in a fixed sample number testing
$\bar{Z}_N$	Log-likelihood ratio in the SPRT
$\hat{Z}_N$	Log-likelihood ratio in the generalized SPRT
α	Nominal value of the probability of false alarm
β	Nominal value of the probability of miss detection
$\delta(\cdot)$	Kronecker delta function
Δ_N^i	Modified value of the threshold at stage N under $\mathcal{H}_i$
ΔZ_N^i	the error term of LLR under $\mathcal{H}_i$
ϵ_0	Tolerance value of the probability of false alarm
ϵ_1	Tolerance value of the probability of miss detection
η	Real-valued parameter
$\Gamma(\cdot)$	Gamma function
$\mathbf{\Lambda}_A$	A set of upper thresholds

$\boldsymbol{\Lambda}_B$	A set of lower thresholds
ν	Degree of freedom for Student's t-distribution
ϕ	Unknown channel phase shift
$\boldsymbol{\psi}$	Specific parameters or statistics quantifying the existence of the desired signal
$\boldsymbol{\Psi}_i$	Spaces under hypothesis $\mathcal{H}_i$
σ_i^2	Variances under hypothesis $\mathcal{H}_i$
σ_w^2	Noise power
$\boldsymbol{\theta}_i$	True parameter under hypothesis $\mathcal{H}_i$
$\boldsymbol{\Theta}_i$	Parameter spaces for hypothesis $\mathcal{H}_i$
$\hat{\boldsymbol{\theta}}_i$	Maximum likelihood estimator of $\boldsymbol{\theta}_i$ assuming $\mathcal{H}_i$ is true
τ	threshold

Bibliography

[1] E. Hossain, D. Niyato, and Z. Han, *Dynamic Spectrum Access and Management in Cognitive Radio Networks*, Cambridge University Press, 2009.

[2] M.A. McHenry and D. McClosky, *Spectrum occupancy measurements, Chicago, Illinois*, Shared Spectrum Company, 2005.

[3] I.F. Akyildiz, W. Lee, M.C. Vuran, and S. Mohanty, "Next generation dynamic spectrum access cognitive wireless networks: A survey," *Computer Networks*, vol. 50, pp. 2127–2159, May 2006.

[4] C. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. Shellhammer, and W. Caldwell, "IEEE 802.22: The first cognitive radio wireless regional area network standard," *IEEE Communications Magazine*, vol. 47, no. 1, pp. 130–138, 2009.

[5] C. Cordeiro, K. Challapali, D. Birru, and N.S. Shankar, "IEEE 802.22: the first worldwide wireless standard based on cognitive radios," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*. IEEE, 2005, pp. 328–337.

[6] M. Sherman, A.N. Mody, R. Martinez, C. Rodriguez, and R. Reddy, "IEEE standards supporting cognitive radio and networks, dynamic spectrum access, and coexistence," *IEEE Communications Magazine*, vol. 46, no. 7, pp. 72–79, 2008.

[7] M. Murroni, R.V. Prasad, P. Marques, B. Bochow, D. Noguet, C. Sun, K. Moessner, and H. Harada, "IEEE 1900.6: spectrum sensing interfaces and data structures for dynamic spectrum access and other advanced radio communication systems standard: technical aspects and future outlook," *IEEE Communications Magazine*, vol. 49, no. 12, pp. 118–127, 2011.

[8] Y.C. Liang, A.T. Hoang, and H.H. Chen, "Cognitive radio on TV bands: a new approach to provide wireless connectivity for rural areas," *IEEE Wireless Communications*, vol. 15, no. 3, pp. 16–22, 2008.

[9] J. Wang, M. Ghosh, and K. Challapali, "Emerging cognitive radio applications: A survey," *IEEE Communications Magazine*, vol. 49, no. 3, pp. 74–81, 2011.

[10] A.J. Goldsmith, L.J. Greenstein, N. Mandayam, and H.V. Poor, *Principles of Cognitive Radio*, Cambridge University Press, 2012.

[11] S.M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, Prentice-Hall, 1998.

[12] M.S. Alouni and M.K. Simon, "On the energy detection of unknown signals over fading channels," *IEEE Transactions on Communications*, vol. 55, pp. 21–24, Jan 2007.

[13] U. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, pp. 523–531, Apr 1967.

[14] R. Tandra and A. Sahai, "SNR walls for signal detection," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, pp. 4–17, 2008.

[15] D. Cabric, A. Tkachenko, and R.W. Brodersen, "Spectrum sensing measurements of pilot, energy, and collaborative detection," in *IEEE Military Communications Conference (MILCOM)*. IEEE, 2006, pp. 1–7.

[16] W.A. Gardner, "Signal interception: a unifying theoretical framework for feature detection," *IEEE Transactions on Communications*, vol. 36, pp. 897–906, 1988.

[17] A.V. Dandawate and G.B. Giannakis, "Statistical tests for presence of cyclostationarity," *IEEE Transactions on Signal Processing*, vol. 42, no. 9, pp. 2355–2369, 1994.

[18] J. Lundén, V. Koivunen, A. Huttunen, and H.V. Poor, "Spectrum sensing in cognitive radios based on multiple cyclic frequencies," in *2nd International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CrownCom)*. IEEE, 2007, pp. 37–43.

[19] J. Lundén, V. Koivunen, A. Huttunen, and H.V. Poor, "Collaborative cyclostationary spectrum sensing for cognitive radio systems," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4182–4195, 2009.

[20] J. Lundén, S.A. Kassam, and V. Koivunen, "Robust nonparametric cyclic correlation-based spectrum sensing for cognitive radio," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 38–52, 2010.

[21] Y. Zeng and Y.C. Liang, "Covariance based signal detections for cognitive radio," in *IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*. IEEE, 2007, pp. 202–207.

[22] Y. Zeng and Y.C. Liang, "Spectrum-sensing algorithms for cognitive radio based on statistical covariances," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 4, pp. 1804–1815, 2009.

[23] Y. Zeng and Y.C. Liang, "Maximum-minimum eigenvalue detection for cognitive radio," in *International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2007, pp. 1–5.

[24] Y. Zeng and Y.C. Liang, "Eigenvalue-based spectrum sensing algorithms for cognitive radio," *IEEE Transactions on Communications*, vol. 57, no. 6, pp. 1784–1793, 2009.

[25] R. Zhang, T. Lim, Y.C. Liang, and Y. Zeng, "Multi-antenna based spectrum sensing for cognitive radios: A GLRT approach," *IEEE Transactions on Communications*, vol. 58, no. 1, pp. 84–88, 2010.

[26] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *IEEE Transactions on Wireless Communications*, vol. 9, no. 2, pp. 814–823, 2010.

[27] P. Wang, J. Fang, N. Han, and H. Li, "Multiantenna-assisted spectrum sensing for cognitive radio," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1791–1800, 2010.

[28] D. Ramírez, G. Vazquez-Vilar, R. López-Valcarce, J. Vía, and I. Santamaría, "Detection of rank-p signals in cognitive radio networks with uncalibrated multiple antennas," *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3764–3774, 2011.

[29] M. Orooji, R. Soosahabi, and M. Naraghi-Pour, "Blind spectrum sensing using antenna arrays and path correlation," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 8, pp. 3758–3767, 2011.

[30] J.K. Tugnait, "On multiple antenna spectrum sensing under noise variance uncertainty and flat fading," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1823–1832, 2012.

[31] J. Sala-Alvarez, G. Vazquez-Vilar, and R. López-Valcarce, "Multiantenna GLR detection of rank-one signals with known power spectrum in white noise with unknown spatial correlation," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 3065–3078, 2012.

[32] P.K. Varshney, *Distributed detection and data fusion*, Springer, 1997.

[33] A. Ghasemi and E.S. Sousa, "Collaborative spectrum sensing for opportunistic access in fading environments," in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*. IEEE, 2005, pp. 131–136.

[34] J. Ma, G. Zhao, and Y. Li, "Soft combination and detection for cooperative spectrum sensing in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, pp. 4502–4507, 2008.

[35] Z. Quan, S. Cui, H.V. Poor, and A.H. Sayed, "Collaborative wideband sensing for cognitive radios," *IEEE Signal Processing Magazine*, vol. 25, no. 6, pp. 60–73, 2008.

[36] Z. Quan, S. Cui, and A.H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE Jounal of Selected Topics in Signal Processing*, vol. 2, pp. 28–40, 2008.

[37] R. Viswanathan and P.K. Varshney, "Distributed detection with multiple sensors i. fundamentals," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 54–63, 1997.

[38] A. Ghasemi and E.S. Sousa, "Opportunistic spectrum access in fading channels through collaborative sensing," *Journal of Communications*, vol. 2, no. 2, pp. 71–82, 2007.

[39] W. Zhang, R. Mallik, and K. Letaief, "Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 5761–5766, 2009.

[40] K.B. Letaief and W. Zhang, "Cooperative communications for cognitive radio networks," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 878–893, 2009.

[41] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio, part I: Two user networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 6, pp. 2204–2213, 2007.

[42] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio, part II: Multiuser networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 6, pp. 2214–2222, 2007.

[43] Y.C. Liang, Y. Zeng, E.C.Y. Peh, and A.T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 4, pp. 1326–1337, 2008.

[44] A. Wald, *Sequential analysis*, Dover Publications, 2004.

[45] A. Wald, "Sequential tests of statistical hypotheses," *The Annals of Mathematical Statistics*, pp. 117–186, 1945.

[46] E. Axell, G. Leus, E.G. Larsson, and H.V. Poor, "Spectrum sensing for cognitive radio: State-of-the-art and recent advances," *IEEE Signal Processing Magazine*, vol. 29, no. 3, pp. 101–116, 2012.

[47] J. Bussgang and D. Middleton, "Optimum sequential detection of signals in noise," *IRE Transactions on Information Theory*, vol. 1, no. 3, pp. 5–18, 1955.

[48] J. Bussgang and M. Marcus, "Truncated sequential hypothesis tests," *IEEE Transactions on Information Theory*, vol. 13, no. 3, pp. 512–516, 1967.

[49] S. Tantaratana and J.B. Thomas, "Truncated sequential probability ratio test," *Information Sciences*, vol. 13, no. 3, pp. 283–300, 1977.

[50] S. Tantaratana and J.B. Thomas, "Relative efficiency of the sequential probability ratio test in signal detection," *IEEE Transactions on Information Theory*, vol. 24, no. 1, pp. 22–31, 1978.

[51] S. Tantaratana and H.V. Poor, "Asymptotic efficiencies of truncated sequential tests," *IEEE Transactions on Information Theory*, vol. 28, no. 6, pp. 911–923, 1982.

[52] C. Lee and J. Thomas, "A modified sequential detection procedure," *IEEE Transactions on Information Theory*, vol. 30, no. 1, pp. 16–23, 1984.

[53] K.W. Choi, W.S. Jeon, and D.G. Jeong, "Sequential detection of cyclostationary signal for cognitive radio systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 9, pp. 4480–4485, 2009.

[54] Q. Zou, S. Zheng, and A.H. Sayed, "Cooperative sensing via sequential detection," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6266–6283, 2010.

[55] S. Chaudhari, V. Koivunen, and H.V. Poor, "Autocorrelation-based decentralized sequential detection of ofdm signals in cognitive radios," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2690–2700, 2009.

[56] S.J. Kim and G.B. Giannakis, "Sequential and cooperative sensing for multichannel cognitive radios," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, pp. 4239–4253, 2010.

[57] Y. Yilmaz, G.V. Moustakides, and X. Wang, "Cooperative sequential spectrum sensing based on level-triggered sampling," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4509–4524, 2012.

[58] J.D. Gibson and J.L. Melsa, *Introduction to nonparametric detection with applications*, IEEE Press New York, 1996.

[59] V.H. Poor, *An introduction to signal detection and estimation*, Springer-Verlag, 1988.

[60] S.M. Kay, *Fundamentals of Statistical signal processing, Volume 2: Detection theory*, Prentice Hall PTR, 1998.

[61] A.M. Zoubir, *Bootstrap techniques for signal processing*, Cambridge University Press, 2004.

[62] A.C. Davison and D.V. Hinkley, *Bootstrap methods and their application*, vol. 1, Cambridge university press, 1997.

[63] B. Efron and R. Tibshirani, *An introduction to the bootstrap*, vol. 57, CRC press, 1993.

[64] P. Hall and S.R. Wilson, "Two guidelines for bootstrap hypothesis testing," *Biometrics*, pp. 757–762, 1991.

[65] B.K. Ghosh and P.K. Sen, *Handbook of sequential analysis*, vol. 118, CRC Press, 1991.

[66] A. Wald, "On cumulative sums of random variables," *The Annals of Mathematical Statistics*, vol. 15, no. 3, pp. 283–296, 1944.

[67] A.M. Zoubir and D.R. Iskander, "Bootstrap methods and applications," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 10–19, 2007.

[68] A.M. Zoubir and B. Boashash, "The bootstrap and its application in signal processing," *IEEE Signal Processing Magazine*, vol. 15, no. 1, pp. 56–76, 1998.

[69] R.F. Brcich, A.M. Zoubir, and P. Pelin, "Detection of sources using bootstrap techniques," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 206–215, 2002.

[70] H.T. Ong and A.M. Zoubir, "Bootstrap-based detection of signals with unknown parameters in unspecified correlated interference," *IEEE Transactions on Signal Processing*, vol. 51, no. 1, pp. 135–141, 2003.

[71] C. Debes, C. Weiss, A.M. Zoubir, and M.G. Amin, "Distributed target detection in through-the-wall radar imaging using the bootstrap," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, 2010, pp. 3530–3533.

[72] L. Arienzo, "Bootstrapping the spectrum in ultra wide-band cognitive radio networks," in *Second IEEE International Workshop on Cognitive Radio and Advanced Spectrum Management (CogART)*. IEEE, 2009, pp. 105–109.

[73] L. Luo, W. Zhou, and H. Meng, "Threshold estimation method for spectrum sensing using bootstrap technique," in *Intelligent Computing Theories*, pp. 362–367. Springer, 2013.

[74] M.A. Martin, "Bootstrap hypothesis testing for some common statistical problems: A critical evaluation of size and power properties," *Computational Statistics & Data Analysis*, vol. 51, no. 12, pp. 6321–6342, 2007.

[75] Y. Hochberg and A.C. Tamhane, *Multiple comparison procedures*, John Wiley & Sons, Inc., 1987.

[76] R.Y. Liu and K. Singh, "Moving blocks jackknife and bootstrap capture weak dependence," *Exploring the limits of bootstrap*, vol. 225, pp. 248, 1992.

[77] H.R. Kunsch, "The jackknife and the bootstrap for general stationary observations," *The Annals of Statistics*, vol. 17, no. 3, pp. 1217–1241, 1989.

[78] R. Tibshirani, "Variance stabilization and the bootstrap," *Biometrika*, vol. 75, no. 3, pp. 433–444, 1988.

[79] J.P. Shaffer, "Multiple hypothesis testing," *Annual review of psychology*, vol. 46, no. 1, pp. 561–584, 1995.

[80] S. Dudoit and M.J.V.D. Laan, *Multiple testing procedures with applications to genomics*, Springer, 2008.

[81] P.H. Westfall and S.S. Young, *Resampling-based multiple testing: Examples and methods for p-value adjustment*, vol. 279, Wiley. com, 1993.

[82] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian journal of statistics*, pp. 65–70, 1979.

[83] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 289–300, 1995.

[84] S.H. Wu, C.Y. Yang, and D.H.T. Huang, "Cooperative sensing of wideband cognitive radio: a multiple-hypothesis-testing approach," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1835–1846, 2010.

[85] P.J. Chung, J.F. Bohme, C.F. Mecklenbrauker, and A.O. Hero, "Detection of the number of signals using the benjamini-hochberg procedure," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2497–2508, 2007.

[86] A.M. Zoubir and J.F. Bohme, "Bootstrap multiple tests applied to sensor location," *IEEE Transactions on Signal Processing*, vol. 43, no. 6, pp. 1386–1396, 1995.

[87] P. Ray and P.K. Varshney, "False discovery rate based sensor decision rules for the network-wide distributed detection problem," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 3, pp. 1785–1799, 2011.

[88] P. Ray and P.K. Varshney, "Distributed detection in wireless sensor networks using dynamic sensor thresholds," *International Journal of Distributed Sensor Networks*, vol. 4, no. 1, pp. 5–12, 2008.

[89] E.B. Ermis and V. Saligrama, "Distributed detection in sensor networks with limited range multimodal sensors," *IEEE Transactions on Signal Processing*, vol. 58, no. 2, pp. 843–858, 2010.

[90] E.B. Ermis and V. Saligrama, "Detection and localization in sensor networks using distributed FDR," in *40th Annual Conference on Information Sciences and Systems*. IEEE, 2006, pp. 699–704.

[91] E.B. Ermis and V. Saligrama, "Adaptive statistical sampling methods for decentralized estimation and detection of localized phenomena," in *Fourth International Symposium on Information Processing in Sensor Networks (IPSN)*. IEEE, 2005, pp. 143–150.

[92] Z. Quan, S. Cui, A.H. Sayed, and H.V. Poor, "Optimal multiband joint detection for spectrum sensing in cognitive radio networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1128–1140, 2009.

[93] G. Atia, E.B. Ermis, and V. Saligrama, "Robust energy efficient cooperative spectrum sensing in cognitive radios," in *14th Workshop on Statistical Signal Processing (SSP)*. IEEE, 2007, pp. 502–506.

[94] Y. Benjamini and Y. Hochberg, "On the adaptive control of the false discovery rate in multiple testing with independent statistics," *Journal of Educational and Behavioral Statistics*, vol. 25, no. 1, pp. 60–83, 2000.

[95] S.C.A. Thomopoulos, R. Viswanathan, and D.C. Bougoulias, "Optimal decision fusion in multiple sensor systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-23, pp. 644–6653, 1987.

[96] A.G. Tartakovsky, "An efficient adaptive sequential procedure for detecting targets," in *IEEE Aerospace Conference Proceedings*. IEEE, 2002, vol. 4, pp. 4–1581.

[97] A.G. Tartakovsky, X.R. Li, and G. Yaralov, "Sequential detection of targets in multichannel systems," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 425–445, 2003.

[98] R.J. Serfling, *Approximation theorems of mathematical statistics*, vol. 162, Wiley. com, 2009.

[99] S.J. Shellhammer, "Spectrum sensing in IEEE 802.22," *IAPR Wksp. Cognitive Info. Processing*, pp. 9–10, 2008.

[100] B. Efron, "Better bootstrap confidence intervals," *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 171–185, 1987.

[101] C. Kirch, "Bootstrapping sequential change-point tests," *Sequential Analysis*, vol. 27, no. 3, pp. 330–349, 2008.

[102] W.Y. Lee and I.F. Akyildiz, "Optimal spectrum sensing framework for cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 7, no. 10, pp. 3845–3857, 2008.

[103] V.P. Draglia, A.G. Tartakovsky, and V.V. Veeravalli, "Multihypothesis sequential probability ratio tests. I. asymptotic optimality," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2448–2461, 1999.

[104] V.P. Dragalin, A.G. Tartakovsky, and V.V. Veeravalli, "Multihypothesis sequential probability ratio tests. II. accurate asymptotic expansions for the expected sample size," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1366–1383, 2000.

[105] A. Tartakovsky and X.R. Li, "Sequential testing of multiple hypotheses in distributed systems," in *Proceedings of the Third International Conference on Information Fusion (FUSION)*. IEEE, 2000, vol. 2, pp. THC4–17.

[106] 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*. IEEE, 2005, vol. 1, pp. 464–469.

[107] S.A. Kassam, *Signal detection in non-Gaussian noise*, Springer-Verlag, 1988.

[108] M. Ghogho, M. Cardenas-Juarez, A. Swami, and T. Whitworth, "Locally optimum detection for spectrum sensing in cognitive radio," in *4th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*. IEEE, 2009, pp. 1–6.

[109] M. Abramovich and I.A. Stegun, *Handbook of mathematical functions*, Dover publications, 1972.

[110] C.C. Lee and J.J. Chao, "Optimum local decision space partitioning for distributed detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, no. 4, pp. 536–544, 1989.

[111] H.V. Poor and J.B. Thomas, "Applications of Ali-Silvey distance measures in the design of generalized quantizers for binary decision systems," *IEEE Transactions on Communications*, vol. COM-25, pp. 893–900, 1977.

[112] S. Haykin, *Communication systems*, John-Wiley, 2001.

Curriculum vitae

Name:	Fiky Y. Suratman
Date of birth:	02.05.1976
Place of birth:	Jakarta, Indonesia
Family status:	Married
Nationality:	Indonesian
e-mail:	fiky.y.suratman@ieee.org

Education

09/2003 – 08/2006	School of Electical Engineering and Informatics (STEI) Bandung Institute of Technology (ITB), Indonesia (Master)
09/1993 – 08/1998	Engineering Physics Bandung Institute of Technology (ITB), Indonesia (Bachelor)

Work experience

10/2007 – Now	Faculty member, Telkom Engineering School, Telkom University, Bandung, Indonesia
02/2005 – 10/2007	Faculty member, Computer Engineering Department, Indonesian Computer University, Bandung, Indonesia.
09/2004 – 07/2006	Research Assistant, School of Electical Engineering and Informatics, ITB, Bandung, Indonesia.
05/2002 – 10/2002	Technical Support, Wira Eka Bhakti, P.T., Bandung, Indonesia.
12/1998 – 01/2001	Test Engineer, Advanced Interconnect Technology (now Unisem), P.T., Batam, Indonesia.

Erklärung laut §9 der Promotionsordnung

Ich versichere hiermit, dass ich die vorliegende Dissertation allein und nur unter Verwendung der angegebenen Literatur verfasst habe. Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Darmstadt, 19.11.2013