Channel Activity Statistics Estimation in Spectrum Sharing Systems Based on Imperfect Spectrum Sensing

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Ogeen Hanna Toma

May 2022

Abstract

As we are stepping into the era of beyond 5G, the demand for frequency bands will increase significantly to accommodate the fast growing tendency in wireless communications technology. Spectrum sharing is one of the promising solutions to overcome the frequency scarcity problem and maximise spectrum utilisation efficiency. Its consideration can be seen in the recent ongoing deployment of 5G as in 5G New Radio Unlicensed (5G NR-U). The harmonious coexistence of several wireless systems in a shared frequency spectrum is highly dependent on making effective decisions for the utilisation of such spectrum. These decisions are usually based on the users' traffic activity within the channel and their statistical information. Therefore, it is crucial for a spectrum sharing system to accurately obtain channel activity statistics. Although spectrum sensing is used in such systems to determine the instantaneous state of the channel, sensing decisions can further be exploited to provide a broad range of statistical information of the channel activity. However, spectrum sensing is imperfect in real world which therefore leads to inaccurate estimation of the channel activity statistics.

In this context, this thesis studies the problem of estimating the channel activity statistics under (realistic) Imperfect Spectrum Sensing (ISS) and it finds mathematical relationships (in closed-form expressions) between the observed channel activity statistics under ISS and their corresponding actual statistical information, as a function of relevant operating parameters including the probability of sensing errors, the employed sensing period and the sample size. Such problem is poorly addressed in the literature, without deep and rigorous mathematical analyses taking into account all the factors that would influence the estimation accuracy of the channel activity statistics. Then, the thesis investigates different approaches that can be used to improve the estimation of the channel activity statistics under ISS, namely the *closed-form expression approach*, which is based on the obtained mathematical expressions for these statistics; the *algorithmic approach*, which is based on reconstruction algorithms; and finally a *Traffic Learning (TL) approach*, which is based on deep learning techniques. It is shown that the proposed estimation methods in this thesis outperform the existing methods in the literature without requiring any prior knowledge of the channel activity. The correctness of the obtained analytical expressions and proposed estimation methods are corroborated with both simulation and experimental results, for which a USRP-based prototype is developed as an experimental platform to validate the theoretical analyses conducted for the estimation of the channel activity statistics in spectrum sharing systems.

Acknowledgements

As I look back on the whole journey of my PhD, especially the strange period of the pandemic, I remember all the ups and downs I went through and the challenges that made me grow as a researcher and as an individual as well. I recall the many people who have guided and supported me throughout this journey.

A special thank you to my supervisor Dr. Miguel López-Benítez for the guidance and advice throughout my PhD. I appreciate the continuous support received from him to build up my analytical skills and the whole learning process of formulating real-world problems mathematically and yet finding solutions.

A thank you must also go to my collaborators Dr. Dhaval Patel from India and Dr. Kenta Umebayashi from Japan for their valuable suggestions, comments and feedback on my research work and on our produced publications.

I am also grateful for the friendships I have made during my time here in Liverpool. To all my wonderful friends either from the university or outside the university I say thank you for your kindness and support.

Last but not least, to my family (my father Hanna, my mother Khairiya, my sister Rouaa and my brother Rivan), you are in every page, I cannot thank you enough, without your love and encouragement this journey would not be possible. I hope I am making you proud.

Contents

List of Figures

List of Tables

List of Abbreviations

Symbol	Description
3GPP	3rd Generation Partnership Project
5G	Fifth Generation
6G	Sixth Generation
AR	Augmented Reality
ASA	Authorised Shared Access
AWGN	Additive White Gaussian Noise
B5G	Beyond Fifth Generation
CBRS	Citizen Broadband Radio System
CDF	Cumulative Distribution Function
CNN	Convolutional Neural Network
CR	Cognitive Radio
CSAT	Carrier Sense Adaptive Transmission
CSS	Cooperative Spectrum Sensing
DC	Duty Cycle
DL	Deep Learning
DSA	Dynamic Spectrum Access
E	Exponential
ED	Energy Detection
ESC	Environmental Sensing Capability
FC	Fusion Centre
FCC	Federal Communications Commission
FR1	Frequency Range 1
FR2	Frequency Range 2
G	Gamma
GAA	General Authorized Access
GP	Generalised Pareto
GUI	Graphical User Interface
IED	Improved Energy Detection
IoT	Internet-of-Things
ISS	Imperfect Spectrum Sensing
KS	Kolmogorov-Smirnov

LBT	Listen Before Talk
LAA	License Assisted Access
LSA	Licensed Shared Access
LTE	Long Term Evolution
LTE-LAA	LTE-License Assisted Access
LTE-U	LTE-Unlicensed
M2M	Machine-to-Machine
MAE	Mean Absolute Error
MLP	Multilayer Perceptron
mmWave	millimeter Wave
MoM	Method of Moments
MSE	Mean Squared Error
NN	Neural Network
NR	New Radio
NR-U	New Radio Unlicensed
NRA	National Regulatory Agency
OOK	ON-OFF Keying
PAL	Priority Access License
PDF	Probability Density Function
PECAS	Prototype for the Estimation of Channel Activity Statistics
PMF	Probability Mass Function
PSS	Perfect Spectrum Sensing
PU	Primary User
QoS	Quality of Service
RNN	Recurrent Neural Network
RSPG	Radio Spectrum Policy Group
SAS	Spectrum Access System
SDR	Software Defined Radio
SNR	Signal-to-Noise Ratio
SS	Spectrum Sharing
SU	Secondary User
TL	Traffic Learning
TVWS	TV White Space
UHD	Ultra High Definition
USRP	Universal Software Radio Peripheral
VANET	Vehicular Ad Hoc Network
VR	Virtual Reality
W	Weibull
ZB	Zeta Byte

Chapter 1

Introduction

1.1 Background

Since the early civilisations, humans have sought innovative ways to communicate with one another. The history of the early communication means can be traced back to the pre-industrial age, where rudimentary techniques such as smoke signals, fires, and semaphore flags were used by the ancient people of China, North America and Europe in order to send important messages over far distances. Despite the simplicity of those techniques back then, they hold a similar sort of engineering trade-offs that are well known in our modern communication systems such as *transmission distance* and *data rate*. More complex systems of communications started to appear in the 18th century when Samuel Morse patented the electric telegraph in 1838 as the first electrical telecommunication system that uses electrical pulses, Morse code, to send text messages over long distances. The latter was gradually replaced after the world's first telephone was invented by Alexander Graham Bell in 1876, commencing a new era of voice communications over electrical wires. However, after the theoretical foundations of electromagnetic waves were presented by James Clerk Maxwell in 1864, it was not very long until Marconi was able to demonstrate the first radio transmission in 1895 between a transmitter and a receiver, hence radio communications was born.

Since then, the interests in radio technology have arisen rapidly to enable wireless communications support larger capacities (i.e., number of users, data rate, etc.). The initial systems of wireless communications were based on a large central transmitter that covers a wide geographical area. The challenges of wireless communications were soon realised, including the capacity limitation of the frequency spectrum. This capacity problem was then overcome when the cellular concept was developed by researchers

from Bell Laboratories [1], in which the fact that the power of the transmitted signals dissipate over distances is exploited in order to reuse the same signals frequency at spatially separated areas (cells) while maintaining minimal interference among them. The development of wireless communications continued from one generation to another to accommodate the increasingly growing number of wireless-connected devices and to support new emerging technologies. As we are stepping into the era of Beyond Fifth Generation (B5G)/6G of wireless communications, spectrum capacity is expected to grow massively to support data-hungry applications (e.g., 4k/8k Ultra High Definition (UHD) video traffic, Virtual Reality (VR), Augmented Reality (AR), real time mobile gaming, etc.) and the exponentially rising number of Internet-of-Things (IoT) devices [2–5]. It is expected that the global mobile traffic volume will exceed 5 ZB (Zeta Byte) per month by the end of 2030 from 0.062 ZB as in 2020, and the number of machine-to-machine (M2M) subscriptions will reach 97 billion in 2030 from 7 billion as of 2020 [6]. Such growing tendency also brings burdens of rising demand for frequency spectrum to support this tremendous number of interconnected terminals. This demand has been tried to be fulfilled in 5G by introducing Frequency Range 2 (FR2) [7] of millimeter wave (mmWave) bands to support the traditional wireless communications bands given by Frequency Range 1 (FR1) [8] (which include sub-6 GHz low and mid-frequency bands). Although there is significant unexploited bandwidth available in mmWave bands, moving to higher frequencies does not solve the problem of spectrum scarcity as it brings other challenges that might restrict its deployment as a standalone system. These challenges are significant in wireless communications, which include high signal attenuation (path loss) and low signal penetration (signal blockage by solid materials) [9], which can therefore limit mmWave's applications and deployments. This indeed makes FR1 still attractive and irreplaceable by introducing higher frequency bands. However, due to its limited bandwidth, the need for efficient and smart spectrum management strategies to maximise the utilisation efficiency and the exploitation of underutilised patterns of low and mid-frequency bands has become a crucial ongoing necessity. Solutions for efficient spectrum utilisation were first explored through Cognitive Radio (CR) concept [10,11] and then evolved into a more recent and wider context of Spectrum Sharing (SS) [12–15] systems.

1.2 Spectrum Sharing

Spectrum sharing was traditionally introduced two decades ago and received popularity in Cognitive Radio (CR) concept to enable Dynamic Spectrum Access (DSA) [16, 17],

which was proposed as a solution to overcome the underutilisation problem in the legacy management policy of fixed spectrum allocation. In DSA/CR systems, spectrum utilisation is improved by allowing unlicensed (secondary) users to exploit the idle spectrum (spectrum holes) opportunistically without causing any harmful interference with the licensed (primary) users. This approach has found its applicability, for example, targeting the television frequency spectrum (56–806 MHz) also known as TV White Space (TVWS), which has been implemented and standardised in the IEEE 802.22 standard [18]. Spectrum sensing is the key enabler of the DSA/CR systems, which enables Secondary Users (SUs) to monitor and access the primary channel based on the sensing observations of the Primary Users (PUs) activity.

The next deployment of spectrum sharing emerged in Europe [19] when the Authorised Shared Access (ASA) concept, also known as Licensed Shared Access (LSA), was introduced by Nokia and Qualcomm in 2011. The LSA concept was defined by the Radio Spectrum Policy Group (RSPG) as [20]:

"A regulatory approach aiming to facilitate the introduction of radio communication systems operated by a limited number of licensees under an individual licensing regime in a frequency band already assigned or expected to be assigned to one or more incumbent users. Under the LSA approach, the additional users are authorised to use the spectrum (or part of the spectrum) in accordance with sharing rules included in their spectrum rights of use, thereby allowing all authorised users, including incumbents, to provide a certain QoS".

In other words, spectrum sharing in LSA is allowed between two tiers of users, incumbents and additional users called LSA licensees (similar to primary and secondary concept), however, certain level of Quality of Service (QoS) and interference protection will be guaranteed for LSA licensees (unlike previous spectrum sharing approaches). The LSA architecture is depicted in Fig. 1.1, which shows the involvement of three stakeholders: incumbent, who has the primary rights to use the frequency band; LSA licensee, who agrees to the sharing conditions of the incumbent with guaranteed interference protection and QoS; and National Regulatory Agency (NRA), which controls and monitors the sharing process between incumbents and LSA licensees [21].

Another important initiative for the deployment of spectrum sharing came from the Federal Communications Commission (FCC) in the USA [22] by enabling the operation of the so-called Citizen Broadband Radio System (CBRS) in the frequency band 3.55–3.7 GHz. This sharing paradigm includes three tiers of users as shown in Fig. 1.2: incumbent, who has the top priority for using the spectrum; Priority Access License (PAL), who has licensed access to the spectrum exclusively in the absence of

Figure 1.1: Baseline LSA architecture [21].

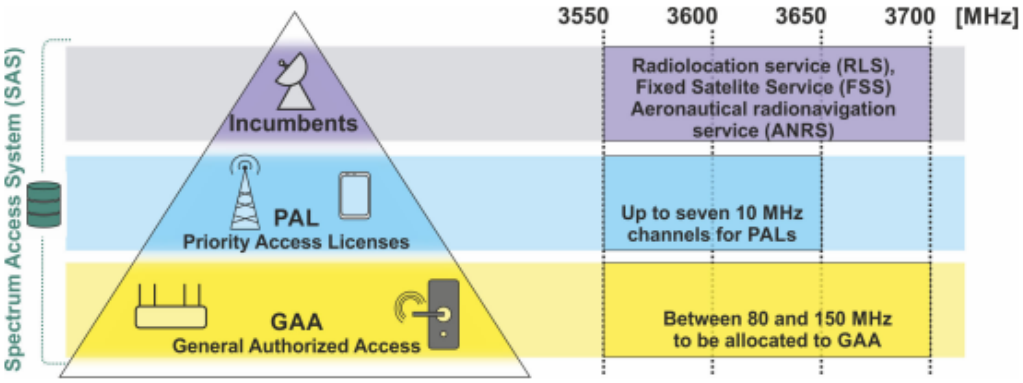


Figure 1.2: Three-tier model of SAS [24].

the incumbent; and General Authorized Access (GAA), who has unlicensed access to the spectrum in the absence of the incumbent assisted by sensing (similar to the traditional CR). These tiers are coordinated and managed by Spectrum Access System (SAS), in which the allocation of the spectrum to the PALs and GAAs is carried out. It is worth mentioning that spectrum sensing concept is adopted in SAS, also referred to as Environmental Sensing Capability (ESC), by which essential information on the incumbents activity and the actual available band is provided to the SAS [23]. A brief comparison between the two concepts (LSA and SAS) of spectrum sharing is provided in Table 1.1.

Note that, spectrum sensing, which was not part of the original LSA concept, has recently been considered in the evolved version of the LSA aiming towards a dynamic

Table 1.1: LSA and SAS comparison

LSA (EU version)	**SAS (USA version)**
Two-tier model: incumbents and LSA licensees.	Three-tier model: incumbents, PAL and GAA.
Static geo-location database (LSA repository), where spectrum usage information is provided by the incumbents as a static priori information.	Dynamic geo-location database, where spectrum usage information is obtained by employing spectrum sensing.
Incumbent protection is provided through database.	Incumbent protection is provided through spectrum sensing.
Frequency band: 2.3-2.4 GHz.	Frequency band: 3.55-3.7 GHz.

operation as proposed in the EU-funded project ADEL (Advanced Dynamic spectrum 5G mobile networks Employing Licensed shared access) [26].

More recently, new interests in spectrum sharing concept have arisen targeting the use of Long Term Evolution (LTE) systems over unlicensed spectrum bands. These were focused on a fair and harmonious coexistence between LTE and Wi-Fi networks in the unlicensed 5 GHz band, aiming to increase the LTE network capacity without degrading the performance of the existing WiFi networks. Two main standards have been introduced in this context: LTE-Unlicensed (LTE-U) proposed by LTE-U Forum [27] and LTE-License Assisted Access (LTE-LAA) proposed by 3rd Generation Partnership Project (3GPP) Release 13 [28]. LTE-U was the first version allowing LTE operation over unlicensed band, in which the duty cycle is adjusted by LTE networks combined with Carrier Sense Adaptive Transmission (CSAT) to allow coexistence between LTE and WiFi in 5 GHz band. However, this was improved through LTE-LAA, which behaves similar to WiFi node by using Listen Before Talk (LBT) protocol, which can achieve a better fairness in the coexistence between LTE and WiFi. In both approaches sensing is an important element for achieving this coexistence mechanism between LTE and WiFi networks. Nowadays, 3GPP efforts in 5G New Radio (NR) have been extended targeting the unlicensed bands at 5 GHz and 6 GHz to introduce the so-called 5G New Radio Unlicensed (5G NR-U) as specified in Release 16 [29]. The future specification will also target the unlicensed millimeter wave (mmWave) at 60 GHz bands [30–32].

Spectrum sharing has been constantly evolving since its traditional DSA/CR concept to meet the ever-increasing demand for the frequency spectrum. However, re-

gardless of the sharing method used, spectrum sensing has been a crucial and essential part of spectrum sharing systems, which has accompanied and supported most of the emerging concepts of spectrum sharing to provide information and awareness about the activity of the spectrum.

1.3 Spectrum Awareness

The function of spectrum awareness in spectrum sharing systems is to provide environmental information of the surrounding radio activity in the spectrum. This awareness includes the knowledge of when (the time instants) and where (the frequency channels) the spectrum would be occupied or unoccupied, which in turn will support spectrum sharing users to utilise the spectrum efficiently. Several techniques can be used to obtain spectrum awareness, including spectrum sensing [33], geolocation databases [34] and beacon signals [35]. Geolocation databases and beacon signals require licensed systems to provide information of their occupancy patterns and therefore the deployment of these methods would require external systems, which may lead to additional cost and compatibility issues with the legacy systems. Spectrum sensing, on the other hand, senses and observes the spectrum occupancy independently without the need of any external system to provide spectrum awareness, thus it is simpler and more attractive for deployment than the other methods [36]. However, the information obtained by means of spectrum sensing can be affected by the practical impairments introduced by the radio channel propagation and therefore it can be less accurate and reliable. In this work, spectrum sensing is considered for spectrum sharing systems as discussed in the next section.

Spectrum Sensing

Spectrum sensing was first introduced in DSA/CR systems to enable SUs detect the activities (i.e., idle/busy states) of the PUs within the frequency channel in order to exploit the unoccupied spectrum in the time and frequency domains (i.e., spectrum holes). It also continues to play the same essential role in the recent emerging spectrum sharing scenarios whereby the unlicensed users can autonomously obtain environmental information of the existing users (e.g., incumbents) in the frequency channels, without any prior negotiation. The main challenge of spectrum sensing is the accuracy of sensing. This accuracy is primarily dependent on the Signal-to-Noise Ratio (SNR) of the users' signal being sensed in the channel. As such, Perfect Spectrum Sensing (PSS) can be

assumed (i.e., without sensing errors) under sufficiently high SNR conditions. However, in practice, spectrum sharing systems are more likely to operate in moderate/low SNR environments and sensing errors are likely to occur due to the wireless channel impairments (e.g., noise, multipath fading and shadowing), thus Imperfect Spectrum Sensing (ISS) is a more realistic scenario. Two types of sensing errors can be identified under ISS: *false alarms*, where an idle state of the channel is sensed as a busy state, and *missed detections*, where a busy state of the channel is sensed as an idle state. There have been significant research efforts in the literature to improve the accuracy of spectrum sensing methods and algorithms. These methods vary in their complexities and sensing capabilities, which include:

- Energy Detection (ED) [37]: ED is the simplest and most widely used method for spectrum sensing, by which the channel is reported as busy (i.e., occupied by other users) if the detected signal energy is higher than a predefined threshold, or as idle otherwise. ED method detects the presence/absence of the users irrespective of their signal characteristics (e.g., modulation type). Therefore, ED is a suitable method when spectrum sharing systems have no prior information about the licensed users or when low complexity sensing is required. However, this method is affected by the uncertainty of the noise power, which results in a degradation in the sensing performance [38]. More accurate and advanced ED algorithms have also been proposed in the literature such as Improved Energy Detection (IED) [39], adaptive threshold energy detection [40] and more recently Machine Learning aided energy detection [41].

- Matched filter [42]: this method provides the optimal detection performance in Additive White Gaussian Noise (AWGN), in which the received SNR can be maximised. However, matched filter detection method requires prior knowledge about the characteristics of the signals being detected in the spectrum, which include the modulation type, pulse shape, packet format, etc. [43].

- Cyclostationary detection [44]: this method exploits the cyclostationarity in the modulated signals in order to identify the presence of the users' signals in the spectrum. Such features result from the embedded sinusoidal carriers, pulse trains, repeating spreading, hopping sequences or cyclic prefixes [45]. As such, this property cannot be found in the stationary noise or interference signal. Therefore, cyclostationary signal detection method performs better than ED under low SNR conditions, however, it is more complex and it requires the cyclic spectral correlation function to be known to detect the licensed signals.

- Covariance-based detection [46–48]: this method does not require any prior knowledge about the licensed users' signals in the spectrum sharing systems. However, it depends on the assumption that users' signal are autocorrelated, which can then be exploited to distinguish these (correlated) signals of the licensed users from the (uncorrelated) noise. This method provides higher detection accuracy than ED when signals are correlated (and they are usually in practice or can be made otherwise), however, signal covariance is computationally complex to find.

In addition to the above discussed methods, spectrum sensing can also be classified into two operational mechanisms: non-cooperative and cooperative sensing.

i. Non-Cooperative Spectrum Sensing: in which the unlicensed user independently senses the activity (i.e., presence/absence) of the licensed users in the frequency channel and makes a local decision (either idle or busy) on the state of the channel.

ii. Cooperative Spectrum Sensing: in which multiple unlicensed users share their local sensing observations with a common receiver, the Fusion Centre (FC), in order to make a better global decision about the presence/absence of the licensed users in the channel. Cooperative spectrum sensing can significantly improve the detection performance by exploiting the spatial diversity of the cooperating users.

Regardless of the spectrum sensing method and the cooperative mechanism used, spectrum sensing objective and output is the same, which is that to provide binary decisions on the state of the channel, either $\mathcal{H}_0$ for idle state or $\mathcal{H}_1$ for busy state.

1.4 Channel Activity Statistics Estimation

Since the activity of the users in a spectrum sharing system would vary over both time and frequency, it is very critical and inefficient for an unlicensed user to coexist and access the spectrum without having a proper knowledge about the activity patterns of the channel. Such knowledge can be acquired from exploiting the statistical information of the spectrum activity obtained from spectrum sensing. Although the main purpose of spectrum sensing is to determine the instantaneous state of the channel (i.e., either idle $\mathcal{H}_0$ or busy $\mathcal{H}_1$), sensing observations can further be exploited to provide a broad range of statistical information about the channel activity. These statistics include, but are not limited to, the minimum, mean and variance of the idle/busy periods, the duty cycle of the channel occupancy as well as the distribution of the idle/busy periods. The idle and busy periods of the channel activity can be estimated from spectrum sensing

decisions such that the elapsed time between any two changes in the sensing decisions ($\mathcal{H}_0/\mathcal{H}_1$) can provide an estimation for the original idle/busy periods' durations, which in turn are used to calculate the statistical parameters of the channel.

Since channel statistics in spectrum sharing systems can be estimated autonomously from the observations of spectrum sensing, they may differ from the original statistics of the channel. Under high SNR conditions (i.e., under PSS), the estimation accuracy of the channel statistics is affected to some extent by the time resolution of the employed sensing period T_s [49]. On the other hand, under low SNR conditions (i.e., under ISS) channel statistics estimation is significantly corrupted due to the presence of sensing errors in the spectrum sensing observations in addition to the sensing resolution impact. Therefore, the impact of the employed parameters to perform spectrum sensing will be crucial on the estimation accuracy of the channel activity statistics. These parameters include the employed sensing period T_s (i.e., sensing resolution), the predefined probability of false alarm P_{fa} and probability of missed detection P_{md} (i.e., sensing errors), as well as the number of sensing decisions or number of idle/busy periods N considered for statistics estimation (i.e., sample size). In this context, this thesis studies the estimation of the channel activity statistics in spectrum sharing systems based on the realistic observations of spectrum sensing (i.e., ISS), regardless of the spectrum sensing method/algorithm used. In addition, it analyses the impact of the spectrum sensing parameters on the estimation accuracy of the channel activity statistics.

1.5 Motivation and Objectives

Channel activity statistics (also called channel traffic statistics) have recently gained increasing attention due to their remarkable role in the performance improvement of spectrum sharing systems, where they can also find a wide range of applications in wireless communication networks. The utilisation of these statistics include, but is not limited to:

- Spectrum prediction [50–53], in which the historical statistical data of the spectrum can be exploited, using various spectrum prediction algorithms (e.g., machine learning [51]), to predict or infer the future behaviour of the licensed users and thus their spectrum activity.
- Spectrum sensing [54], in which the statistical information of the channel can help in selecting a more accurate threshold for spectrum sensing algorithms (e.g., energy detection algorithm).

- Channel selection [55–58], in which spectrum sharing systems can exploit the statistical parameters of the channels' activity in order to select the most appropriate channel that can be offered to the unlicensed users.

- Radio resource management [59–61], in which the statistical information can also help spectrum sharing systems to make crucial decisions to mitigate the interference between licensed and unlicensed users, optimise the system performance and enhance the utilization of the spectrum.

From above, it is evident that channel activity statistics play an effective role in the performance of various applications in spectrum sharing systems. As a result, it is significantly crucial for such systems to obtain accurate estimation for these statistics especially under (a realistic) ISS scenario. In the literature, channel activity statistics have been analysed under PSS more comprehensively than under ISS as in [49, 62–68]. The majority of existing work where ISS is considered has mainly focused on the estimation of the channel duty cycle as in [69–71], paying less attention to other equally important statistical properties of the channel activity. Few studies (e.g., [72–74]) have also considered the mean of the channel idle/busy periods, but they are typically constrained to an exponential distribution to model the idle/busy periods, which is not a realistic assumption in practice [78]. Another approach of reconstruction method in the form of algorithms has been proposed in [75–77] to correct the estimation of the channel activity statistics under ISS. These works however suffer from the following limitations: 1) no closed-form expressions are provided for these statistics, only heuristic estimation methods in the form of algorithms, and 2) the employed reconstruction algorithms assume perfect knowledge of some of the channel statistical parameters (e.g., the minimum idle/busy period). Fig. 1.3 summarises all the possible approaches (including the proposed ones) that can be used for the estimation of the channel activity statistics.

The limitations of the previous works and the lack for a comprehensive study that analyses a wider range of channel activity statistics under ISS, without making any assumption about the activity of the channel, have motivated this work. Therefore, a detailed mathematical analysis for various statistical parameters of the channel is carried out under ISS. In addition, a relationship between the estimated statistics under ISS and the original statistics is provided in closed-form expressions without introducing any constraints or requiring any prior-knowledge on the channel activity. This mathematical analysis provides the basis for the formulation of novel methods for an accurate estimation of the channel activity statistics under realistic ISS operating conditions.

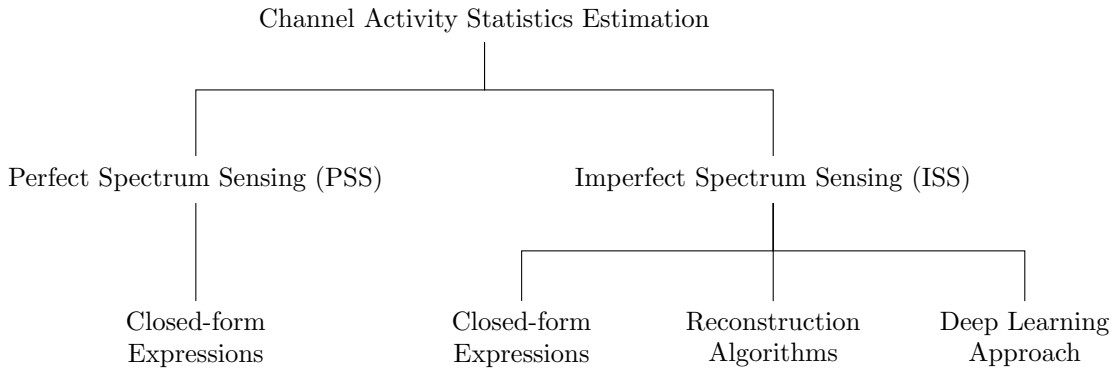


Figure 1.3: Estimation of the channel activity statistics using different approaches.

1.6 Thesis Contribution

This thesis addresses a significant and a highly challenging problem in spectrum sharing systems, which is that of obtaining accurate statistical information about the activity of the spectrum based on realistic spectrum sensing (i.e., ISS). Such problem is poorly addressed in the literature without deep and rigorous mathematical analyses taking into account all the factors that would influence the estimation accuracy of the channel activity statistics. In addition, it aims to find solutions through proposing new methods and approaches that can correct the estimation of the channel activity statistics under any operational scenario of spectrum sensing (ISS or PSS). Hence, the contribution of this work can be summarised as follows (ordered based on significance):

1. A set of closed-form expressions is developed for the statistics calculated under ISS as a function of the real channel statistics (obtained from the actual channel activity), probabilities of sensing error, and the employed sensing period. The obtained expressions are useful for spectrum sharing systems since they can provide insights into how spectrum sensing configurations can affect the estimation of statistics in the presence of sensing errors. This contribution is studied in Chapter 3 and published in the journal number 2 and conference papers number 7 and 8 (from the thesis's list of publications in Section 1.8).

2. A set of novel estimation methods is proposed based on the achieved mathematical analyses, which can provide accurate estimation for the channel statistics even under high probabilities of sensing error. The proposed estimators outperform the conventional methods proposed in the literature (e.g., [69–71]) to estimate

the statistics of the channel activity. This contribution is studied in Chapter 4 and published in the journal number 2 and conference papers number 5 and 8 (from the thesis's list of publications in Section 1.8).

3. The impact of the sample size on the estimation of the channel activity statistics under ISS scenario is also studied, in which closed-form expressions for the required sample size of the idle/busy periods under ISS to achieve a targeted level of accuracy are found. This contribution is studied in Chapter 3 and published in the conference papers number 3 and 6 (from the thesis's list of publications in Section 1.8).

4. A novel reconstruction algorithm assisted by a closed-form expression (obtained for the estimation of the mean period) is proposed, which can reach the performance of the state of the art algorithm presented in the literature [77] and yet without requiring any additional knowledge about the channel activity, unlike all the existing reconstruction algorithms in the literature (e.g., [75–77]). This contribution is studied in Chapter 4 and published in the conference paper number 4 (from the thesis's list of publications in Section 1.8).

5. This thesis also introduces a new approach, named Traffic Learning (TL), as a deep learning approach to learn from the channel traffic observations under realistic ISS scenario in order to predict the actual statistical information of the channel traffic activity in spectrum sharing systems. This contribution is studied in Chapter 4 and published in the journal number 1 (from the thesis's list of publications in Section 1.8).

6. A simple yet efficient cooperative spectrum sensing approach is proposed, which takes into consideration not only the impact of sensing errors but also the impact of sensing resolution. Such consideration, to the best of the author's knowledge, has not been presented in the literature which can outperform the conventional approach of cooperative spectrum sensing (e.g., [79–81]) for achieving minimum interference and maximum utilisation in spectrum sharing systems. It is worthy to mention that this contribution was inspired while analysing the impact of the sensing resolution and sensing errors on the observations of the channel activity and, even though it does not particularly target the estimation of the channel activity statistics, can effectively contribute to more accurate estimations of the channel activity through a better performance of spectrum sensing. This contribution is studied in Chapter 2 and published in the conference paper number 1

(from the thesis's list of publications in Section 1.8).

7. A USRP-based prototype is developed to support experimental validation for the conducted theoretical analyses of the channel activity statistics under real-world conditions, in which sensing energy of the existing signals in the channel can be monitored instantaneously and real-time estimation of the channel activity statistics can be provided accordingly. This prototype demonstrates the feasibility of the contributions of this research in practical system implementations. This contribution is studied in Chapter 2 and applied to validate the analysis in Chapter 3. The contribution is also published in the conference paper number 2 (from the thesis's list of publications in Section 1.8).

1.7 Thesis Outline

The remainder of this thesis is organised as follows. First, Chapter 2 presents two preliminary research contributions that do not target specifically the problem of estimating the channel activity statistics but can contribute to provide more accurate estimations and performance assessment of the estimation accuracy in realistic scenarios. The first contribution investigates the problem of cooperative spectrum sensing and introduces a novel approach that considers the impact of both sensing errors and sensing resolution to achieve minimum interference and maximum utilisation in spectrum sharing systems. The proposed cooperative approach is only presented in Chapter 2 as a preliminary contribution to show how spectrum sensing can be enhanced, which can indirectly contribute to a more accurate estimation of the channel activity statistics studied in the other chapters. Then the second contribution of the chapter introduces a USRP-based prototype developed to support a wide range of experiments and provide an experimental validation for the theoretical analyses and methods proposed in this work for an accurate estimation of the channel activity statistics. This experimental platform will be widely used in Chapter 3 to demonstrate the accuracy and correctness of the obtained mathematical results. This is achieved by deploying experiments under realistic conditions of wireless communications and then apply the obtained estimation methods to validate them experimentally.

Chapter 3 analyses a broad range of channel activity statistics and finds a set of closed-form expressions for the statistics calculated under ISS as a function of the original statistics and relevant operating parameters including the probability of sensing errors (false alarms and missed detections) and the employed sensing period. In addi-

tion, the impact of the sample size on the estimation of the channel activity statistics under ISS is also studied, in which closed-form expressions for the required sample size under ISS to achieve a targeted level of accuracy of the idle/busy periods' statistics are found. The achieved mathematical expressions are validated by means of both simulations and experimental results.

Chapter 4, on the other hand, investigates three different approaches to correct the estimation of the channel activity statistics under ISS. The first approach is based on the closed-form expressions obtained in Chapter 3 in order to derive novel estimation methods for the channel activity statistics under ISS, without making any assumption about the activity of the channel. The second approach investigates the reconstruction method, for which a new reconstruction algorithm is introduced to correct the observation of the channel activity statistics under ISS, which can achieve the same accuracy of the latest reconstruction method in the literature without requiring any prior knowledge of the channel activity. Finally, this chapter introduces a novel approach, Traffic Learning (TL), as deep learning approach for obtaining accurate statistical information of the channel traffic in spectrum sharing systems. This approach learns from the imperfect observations of the channel activity statistics in order to predict their accurate estimations. All these approaches are compared for the estimation of the channel activity statistics under ISS in terms of their performance and complexity.

Finally, Chapter 5 summarises all the findings achieved in this thesis and draws a comprehensive conclusion about the outcomes of this study on the estimation of the channel activity statistics under ISS. Suggestions for future work are provided as well.

1.8 List of Publications

Journal Publications

1. **O. H. Toma** and M. López-Benítez, "Traffic Learning: a Deep Learning Approach for Obtaining Accurate Statistical Information of the Channel Traffic in Spectrum Sharing Systems," in *IEEE Access*, vol. 9, pp. 124324–124336, 2021.
   (This publication is related to Chapter 4 findings).

2. **O. H. Toma**, M. López-Benítez, D. K. Patel and K. Umebayashi, "Estimation of Primary Channel Activity Statistics in Cognitive Radio Based on Imperfect Spectrum Sensing," in *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2016–2031, April 2020.
   (This publication is related to Chapter 3 and Chapter 4 findings).

Conference Publications

1. **O. H. Toma** and M. López-Benítez, "Cooperative Spectrum Sensing: A New Approach for Minimum Interference and Maximum Utilisation," *2021 IEEE International Conference on Communications Workshops (ICC)*, Montreal, Canada, 2021, pp. 1–6.
   (This publication is related to Chapter 2 findings).

2. **O. H. Toma** and M. López-Benítez, "USRP-Based Prototype for Real-Time Estimation of Channel Activity Statistics in Spectrum Sharing," *2021 17th International Symposium on Wireless Communication Systems (ISWCS)*, Berlin, Germany, 2021, pp. 1–6.
   (This publication is related to Chapter 2 findings).

3. **O. H. Toma**, M. López-Benítez and D. K. Patel, "Analysis of the Sample Size Required for an Accurate Estimation of Primary Channel Activity Statistics under Imperfect Spectrum Sensing," *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, London, UK, 2020, pp. 1–6.
   (This publication is related to Chapter 3 findings).

4. **O. H. Toma**, M. López-Benítez, D. K. Patel and K. Umebayashi, "Reconstruction Algorithm for Primary Channel Statistics Estimation Under Imperfect Spectrum Sensing," *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, Korea (South), 2020, pp. 1–5.
   (This publication is related to Chapter 4 findings).

5. M. López-Benítez, **O. H. Toma**, D. K. Patel and K. Umebayashi, "Methods for Fast Estimation of Primary Activity Statistics in Cognitive Radio Systems," *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, Korea (South), 2020, pp. 1–5.
   (This publication is related to Chapter 4 findings).

6. M. López-Benítez, **O. H. Toma**, D. K. Patel and K. Umebayashi, "Sample Size Analysis of Energy Detection under Fading Channels," *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, Korea (South), 2020, pp. 1–6.
   (This publication is related to Chapter 3 findings).

7. M. López-Benítez, **O. H. Toma** and D. K. Patel, "Mathematical Models for the Accuracy of the Estimated Distribution of Primary Activity Times in Dynamic Spectrum Access Systems," *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, Korea (South), 2020, pp. 1–5.
(This publication is related to Chapter 3 findings).

8. **O. H. Toma**, M. López-Benítez, D. K. Patel and K. Umebayashi, "Primary Channel Duty Cycle Estimation under Imperfect Spectrum Sensing Based on Mean Channel Periods," *2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, 2019, pp. 1–6.
(This publication is related to Chapter 3 and Chapter 4 findings).

Chapter 2

Cooperative Spectrum Sensing and Experimental Prototype

2.1 Introduction

This chapter presents two preliminary research contributions that do not target specifically the problem of estimating the channel activity statistics but can contribute to provide more accurate estimations and performance assessment of the estimation accuracy in realistic scenarios. The first contribution investigates the problem of cooperative spectrum sensing and introduces a novel approach which can achieve minimum interference and maximum utilisation in spectrum sharing systems. The proposed cooperative approach[1] can effectively contribute to a more accurate estimation of the channel activity statistics through an enhanced performance of spectrum sensing methods. The second contribution of the chapter introduces a USRP-based experimental system prototype developed to support a wide range of experiments in the context of channel activity statistics estimation. This experimental platform will be widely used in Chapter 3 to demonstrate the accuracy and correctness of the obtained mathematical results as well as the improved estimation accuracy that can be obtained with the methods proposed in this thesis.

[1]Since cooperative spectrum sensing is not the scope of this thesis, it is only presented in this chapter as a preliminary research to show how spectrum sensing can be improved, which can eventually enhance the estimation of statistics studied in the other chapters.

2.2 Cooperative Spectrum Sensing

Cooperative Spectrum Sensing (CSS) [79–81] is a widely studied topic in spectrum sharing systems, which is capable of improving the detection accuracy of the users activity in the spectrum through taking advantage of spatial diversity of multiple cooperating unlicensed users performing spectrum sensing. In the context of DSA/CR systems, Secondary Users' (SUs) local sensing decisions are shared with a common receiver, the Fusion Centre (FC), to make a global decision about the presence/absence of a Primary User (PU) within a particular licensed channel. The global decision is made by the FC after combining the sensing data forwarded by the SUs. The combining methods in CSS are classified into two types: hard and soft combining. The hard combining approach is based on the binary local decisions of the SUs, while the soft combining approach is based on the detected signal energy itself at the SUs [82]. Soft combining usually provides a better performance at the expense of an increased signaling overhead. Regardless of which combining approach is used, CSS improves the accuracy of sensing of the PU's activity within a particular frequency channel by reducing the impact of sensing errors. CSS has been extensively researched in the literature for a long time and the majority of the works in the literature (e.g., [79, 82, 83]) focus on exploiting SUs' observations to reduce the impact of sensing errors, thus producing more accurate decisions about the state of the primary channel. The main aim of reducing the impact of sensing errors is to avoid collision between SUs and PUs as well as to maintain high spectral utilisation. In this section a different approach is investigated for exploiting SUs' observations in CSS. The new approach suggests that by performing spectrum sensing in every cooperating SU at a different time instant (i.e., rather than to be synchronised among all the SUs as in the conventional approach) will help reducing the impact of the overall sensing resolution. The proposed approach, to the best of the author's knowledge, has not been presented in the literature, which outperforms the conventional approach of CSS for achieving minimum collision between secondary and primary users and maximum spectral utilisation (i.e., minimum missed opportunities).

2.2.1 Collision Ratio

Before delving into pursuing the efficient way of utilising SUs' observations in cooperative spectrum sensing, a new metric $\mathcal{C}$ is defined to represent the collision ratio between a SU and a PU [85], which can be given by:

$$\mathcal{C} = \frac{T_c}{T_1}, \tag{2.1}$$

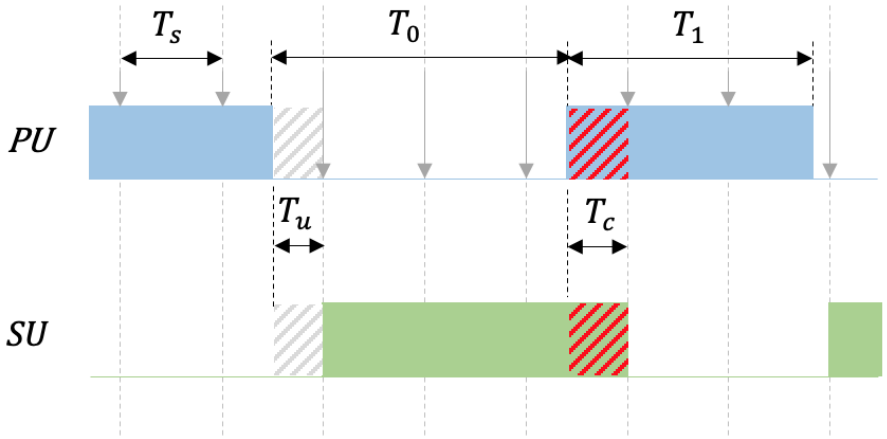


Figure 2.1: Collision and missed opportunity in PU and SU coexistence.

where T_c denotes the collision time between a SU and a PU, and T_1 denotes the busy time of the PU as illustrated in Fig. 2.1. The collision ratio $\mathcal{C}$ represents the fraction of time that a PU transmission is under interference from a SU. Note that $\mathcal{C} = 1$ when the collision time T_c equals the busy time T_1, which means there is 100% interference between the SU and the PU (which could only happen when the probability of missed detection is $P_{md} = 1$), while $\mathcal{C} = 0$ when the collision time is $T_c = 0$, which means there is no interference at all between the SU and the PU (which could only happen when $P_{md} = 0$ and the sensing period is $T_s = 0$).

Under PSS, collision between a SU and a PU results from a late detection of the PU's busy periods, which depends on the resolution of the sensing period T_s. Therefore, collision time could vary uniformly between 0 and T_s (i.e., $T_c \sim \mathcal{U}(0, T_s)$) and its expectation is $\mathbb{E}(T_c) = T_s/2$. As a result, the total collision ratio for a given set $\{T_{1,n}\}_{n=1}^{N}$ of N busy periods under PSS can be found as:

$$\mathcal{C}_{pss} = \frac{\sum_{n=1}^{N} T_{c,n}}{\sum_{n=1}^{N} T_{1,n}} = \frac{N\frac{T_s}{2}}{N\mathbb{E}(T_1)} = \frac{T_s}{2\mathbb{E}(T_1)}, \tag{2.2}$$

where $\mathbb{E}(T_1)$ represents the mean of the busy periods.

On the other hand, under ISS, collision between a SU and a PU results from the resolution of the sensing period T_s as well as the missed detection errors. Every missed detection error increases the collision time by T_s, except when a missed detection occurs at the end edge of a busy period where it only increases the collision time by an average

of $T_s/2$. Therefore, for N busy periods, the collision time under ISS can be found as:

$$\sum_{n=1}^{N} T_{c,n} = N\frac{T_s}{2} + N_{md}T_s - NP_{md}\frac{T_s}{2}, \tag{2.3}$$

where N_{md} denotes the number of the missed detection errors within N busy periods, and it can be found as:

$$N_{md} = \frac{N\mathbb{E}(T_1)}{T_s} \cdot P_{md}. \tag{2.4}$$

Thus, (2.3) can be written as:

$$\sum_{n=1}^{N} T_{c,n} = N\frac{T_s}{2} + N\mathbb{E}(T_1)P_{md} - NP_{md}\frac{T_s}{2}. \tag{2.5}$$

Finally, the collision ratio $\mathcal{C}$ under ISS can be found as:

$$\begin{aligned}
\mathcal{C} &= \frac{N\frac{T_s}{2} + N\mathbb{E}(T_1)P_{md} - NP_{md}\frac{T_s}{2}}{N\mathbb{E}(T_1)} \\
&= \frac{T_s}{2\mathbb{E}(T_1)} + P_{md} - \frac{P_{md}T_s}{2\mathbb{E}(T_1)} \\
&= \underbrace{\frac{T_s}{2\mathbb{E}(T_1)}}_{\text{Due to sensing period}} + \underbrace{P_{md}\left(1 - \frac{T_s}{2\mathbb{E}(T_1)}\right)}_{\text{Due to missed detections}}.
\end{aligned} \tag{2.6}$$

As it can be noticed from (2.6), collision ratio under ISS has two components contributed by the effect of the finite sensing period and the presence of missed detections, and they are directly proportional to the overall resulting collision ratio. Note that when $P_{md} = 0$ in expression (2.6), collision ratio $\mathcal{C}$ will be the same as (2.2) for PSS where only the sensing period has an impact. Therefore, in this work, (2.6) can be used as a general form expression for calculating the collision ratio $\mathcal{C}$. The correctness of the obtained collision ratio expression in (2.6) can be validated by means of simulations as discussed in Section 2.2.5.

2.2.2 Missed-Opportunity Ratio

Another metric is introduced for calculating the utilisation of the available opportunities in a shared spectrum. Missed-opportunity ratio $\mathcal{M}$ is here used to represent the fraction of the opportunistic periods that has not been exploited or has been missed by the SUs [85]. This fraction can be found as the ratio of the unexploited time to the available

idle time in a primary channel as:

$$\mathcal{M} = \frac{T_u}{T_0}, \tag{2.7}$$

where T_u denotes the unexploited time by an SU and T_0 denotes the idle time of the PU as illustrated in Fig. 2.1. Note that $\mathcal{M} = 1$ when the unexploited time T_u equals the idle time T_0, which means that SUs have not utilised any of the available opportunities in the primary channel and therefore they are 100% unexploited (which could only happen when the probability of false alarm is $P_{fa} = 1$), while $\mathcal{M} = 0$ when the unexploited time is $T_u = 0$, which means there is no missed opportunity at all or all the available opportunities have been exploited by the SUs (which could only happen when $P_{fa} = 0$ and $T_s = 0$). Also note that minimum missed-opportunity ratio $\mathcal{M}$ refers to maximum utilisation.

Under PSS, the available opportunities can be missed by a SU due to the late detection of the PU's idle periods, which depends on the resolution of the sensing period T_s. Therefore, the unexploited time could vary uniformly between 0 and T_s (i.e., $T_u \sim \mathcal{U}(0, T_s)$) and its expectation is $\mathbb{E}(T_u) = T_s/2$. On the other hand, under ISS, missed opportunities result from the resolution of the sensing period T_s as well as the false alarm errors. Every false alarm error increases the unexploited time by T_s, except when a false alarm occurs at the end edge of an idle period where it only increases the unexploited time by an average of $T_s/2$. As a result, the missed-opportunity ratio $\mathcal{M}$ can be found following the same analysis as for the collision ratio $\mathcal{C}$, by using P_{fa} and T_0 parameters instead of P_{md} and T_1 respectively, which yields:

$$\mathcal{M} = \underbrace{\frac{T_s}{2\mathbb{E}(T_0)}}_{\text{Due to sensing period}} + \underbrace{P_{fa}\left(1 - \frac{T_s}{2\mathbb{E}(T_0)}\right)}_{\text{Due to false alarms}}. \tag{2.8}$$

As it can be noticed from (2.8), missed-opportunity ratio under ISS has two components contributed by the effect of the finite sensing period and the presence of false alarms, and they are directly proportional to the overall resulting missed-opportunity ratio. The correctness of the obtained missed-opportunity ratio expression in (2.8) can also be validated by means of simulations as discussed in Section 2.2.5.

One can understand from both (2.6) and (2.8) that the impact of the collision ratio $\mathcal{C}$ and the missed-opportunity ratio $\mathcal{M}$ in DSA/CR systems can be reduced by adjusting T_s, P_{md} and P_{fa} parameters, while $\mathbb{E}(T_1)$ and $\mathbb{E}(T_0)$ are non-adjustable since they depend on the PU activity within the primary channel, which is assumed to be unknown to the DSA/CR system.

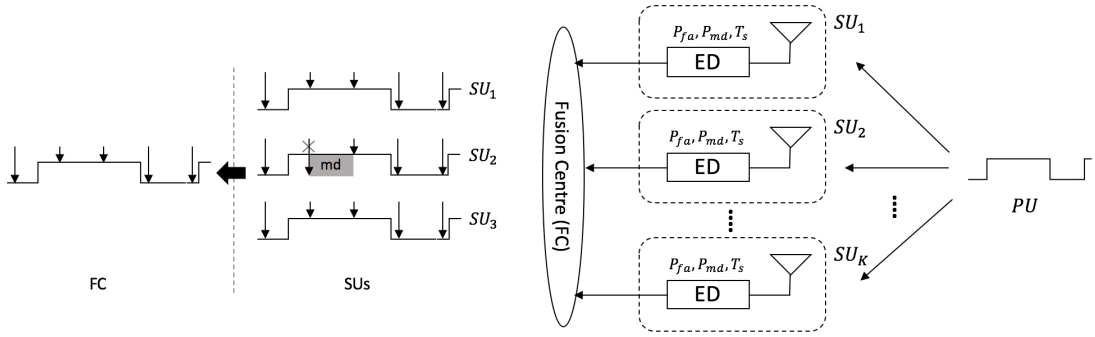


Figure 2.2: Conventional cooperative spectrum sensing approach.

Since the aim of this section is to investigate how SUs' observations in cooperative spectrum sensing can be exploited in such a way that minimum collision ratio and minimum missed-opportunity ratio (i.e., maximum utilisation) can be reached, the conventional approach of CSS is first introduced and its collision ratio and missed-opportunity ratio are analysed.

2.2.3 Conventional CSS Approach

Consider a single primary channel which is occupied by a single PU. A group of K SUs on the other hand perform spectrum sensing to monitor the activity of the primary channel. Spectrum sensing using, for example, Energy Detection (ED) method can be applied at each SU based on a predefined probability of error (i.e., P_{fa} and P_{md}), which can be assumed to be the same for all K SUs. In addition, it is assumed that the performed sensing events at the SUs are synchronised with a periodic sensing interval T_s. In the centralized common receiver FC, the sensing data forwarded by the SUs are combined to make a global decision about the presence of the PU. Either hard or soft combining method can be applied to combine SUs observations. Both combining methods aim to increase the accuracy of the final decision taken by the FC about the presence of the PU. This approach (shown in Fig. 2.2) is the widely considered approach in the literature for CSS and for which the collision ratio $\mathcal{C}$ and the missed-opportunity ratio $\mathcal{M}$ will be analysed.

Analysis of $\mathcal{C}$ and $\mathcal{M}$ Ratios for the Conventional CSS Approach

In this section, the hard combining method (using "n out of K" rule [84]) is considered to anlyse the collision and missed-opportunity ratios (similar analysis can also be applied for soft combining method). In hard combining, each SU produces a binary

decision about the status of the primary channel at each sensing event (where sensing events are synchronised for all SUs). Then a one-bit decision D_i for each sensing event is forwarded to the FC (where $D_i = 1$ stands for busy state and $D_i = 0$ for idle state of the PU). Since there are K SUs, the FC will receive K one-bit decisions made for the same sensing event from different SUs. Based on which a global decision can be made as hypothesis $\mathcal{H}_1$ if at least n out of K are 1s and hypothesis $\mathcal{H}_0$ otherwise [84]:

$$Y = \sum_{i=1}^{K} D_i \begin{cases} \geq n, & \mathcal{H}_1 \\ < n, & \mathcal{H}_0 \end{cases} \tag{2.9}$$

The overall probability of false alarm Q_{fa} and missed detection Q_{md} of a cooperative spectrum sensing scheme using such rule are found as [84]:

$$Q_{fa} = \sum_{l=n}^{K} \binom{K}{l} P_{fa}^{l} (1 - P_{fa})^{K-l}, \tag{2.10}$$

$$Q_{md} = 1 - \sum_{l=n}^{K} \binom{K}{l} P_{d}^{l} (1 - P_{d})^{K-l}, \tag{2.11}$$

where $P_d = 1 - P_{md}$, and the optimum n is found as [84]:

$$n_{opt} = \left\lceil \frac{K}{1+\alpha} \right\rceil, \quad \text{where } \alpha = \frac{\ln \frac{P_{fa}}{1-P_{md}}}{\ln \frac{P_{md}}{1-P_{fa}}}. \tag{2.12}$$

If $P_{fa} = P_{md}$, then $\alpha = 1$ and $n_{opt} = \lceil \frac{K}{2} \rceil$.

In comparison with P_{fa} and P_{md} predefined at each SU, the overall probabilities of false alarm and missed detection (i.e., Q_{fa} and Q_{md}) are significantly decreased as the number of the SUs (i.e., K) increases. As a result, the overall collision ratio $\mathcal{C}$ resulting from CSS using the conventional approach can be written based on (2.6) as:

$$\mathcal{C} = \frac{T_s}{2\mathbb{E}(T_1)} + Q_{md} \left(1 - \frac{T_s}{2\mathbb{E}(T_1)} \right), \tag{2.13}$$

where $Q_{md} \ll P_{md}$ for $K \gg 1$ and as a result the collision ratio in (2.13) is lower than that in (2.6).

Similarly, the overall missed-opportunity ratio $\mathcal{M}$ resulting from CSS using the

conventional approach can be written based on (2.8) as:

$$\mathcal{M} = \frac{T_s}{2\mathbb{E}(T_0)} + Q_{fa}\left(1 - \frac{T_s}{2\mathbb{E}(T_0)}\right), \tag{2.14}$$

where $Q_{fa} \ll P_{fa}$ for $K \gg 1$ and as a result the missed-opportunity ratio in (2.14) is lower than that in (2.8).

2.2.4 Proposed CSS Approach

As it can be noticed from (2.13) with reference to (2.6) and from (2.14) with reference to (2.8), CSS given by the conventional approach can only reduce the collision ratio $\mathcal{C}$ and the missed-opportunity ratio $\mathcal{M}$ by reducing the impact of sensing error (probability of missed detection in $\mathcal{C}$ and probability of false alarm in $\mathcal{M}$). As a result, the interest in the following two questions is motivated:

- Q1: Can CSS be exploited to reduce the collision ratio and the missed-opportunity ratio caused by the time resolution resulting from the employed sensing period T_s while keeping constant P_{fa}, P_{md} and T_s used at each SU? If so, what would be the method for such scheme?
- Q2: Which scheme would provide a lower collision ratio and a lower missed-opportunity ratio?

It is possible to reduce the impact of the collision ratio and the missed-opportunity ratio caused by the employed sensing period T_s (answer to Q1) by letting each SU to start sensing at a different time within T_s (i.e., unlike the previous approach, SUs' sensing events are not synchronised) [85]. A time difference of T_s/K can be allowed among the SUs' sensing time instants. In addition, a combining method will not be required at the FC since each received report from each SU represents new sensing information about the presence of the PU at a different time instant, which also reduces the complexity and computational requirements of the FC. Fig. 2.3 shows the proposed approach of the CSS using asynchronous sensing events[2] at the SUs, which is capable of reducing the collision ratio and the missed-opportunity ratio caused by the employed sensing period T_s. Note that τ_i represents the relative sensing time instants across the SUs. If SU_1 starts sensing at time $\tau_1 = 0$, then SU_2 starts at $\tau_2 = T_s/K$ and SU_3 starts at $\tau_3 = 2T_s/K$, where $K = 3$ in this example.

[2] Although synchronisation accuracy of the CSS is out of the scope of this work, it is worth mentioning that the conventional approach is more sensitive to the synchronisation error than the proposed one since its sensing events have to take place at the same time instant along with the other SUs.

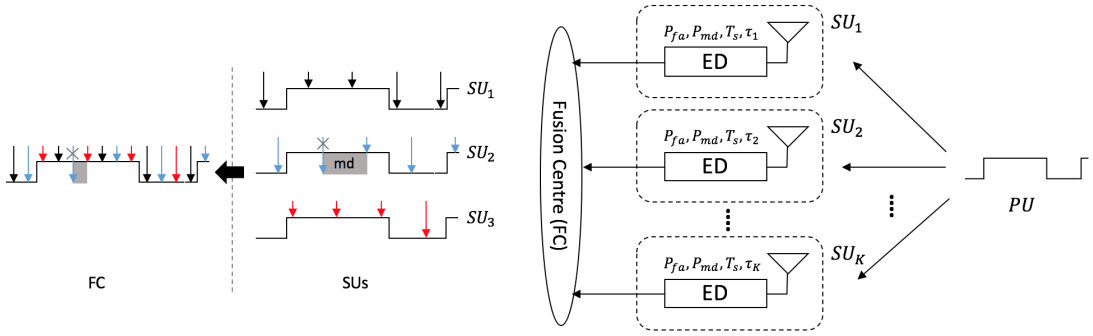


Figure 2.3: Proposed cooperative spectrum sensing approach.

Analysis of $\mathcal{C}$ and $\mathcal{M}$ Ratios for the Proposed CSS Approach

The proposed approach can reduce the overall resolution error of the sensing period T_s to T_s/K, which in turn will reduce the collision ratio $\mathcal{C}$ based on (2.6) to:

$$\mathcal{C} = \frac{T_s}{2K\mathbb{E}(T_1)} + P_{md}\left(1 - \frac{T_s}{2K\mathbb{E}(T_1)}\right), \tag{2.15}$$

and also will reduce the missed-opportunity ratio $\mathcal{M}$ based on (2.8) to:

$$\mathcal{M} = \frac{T_s}{2K\mathbb{E}(T_0)} + P_{fa}\left(1 - \frac{T_s}{2K\mathbb{E}(T_0)}\right). \tag{2.16}$$

As it can be noticed from (2.15) and (2.16), the proposed approach cannot decrease the impact of the sensing errors probabilities P_{md} and P_{fa} (opposite to the previous approach in (2.13) and (2.14) where P_{md} and P_{fa} were decreased to Q_{md} and Q_{fa}, respectively). This leads us to ask the second important question (Q2): which parameter is more significant to be decreased, sensing error or sensing resolution? The answer to this question is dependent on the values of T_s, P_{md} and P_{fa} themselves as well as the number of SUs K used in the CSS. Under PSS scenario, the conventional approach fails to mitigate the collision and missed-opportunity ratios since the probabilities of sensing errors are already zero under PSS and cannot be further reduced by increasing the number K of cooperating SUs. Meanwhile increasing the number of cooperating users in such a case under the proposed approach would still reduce the collision and missed-opportunity ratios resulting from the time resolution imposed by the sensing period T_s. On the other hand, the conventional approach would perform better in some ISS scenarios as P_{md} and P_{fa} increase since their impact on the collision ratio and missed-opportunity ratio becomes more severe than the time resolution imposed by

the sensing period T_s. However, a threshold can be obtained to decide which approach is more efficient to exploit SUs' observations in CSS in order to provide the lowest achievable collision ratio $\mathcal{C}$ and missed-opportunity ratio $\mathcal{M}$ (answer to Q2) as given by the following two selection criteria:

$$\mathcal{C}_c \underset{Conventional}{\overset{Proposed}{\gtrless}} \mathcal{C}_p, \tag{2.17}$$

$$\mathcal{M}_c \underset{Conventional}{\overset{Proposed}{\gtrless}} \mathcal{M}_p, \tag{2.18}$$

where $\mathcal{C}_c$ and $\mathcal{M}_c$ are the collision ratio and missed-opportunity ratio of the conventional approach based on (2.13) and (2.14), respectively, while $\mathcal{C}_p$ and $\mathcal{M}_p$ are the collision ratio and missed-opportunity ratio of the proposed approach based on (2.15) and (2.16), respectively. If $\mathcal{C}_c > \mathcal{C}_p$, the proposed approach should be selected. Otherwise, the conventional approach should be selected. The same rule applies when $\mathcal{M}_c > \mathcal{M}_p$.

2.2.5 Simulation Results

First of all, the obtained expressions (2.6) and (2.8) for collision ratio $\mathcal{C}$ and missed-opportunity ratio $\mathcal{M}$, respectively, are validated by means of simulation. In order to calculate the collisions and the missed opportunities in simulation a large number (10^6) of idle/busy periods of a PU is generated. The duration of these periods are modeled to follow a Generalised Pareto (GP) distribution, which provides the best representation for PU periods according to the experimental measurements in [78]. The distribution parameters of GP are configured as: location $\mu = 10$ t.u. (time units), scale $\lambda = 30$ t.u., and shape $\alpha = 0.25$. This configuration results in a sequence of PU periods that have a busy mean period $\mathbb{E}(T_1) = 50$ t.u., an idle mean period of $\mathbb{E}(T_0) = 50$ t.u., a minimum busy period of $\mu_1 = 10$ t.u. and a minimum idle period of $\mu_0 = 10$ t.u.. Spectrum sensing can then be performed on the generated periods using a sensing period T_s in order to obtain the sensing decisions that would be observed by a SU. Based on these decisions, SU's accessing/waiting periods can be computed. Therefore, the collision between SU (accessing/waiting periods) and PU (idle/busy periods) can be calculated in the simulation and compared with the theoretical expression obtained in (2.6). As shown in Fig. 2.4 (left), the calculated collision ratio $\mathcal{C}$ using (2.6) perfectly matches the simulation results (for both PSS and ISS) for different T_s values. In the same way, the missed-opportunity ratio $\mathcal{M}$ can be calculated from the simulation and compared with the theoretical expression obtained in (2.8) as shown in Fig. 2.4 (right), where a

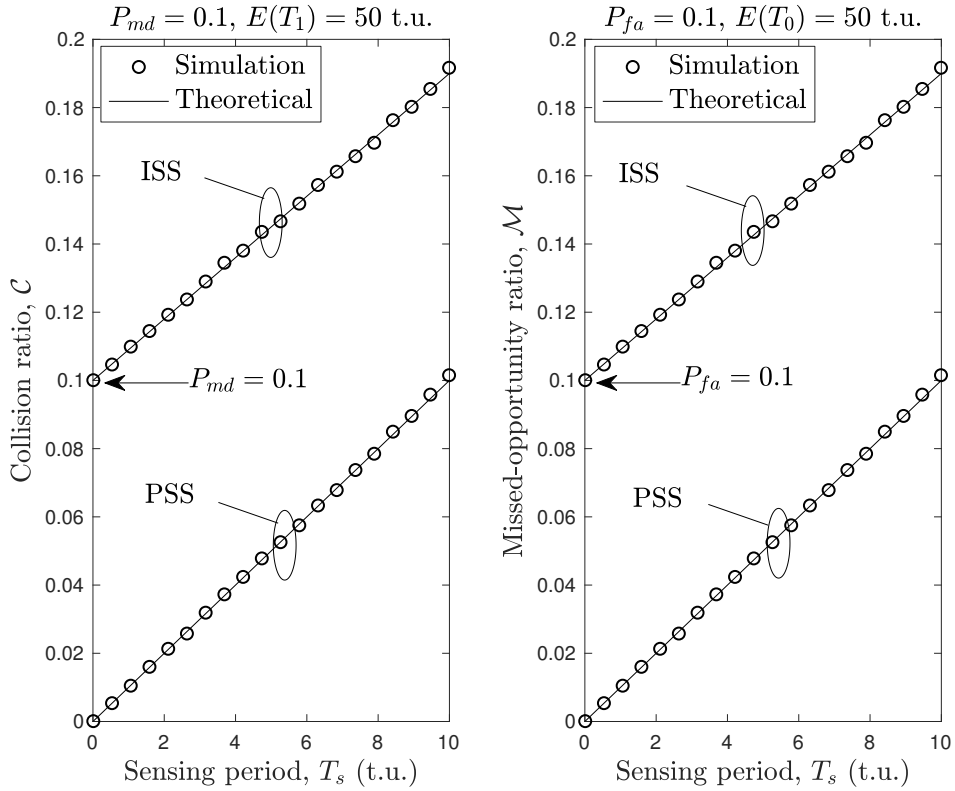


Figure 2.4: Collision ratio (left) and missed-opportunity ratio (right) as a function of T_s under PSS and ISS.

perfect match can also be observed, thus validating our analysis for $\mathcal{M}$ as well. Note that $\mathcal{C}$ and $\mathcal{M}$ in Fig. 2.4 show similar trends because the parameters P_{md} and $\mathbb{E}(T_1)$ (which control on the $\mathcal{C}$ ratio) are set similar to the ones P_{fa} and $\mathbb{E}(T_0)$ (which control on the $\mathcal{M}$ ratio), however, they are not necessarily to be the same in general.

On the other hand, to evaluate the collision ratio $\mathcal{C}$ and the missed-opportunity ratio $\mathcal{M}$ for the proposed approach of CSS with respect to the conventional approach, consider a CSS system with $K = 10$ SUs monitoring the idle/busy periods of the PU. The collision ratio $\mathcal{C}_c$ calculated in (2.13) based on the conventional approach, and the collision ratio $\mathcal{C}_p$ calculated in (2.15) based on the proposed approach can be evaluated and compared over different values of T_s and different values of P_{md} as shown in Fig. 2.5. As it can be noticed from this figure, when P_{md} is low or approaching zero, the collision ratio resulting from the proposed approach (the plane in blue colour) is significantly lower than the conventional approach (the plane in gray colour) for all T_s values. In contrast, the conventional approach performs better when P_{md} increases.

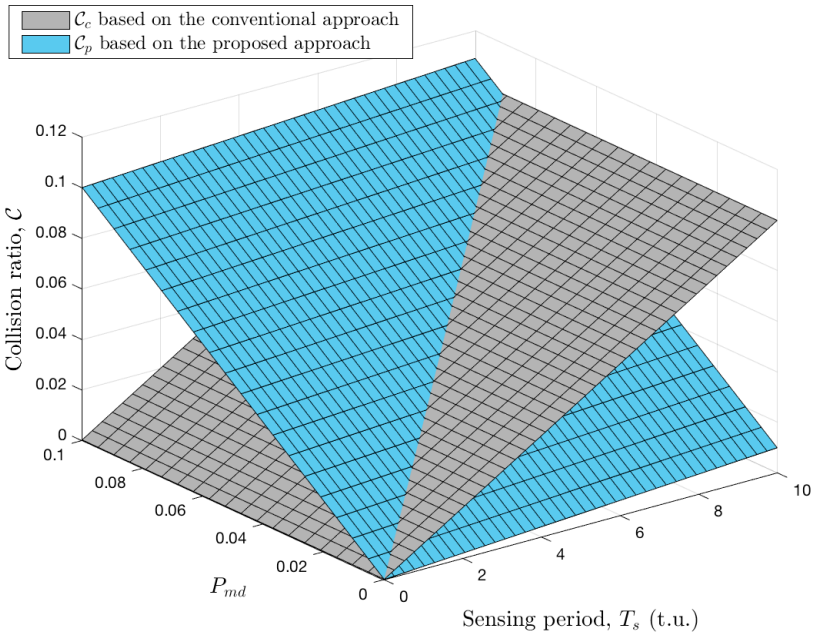


Figure 2.5: Collision ratio as a function of T_s and P_{md}, when $K = 10, \mathbb{E}(T_1) = 50$ t.u. and $\mu_1 = 10$ t.u..

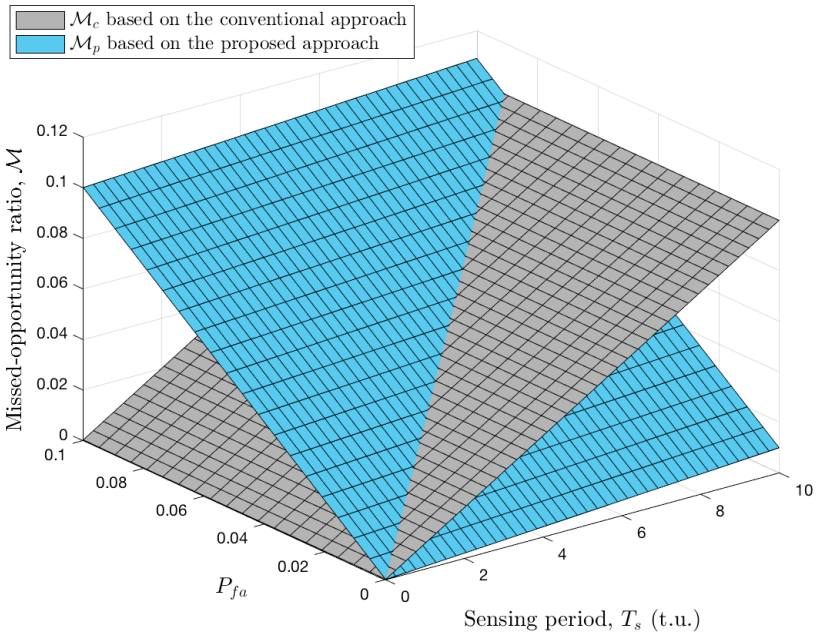


Figure 2.6: Missed-opportunity ratio as a function of T_s and P_{fa}, when $K = 10, \mathbb{E}(T_0) = 50$ t.u. and $\mu_0 = 10$ t.u..

In addition, when P_{md} is somewhere in the middle (e.g., 0.05), the collision ratio will be lower for the proposed approach when T_s is high, and will be higher otherwise. It can also be noticed that the conventional approach is not useful at all under PSS (i.e., when $P_{md} = 0$) because P_{md} can not be reduced any further, whereas the impact of the sensing period T_s can still be reduced through the proposed approach. Similar trends can also be observed for the calculated missed-opportunity ratio $\mathcal{M}_c$ based on the conventional approach and the calculated missed-opportunity ratio $\mathcal{M}_p$ based on the proposed approach over different values of T_s and different values of P_{fa} as shown in Fig. 2.6. It is worth mentioning that when P_{md} and P_{fa} are both low, the proposed approach can perform better in reducing both $\mathcal{C}$ and $\mathcal{M}$. In contrast, when P_{md} and P_{fa} are both high, the conventional approach would then perform better. In some scenarios, when P_{md} is low and P_{fa} is high (or vice versa), one approach would perform better than the other in reducing only one of the metrics ($\mathcal{C}$ or $\mathcal{M}$). As such, an approach can be selected based on what would be of most interest to a system to reduce (i.e., reducing $\mathcal{C}$ or $\mathcal{M}$). As a result, based on the parameters that are selected by the CSS (K, P_{md}, P_{fa} and T_s), it can easily be decided (using (2.17) and (2.18)) which approach is the most efficient one for mitigating the interference and maximising the utilisation of the spectrum in spectrum sharing systems.

2.3 Prototype for Channel Activity Statistics Estimation

In this section, a prototype is introduced that is used as a platform for carrying out experiments on the estimation of the channel activity statistics, which will serve as a proof-of-concept for the conducted theoretical analyses and proposed estimation methods in the following chapters. Channel activity statistics have been studied and analysed in the literature mainly theoretically. The work in [86] has presented a simple platform, Prototype for the Estimation of Channel Activity Statistics (PECAS), to validate the mathematical analyses of the channel activity statistics experimentally. PECAS model, however, has several hardware limitations in its transmitter and receiver. Its transmitter, which acts as a PU, is based on a Raspberry Pi with an ON-OFF Keying (OOK) modulator connected to it. The used OOK modulator can only operate at a central frequency of 433.92 MHz, thus limiting the experiments to such frequency band and making it impossible to conduct a wider range of experiments for different research purposes. In addition, the maximum modulation frequency supported by the modulator is 10 kHz, which limits the time resolution of the generated (PU activity) idle/busy periods to 0.1 ms. This limited time resolution is inconvenient since a reliable study

requires the transmission and reception of a sufficiently large number of idle/busy periods, which may take unreasonably long times for many experiments under such limited time resolution. The PECAS receiver, on the other hand, which acts as a SU, is based on the RTL-SDR platform, which supports a limited frequency range of 24 MHz–1766 MHz. Therefore, experiments on a higher frequency channel (e.g., 2.5 GHz and 5 GHz of the WLAN frequency bands) are not possible. In addition, the maximum sample rate that RTL-SDR can provide is 3.2 MS/s, which might not be sufficient when fast spectrum sensing is required.

While PECAS prototype is suitable for low-cost experiments on channel activity statistics, it might not be applicable on a wider range and sophisticated experiments. In addition, it has no capability of monitoring the channel activity statistics while the experiment is running in real-time (it only provides the statistical information after the execution of the experiment). In this context, this section presents a solution that overcomes the aforementioned limitations of the PECAS prototype by proposing a new sophisticated prototype based on the Universal Software Radio Peripheral (USRP) [87]. USRP is a readily available and widely used platform in the community, which enables other researchers and engineers to easily implement and reproduce our proposed system and benefit from its advantages (we provide free open source code in [88]). The new prototype enables monitoring the instantaneous sensing energy of the channel and provides real-time estimation of the channel activity statistics. It provides a high sample rate up to 56 MS/s, which enables a better time resolution for generating idle/busy periods at the transmitter (with 17.8 ns compared to 0.1 ms for PECAS) and a faster spectrum sensing at the receiver (30 times faster than PECAS). It can operate in a wide frequency range of 70 MHz–6 GHz (or even larger for some USRP models), which supports a wide range of experiments such as those in 5G wireless communications, while PECAS prototype is restricted to the experiments at a central frequency of 433.92 MHz. In addition, a Graphical User Interface (GUI) is developed for both transmitter and receiver to easily configure the USRP used in this prototype, which will also help other researchers to conduct various experiments without the need to modify the source code.

2.3.1 Prototype Overview

The proposed prototype consists of a transmitter and a receiver. For each, a host computer (PC) and a USRP are used as illustrated in the block diagram in Fig. 2.7 and its actual setup in Fig. 2.8. The transmitter acts as a PU, which transmits a sequence of

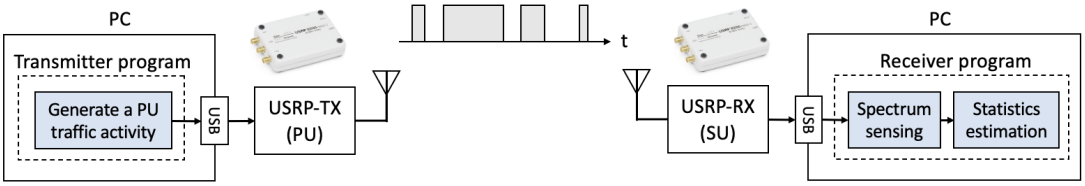


Figure 2.7: Block diagram of the proposed prototype.

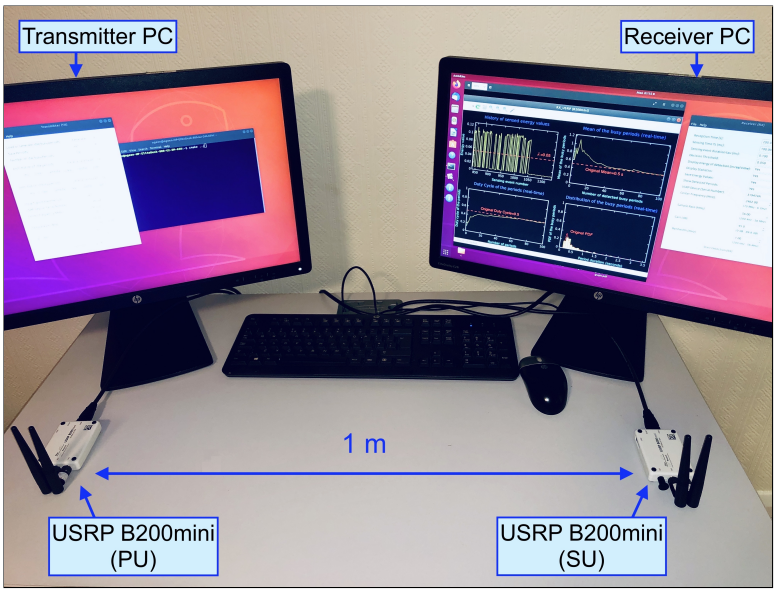


Figure 2.8: Experimental setup for the proposed prototype.

idle/busy periods (with known statistical parameters) in a particular frequency channel in order to generate a PU channel activity. The receiver, on the other hand, acts as a SU, which performs spectrum sensing (using ED algorithm) in the same channel with a periodic sensing time T_s. The energy of each sensing event is compared with a threshod to decide whether the channel is idle or busy. Sensing decisions can then be used to calculate the durations of the PU's idle/busy periods, based on which PU activity statistics can be estimated. By comparing the statistics of the generated periods at the transmitter (PU) with the estimated ones at the receiver (SU), it is possible to validate the accuracy of the new estimation methods and algorithms proposed in this thesis (including the previous ones in the literature such as [49, 63, 68, 75]) under a realistic conditions of wireless channel impairments (i.e., noise, path loss, shadowing and fading) and hardware limitations of the transmitter and receiver.

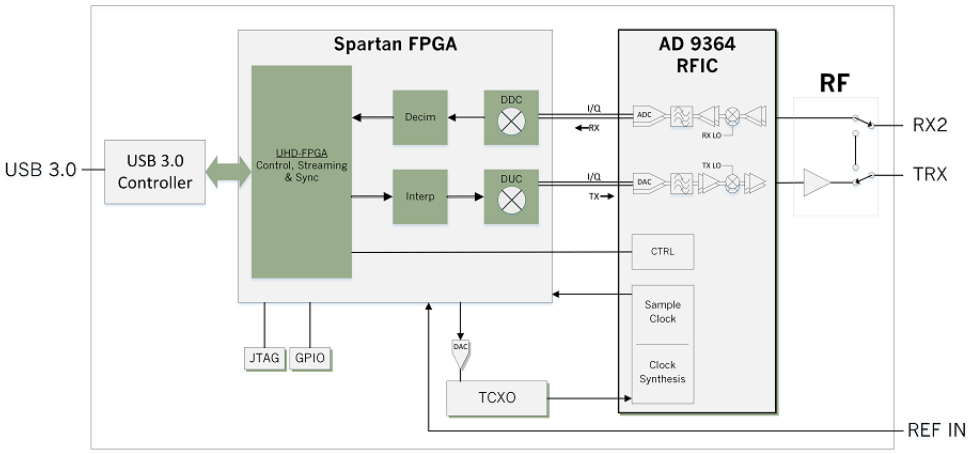


Figure 2.9: USRP B200mini block diagram [89].

2.3.2 Hardware implementation

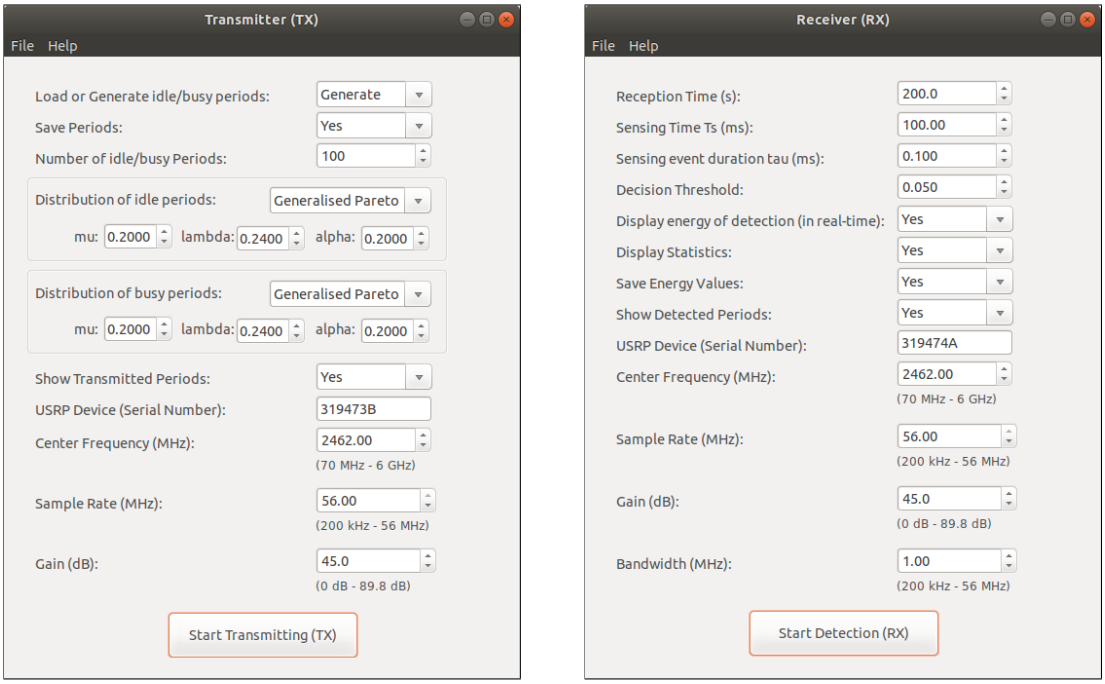
The hardware implementation of the proposed prototype comprises a USRP and a host PC at both sides of the system (transmitter and receiver). USRP is a Software Defined Radio (SDR) platform that is widely used to implement and prototype sophisticated radio communication systems. In this prototype, USRP B200mini Series [89] is adopted, which is a small form factor and easily portable USB-powered USRP with 1 Tx and 1 Rx front ends as shown in its block diagram in Fig. 2.9. This USRP supports a wide frequency range from 70 MHz to 6 GHz, which enables a wide range of experiments (e.g., FM and TV broadcast, cellular, Wi-Fi and etc.). The USRP front end filter has an adjustable bandwidth of 200 kHz - 56 MHz and an available gain up to 89.8 dB (for the transmit front end) and 76 dB (for the receive front end). In addition, the ADC/DAC of this USRP can provide a maximum sample rate (master clock) of 61.44 MS/s, however, rates above 56 MS/s are possible, but not recommended. The I/Q samples of the USRP are streamed to/from a host computer PC for additional processing through a high-speed USB 3.0 bus (which has a transmission speed of up to 5 Gbit/s). Meanwhile, the host PC adopted in this prototype has the following specifications: Ubuntu 18.04.3 LTS operating system, Intel Core i5-6500 CPU processor @ 3.20GHz and 8 GB memory. This host PC provides enough computational power to run a broad range of complex experiments in real-time with the developed software implementation, which is described in the following section.

2.3.3 Software implementation

Transmitter software

The software of the transmitter aims to configure the USRP in order to operate as a PU. A program is developed using C language on the host PC to communicate with the USRP. First, the USRP Hardware Driver (UHD) is required to be installed on the PC in order to provide all the necessary controls and libraries used to transport I/Q samples to/from USRP hardware. Then the USRP-transmitter is programmed such that it generates a sequence of idle/busy periods in a frequency channel to represent PU activity. The busy period durations T_1 can be generated by letting the USRP transmit a signal for a desired duration of time T_1. Any modulation scheme can be used for the transmitted signal since signal modulation is irrelevant when energy detection method is used at the SU [37] (the purpose of this signal is to generate an energy activity in the channel rather than to transmit data). However, for simplicity, ON-OFF Keying (OOK) modulation is used by streaming a binary 1 data (i.e., I=1 and Q=0) for a duration of T_1. The code implementation allows to easily add sequences of random bits and more sophisticated modulations if desired. The maximum sample rate (56 MS/s) of USRP B200mini allows representing continuous values of busy periods with a time resolution of 17.8 ns (which significantly improves the accuracy of the generated periods by using PECAS prototype [86] where the time resolution was 0.1 ms). The idle periods T_0, on the other hand, are produced by halting the transmission of the USRP for a duration of time T_0 using `nanosleep` function in C, in which an idle duration with nano seconds resolution is used to hold up the transmission before the next busy period.

The duration values of the idle/busy periods that are wanted to be transmitted can either be imported from a plain text file containing a list of pre-generated period durations or randomly generated by the transmitter program in real-time (in both cases the number of periods can be specified). If the latter is selected, the program will generate random durations of idle/busy periods based on a distribution selected from a list of distributions that provide an accurate representation for the empirical data in a real system [78]. This list includes Exponential, Generalised Exponential, Pareto, Generalised Pareto, Log-Normal, Gamma and Weibull distributions as shown in [78, Table I]. Note that a random value from any distribution can be obtained (based on the inversion method in [90, p. 28]) using uniform random number generator (e.g., `rand` function in C) and the inverse CDF [86]. A test mode is also included in the list of distributions to transmit a test sequence of 1 second idle and 1 second busy periods.



(a) Transmitter GUI (b) Receiver GUI

Figure 2.10: Designed GUI for the proposed USRP-based prototype.

Finally, a Graphical User Interface (GUI) is designed as shown in Fig. 2.10a to ease the configuration of the USRP transmitter and the created PU activity without the need to modify the original source code, which will also help other researchers to easily conduct relevant experiments on such platform for different research purposes.

Receiver software

Receiver software aims to configure the USRP in order to operate as a SU. Similar to the transmitter, a C program is developed on the receiver PC to control on the USRP via UHD library. The USRP-receiver is programmed such that it senses the energy activity of a frequency channel periodically every T_s sensing time (using ED algorithm); and then make a binary decision on whether the channel is idle or busy. At every sensing event a set of samples are captured from the desired channel for a time slot of τ as shown in Fig. 2.11. Note that τ must significantly be shorter than T_s such that the remaining time of $T_s - \tau$ would be reasonable to exploit in spectrum sharing systems (when the channel is idle). The number of samples N that can be captured during τ time slot depends on the sample rate of the USRP hardware and it

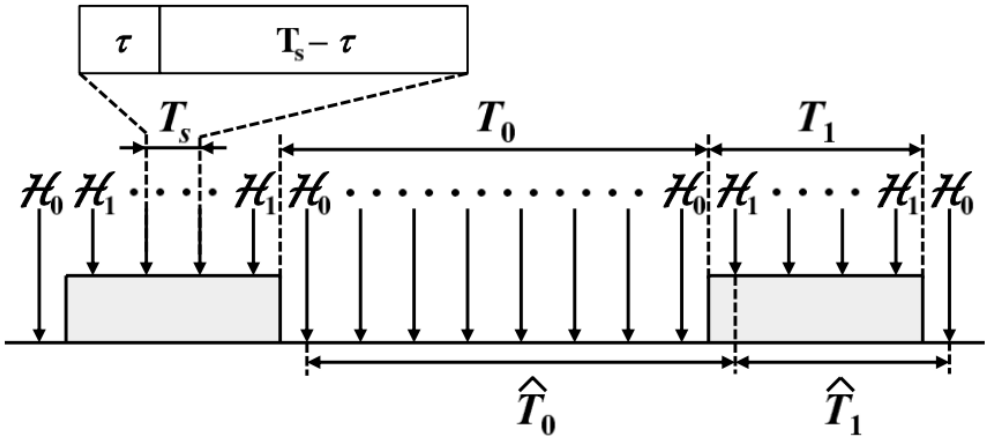


Figure 2.11: Spectrum sensing decisions [86].

is given by $N = \lceil \tau f_s \rceil$, where f_s is the sample rate configured for the USRP. Using the USRP's maximum sample rate of 56 MS/s enables capturing 1000 samples in $\tau = 17.8$ μs time (whereas PECAS model would require a time of $\tau = 312.5$ μs to capture the same number of samples, thus slowing down the process of spectrum sensing and the whole experiment, in particular when a large number of idle/busy periods is required). The total energy of these samples is then calculated and compared with a predefined threshold as:

$$E_x = \sum_{n=1}^{N} |x[n]|^2 \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \lambda \tag{2.19}$$

If the energy of the N samples is greater than the threshold λ a binary decision of $\mathcal{H}_1$ is made to indicate the channel is busy, otherwise $\mathcal{H}_0$ is made to indicate the channel is idle. Selecting threshold value will determine the operation of the system. The prototype can operate under PSS by selecting a threshold energy midway between idle and busy energies after adjusting the gain of the receiver to be sufficiently high in order to easily differentiate the energy between the two states without causing any sensing error. On the other hand, under ISS scenario, threshold value λ is selected to meet a predefined probability of false alarm P_{fa}, where only receiver's noise need to be known [39]. This threshold can be selected by first saving the energy values of the sensing events when there is no signal in the channel but only noise (i.e., transmitter is off) into a file for post-processing. These energy values are then used to select a threshold that would cause a false decision with a probability P_{fa} for which the experiment wanted to be tested under ISS.

After obtaining the sensing decisions, they can be further exploited to provide

statistical information of the channel activity as discussed previously. The program first estimates the durations of the idle/busy periods observed in the channel by computing the time difference between any two changes in the decision of sensing as shown in Fig. 2.11. The estimated periods can then be printed on the terminal window of the program in real-time (while the experiment is running) such that their accuracy can be examined instantaneously in comparison with the transmitted periods. They can also be saved into a text file for post-processing. Subsequently, the statistical parameters of the detected periods can then be calculated to find, for example, the minimum/maximum period, the mean and variance of periods, duty cycle and distribution of periods. These statistics are valuable information to enhance the performance of spectrum sharing systems. Receiver program is developed such that it can provide a real-time graphical illustration for the detected energy of the idle/busy periods as well as a real-time estimation for the statistical parameters of the channel activity. Every time the program detects a new period it updates the estimation of the statistics instantaneously. In addition, similar to the transmitter, a GUI is also designed for the receiver as shown in Fig. 2.10b to ease the configuration of the USRP that is used to perform spectrum sensing and process the sensing decisions.

By comparing the estimated statistics at the receiver with the statistics used to generate the idle/busy periods at the transmitter, it will be possible to use this platform to evaluate the performance of the new estimation methods and algorithms proposed in the following chapters (including those in the literature) under realistic conditions of wireless communication system imposed by the channel impairments (noise, path loss, shadowing and fading) and hardware limitations of transmitter and receiver. In addition, the impact of the parameters employed by spectrum sensing, which include sensing period T_s, sensing time slot τ, probability of sensing errors and sample size N, on the estimation of the channel activity statistics can easily be examined using this prototype in real-time and under realistic conditions.

2.3.4 Illustrative Experiment and Results

In this section, the operation of the proposed prototype is demonstrated by carrying out an illustrative experiment to show the whole process which involves generating PU activity in a frequency channel (i.e., transmitting idle/busy periods), detecting the energy of the channel activity (i.e., spectrum sensing), estimating the idle/busy periods durations, and finally estimating their statistical information. All these operations take place in real-time while the experiment is running.

First, the USRP of the transmitter and the receiver are configured (using the designed GUI shown previously) such that they both operate on the same frequency channel. In this context, we run the experiment on 2.5 GHz WiFi band using channel 11 (2.451 GHz - 2.473 GHz) centred at 2.462 GHz. It is worthy to mention that such experiment would not be possible to carry out using PECAS [86], which can only operate at a central frequency of 433.92 MHz. In addition, channel 11 is selected as such was less crowded in the WiFi environment where our experiment was tested. Selecting the less crowded WiFi channel (non-overlapping) allows us to generate our own PU traffic with known statistics, which then will be compared and validated with the statistics estimated at the SU. Therefore, in the GUI of the transmitter and receiver, the centre frequency of the USRPs is set to 2.462 GHz as shown in Fig. 2.10. In addition, the full functionality of the USRP is used to set the sample rate to its maximum 56 MS/s, which will help providing high resolution idle/busy periods at the transmitter and fast energy detection at the receiver. Placing the SU 1 meter apart from the PU and using a gain of 45 dB at both sides will be sufficient to detect the transmitted signal. Note that if a longer distance is chosen between the PU and SU, a higher gain might be needed for the experiments conducted under PSS (i.e., without sensing errors), or an adjusted gain might be needed for the experiments conducted under ISS (with predefined probability of sensing error).

The transmitter is configured to generate and transmit 200 periods (100 idle and 100 busy). These periods are produced from the Generalised Pareto distribution (which is the best desecription for the empirical data in real system [78]). This distribution can be selected from the GUI and its parameters (for both idle and busy periods) can be configured as: location $\mu = 0.2$ s, scale $\lambda = 0.24$ s, and shape $\alpha = 0.2$. Based on which, the generated idle/busy periods will have the following statistical characteristics: a minimum period $\min(T_0) = \min(T_1) = 0.2$ s, a mean period $\mathbb{E}(T_0) = \mathbb{E}(T_1) = 0.5$ s, and a duty cycle $\Psi = 0.5$. Since these generated periods require 100 seconds to be transmitted, the reception time of the receiver is adjusted to be sufficiently high to detect the whole transmitted sequence. For example, a reception time of 200 s (with extra 50 s before starting the transmitter and extra 50 s after) can be used to guarantee all the periods will be detected properly.

The receiver senses the channel periodically using a sensing time $T_s = 100$ ms, where T_s has to be smaller than the minimum transmitted period which is $\min(T) = 0.2$ s. Note that the minimum T_s that can be configured by the proposed prototype using the host PC (with Intel Core i5 processor) while still maintaining real-time operation is 0.33 ms (which is 30 times faster than PECAS [86] where its minimum T_s is 10 ms). At every

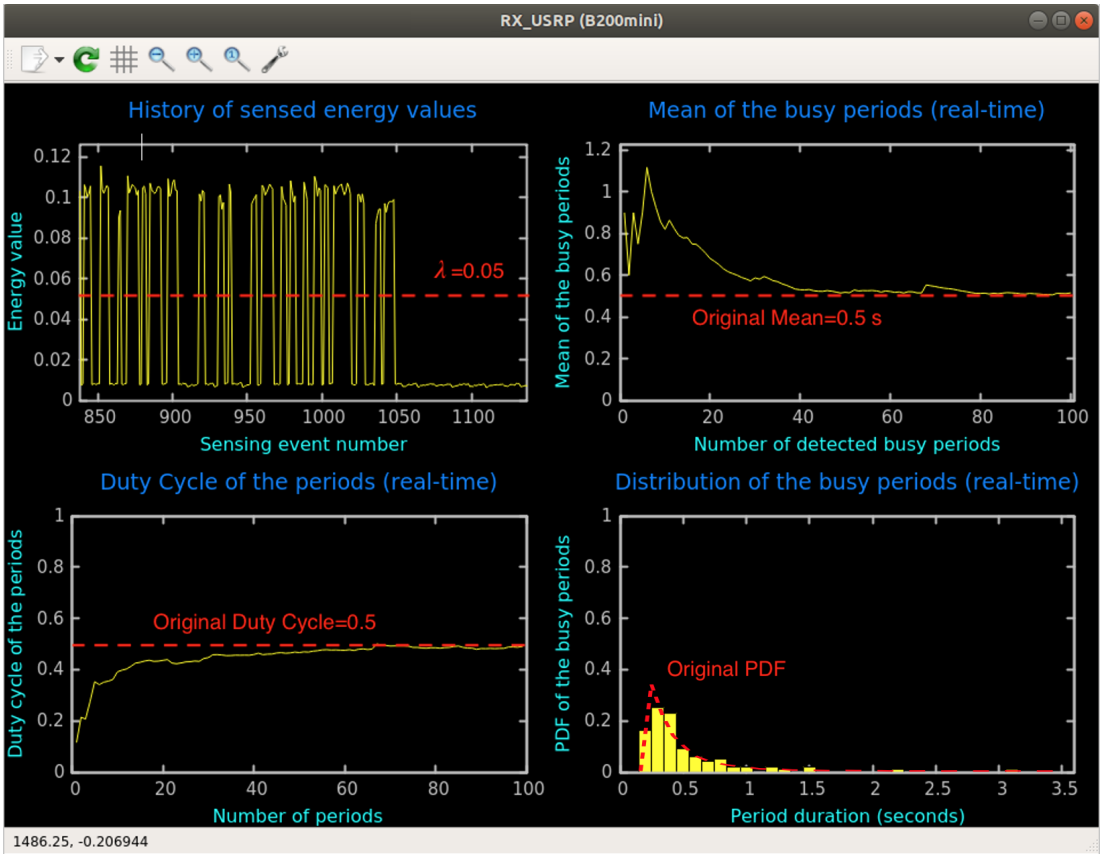


Figure 2.12: Real-time energy detection and statistics estimation using the proposed prototype.

sensing event a set of samples are captured for a time slot of $\tau = 0.1$ ms. The number of these samples is found as $N = \lceil \tau f_s \rceil = \lceil 0.1 \text{ ms} \times 56 \text{ MS/s} \rceil = 5600$ samples. Based on which, the energy of each sensing event can be calculated and plotted instantaneously as shown in Fig. 2.12 (top-left), which shows the real-time energy detection of the idle/busy periods in the frequency channel 2.462 GHz. Comparing the energy values with a predefined threshold $\lambda = 0.05$ (which is selected to be in the middle for PSS operation), binary decisions can be made about the state of the channel (idle $\mathcal{H}_0$ or busy $\mathcal{H}_1$). Based on these decisions, the idle/busy periods durations can be estimated, which in turn will be used to provide an estimation for the channel activity statistics such as the mean period as shown in Fig. 2.12 (top-right), duty cycle (bottom-left) and distribution (bottom-right). These statistics are shown for busy periods, however, similar tendency can also be observed for the idle periods.

As it can be appreciated from Fig. 2.12, the larger the number of periods (sample

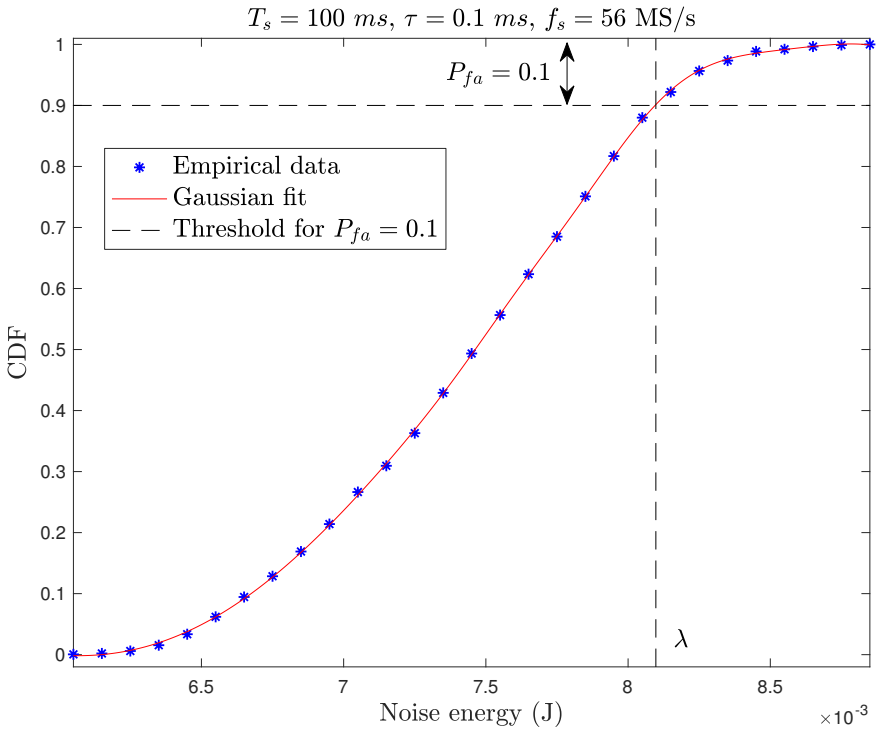


Figure 2.13: Threshold selection from the CDF of the noise energy.

size) used to estimate the statistics the closer the estimation approaches the original statistics. This can help to determine how many periods are required to provide an accurate estimation for channel activity statistics, which can therefore validate the analysis conducted in [63,68]. Notice that the statistics shown in Fig. 2.12 are recalculated and updated in real-time every time a new period is observed.

The prototype can also be configured to operate under ISS by selecting a threshold value that satisfies a predefined P_{fa}. This threshold can be found by first running the receiver to save a large number of (noise-only) sensing energies when the transmitter is off, then selecting the point where the Cumulative Distribution Function (CDF) of these energies is equal to $1 - P_{fa}$, i.e., $\lambda = F_{E_x}^{-1}(1 - P_{fa})$. As shown in Fig. 2.13, the CDF of the energy values fits well with the Gaussian CDF, from which a threshold $\lambda = 0.0081$ J is found to run the prototype under ISS with $P_{fa} = 0.1$.

This prototype will be used to serve as a proof-of-concept for the conducted mathematical analysis on the channel activity statistics in the subsequent chapters of this thesis, where experimental validation for the obtained closed-form expressions and novel estimation methods will be provided under realistic conditions of the wireless channel

and hardware limitations. This platform will also help other researchers and engineers to support their ongoing research on the channel activity statistics in spectrum sharing systems and other works in the future.

2.4 Summary

This chapter has presented two preliminary research contributions, which are not specifically targeting the problem of estimating the channel activity statistics but can contribute to provide more accurate estimations and performance assessment of the estimation accuracy in realistic scenarios. The first contribution has investigated the problem of cooperative spectrum sensing and introduced a novel approach which can achieve minimum interference and maximum utilisation in spectrum sharing systems. The proposed cooperative approach outperforms the conventional approach by taking into account both the impact of sensing resolution and sensing error. The second contribution of the chapter has introduced a USRP-based prototype as an experimental platform to support a wide range of experiments in the context of channel activity statistics estimation, which outperforms the functionality and applicability of the existing platform in the literature (PECAS). This platform will be widely used in the subsequent chapters to demonstrate the accuracy and correctness of the obtained mathematical results as well as the improved estimation accuracy that can be obtained with the methods proposed in this thesis.

Chapter 3

Mathematical Analysis of the Channel Activity Statistics

3.1 Introduction

Channel activity statistics can play a significant role in enhancing the performance of spectrum sharing systems. However, in practice these statistics can be severely corrupted by sensing errors under ISS to the extent that they would not be of any practical use in spectrum sharing systems. In the literature, channel activity statistics have been analysed under PSS more comprehensively than under ISS as in [49, 62–68]. Few studies (e.g., [72–74]), which consider the ISS scenario, are usually constrained to a particular distribution (exponential) to model the idle/busy periods, which can be an unrealistic assumption according to the field measurements in [78]. Moreover, the majority of these works have been mainly focused on the estimation of the channel duty cycle as in [69–71], paying less attention to other equally important statistical properties of the channel activity.

In this context, this chapter addresses a significant and highly challenging problem in spectrum sharing systems, which is that of finding mathematical relationships in closed-form between the observed channel activity statistics under ISS and their corresponding original statistics, as a function of relevant operating parameters including the probability of sensing errors (false alarms and missed detections) and the employed sensing period. Such problem is poorly addressed in the literature without deep and rigorous mathematical analyses taking into account all the factors that influence the estimation accuracy of the channel activity statistics. In addition, the impact of the sample size, i.e., the number of periods used to estimate the statistical metrics of the

channel under ISS, is also analysed and closed-form expressions for the required sample size to achieve a targeted level of accuracy for the estimation of the channel activity statistics under ISS are found. The obtained closed form expressions will then be validated by means of simulations and experiments, using the experimental platform proposed previously in Chapter 2.

3.2 System Model and Problem Formulation

Consider, without loss of generality, a single primary channel which is allocated to a single PU. The activity of the PU within this channel is represented by a sequence of idle/busy periods in the time domain. The time durations of these periods are continuous random variables, which can be modelled following a particular distribution. The experimental measurements in [78] show that the Generalised Pareto (GP) distribution is the best fit representation for those periods. In our analysis, however, this information will be considered to be unknown to the DSA/CR system (i.e., no prior knowledge is required), which makes this study independent and applicable to any distribution type.

A SU, on the other hand, monitors the activity patterns of the PU based on spectrum sensing. In spectrum sensing, periodic sensing events are performed at a constant time interval referred to as the sensing period T_s. At each sensing event the instantaneous state of the channel is detected and reported as idle ($\mathcal{H}_0$) or busy ($\mathcal{H}_1$). The observations of spectrum sensing, therefore, will be a set of binary decisions $\mathcal{H}_0/\mathcal{H}_1$, based on which the durations of idle/busy periods can be calculated. The time duration elapsed between any two adjacent changes in the sensing decisions is considered as an estimation of the original period (i.e., the actual duration of the channel activity).

As shown in Fig. 3.1(a), the original idle/busy periods T_i (where i refers to the type of the period, $i = 0$ for idle and $i = 1$ for busy) are estimated as $\widehat{T}_i$ under PSS (i.e., without sensing errors). The accuracy of this estimation (i.e., under PSS) is only affected by the time resolution imposed by the employed sensing period T_s when no sensing errors are assumed. In practice, however, spectrum sensing is imperfect due to low SNR conditions and sensing errors are likely to occur in the sensing events $\mathcal{H}_i$. Two types of sensing errors can be identified: false alarms (where an idle state of the channel is sensed as a busy state) and missed detections (where a busy state of the channel is sensed as an idle state). The estimated periods in the presence of the sensing errors are denoted as $\breve{T}_i$ to represent the ISS estimation of the real periods T_i as shown in Fig. 3.1(b), where the impact of a false alarm and a missed detection is illustrated.

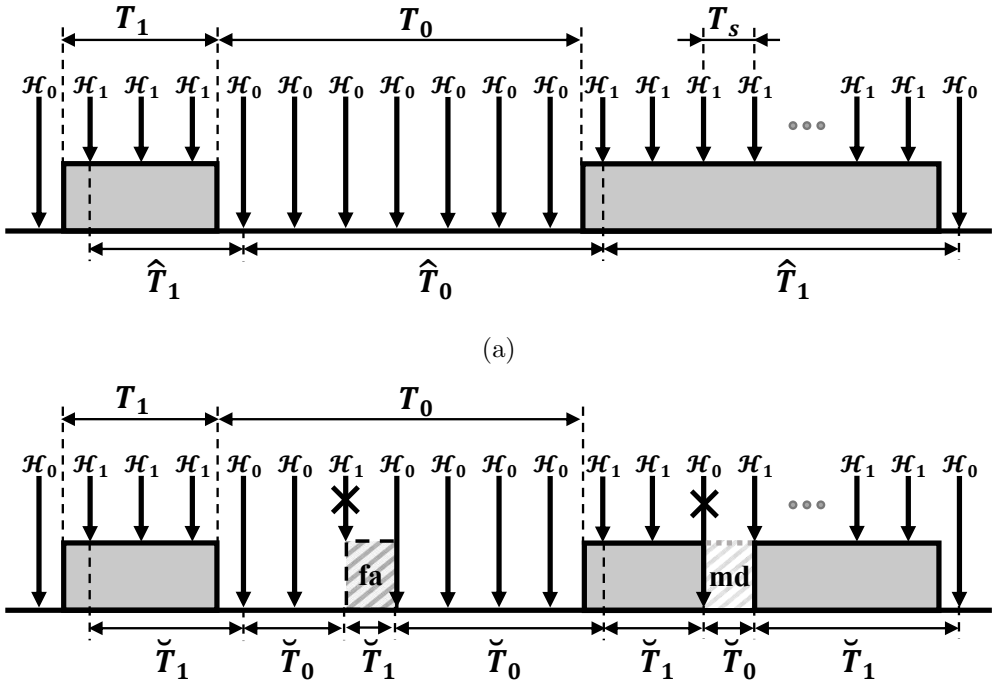


Figure 3.1: Estimation of idle/busy periods based on spectrum sensing: (a) under perfect spectrum sensing (PSS), (b) under imperfect spectrum sensing (ISS).

Sensing errors are random variables and can be modelled with a given probability of false alarm P_{fa} and probability of missed detection P_{md}, which is a common modelling approach in the DSA/CR literature. The value of these probabilities in a practical scenario will depend on the system's operating conditions as well as the configuration of the employed spectrum sensing algorithm, which is an algorithm-specific problem and therefore is out of the scope of this work. However, the system designer can always know how a particular sensing algorithm will behave in terms of these error probabilities (e.g., in an energy detector these probabilities can be easily related to the selected energy decision threshold). Therefore, these probabilities can be assumed to be known in practice [39, 69]. The characterisation of the spectrum sensing algorithm performance in terms of its false alarm and missed detection probabilities makes the analysis presented in this chapter valid irrespective of the specific spectrum sensing algorithm implemented in the DSA/CR system or spectrum sharing systems in general.

3.3 Analysis of the Channel Activity Statistics Under ISS

DSA/CR systems can estimate the idle/busy periods of the primary channel from the observations of the spectrum sensing decisions. Based on a sufficiently large set of these periods, DSA/CR systems can also obtain the statistical information of the primary channel activity [92]. Since this work is concerned with the estimation of the primary activity statistics in the presence of sensing errors (i.e., considering a realistic spectrum sensing scenario), the set of idle/busy periods observed under ISS, as shown in Fig. 3.1(b), is the one considered in this work for statistics calculations. Therefore, a set $\{\breve{T}_{i,n}\}_{n=1}^{N_{iss}}$ of N_{iss} periods observed under ISS is used to estimate the statistical parameters of the primary channel.

The idle/busy period durations $\breve{T}_i$ observed under ISS are affected by several configuration parameters of the spectrum sensing algorithm. These parameters include the employed sensing period T_s and the selected value of probabilities of sensing error (i.e., P_{fa} and P_{md}), which in the case of ED depend on the selected energy decision threshold [37,39]. The estimation of the primary activity statistics under ISS, therefore, can be analysed based on these parameters and a relationship between the statistics estimated under ISS and the original statistics can be found as a function of T_s, P_{fa} and P_{md}. Note that the impact of the sample size parameter N_{iss} on the estimation of the channel activity statistics under ISS will be considered later on in Section 3.7, therefore for now, N_{iss} will be assumed to be sufficiently large to provide accurate results.

In order to find a mathematical relationship between the statistics estimated under ISS and the original statistics of the primary channel activity, three stages of analysis are required:

1. Finding a relationship between the statistics estimated under ISS and the statistics estimated under PSS.

2. Finding a relationship between the statistics estimated under PSS and the original statistics.

3. Combining the second relationship with the first one results in the desired relationship, i.e., the relationship between the statistics estimated under ISS and the original statistics.

The relationship obtained in the first stage of this analysis should be as a function of the sensing period T_s, P_{fa}, and P_{md} as mentioned above, while the relationship in the second stage should only be as a function of the sensing period T_s since PSS is

assumed with no sensing errors (i.e., $P_{fa} = P_{md} = 0$). As discussed in the motivation of this thesis in Chapter 1, channel activity statistics have been analysed in the literature mainly under PSS as in [49], in which the second relationship between the statistics estimated under PSS and the original statistics is provided. In contrast, the analysis of the channel activity statistics under ISS (which is the more realistic scenario) has not been investigated comprehensively in the literature. Therefore, in this chapter the analysis of the first relationship (i.e., between the statistics estimated under ISS and the statistics estimated under PSS) will be obtained first. Then, the final relationship between the statistics estimated under ISS and the original statistics will be achieved by combining the analysis presented in [49] with the analysis presented in this chapter, which considers the same set of statistical metrics as in [49], namely the minimum period, mean of periods, duty cycle, and distribution of the estimated periods.

3.4 Analysis of the Minimum Period

The minimum period μ_i of a primary channel is the shortest time that a primary channel can be active or inactive. This parameter can help determine the minimum opportunistic time that can be exploited by the SUs, or the minimum time that the SUs need to wait until a new opportunity becomes available in the primary channel. DSA/CR systems can estimate the minimum period μ_i of the channel from spectrum sensing observations. Although this parameter has already been studied in [49] under PSS and in [75] under ISS, we summarise the analysis here in order to make this exposition self-contained.

Based on a given set $\{\widehat{T}_{i,n}\}_{n=1}^{N_{pss}}$ of N_{pss} periods observed under PSS, [49] found that the estimated minimum period $\widehat{\mu}_i$ under PSS can be expressed as a function of the original minimum period μ_i as [49, eq. (7)]:

$$\widehat{\mu}_i = \min\left(\{\widehat{T}_{i,n}\}_{n=1}^{N_{pss}}\right) = \left\lfloor \frac{\mu_i}{T_s} \right\rfloor T_s, \quad T_s \leq \mu_i, \tag{3.1}$$

where $\lfloor \cdot \rfloor$ denotes the floor operator. As it can be appreciated from (3.1), the minimum period can be correctly estimated under PSS (i.e., $\widehat{\mu}_i = \mu_i$) when T_s is an integer submultiple of the true minimum (i.e., when $T_s = \mu_i/k$, $k \in \mathbb{N}^+$), otherwise $\widehat{\mu}_i < \mu_i$ [49]. On the other hand, the minimum period cannot be estimated correctly under ISS because the minimum observed period would be equal to the duration of a single sensing error, which in turn is equal to the sensing period T_s [75], no matter how low the probability of sensing error is. Therefore, from a given set $\{\breve{T}_{i,n}\}_{n=1}^{N_{iss}}$ of N_{iss} periods

observed under ISS, the minimum estimated period would be [75]:

$$\breve{\mu}_i = \min\left(\{\breve{T}_{i,n}\}_{n=1}^{N_{iss}}\right) = T_s. \tag{3.2}$$

As it can be noticed from the observed periods under ISS in Fig. 3.1(b), the minimum busy period $\breve{T}_1$ would be the duration of the single false alarm, which is equal to T_s (the same applies to the minimum idle period $\breve{T}_0$ where a single missed detection would be the minimum observed period). Since (3.2) shows that the estimated minimum period under ISS is solely dependent on the sensing period T_s (regardless of the original value of the minimum period), it is impossible to find, from this expression, an exact relationship between the estimated minimum period $\breve{\mu}_i$ under ISS and the original minimum μ_i.

3.5 Analysis of the Mean Period

One of the main statistical moments of the primary channel activity is the mean of the idle/busy periods. For a given set $\{\widehat{T}_{i,n}\}_{n=1}^{N_{pss}}$ of N_{pss} periods estimated under PSS, the mean $\mathbb{E}(\widehat{T}_i)$ of the observed periods can be found by using the conventional (unbiased) sample mean estimator $\widehat{m}_i$:

$$\mathbb{E}(\widehat{T}_i) \approx \widehat{m}_i = \frac{1}{N_{pss}} \sum_{n=1}^{N_{pss}} \widehat{T}_{i,n}. \tag{3.3}$$

The analysis in [49] has shown that the estimated mean under PSS is approximately equal to the true mean of the channel periods (i.e., $\widehat{m}_i \approx \mathbb{E}(T_i)$). However, this does not apply to the estimated mean under ISS because when an error occurs in the sensing decisions (either false alarm or missed detection), it will divide the original period duration T_i into shorter fragments [93]. As it can be noticed in Fig. 3.1(b), a single false alarm error could corrupt the estimation of an idle period T_0 period by dividing it into three new shorter periods, which are $\breve{T}_0, \breve{T}_1$, and $\breve{T}_0$. The duration of the $\breve{T}_1$ fragment is equal to the sensing period T_s, while the durations of the two $\breve{T}_0$ fragments are random, depending on the position of the error itself within T_0.

Due to this phenomenon, the number of the observed periods under ISS (N_{iss}) would not be the same as the number of the periods observed under PSS (N_{pss}). As shown in Fig. 3.1(b), with a single false alarm, the original T_0 period is estimated as two $\breve{T}_0$ periods and one $\breve{T}_1$ period. If there were two false alarms within T_0, then they would result in three $\breve{T}_0$ and two $\breve{T}_1$ periods, and so on. Therefore, each false alarm would produce an additional estimated idle period $\breve{T}_0$ and an additional estimated busy

period $\breve{T}_1$ (a similar effect would be observed with missed detections). As a result, the number of periods N_{iss} observed under ISS will be greater than the actual number of the periods N (unlike under PSS where no additional periods are produced during the spectrum sensing process and thus $N_{pss} = N$). Therefore, from a given set $\{\breve{T}_{i,n}\}_{n=1}^{N_{iss}}$ of N_{iss} periods estimated under ISS, calculating the mean using the following conventional mean estimator:

$$\mathbb{E}(\breve{T}_i) \approx \breve{m}_i = \frac{1}{N_{iss}} \sum_{n=1}^{N_{iss}} \breve{T}_{i,n} \tag{3.4}$$

would be highly inaccurate (indeed, much lower than the original value of the mean). In order to find the relationship between the mean calculated under ISS $\mathbb{E}(\breve{T}_i)$ and the original value of the mean $\mathbb{E}(T_i)$ of the channel periods, we first find its relationship with the mean calculated under PSS $\mathbb{E}(\widehat{T}_i)$ as discussed in Section 3.3. The analysis will consider first, without loss of generality, the sample mean for idle periods (i.e., $\breve{m}_0$, $\breve{m}_i$ with $i = 0$), and will be later on generalised to both idle and busy periods. Consider the idle period T_0 illustrated in Fig. 3.1(b), the mean of the observed idle periods under ISS within T_0 (which are caused by a single false alarm in this case) can be found as:

$$\begin{aligned} \breve{m}_0 &= \frac{\breve{T}_{0,1} + \breve{T}_{0,2}}{2} \\ &= \frac{\widehat{T}_0 - \breve{T}_1}{2} \\ &= \frac{\widehat{T}_0 - T_s}{2}. \end{aligned}$$

This mean represents the summation of the two idle fragments ($\breve{T}_{0,1}$ and $\breve{T}_{0,2}$) divided by 2. This summation is equivalent to subtracting $\breve{T}_1$ from the estimated period $\widehat{T}_0$ under PSS, knowing that the produced $\breve{T}_1$ period from the false alarm error is equal to the sensing period T_s. In addition, the denominator 2, which represents the number of the estimated idle periods under ISS (i.e., N_{iss}), can be substituted with the number of estimated idle periods under PSS plus one for the single false alarm (i.e., $N_{iss} = N_{pss} + 1$). This analysis of a single false alarm error within a single idle period can be extended to a general form for any arbitrary number of false alarm errors within the whole set of idle periods as:

$$\breve{m}_0 = \frac{\sum_{n=1}^{N_{pss}} \widehat{T}_{0,n} - N_{fa} T_s}{N_{pss} + N_{fa}}, \tag{3.5}$$

where N_{fa} represents the total number of false alarm errors in the entire set of observed periods, which can be found by multiplying the entire number of $\mathcal{H}_0$ events (i.e., idle sensing decisions) by the probability of false alarm P_{fa} as:

$$N_{fa} = \frac{\sum_{n=1}^{N_{pss}} \widehat{T}_{0,n}}{T_s} \cdot P_{fa}. \tag{3.6}$$

The above analysis does not consider the impact of missed detection errors. Missed detections would also lead to additional idle periods in the observed set of periods, where every single missed detection produces a single idle period of a duration equal to T_s. In order to find $\breve{m}_0$ by taking into consideration the missed detections as well, the same concept can be followed so that (3.5) can be rewritten to include both sensing error types:

$$\breve{m}_0 = \frac{\sum_{n=1}^{N_{pss}} \widehat{T}_{0,n} - N_{fa}T_s + N_{md}T_s}{N_{pss} + N_{fa} + N_{md}}, \tag{3.7}$$

where N_{md} represents the total number of missed detection errors in the entire set of observed periods, which can be found by multiplying the entire number of $\mathcal{H}_1$ events (i.e., busy sensing decisions) by the probability of missed detection P_{md} as:

$$N_{md} = \frac{\sum_{n=1}^{N_{pss}} \widehat{T}_{1,n}}{T_s} \cdot P_{md}. \tag{3.8}$$

By substituting (3.6) and (3.8) in (3.7):

$$\breve{m}_0 = \frac{\sum_{n=1}^{N_{pss}} \widehat{T}_{0,n} - \frac{\sum_{n=1}^{N_{pss}} \widehat{T}_{0,n}}{T_s} \cdot P_{fa}T_s + \frac{\sum_{n=1}^{N_{pss}} \widehat{T}_{1,n}}{T_s} \cdot P_{md}T_s}{N_{pss} + \frac{\sum_{n=1}^{N_{pss}} \widehat{T}_{0,n}}{T_s} \cdot P_{fa} + \frac{\sum_{n=1}^{N_{pss}} \widehat{T}_{1,n}}{T_s} \cdot P_{md}}. \tag{3.9}$$

Note that from (3.3), the term $\sum_{n=1}^{N_{pss}} \widehat{T}_{i,n}$ can be written as:

$$\sum_{n=1}^{N_{pss}} \widehat{T}_{i,n} = N_{pss}\widehat{m}_i. \tag{3.10}$$

Therefore, using (3.10), expression (3.9) can be further simplified to:

$$\breve{m}_0 = \frac{\widehat{m}_0(1 - P_{fa}) + \widehat{m}_1 P_{md}}{1 + \frac{\widehat{m}_0}{T_s}P_{fa} + \frac{\widehat{m}_1}{T_s}P_{md}}. \tag{3.11}$$

Although this equation can provide a mathematical relationship between the calculated mean under ISS, $\breve{m}_0$, and the calculated mean under PSS, $\widehat{m}_0$, it lacks some accuracy.

The reason is that, based on our analysis, there are two special cases where sensing errors (either false alarms or missed detections) will not produce additional estimated idle periods ($\breve{T}_0$) and additional estimated busy periods ($\breve{T}_1$). This will be analysed exhaustively in the following two cases.

3.5.1 Case I

When a sensing error occurs at the first (or last) sensing event within a period, it will be observed as a part of the previous (or next) period attached to it, thus causing no fragments or additional periods. As shown in Fig. 3.2, the false alarm at the left edge of the idle period is combined with the adjacent busy period and both together are estimated as a single $\breve{T}_1$ period. Therefore, there will be no additional $\breve{T}_0$ or $\breve{T}_1$ fragments produced from such false alarm. The previous analysis has assumed that all the sensing errors will produce an additional idle period and an additional busy period without considering the case explained here. To include this effect, the denominator in (3.7), which represents the number of the estimated periods under ISS ($N_{iss} = N_{pss}+N_{fa}+N_{md}$), should not count the cases when the sensing errors occur at the edges of the periods since there will be no additional periods produced by them. This can be attained by knowing that each estimated idle/busy period will have two edges and these edges are actually represented by sensing events $\mathcal{H}_0/\mathcal{H}_1$. Therefore, the problem in the denominator of (3.7) and the resulting (3.11) can be solved by subtracting the two sensing events for both edges from the total number of sensing events within a single period (or $2N_{pss}$ from the entire number of the events within N_{pss} periods), and thus (3.11) can be corrected to yield:

$$\breve{m}_0 = \frac{\widehat{m}_0(1-P_{fa}) + \widehat{m}_1 P_{md}}{1 + \left(\frac{\widehat{m}_0}{T_s} - 2\right) P_{fa} + \left(\frac{\widehat{m}_1}{T_s} - 2\right) P_{md}}. \tag{3.12}$$

3.5.2 Case II

It is also possible that some sensing errors will not produce additional periods when they occur in bursts (i.e., they are consecutive to other errors). Fig. 3.3 shows how two false alarms could occur consecutively in the sensing decisions of an idle period. Although most of the sensing errors could occur as individual periods (with a duration equal to T_s), it is still possible, depending on the probability of sensing error, to observe some occasional consecutive errors in the sensing decisions. However, the probability of having two consecutive errors is higher than that of having three or more consecutive

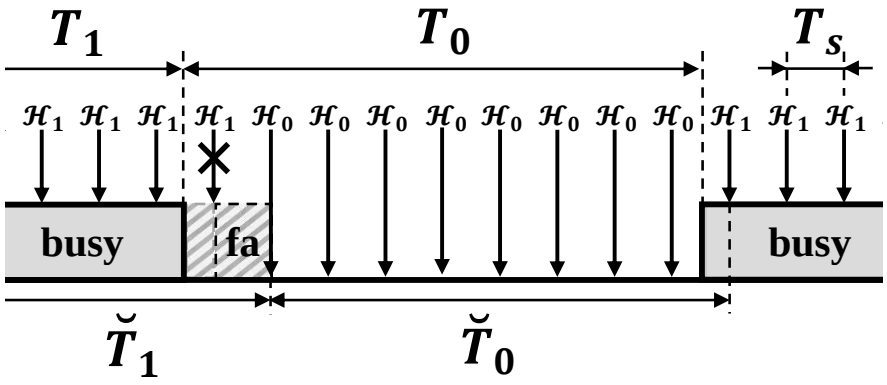


Figure 3.2: Case I: A single sensing error at the edge of a period.

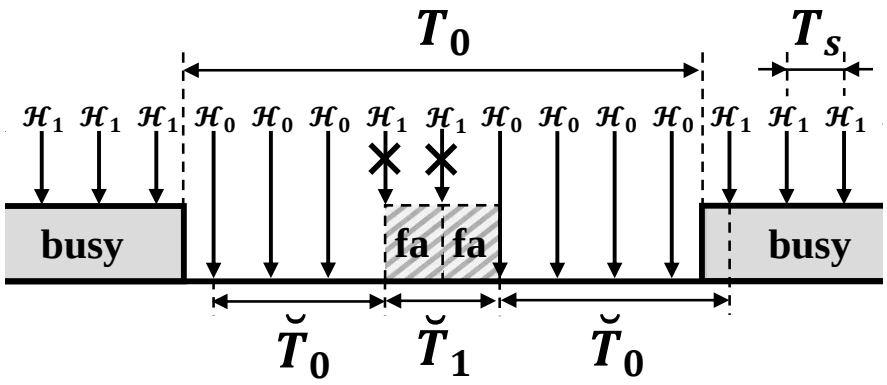


Figure 3.3: Case II: Two consecutive sensing errors in the middle of a period.

errors as also illustrated in [77, Fig. 3]. Consecutive errors will have the same effect of a single error in terms of the number of produced fragments. For example, the resulting fragments of the two consecutive false alarms in Fig. 3.3 are $\breve{T}_0$, $\breve{T}_1$, and $\breve{T}_0$, which is the same number of the resulting fragments from a single false alarm observed in the example of Fig. 3.1(b). Since there are no additional fragments resulting from consecutive errors, the denominator of (3.7) should therefore not count these errors. This can be attained by subtracting the probability of having consecutive errors from the probability of sensing error itself. As a result, P_{fa} and P_{md} used in (3.6) and (3.8)

to find N_{fa} and N_{md} should be modified as follows:

$$\begin{aligned} \grave{P}_{fa} &= P_{fa} - \sum_{j=2}^{\infty} P_{fa}^{j} \\ &= P_{fa}\left(\frac{1-2P_{fa}}{1-P_{fa}}\right), \end{aligned} \tag{3.13}$$

where $\grave{P}_{fa}$ represents the probability of having false alarms as individual periods, irrespective of being consecutive or isolated errors, and the relation $\sum_{j=2}^{\infty} a^j = \frac{a^2}{1-a}$, when $|a| < 1$ is used to obtain the final expression in (3.13). This also applies to the consecutive missed detection errors, hence:

$$\begin{aligned} \grave{P}_{md} &= P_{md} - \sum_{j=2}^{\infty} P_{md}^{j} \\ &= P_{md}\left(\frac{1-2P_{md}}{1-P_{md}}\right). \end{aligned} \tag{3.14}$$

3.5.3 Final Closed-Form Expression of the Mean Period

Taking into account the previous two special cases, the final expression of the estimated mean idle period under ISS, $\breve{m}_0$, is obtained by introducing (3.13) and (3.14) into the denominator of (3.12), which yields:

$$\breve{m}_0 = \frac{\widehat{m}_0(1-P_{fa}) + \widehat{m}_1 P_{md}}{1 + \left(\frac{\widehat{m}_0}{T_s} - 2\right)\grave{P}_{fa} + \left(\frac{\widehat{m}_1}{T_s} - 2\right)\grave{P}_{md}}. \tag{3.15}$$

The achieved expression in (3.15) provides the relationship between the mean $\breve{m}_0$ estimated under ISS (for idle periods) and the means $\widehat{m}_i$ estimated under PSS (for both idle and busy periods), which satisfies the first stage of the analysis as mentioned in Section 3.3. On the other hand, the relationship between the mean $\widehat{m}_i$ estimated under PSS and the original mean m_i was analysed in [49], which shows that the population mean of the periods observed under PSS is exactly equal to the population mean of the original periods (i.e., $\mathbb{E}(\widehat{T}_i) = \mathbb{E}(T_i)$), thus providing the second stage of the analysis as mentioned in Section 3.3. Therefore, by assuming a sufficiently large sample size (i.e., number of periods), the sample means in (3.15) can be substituted by population means so that the second relationship can be combined with the first one to obtain the desired relationship between the estimated mean under ISS and the original mean,

which completes the third stage of the analysis mentioned in Section 3.3 as:

$$\mathbb{E}(\breve{T}_0) = \frac{\mathbb{E}(T_0)(1-P_{fa}) + \mathbb{E}(T_1)P_{md}}{1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right)\grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right)\grave{P}_{md}}. \tag{3.16}$$

A similar analysis can also be followed to find the estimated mean of the busy periods $\mathbb{E}(\breve{T}_1)$ under ISS as:

$$\mathbb{E}(\breve{T}_1) = \frac{\mathbb{E}(T_1)(1-P_{md}) + \mathbb{E}(T_0)P_{fa}}{1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right)\grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right)\grave{P}_{md}}. \tag{3.17}$$

Therefore, the final closed-form expression for the estimated mean of the idle/busy periods $\mathbb{E}(\breve{T}_i)$ under ISS can be expressed in a compact form as:

$$\mathbb{E}(\breve{T}_i) = \frac{\mathbb{E}(T_i) - (-1)^i\,\mathbb{E}(T_0)P_{fa} + (-1)^i\,\mathbb{E}(T_1)P_{md}}{1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right)\grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right)\grave{P}_{md}}, \tag{3.18}$$

which provides the final expression for the estimated mean $\mathbb{E}(\breve{T}_i)$ under ISS as a function of the original mean $\mathbb{E}(T_i)$, probabilities of sensing error P_{fa} and P_{md}, and sensing period T_s.

Note that a novel estimation method can be derived from 3.18 to accurately estimate the original mean period under ISS as well as the duty cycle of the periods, where duty cycle $\Psi = \frac{\mathbb{E}(T_1)}{\mathbb{E}(T_1)+\mathbb{E}(T_0)}$. This will be discussed in detail in Chapter 4.

3.6 Analysis of the Distribution of the Periods

The idle/busy periods observed from sensing decisions are integer multiples of the sensing period (i.e., $\breve{T}_i = kT_s$, $k \in \mathbb{N}^+$) and, as discussed in Section 3.4, the minimum estimated period under ISS is T_s, which is due to the presence of individual sensing errors. As a result, the distribution of the idle/busy periods estimated under ISS is discretely shaped with a discrete step of T_s, starting from the minimum period T_s up to the maximum multiple integer of T_s observed in the channel. In order to find a closed-form expression for the Probability Mass Function (PMF) of the idle/busy periods estimated under ISS as a function of the original Probability Density Function (PDF), probabilities of sensing error and sensing period, this analysis considers, without loss of generality, the case of idle periods, introducing the false alarms first and missed detections later on.

3.6.1 Impact of False Alarms

False alarms occur in the sensing decisions of the idle periods with a probability $P_{fa} > 0$. This means that any sensing decision $\mathcal{H}_0$ within a T_0 idle period as shown in Fig. 3.1(b) will have a probability of P_{fa} to be a false alarm, and a probability of $1 - P_{fa}$ not to be a false alarm (i.e., to be a correct decision). Consequently, to find the PMF of idle periods under ISS, denoted as $f_{\breve{T}_0}(\breve{T}_0 = kT_s)$, we need to consider all the possible cases in which the observed idle periods under ISS, $\breve{T}_0 = kT_s$, could be affected by the presence of sensing errors. This can be summarised into three exhaustive cases, which are analysed in detail below.

Case I

In this case, an idle period is observed between two busy periods without having any false alarms as shown in Fig. 3.4. The probability of having such case can be calculated as:

$$P(\breve{T}_0 = kT_s \mid \widehat{T}_0 = kT_s) = (1 - P_{fa})^k, \tag{3.19}$$

which is the probability of having k correct decisions in k sensing events within an idle period (i.e., k non-false alarms). To find the unconditional PMF of such periods from a set $\{\widehat{T}_{0,n}\}_{n=1}^{N_{pss}}$ of N_{pss} idle periods observed under PSS, such probability should be multiplied by the ratio N_{pss}^I/N_{iss} as follows:

$$f_{\breve{T}_0}^I(\breve{T}_0 = kT_s) = \frac{(1 - P_{fa})^k N_{pss}^I}{N_{iss}}, \tag{3.20}$$

where N_{pss}^I represents the number of periods observed under PSS with a duration equal to kT_s (i.e., $\widehat{T}_0 = kT_s$), which is given by $f_{\widehat{T}_0}(\widehat{T}_0 = kT_s)N_{pss}$. Thus, (3.20) can be written as:

$$\begin{aligned} f_{\breve{T}_0}^I(\breve{T}_0 = kT_s) &= \frac{(1 - P_{fa})^k f_{\widehat{T}_0}(\widehat{T}_0 = kT_s)N_{pss}}{N_{iss}} \\ &= \beta(1 - P_{fa})^k f_{\widehat{T}_0}(\widehat{T}_0 = kT_s), \end{aligned} \tag{3.21}$$

where β is defined to be the ratio $N_{pss}/N_{iss} < 1$, and it can be found as (see Section 3.7 for the analysis of the sample size under ISS):

$$\beta = \frac{N_{pss}}{N_{iss}} = \frac{1}{1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right)\grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right)\grave{P}_{md}}. \tag{3.22}$$

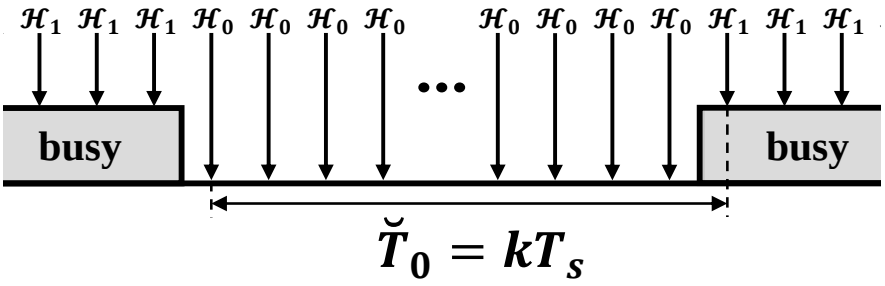


Figure 3.4: Case I: An idle period $\breve{T}_0 = kT_s$ observed between two busy periods without sensing errors.

The expression obtained in (3.21) represents the PMF of the observed ISS idle periods resulting from Case I shown in Fig. 3.4.

Case II

Another case where an idle period can be observed under ISS occurs between a single false alarm and the edge of the adjacent busy period as shown in Fig. 3.5. The probability of having such case is:

$$P(\breve{T}_0 = kT_s \mid \widehat{T}_0 \geq (k+1)T_s) = 2P_{fa}(1-P_{fa})^k, \tag{3.23}$$

which is the probability of having a single false alarm and k non-false alarms in at least $k+1$ sensing events within an idle period. Note the presence of the factor of 2 because this case could occur at the left and right ends of an idle period next to a busy period. Therefore, following the same principle as in Case I, the unconditional PMF of such periods can be found by multiplying (3.23) by the ratio N_{pss}^{II}/N_{iss} as:

$$f_{\breve{T}_0}^{II}(\breve{T}_0 = kT_s) = \frac{2P_{fa}(1-P_{fa})^k N_{pss}^{II}}{N_{iss}}, \tag{3.24}$$

where N_{pss}^{II} represents the number of periods observed under PSS with a duration $(\widehat{T}_0 \geq (k+1)T_s)$, which is given by $N_{pss}\left(1 - F_{\widehat{T}_0}(kT_s)\right)$ where $F_{\widehat{T}_0}(kT_s)$ represents the Cumulative Distribution Function (CDF) of the idle periods observed under PSS.

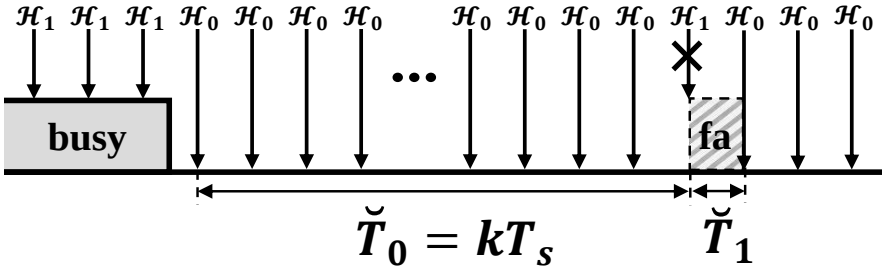


Figure 3.5: Case II: An idle period $\breve{T}_0 = kT_s$ observed between a busy period and a false alarm.

Thus, (3.24) can be written as:

$$f^{II}_{\breve{T}_0}(\breve{T}_0 = kT_s) = \frac{2P_{fa}(1-P_{fa})^k N_{pss}\left(1 - F_{\widehat{T}_0}(kT_s)\right)}{N_{iss}} = 2\beta P_{fa}(1-P_{fa})^k\left(1 - F_{\widehat{T}_0}(kT_s)\right). \quad (3.25)$$

The expression obtained in (3.25) represents the PMF of the observed ISS idle periods resulting from Case II as shown in Fig. 3.5.

Case III

The last case where an idle period can be observed under ISS occurs between two false alarms within the original idle period as shown in Fig. 3.6. The probability of having such case is:

$$P(\breve{T}_0 = kT_s \mid \widehat{T}_0 \geq (k+2)T_s) = \left(\frac{\widehat{T}_0}{T_s} - (k+1)\right) P^2_{fa}(1-P_{fa})^k, \quad (3.26)$$

which is the probability to have two false alarms and k non-false alarms in at least $k+2$ sensing events within an idle period. Note the presence of the factor $\left(\frac{\widehat{T}_0}{T_s} - (k+1)\right)$ as there are $\left(\frac{\widehat{T}_0}{T_s} - (k+1)\right)$ different ways for such case to occur within an idle period. For example, if $\widehat{T}_0 = (k+2)T_s$, the result of the above probability would simply be $P^2_{fa}(1-P_{fa})^k$. Therefore, following the same principle as in Cases I and II, the

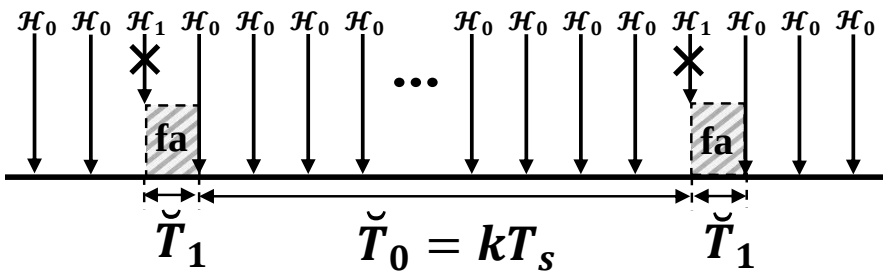


Figure 3.6: Case III: An idle period $\breve{T}_0 = kT_s$ observed between two false alarms.

unconditional PMF of such periods can be found as:

$$f_{\breve{T}_0}^{III}(\breve{T}_0 = kT_s) = \frac{\left(\frac{\sum_{n=1}^{N_{pss}} \widehat{T}_{0,n}^{III}}{T_s} - (k+1)N_{pss}^{III}\right) P_{fa}^2 (1-P_{fa})^k}{N_{iss}}, \tag{3.27}$$

where N_{pss}^{III} represents the number of periods observed under PSS with a duration ($\widehat{T}_0 \geq (k+2)T_s$) and $\sum_{n=1}^{N_{pss}} \widehat{T}_{0,n}^{III}$ is their summation. Thus, (3.27) can be further solved as in (3.28).

The expression obtained in (3.28) represents the PMF of the observed ISS idle periods resulting from Case III shown in Fig. 3.6.

The analysis presented so far has considered all the possible cases that can lead to the observation of an idle period under ISS (due to false alarms only), and the corresponding PMF was obtained for each case separately. Therefore, a general expression for the PMF that jointly considers the three cases can be obtained by combining (3.21), (3.25) and (3.28) as in (3.29).

The expression $f_{\breve{T}_0}^{fa}(\breve{T}_0 = kT_s)$ obtained in (3.29) represents the PMF of the periods observed under ISS as a function of the corresponding PMF, CDF, and mean period that would be estimated under PSS as well as the probability of false alarm and the sensing period. Since the missed detections were not considered in the previous analysis (i.e., $P_{md} = 0$), next section studies the effect of the missed detections on the calculation of the estimated distribution.

$$f^{III}_{\breve{T}_0}(\breve{T}_0 = kT_s) = \frac{\left(\frac{\sum_{m=k+2}^{\infty}\left(mT_s f_{\widehat{T}_0}(\widehat{T}_0=mT_s)\right)N_{pss}}{T_s} - (k+1)N_{pss}\left(1-F_{\widehat{T}_0}\left((k+1)T_s\right)\right)\right)P^2_{fa}(1-P_{fa})^k}{N_{iss}}$$

$$= \frac{N_{pss}P^2_{fa}(1-P_{fa})^k}{N_{iss}}\left(\frac{\mathbb{E}(\widehat{T}_0) - \sum_{m=1}^{k+1}\left(mT_s f_{\widehat{T}_0}(\widehat{T}_0 = mT_s)\right)}{T_s} - (k+1)\left(1-F_{\widehat{T}_0}\left((k+1)T_s\right)\right)\right)$$

$$= \beta P^2_{fa}(1-P_{fa})^k\left(\frac{\mathbb{E}(\widehat{T}_0)}{T_s} - \sum_{m=1}^{k}\left(mf_{\widehat{T}_0}(\widehat{T}_0 = mT_s)\right) - (k+1)\left(1-F_{\widehat{T}_0}(kT_s)\right)\right). \tag{3.28}$$

$$f^{fa}_{\breve{T}_0}(\breve{T}_0 = kT_s) = \beta(1-P_{fa})^k\Bigg[f_{\widehat{T}_0}(\widehat{T}_0 = kT_s)$$

$$+ 2P_{fa}\left(1-F_{\widehat{T}_0}(kT_s)\right) + P^2_{fa}\left(\frac{\mathbb{E}(\widehat{T}_0)}{T_s} - \sum_{m=1}^{k}\left(mf_{\widehat{T}_0}(\widehat{T}_0 = mT_s)\right) - (k+1)\left(1-F_{\widehat{T}_0}(kT_s)\right)\right)\Bigg]. \tag{3.29}$$

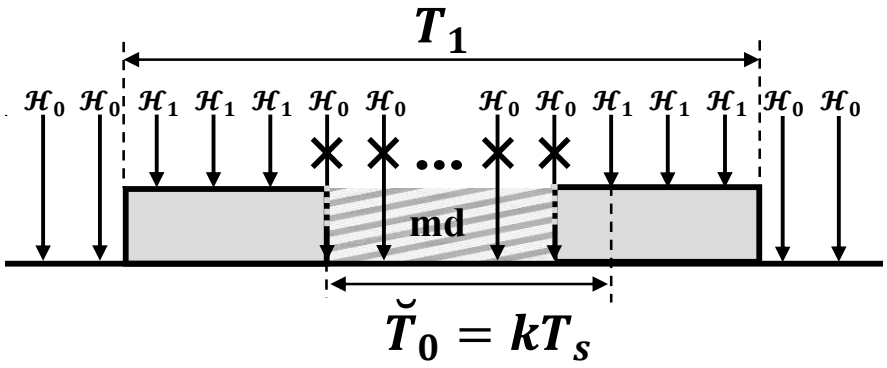


Figure 3.7: An idle period $\breve{T}_0 = kT_s$ observed within a busy period because of missed detections.

3.6.2 Impact of Missed Detections

Under imperfect sensing, missed detections can also occur in the sensing decisions of the busy states of the channel, so that a busy state $\mathcal{H}_1$ can be incorrectly reported as an idle state $\mathcal{H}_0$. Therefore, any $\mathcal{H}_1$ sensing event within a T_1 period will have a probability of P_{md} to be misdetected, and a probability of $1 - P_{md}$ not to be misdetected. Since there will be additional idle periods $\breve{T}_0$ resulting from missed detections as shown in Fig. 3.7, these periods need to be considered as well when calculating the PMF of the idle periods. Therefore, the probability of observing an idle period within a busy period due to missed detections can be calculated as:

$$P(\breve{T}_0 = kT_s \mid \widehat{T}_1 \geq (k+2)T_s) = \left(\frac{\widehat{T}_1}{T_s} - (k+1)\right) P_{md}^k (1 - P_{md})^2, \tag{3.30}$$

which is the probability to have k consecutive missed detections between two non-missed detections within at least $k+2$ sensing events of a busy period. Therefore, to find the PMF of such periods from a set $\{\widehat{T}_{1,n}\}_{n=1}^{N_{pss}}$ of N_{pss} busy periods the following can be yield:

$$\begin{aligned} f_{\breve{T}_0}^{md}(\breve{T}_0 = kT_s) &= \frac{\left(\frac{\sum_{n=1}^{N_{pss}} \widehat{T}_{1,n}}{T_s} - (k+1)N_{pss}\right) P_{md}^k (1 - P_{md})^2}{N_{iss}} \\ &= \beta \left(\frac{\mathbb{E}(\widehat{T}_1)}{T_s} - (k+1)\right) P_{md}^k (1 - P_{md})^2. \end{aligned} \tag{3.31}$$

Note that the idle periods resulting from missed detections are more likely to occur as short periods due to single missed detection errors (i.e., $k = 1$) than consecutive errors, and it is very unlikely that the whole sensing events of a busy period are missed detected. The obtained expression $f^{md}_{\breve{T}_0}(\breve{T}_0 = kT_s)$ in (3.31) represents the PMF of the idle periods observed under ISS (due to missed detections only).

3.6.3 Final Closed-Form Expression of the Distribution of the Periods

After analysing the impact of false alarms and missed detections on the estimation of the PMF of the idle periods observed under ISS, the final closed-form expression can be obtained by combining (3.29) with (3.31) as:

$$\begin{aligned} f_{\breve{T}_0}(\breve{T}_0 = kT_s) =& \beta(1-P_{fa})^k \Bigg[f_{\widehat{T}_0}(\widehat{T}_0 = kT_s) + 2P_{fa}\left(1 - F_{\widehat{T}_0}(kT_s)\right) \\ &+ P_{fa}^2 \left(\frac{\mathbb{E}(\widehat{T}_0)}{T_s} - \sum_{m=1}^{k} \left(m f_{\widehat{T}_0}(\widehat{T}_0 = mT_s) \right) - (k+1)\left(1 - F_{\widehat{T}_0}(kT_s)\right) \right) \\ &+ \left(\frac{\mathbb{E}(\widehat{T}_1)}{T_s} - k - 1 \right) \frac{P_{md}^k (1-P_{md})^2}{(1-P_{fa})^k} \Bigg]. \end{aligned} \tag{3.32}$$

After simplifying (3.32), the following expression can finally be obtained:

$$\begin{aligned} f_{\breve{T}_0}(\breve{T}_0 = kT_s) =& \beta P_{fa}^2 (1-P_{fa})^k \Bigg[f_{\widehat{T}_0}(\widehat{T}_0 = kT_s) \left(\frac{P_{fa}^2 - 2P_{fa} + 1}{P_{fa}^2} \right) \\ &- \sum_{m=1}^{k-1} f_{\widehat{T}_0}(\widehat{T}_0 = mT_s) \left(m - k + \frac{2 - P_{fa}}{P_{fa}} \right) + \frac{\mathbb{E}(\widehat{T}_0)}{T_s} + \frac{2 - P_{fa}}{P_{fa}} - k \\ &+ \left(\frac{\mathbb{E}(\widehat{T}_1)}{T_s} - k - 1 \right) \frac{P_{md}^k (1-P_{md})^2}{P_{fa}^2 (1-P_{fa})^k} \Bigg]. \end{aligned} \tag{3.33}$$

The analytical result in (3.33) provides a closed-form relationship between the PMF of the idle periods observed under ISS, $f_{\breve{T}_0}(\breve{T}_0 = kT_s)$, and the PMF $f_{\widehat{T}_0}(\widehat{T}_0 = kT_s)$ and the mean $\mathbb{E}(\widehat{T}_i)$ of the periods that would be observed under PSS as well as the probabilities of sensing errors P_{fa} and P_{md}, and sensing period T_s. In other words, (3.33) satisfies the first stage of the analysis procedure mentioned in Section 3.3. On the other hand, the expression of the estimated PMF under PSS, $f_{\widehat{T}_0}(\widehat{T}_0 = kT_s)$, as a function of the original PDF was provided in [49], which satisfies the second stage of the analysis mentioned in Section 3.3. Therefore, by combining the second relationship

with the first one, the relationship between the estimated PMF under ISS and the original PDF can then be achieved, which completes the third stage of the analysis mentioned in Section 3.3.

3.7 Analysis of the Sample Size

The previous sections have analysed the channel activity statistics without any constraints on the sample size used to estimate these statistics (i.e., an arbitrarily large sample size as large as required to achieve the best attainable estimation accuracy). Although the work in [68] has analysed the impact of the sample size on the statistics estimation, it was conducted under the assumption that spectrum sensing is perfect (i.e., PSS). Since DSA/CR receivers are more likely to operate under low SNR conditions where sensing errors are likely to occur, this section analyses the impact of the sample size on the channel statistics estimation under (a realistic) ISS scenario [94]. In addition, it finds closed-form expressions for the required sample size of idle/busy periods under ISS to achieve a targeted level of accuracy.

Consider a set $\{T_{i,n}\}_{n=1}^{N}$ to represent N samples of idle/busy periods of a PU within a particular channel. A SU on the other hand monitors the activity of the primary channel using spectrum sensing. The observed duration of the idle/busy periods can easily be corrupted by the presence of sensing errors. Sensing errors divide the original duration of the idle/busy periods into shorter fragments. The number of these fragments as a result is higher than the original number of periods N. In other words, the number of the observed idle/busy periods under ISS, N_{iss}, would be greater than (depending on the probability of spectrum sensing error) the original number of periods N (i.e., $N_{iss} > N$). These N_{iss} periods are short fragments of the original N periods. If probability of sensing error is zero, PSS can be assumed such that the original periods are observed without sensing errors and they are therefore not divided into fragments (i.e., $N_{pss} = N$). The only difference between the idle/busy periods observed under PSS and the original ones is the accuracy of calculating the duration of these periods, which depends on the resolution of the employed sensing period T_s. Since ISS is a more realistic scenario for the DSA/CR systems, it is very important to find a mathematical expression for the number of idle/busy periods observed under ISS, N_{iss}, as a function of the original number of periods N when probabilities of sensing errors, P_{fa} and P_{md}, are predefined by DSA/CR system.

To find the sample size N_{iss} as a function of the original sample size N, consider a single idle period T_0 and a single busy period T_1, which are observed as $\widehat{T}_0$ and $\widehat{T}_1$,

respectively, under PSS (based on K sensing events within each period, assuming they have same duration). If a single false alarm occurs within the K sensing events of the idle period and a single missed detection occurs within the K sensing events of the busy period as shown in Fig. 3.1(b), the total number of idle periods becomes three (the same applies to the busy periods as well). This is because the single false alarm divides the original idle period into two fragments of idle periods and the single missed detection produces another new idle period. As a result, the total number of idle periods becomes three (the two fragments plus the new one). Therefore, one can say:

$$N_{iss} = N + N_{fa} + N_{md}, \tag{3.34}$$

where N_{fa} and N_{md} represent the number of false alarms and missed detections, respectively. They can be found by multiplying the number of sensing events by the probabilities of sensing errors as:

$$\begin{aligned} N_{fa} = KP_{fa} &= \frac{\sum_{n=1}^{N} \widehat{T}_{0,n}}{T_s} \cdot P_{fa} \\ &= \frac{N\mathbb{E}(\widehat{T}_0)}{T_s} P_{fa} = \frac{N\mathbb{E}(T_0)}{T_s} P_{fa}. \end{aligned} \tag{3.35}$$

$$\begin{aligned} N_{md} = KP_{md} &= \frac{\sum_{n=1}^{N} \widehat{T}_{1,n}}{T_s} \cdot P_{md} \\ &= \frac{N\mathbb{E}(\widehat{T}_1)}{T_s} P_{md} = \frac{N\mathbb{E}(T_1)}{T_s} P_{md}. \end{aligned} \tag{3.36}$$

Note that the estimated mean under PSS is equal to the true mean $\mathbb{E}(\widehat{T}_i) = \mathbb{E}(T_i)$ as discussed in [49].

Expression (3.34) is true if all false alarms and missed detecions occur as solo sensing errors within the idle/busy periods. This, however, is not the case as sensing errors can also appear attached to other periods or consecutive to other sensing errors (as discussed in Cases I and II of Section 3.5). Therefore, (3.34) can be corrected as:

$$N_{iss} = N + \grave{N}_{fa} + \grave{N}_{md}, \tag{3.37}$$

where $\grave{N}_{fa}$ and $\grave{N}_{md}$ are found after taking Cases I and II of Section 3.5 into account

as:

$$\grave{N}_{fa} = N\left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right)\grave{P}_{fa}. \tag{3.38}$$

$$\grave{N}_{md} = N\left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right)\grave{P}_{md}. \tag{3.39}$$

After considering these, the final expression of the sample size under ISS as a function of the original sample size and probabilities of sensing errors can be written as:

$$N_{iss} = N\left(1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right)\grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right)\grave{P}_{md}\right). \tag{3.40}$$

From this, the original sample size based on the observed sample size under ISS can also be found as:

$$N = \frac{N_{iss}}{\left(1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right)\grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right)\grave{P}_{md}\right)}. \tag{3.41}$$

3.7.1 Required Sample Size for the Mean Estimation under ISS

Given a set $\{\widehat{T}_{i,n}\}_{n=1}^{N}$ of N idle/busy periods observed under PSS, the mean $\mathbb{E}(\widehat{T}_i)$ of the observed periods can be found based on the sample mean estimator $\widehat{m}_i$:

$$\mathbb{E}(\widehat{T}_i) \approx \widehat{m}_i = \frac{1}{N}\sum_{n=1}^{N}\widehat{T}_{i,n}. \tag{3.42}$$

The maximum relative error of the mean estimator $\widehat{m}_i$ under PSS as a function of the sample size N is found as [68]:

$$\varepsilon_{max}^{\widehat{m}_i} \approx \frac{\kappa}{\mathbb{E}(T_i)}\left[\frac{1}{N}\left(\mathbb{V}(T_i) + \frac{T_s^2}{6}\right)\right]^{\frac{1}{2}}, \tag{3.43}$$

where $\mathbb{V}(T_i)$ denotes the variance of the idle/busy periods, and κ is standard deviation interval defined by the confidence level ρ. For a given confidence level ρ, κ can be found from concentration inequalities as explained in [68]. However, concentration inequalities may lead to loose upper bounds of the maximum relative error. A more accurate result can be achieved by applying the central limit theorem on the mean estimator $\widehat{m}_i$ where

κ can be obtained for a certain confidence level ρ as [68]:

$$\kappa \geq \sqrt{2}\ \mathrm{erf}^{-1}(\rho), \tag{3.44}$$

where erf denotes the Gauss error function [68].

In order to find the maximum relative error of the mean estimator $\breve{m}_i$ under ISS as a function of the sample size N_{iss}, the obtained expression in (3.41), which defines the original sample size as a function of the ISS sample size, is used along with (3.43) as:

$$\varepsilon_{max}^{\breve{m}_i} \approx \frac{\kappa}{\mathbb{E}(T_i)} \left[\frac{\left(1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right) \grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right) \grave{P}_{md}\right)}{N_{iss}} \times \left(\mathbb{V}(T_i) + \frac{T_s^2}{6}\right) \right]^{\frac{1}{2}}. \tag{3.45}$$

Note that if probabilities of sensing errors (P_{fa} and P_{md}) are zero, (3.45) will be the same as (3.43) for the calculated relative error under PSS. Therefore, the result shown in (3.45) can be considered as a general form to find the maximum relative error of the mean estimator based on the sample size under both scenarios.

The required sample size of the idle/busy periods observed under ISS to achieve a targeted maximum relative error of the mean estimator can be found from (3.45) as:

$$N_{iss}^{\breve{m}_i} \approx \left(\frac{\kappa}{\varepsilon_{max}^{\breve{m}_i} \mathbb{E}(T_i)}\right)^2 \left(\mathbb{V}(T_i) + \frac{T_s^2}{6}\right) \left(1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right) \grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right) \grave{P}_{md}\right). \tag{3.46}$$

3.7.2 Required Sample Size for the Duty Cycle Estimation under ISS

The duty cycle of the idle/busy periods of the primary channel is one of the most important statistical information for DSA/CR systems, which helps determine the amount of opportunities available in the primary channels. The channel duty cycle Ψ can be estimated based on the observed sample size under PSS as:

$$\widehat{\Psi} = \frac{\widehat{m}_1}{\widehat{m}_1 + \widehat{m}_0}, \tag{3.47}$$

where $\widehat{m}_1$ and $\widehat{m}_0$ are the sample mean of the busy and idle periods, respectively. The maximum relative error of the duty cycle estimator $\widehat{\Psi}$ under PSS as a function of the

sample size N is found as [68]:

$$\varepsilon_{max}^{\widehat{\Psi}} \approx \frac{\kappa}{\Psi}\left[\frac{1}{N\left[\mathbb{E}(T_0)+\mathbb{E}(T_1)\right]^4}\left\{\left[\mathbb{E}(T_1)\right]^2\left(\mathbb{V}(T_0)+\frac{T_s^2}{6}\right)+\left[\mathbb{E}(T_0)\right]^2\left(\mathbb{V}(T_1)+\frac{T_s^2}{6}\right)\right\}\right]^{\frac{1}{2}}. \tag{3.48}$$

In order to find the maximum relative error of the duty cycle estimator $\breve{\Psi}$ under ISS as a function of the sample size N_{iss}, the obtained expression in (3.41), which defines the original sample size as a function of the ISS sample size, is used along with (3.48) as:

$$\varepsilon_{max}^{\breve{\Psi}} \approx \frac{\kappa}{\Psi}\left[\frac{\left(1+\left(\frac{\mathbb{E}(T_0)}{T_s}-2\right)\grave{P}_{fa}+\left(\frac{\mathbb{E}(T_1)}{T_s}-2\right)\grave{P}_{md}\right)}{N_{iss}\left[\mathbb{E}(T_0)+\mathbb{E}(T_1)\right]^4}\times \right.$$
$$\left.\left\{\left[\mathbb{E}(T_1)\right]^2\left(\mathbb{V}(T_0)+\frac{T_s^2}{6}\right)+\left[\mathbb{E}(T_0)\right]^2\left(\mathbb{V}(T_1)+\frac{T_s^2}{6}\right)\right\}\right]^{\frac{1}{2}}. \tag{3.49}$$

The required sample size of the idle/busy periods observed under ISS to achieve a targeted maximum relative error of the duty cycle estimator can be found from (3.49) as:

$$N_{iss}^{\breve{\Psi}} \approx \left(\frac{\kappa}{\varepsilon_{max}^{\breve{\Psi}}\Psi}\right)^2 \times \frac{\Psi^2\left(\mathbb{V}(T_0)+\frac{T_s^2}{6}\right)+(1-\Psi)^2\left(\mathbb{V}(T_1)+\frac{T_s^2}{6}\right)}{\left[\mathbb{E}(T_0)+\mathbb{E}(T_1)\right]^2}$$
$$\times\left(1+\left(\frac{\mathbb{E}(T_0)}{T_s}-2\right)\grave{P}_{fa}+\left(\frac{\mathbb{E}(T_1)}{T_s}-2\right)\grave{P}_{md}\right) \tag{3.50}$$

3.7.3 Required Sample Size for the Distribution Estimation under ISS

The distribution of the idle/busy periods can also be estimated based on the outcomes of spectrum sensing. Since the experimental measurements in [78] have shown that Generalised Pareto (GP) distribution is the best description for the original idle/busy periods of a PU, the CDF of these periods can be estimated based on the PSS observation as [49]:

$$F_{\widehat{T}_i}(T) = 1-\left[1+\frac{\widehat{\alpha}_i(T-\widehat{\mu}_i)}{\widehat{\lambda}_i}\right]^{-1/\widehat{\alpha}_i}, \quad T \geq \widehat{\mu}_i, \tag{3.51}$$

where $\widehat{\mu}_i$, $\widehat{\lambda}_i$ and $\widehat{\alpha}_i$ are the location, scale and shape parameters of the GP distribution. The location $\widehat{\mu}_i$ also represents the minimum period duration and it can be assumed

to be known $\widehat{\mu}_i \approx \mu_i$, while $\widehat{\lambda}_i$ and $\widehat{\alpha}_i$ can be found as [49]:

$$\widehat{\lambda}_i = \frac{1}{2}\left(1 + \frac{(\widehat{m}_i - \widehat{\mu}_i)^2}{\widetilde{v}_i}\right)(\widehat{m}_i - \widehat{\mu}_i) \tag{3.52a}$$

$$\widehat{\alpha}_i = \frac{1}{2}\left(1 + \frac{(\widehat{m}_i - \widehat{\mu}_i)^2}{\widetilde{v}_i}\right), \tag{3.52b}$$

where $\widehat{m}_i$ and $\widetilde{v}_i$ are the sample mean and variance estimators, respectively. Note that the variance estimator is found in [49] as $\widetilde{v}_i = \widehat{v}_i - \frac{T_s^2}{6}$. The accuracy of the CDF estimator $F_{\widehat{T}_i}(T)$ in (3.51) as a function of the sample size N under PSS is found in terms of the Kolmogorov-Smirnov (KS) distance as [68]:

$$D_{KS}^{F_{\widehat{T}}} = \kappa(1+\alpha_i)^{-\frac{1}{\alpha_i}-1}\left[\frac{1}{N}\left(\frac{1}{\lambda_i^2}\Omega(T_i) + \frac{[(1+\alpha_i)\ln(1+\alpha_i) - \alpha_i]^2}{\alpha_i^4}\Upsilon(T_i)\right)\right]^{\frac{1}{2}}, \tag{3.53}$$

where $\Omega(T_i)$ and $\Upsilon(T_i)$ are given in [68, eq. (45)].

KS distance quantifies the maximum difference or error between the empirical cumulative distribution of the samples and the reference distribution, which can be applied here to find the maximum error between the estimated (discrete) and actual (continuous) distributions. In order to find the KS distance of the CDF estimator $F_{\check{T}_i}(T)$ under ISS as a function of the sample size N_{iss}, the obtained expression in (3.41), which defines the original sample size as a function of the ISS sample size, is used along with (3.53) as:

$$D_{KS}^{F_{\check{T}}} = \kappa(1+\alpha_i)^{-\frac{1}{\alpha_i}-1}\left[\frac{\left(1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right)\grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right)\grave{P}_{md}\right)}{N_{iss}}\right.$$
$$\left.\times\left(\frac{1}{\lambda_i^2}\Omega(T_i) + \frac{[(1+\alpha_i)\ln(1+\alpha_i) - \alpha_i]^2}{\alpha_i^4}\Upsilon(T_i)\right)\right]^{\frac{1}{2}}. \tag{3.54}$$

The required sample size of the idle/busy periods observed under ISS to achieve a targeted KS distance of the CDF estimator can be found from (3.54) as:

$$N_{iss}^{F_{\check{T}}} = \left(\frac{\kappa}{D_{KS}^{F_{\check{T}}}}\right)^2 (1+\alpha_i)^{-\frac{2}{\alpha_i}-2}\left(1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right)\grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right)\grave{P}_{md}\right)$$
$$\times\left(\frac{1}{\lambda_i^2}\Omega(T_i) + \frac{[(1+\alpha_i)\ln(1+\alpha_i) - \alpha_i]^2}{\alpha_i^4}\Upsilon(T_i)\right). \tag{3.55}$$

3.8 Numerical, Simulation and Experimental Results

In order to validate the analyses carried out in this chapter, the numerical results obtained from the derived closed-form expressions are compared with the counterpart obtained by means of both simulations and hardware experiments. Simulations are based on Matlab following a similar procedure as in [75]. A sequence of a sufficiently high number of idle/busy periods T_i is generated with random durations drawn from a Generalised Pareto distribution (using $\mu_i = 10$ t.u., $\lambda_i = 30$ t.u. and $\alpha_i = 0.25$ as the values for the location, scale and shape parameters). Then spectrum sensing is performed on the generated periods by employing a sensing period of T_s, using different values within the interval $(0, \mu_i)$. The calculated idle/busy periods from the sensing decisions represent the corresponding sequence of periods $\widehat{T}_i$ that would be observed by a DSA/CR system under PSS. Introducing sensing errors on the PSS decisions, based on the selected value of P_{fa} and P_{md}, leads to the corresponding sequence of idle/busy periods $\breve{T}_i$ that would be observed under ISS. Finally, the statistics of the periods $\breve{T}_i$ resulting from ISS can be calculated and compared with the original statistics of the generated periods T_i. The experimental results, on the other hand, are obtained by using the USRP-based prototype presented in Chapter 2. The transmitter, which represents the PU, generates a sequence of idle/busy periods from a Generalised Pareto distribution (similar to the simulations settings) and transmits 10^5 pairs of these periods using a 2.462 GHz channel. The receiver on the other hand, which represents the SU, performs spectrum sensing to observe the activity of the signals (idle/busy periods) in the same channel. Different sensing periods T_s (0 ms $< T_s <$ 10 ms) are employed by the USRP to sense the channel activity periodically. The receiver is placed sufficiently far away from the transmitter to ensure that the desired probability of missed detection is reached (further distance for higher probability of missed detection). The probability of false alarm is tuned by adjusting the energy decision threshold at the receiver. At each sensing event, samples are taken from the detected signal at a sampling rate of 56 MS/s for a duration of $\tau = 0.1$ ms. These samples are processed by an energy detection algorithm to decide the state of the channel (either idle or busy). Using the outcomes of these sensing decisions, the duration of the idle/busy periods can be estimated at the receiver side and their statistics can therefore be calculated. By comparing these statistics with the original statistics of the periods generated at the transmitter, the accuracy of the analytical results achieved in this chapter can be verified experimentally under realistic conditions and practical limitations of both transmitter and receiver. It is worth mentioning that the unit of time used in the evaluation of the analytical

expressions is given in a general form of time unit (t.u.). In the experimental scenario, however, where a specific unit needs to be selected, 1 ms is used as the reference time unit (i.e., 1 t.u. = 1 ms).

3.8.1 Estimation of the Minimum Period

The accuracy of the result in (3.2) for the estimated minimum period $\breve{\mu}_i$ under ISS was evaluated in [75] by means of simulations. This evaluation is included here for completeness and to corroborate the simulation validation with experimental results as well. The accuracy is evaluated by calculating the relative error of the estimated minimum idle period $\breve{\mu}_0$ under ISS with respect to its original value μ_0 as $|\breve{\mu}_0 - \mu_0| / \mu_0$. Assuming $\mu_0 = 10$ t.u. and $P_{fa} = P_{md} = 0.1$, the relative error can be found for different T_s values as shown in Fig. 3.8. The obtained results from simulations and experiments perfectly match the analytical expression given by (3.2), which thus corroborate its correctness and accuracy. It can also be noticed that as T_s increases the relative error decreases. This is because the estimated minimum period under ISS is $\breve{\mu}_0 = T_s$ and its value will approach the true minimum μ_0 as T_s tends to μ_0, thus making the relative error tend to zero accordingly. Same observations can also be noticed for any non-zero value of P_{fa} and P_{md} and they are also valid for the estimated minimum busy period $\breve{\mu}_1$ under ISS.

3.8.2 Estimation of the Mean Period

This section validates the analysis presented in Section 3.5 for the estimation of the mean period under ISS. This validation is provided in terms of the relative error of the estimated mean period under ISS $\mathbb{E}(\breve{T}_i)$ with respect to the original mean period $\mathbb{E}(T_i)$ as $\left|\mathbb{E}(\breve{T}_i) - \mathbb{E}(T_i)\right| / \mathbb{E}(T_i)$. Fig. 3.9 shows the calculated relative error based on the achieved mathematical model for the mean period in (3.18) as well as the estimated mean from the simulations and hardware experiments, with respect to the employed sensing period T_s and for both low ($P_{fa} = P_{md} = 0.01$) and high ($P_{fa} = P_{md} = 0.1$) probabilities of sensing errors. As it can be appreciated, there exists a perfect agreement among all three curves, thus demonstrating the accuracy of the closed-form expression obtained in (3.18) for the estimated mean period under ISS as a function of the original mean, sensing period and probabilities of sensing error. When the sensing period T_s increases there will be less sensing events within the observed idle/busy periods, and eventually less number of sensing errors, which means less corrupted fragments resulting from these sensing errors. For this, the relative error of the estimated mean

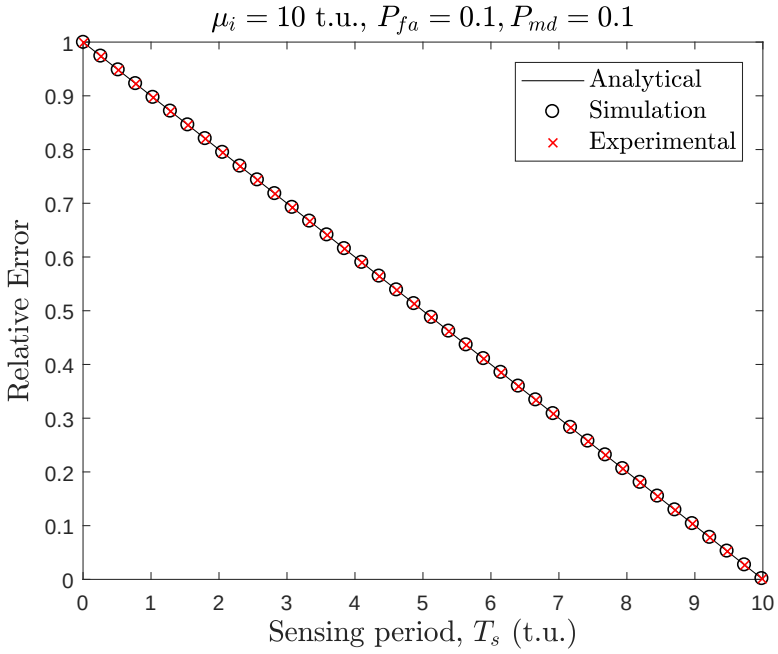


Figure 3.8: Relative error of the calculated minimum period $\breve{\mu}_i$ under ISS.

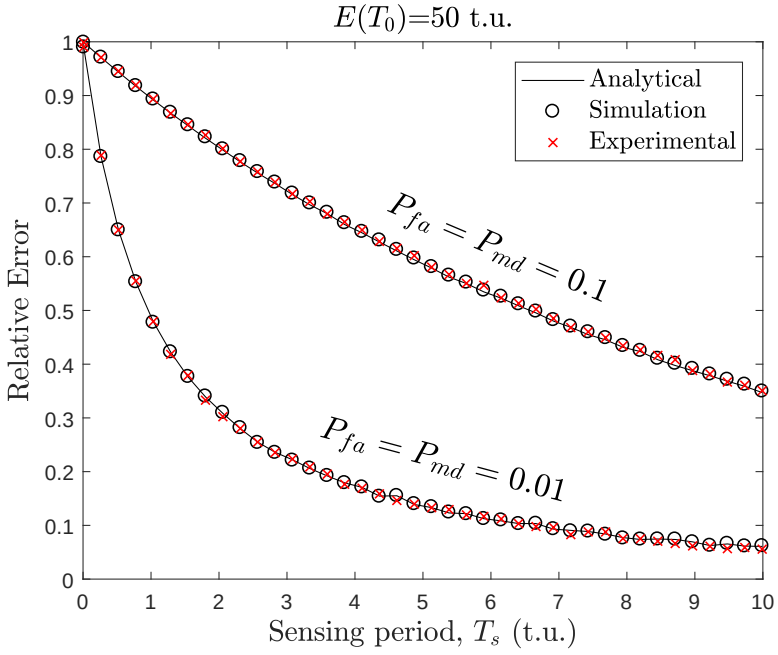


Figure 3.9: Relative error of the calculated mean $\mathbb{E}(\breve{T}_0)$ under ISS.

period as shown in Fig. 3.9 reduces for higher T_s values. On the other hand, when the probabilities of sensing error (P_{fa} and P_{md}) increase the number of sensing errors increases as well, thus causing higher relative error in the estimation of the mean period.

3.8.3 Estimation of the Distribution

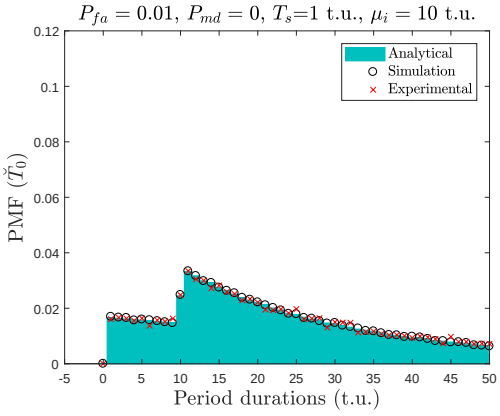
This section validates the analysis presented in Section 3.6 for the estimation of the distribution of the periods under ISS. Fig. 3.10 compares the theoretical expression in (3.33) for the PMF of the periods observed under ISS with the equivalent results obtained from simulations and hardware experiments. As it can be observed, there is a perfect agreement for all the considered cases shown in Fig. 3.10. A more quantitative comparison can be performed based on the well-known Kolmogorov-Smirnov (KS) distance [91], which in the context of this work is defined as the maximum absolute difference between the CDF of the periods observed under ISS, and the CDF of the original periods:

$$D_{KS} = \sup_{k} \mid F_{T_i}(kT_s) - F_{\breve{T}_i}(kT_s) \mid \text{, where } k \in \mathbb{N}^+. \tag{3.56}$$

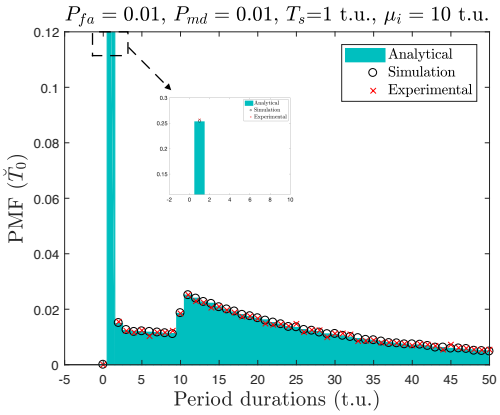
While the original period durations T_i and their CDF $F_{T_i}(\cdot)$ can be assumed to be continuous in general, the periods observed under ISS $\breve{T}_i$ are integer multiple values of the employed sensing period and therefore their CDF $F_{\breve{T}_i}(\cdot)$ is discrete. In order to enable the comparison between these continuous and discrete distributions based on (3.56), the continuous distribution $F_{T_i}(\cdot)$ is evaluated at discrete points for which $F_{\breve{T}_i}(\cdot)$ is defined (i.e., $T_i = kT_s$, $k \in \mathbb{N}^+$). The KS distance as defined in the context of this work in (3.56) is first evaluated numerically based on (3.33) and then compared to the corresponding KS distance obtained from simulations and hardware experiments. The results are shown in Fig. 3.11 with respect to the sensing period T_s and using high and low probabilities of sensing errors. It is observed that as the probability of sensing error increases as the distance of the calculated distribution under ISS from its actual value increases as well (i.e., lower accuracy). It can also be noticed that there is a perfect agreement among analytical, simulation and experimental results, which validates the theoretical expression obtained in (3.33).

3.8.4 Sample Size Evaluation

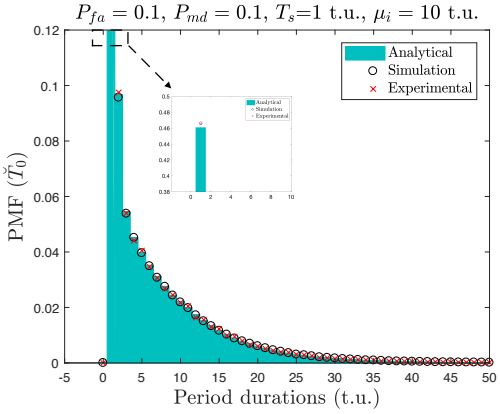
In this section the obtained analytical results for the sample size are validated by means of simulations and experiments. First of all, the obtained expression in (3.41), which



(a) $P_{fa} = 0.01$, $P_{md} = 0$.



(b) $P_{fa} = 0.01$, $P_{md} = 0.01$.



(c) $P_{fa} = 0.1$, $P_{md} = 0.1$.

Figure 3.10: Estimating the PMF of the idle periods under ISS using different probabilities of sensing error and when $T_s = 1$ t.u..

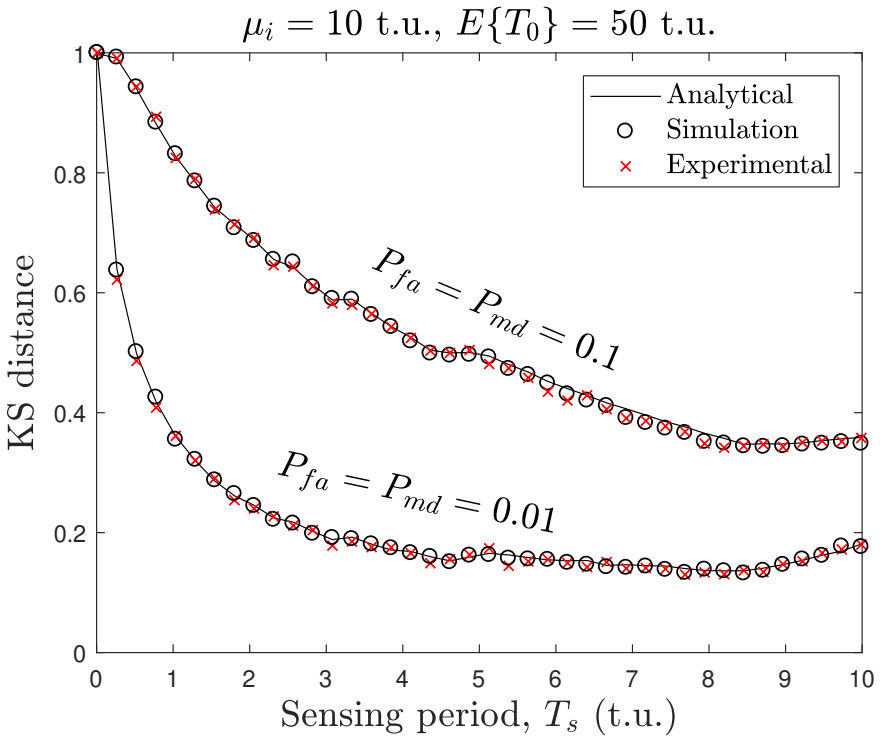


Figure 3.11: KS distance of the calculated distribution under ISS.

provides a mathematical relationship between the original sample size of the idle/busy periods and the observed sample size under ISS, is tested. For which, a large sample size of $N = 10^4$ of idle/busy periods is generated and then observed under ISS using different probabilities of sensing error ($P_{fa}, P_{md} \in \{0.001, 0.01, 0.1\}$). The simulation results of the observed sample size under ISS can then be compared with the calculated one using the achieved expression in (3.41). As shown in Fig. 3.12, the analytical results match the simulation results for the calculated sample size under ISS for different scenarios of probabilities of sensing error and with respect to T_s. It can also be noticed that the number of the observed periods under ISS increases as the probabilities of sensing error increase as well, this is because the presence of more sensing errors result in a larger number of short fragments of the original periods.

After validating (3.41), the obtained maximum relative error expressions, (3.45) and (3.49) for the mean and duty cycle estimations, respectively, as well as the KS distance expression (3.54) for the CDF estimation can then be validated. In the simulation, the impact of the sample size on the mean, duty cycle, and distribution estimations under ISS is evaluated with respect to the obtained analytical expressions (using a confidence

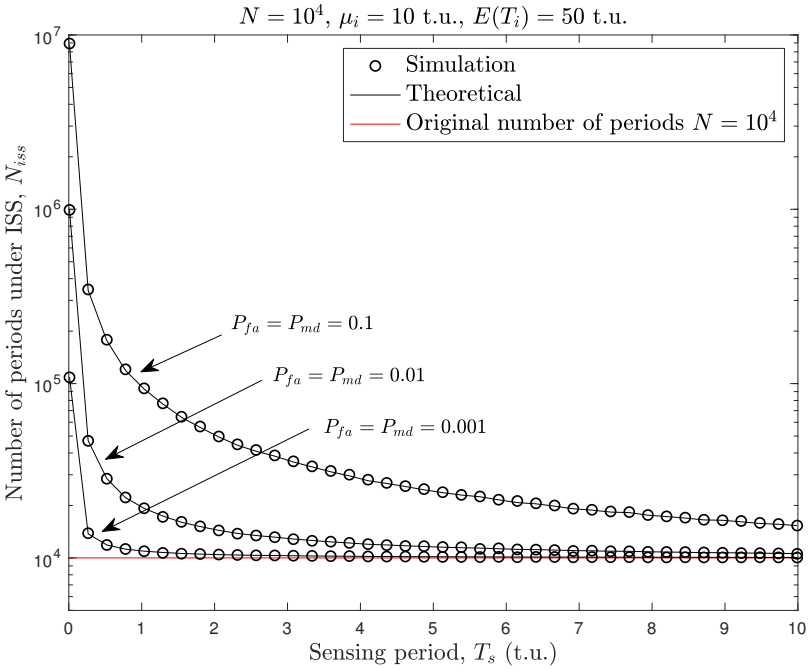


Figure 3.12: The sample size under ISS as a function of the sensing period T_s, when the original sample size $N = 10^4$.

level $\rho = 0.95$). These analyses are also validated experimentally using the USRP-based prototype proposed in Chapter 2. As shown in Fig. 3.13, the relative errors of estimating the mean and duty cycle of the periods as well as the KS distance of estimating the CDF of the periods decrease as the sample size of the ISS observations increases. It can also be noted that the analytical results obtained in this section reproduce accurately the sample sizes required to achieve the desired estimation accuracies and can therefore be useful in DSA/CR system designs under any scenario of spectrum sensing (especially the realistic ISS scenario).

3.9 Summary

This chapter has addressed a highly challenging problem in spectrum sharing systems, which is the problem of observing inaccurate channel activity statistics under ISS. Such problem is poorly addressed in the literature without deep and rigorous mathematical analyses taking into account all the factors that influence the estimation accuracy of the channel activity statistics. The performance of spectrum sharing systems can be improved significantly by knowing the statistical information of the channel. Therefore,

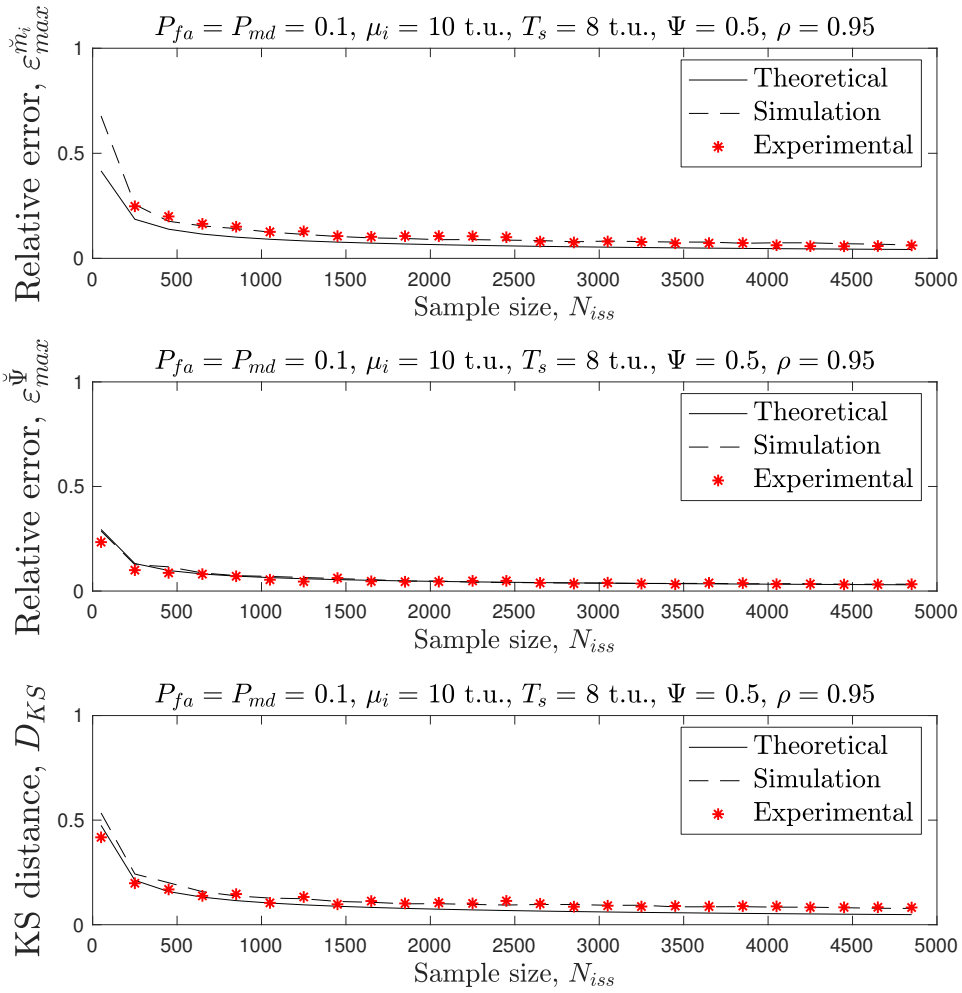


Figure 3.13: (Top) Relative error of the mean estimator, (Middle) Relative error of the duty cycle estimator, and (Bottom) KS distance of the CDF estimator with respect to the sample size under ISS, when confidence level $\rho = 0.95$.

it is important to obtain this information accurately especially under realistic SNR conditions (i.e., ISS). In this context, this chapter has studied analytically the impact of the sensing errors on statistics observation, for which mathematical relationships in closed-form between the observed channel activity statistics under ISS and their corresponding original statistics have been found. A set of closed-form expressions for several statistical metrics has been obtained as a function of the parameters used to configure spectrum sensing operation (i.e., probability of sensing error and sensing period). In addition, the impact of the sample size on the estimation of these statistics has been analysed, and closed-form expressions for the required sample size under ISS

to achieve a targeted level of accuracy have been obtained. These analytical results have been validated by means of simulations and experiments. The obtained closed-form expressions in this chapter will serve as the basis to propose novel estimation methods to overcome the degrading effects of sensing errors in the next chapter, which can provide accurate estimations of the channel activity statistics even under severe ISS conditions.

Chapter 4

Proposed Estimation Methods for Channel Activity Statistics

4.1 Introduction

Channel activity statistics are conventionally estimated directly from the sensing decisions of spectrum sensing. However, this approach can be highly sensitive to the presence of sensing errors as it has been shown in the previous chapter. Another approach based on reconstruction methods in the form of algorithms as in [75–77] has been proposed to correct the estimation of the channel activity statistics under ISS. These works however suffer from the following limitations: i) no closed-form expressions are provided for these statistics, only heuristic estimation methods in the form of algorithms, and ii) the employed reconstruction algorithms assume perfect knowledge of some of the channel statistical parameters (e.g., the minimum idle/busy period).

Given the limitations of the previous works, this chapter investigates three different approaches to improve the estimation of the channel activity statistics under ISS. The first approach proposes novel estimation methods for the channel activity statistics under ISS based on the closed-form expressions achieved in Chapter 3, without making any assumption about the activity of the channel. The proposed methods provide highly accurate estimation for the channel activity statistics under ISS, outperforming the existing methods in the literature. The second approach investigates the reconstruction method, for which a new reconstruction algorithm is introduced to correct the observation of the channel activity statistics under ISS, which can achieve the same accuracy of the latest reconstruction method in the literature without requiring any prior knowledge of the channel activity, opposite to all the previous reconstruction

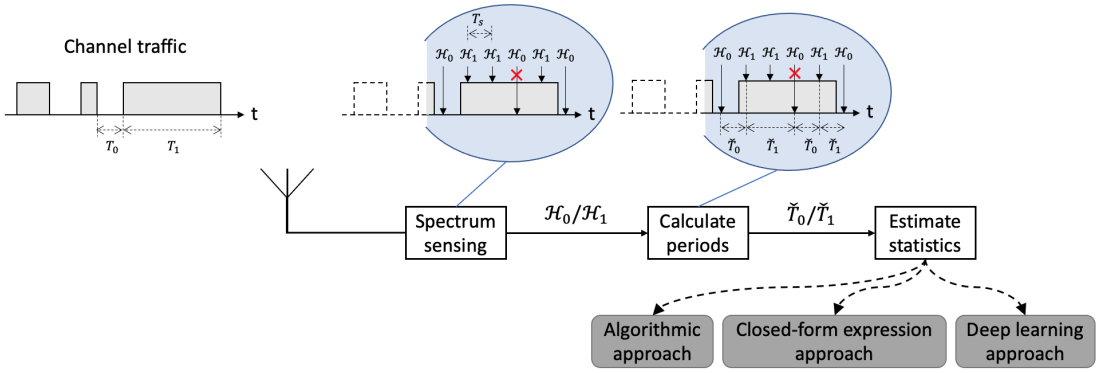


Figure 4.1: Channel activity statistics estimation in spectrum sharing system.

methods. Finally, this chapter introduces a novel approach, Traffic Learning (TL), as deep learning approach for obtaining accurate statistical information of the channel traffic in spectrum sharing systems. This approach learns from the imperfect observations of the channel activity statistics in order to predict their accurate estimations. All these approaches, which can be illustrated as in Fig. 4.1, will be investigated in detail and their performance will be evaluated and compared among each other in this chapter.

4.2 Estimation Methods Based on Closed-Form Expressions

In this section the mathematical expressions obtained in the previous chapter will be exploited to derive novel estimation methods that are able to provide accurate statistical information of the channel activity under ISS.

4.2.1 Estimation Method for the Minimum Period

As it was discussed in Section 3.4, the presented expression in (3.2), $\breve{\mu}_i = T_s$, shows that the estimated minimum period under ISS $\breve{\mu}_i$ is solely dependent on the sensing period T_s (regardless of the original value of the minimum period). Therefore, it is impossible to derive an estimation method from this expression in order to provide an accurate estimation for the original minimum period μ_i. However, another approach will be proposed in Section 4.4 using Deep Learning method to solve such problem through learning from the ISS observations in order to predict the accurate estimation of the minimum period.

4.2.2 Estimation Method for the Mean Period

The closed-form expression obtained in (3.18), which provides a mathematical relationship between the estimated mean under ISS $\mathbb{E}(\breve{T}_i)$ and the original mean $\mathbb{E}(T_i)$, suggests a novel method to accurately find the original mean of the channel periods from the outcomes of the ISS estimates. In this section, therefore, a novel estimation method is proposed to accurately estimate the original value of the mean of the idle/busy periods based on the ISS observations. The analytical result in (3.18) summarises two expressions, namely the estimated mean of idle periods under ISS (when $i = 0$) and the estimated mean of busy periods under ISS (when $i = 1$), which can be written here as:

$$\mathbb{E}(\breve{T}_0) = \frac{\mathbb{E}(T_0) - \mathbb{E}(T_0)P_{fa} + \mathbb{E}(T_1)P_{md}}{1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right)\grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right)\grave{P}_{md}}, \tag{4.1}$$

$$\mathbb{E}(\breve{T}_1) = \frac{\mathbb{E}(T_1) + \mathbb{E}(T_0)P_{fa} - \mathbb{E}(T_1)P_{md}}{1 + \left(\frac{\mathbb{E}(T_0)}{T_s} - 2\right)\grave{P}_{fa} + \left(\frac{\mathbb{E}(T_1)}{T_s} - 2\right)\grave{P}_{md}}. \tag{4.2}$$

The above two expressions can be solved for the original means of periods, i.e., $\mathbb{E}(T_0)$ and $\mathbb{E}(T_1)$, as shown in (4.3) and (4.4), respectively.

$$\mathbb{E}(T_0) = \mathbb{E}(\breve{T}_0)\frac{1 - 2\grave{P}_{fa} - 2\grave{P}_{md}}{1 - P_{fa} - \frac{\grave{P}_{fa}}{T_s}\mathbb{E}(\breve{T}_0)} + \mathbb{E}(T_1)\frac{\frac{\grave{P}_{md}}{T_s}\mathbb{E}(\breve{T}_0) - P_{md}}{1 - P_{fa} - \frac{\grave{P}_{fa}}{T_s}\mathbb{E}(\breve{T}_0)}, \tag{4.3}$$

$$\mathbb{E}(T_1) = \mathbb{E}(\breve{T}_1)\frac{1 - 2\grave{P}_{fa} - 2\grave{P}_{md}}{1 - P_{md} - \frac{\grave{P}_{md}}{T_s}\mathbb{E}(\breve{T}_1)} + \mathbb{E}(T_0)\frac{\frac{\grave{P}_{fa}}{T_s}\mathbb{E}(\breve{T}_1) - P_{fa}}{1 - P_{md} - \frac{\grave{P}_{md}}{T_s}\mathbb{E}(\breve{T}_1)}. \tag{4.4}$$

By substituting (4.4) in (4.3), a new expression can be derived in (4.5), denoted as $\mathbb{E}(\widetilde{T}_0)$, to represent the accurate estimation of the original mean $\mathbb{E}(T_0)$ as a function of the estimated mean under ISS (i.e., $\mathbb{E}(\breve{T}_0)$ and $\mathbb{E}(\breve{T}_1)$), probability of sensing error (i.e., P_{fa} and P_{md}), and sensing period (i.e., T_s). Similarly, the estimator for the mean of the busy periods $\mathbb{E}(\widetilde{T}_1)$ can be derived by substituting (4.3) in (4.4) to obtain (4.6). The final expression for the mean estimator of the idle/busy periods $\mathbb{E}(\widetilde{T}_i)$ can be written in a compact form as shown in (4.7).

The result in (4.7) represents a novel method to accurately estimate the original value of the mean of the channel periods based on the estimated mean under ISS, probabilities of sensing error, and sensing period. It is worth mentioning that the probabilities of sensing error P_{fa} and P_{md}, and sensing period T_s are all configured based on the spectrum sensing algorithm used by the DSA/CR system and are known.

$$\mathbb{E}(T_0) \approx \mathbb{E}(\widetilde{T_0}) = \frac{\Big(\mathbb{E}(\breve{T_0})(1-P_{md}) - \mathbb{E}(\breve{T_1})P_{md}\Big)\Big(1-2\grave{P}_{fa}-2\grave{P}_{md}\Big)}{\Big(1-P_{fa}-\frac{\grave{P}_{fa}}{T_s}\mathbb{E}(\breve{T_0})\Big)\Big(1-P_{md}-\frac{\grave{P}_{md}}{T_s}\mathbb{E}(\breve{T_1})\Big)-\Big(\frac{\grave{P}_{fa}}{T_s}\mathbb{E}(\breve{T_1})-P_{fa}\Big)\Big(\frac{\grave{P}_{md}}{T_s}\mathbb{E}(\breve{T_0})-P_{md}\Big)} \tag{4.5}$$

$$\mathbb{E}(T_1) \approx \mathbb{E}(\widetilde{T_1}) = \frac{\Big(\mathbb{E}(\breve{T_1})(1-P_{fa}) - \mathbb{E}(\breve{T_0})P_{fa}\Big)\Big(1-2\grave{P}_{fa}-2\grave{P}_{md}\Big)}{\Big(1-P_{fa}-\frac{\grave{P}_{fa}}{T_s}\mathbb{E}(\breve{T_0})\Big)\Big(1-P_{md}-\frac{\grave{P}_{md}}{T_s}\mathbb{E}(\breve{T_1})\Big)-\Big(\frac{\grave{P}_{fa}}{T_s}\mathbb{E}(\breve{T_1})-P_{fa}\Big)\Big(\frac{\grave{P}_{md}}{T_s}\mathbb{E}(\breve{T_0})-P_{md}\Big)} \tag{4.6}$$

$$\mathbb{E}(T_i) \approx \mathbb{E}(\widetilde{T_i}) = \frac{\Big(\mathbb{E}(\breve{T_i})(1-P_{md}^{1-i}P_{fa}^{i}) - \mathbb{E}(\breve{T}_{1-i})P_{md}^{1-i}P_{fa}^{i}\Big)\Big(1-2\grave{P}_{fa}-2\grave{P}_{md}\Big)}{\Big(1-P_{fa}-\frac{\grave{P}_{fa}}{T_s}\mathbb{E}(\breve{T_0})\Big)\Big(1-P_{md}-\frac{\grave{P}_{md}}{T_s}\mathbb{E}(\breve{T_1})\Big)-\Big(\frac{\grave{P}_{fa}}{T_s}\mathbb{E}(\breve{T_1})-P_{fa}\Big)\Big(\frac{\grave{P}_{md}}{T_s}\mathbb{E}(\breve{T_0})-P_{md}\Big)} \tag{4.7}$$

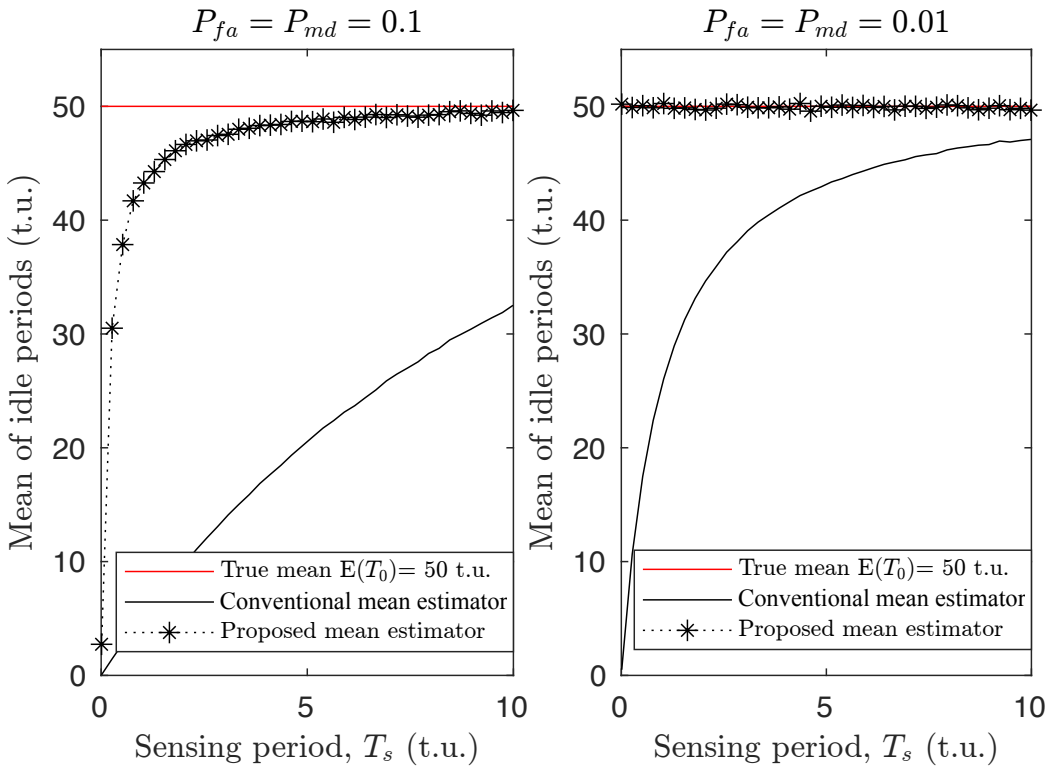


Figure 4.2: Accuracy of the proposed mean estimator $\mathbb{E}(\widetilde{T}_i)$.

Therefore, this method is applicable in real hardware implementations, as opposed to most previous estimation methods proposed in the existing literature [75, 77].

The accuracy of the proposed mean estimation method in (4.7) can be examined following the same simulation procedure in Section 3.8, where a large number of idle/busy periods T_i are generated and then observed under ISS as $\breve{T}_i$. The proposed estimation methods can then be applied to find the original statistics of T_i from the observed ones $\breve{T}_i$ under ISS. Fig. 4.2 shows the accuracy of the achieved mean estimator in (4.7) in comparison with that attained by the conventional mean estimator given in (3.4). As it can be noticed, when the mean period is directly estimated from the channel periods observed under ISS based on (3.4), the resulting estimate is highly inaccurate. On the other hand, the proposed estimator in (4.7) provides a nearly perfect estimation under low probabilities of sensing errors ($P_{fa} = P_{md} = 0.01$) and even a significantly high accuracy under high probabilities of sensing errors ($P_{fa} = P_{md} = 0.1$), which approaches the exact mean period as the employed sensing period increases. The proposed mean estimator using this approach (i.e., closed-form expression) will also be compared with other approaches investigated later on in this chapter.

4.2.3 Estimation Method for the Duty Cycle

The Duty Cycle (DC), also referred to as the channel occupancy rate or the channel load, is one of the most widely used statistical metrics in DSA/CR systems due to its simplicity and applicability in enhancing the efficiency of spectrum utilisation. The DC of the channel, denoted as Ψ, is traditionally estimated from the spectrum sensing observations by dividing the number of busy sensing events over the entire number of the sensing events [95–97]. Although this is the most widely used approach in the literature, it is highly sensitive to the presence of the sensing errors. On the other hand, and in the context of PSS, another method was proposed in [62] to estimate the DC of the channel activity based on the mean of the idle/busy periods as:

$$\Psi = \frac{\mathbb{E}(T_1)}{\mathbb{E}(T_1) + \mathbb{E}(T_0)}. \tag{4.8}$$

The observed idle/busy periods $\widehat{T}_i$ under PSS can serve to obtain an accurate estimation for the mean $\mathbb{E}(\widehat{T}_i)$ and thus an accurate estimation of the DC of the primary channel as well. However, in the ISS scenario, the observed idle/busy periods $\breve{T}_i$ could be significantly corrupted because of the sensing errors, as explained in Chapter 3, and the estimated mean of these periods could be highly inaccurate. Therefore, estimating the DC of the channel under ISS as given in (4.9), which depends solely on the mean of the observed periods, would be highly inaccurate (i.e., $\breve{\Psi} \neq \Psi$).

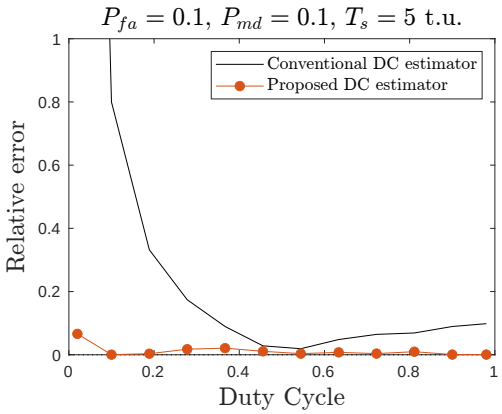
$$\breve{\Psi} = \frac{\mathbb{E}(\breve{T}_1)}{\mathbb{E}(\breve{T}_1) + \mathbb{E}(\breve{T}_0)} \neq \Psi. \tag{4.9}$$

An alternative approach is here proposed based on the analysis presented in Section 4.2.2. The proposed mean estimator $\mathbb{E}(\widetilde{T}_i)$ in (4.7), which can be used to estimate the mean of the idle/busy periods accurately under ISS, can also be exploited to find the DC of the channel under ISS. Therefore, by substituting the mean estimator of (4.7) in (4.8), a new DC estimator $\widetilde{\Psi}$ is obtained as:

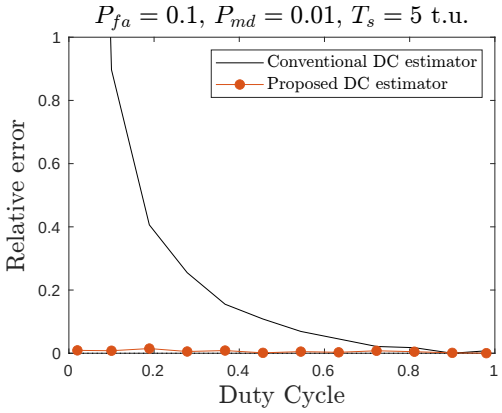
$$\widetilde{\Psi} = \frac{\mathbb{E}(\widetilde{T}_1)}{\mathbb{E}(\widetilde{T}_1) + \mathbb{E}(\widetilde{T}_0)} \approx \Psi, \tag{4.10}$$

where $\mathbb{E}(\widetilde{T}_0)$ and $\mathbb{E}(\widetilde{T}_1)$ are the accurate estimations of $\mathbb{E}(T_0)$ and $\mathbb{E}(T_1)$ provided by (4.5) and (4.6), respectively, and thus the obtained $\widetilde{\Psi}$ provides an accurate estimation of the original DC Ψ.

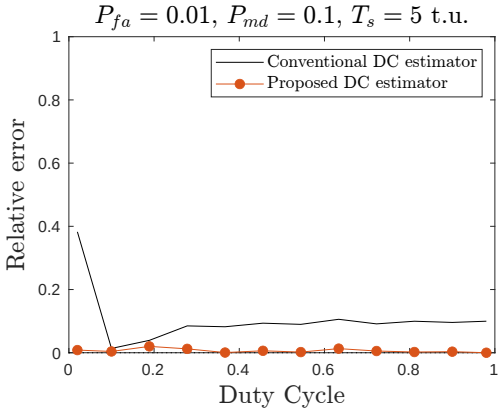
The accuracy of the DC estimation method can also be validated as shown in Fig. 4.3.



(a)



(b)



(c)

Figure 4.3: Relative error of the proposed DC estimator $\widetilde{\Psi}$ for different P_{fa} and P_{md} ($T_s = 5$ t.u.).

This accuracy is presented in terms of the relative error ($|\widetilde{\Psi} - \Psi|/\Psi$) of the estimation methods for the conventional DC estimator (which is found directly by taking the ratio of the busy sensing decisions to the total number of sensing decisions) as well as the proposed DC estimator (which is found by (4.10) based on the mean period estimator). The results are shown for different combinations of low/high probabilities of sensing errors and for the whole range of possible DC values. A sensing period of $T_s = 5$ t.u. is here considered for illustrative purposes but similar results are obtained for other values of the sensing period as well. As it can be clearly observed, the proposed DC estimator significantly outperforms the conventional method widely used in the literature to estimate the DC when a realistic scenario of ISS is considered. The relative error is almost zero in all cases, even when high probabilities of sensing errors are considered (e.g., $P_{fa} = P_{md} = 0.1$ in Fig. 4.3(a)). The excellent level of accuracy achieved by the proposed DC estimation method, even in the presence of severe probabilities of sensing errors, highlights its practical utility in realistic scenarios.

4.2.4 Estimation Method for the Distribution

The closed-form expression in (3.33) suggests a novel method to accurately estimate the PMF of the channel periods under PSS, and therefore the original PDF, from the outcomes of the ISS estimates. Consequently, the analytical result in (3.33) can be solved for the PMF obtained under PSS $f_{\widehat{T}_0}(\widehat{T}_0 = kT_s)$ as a function of the PMF obtained under ISS $f_{\breve{T}_0}(\breve{T}_0 = kT_s)$, $\mathbb{E}(\breve{T}_i)$, P_{fa}, P_{md} and T_s. This can be achieved by simplifying (3.33) as shown below:

$$f_{\breve{T}_0}(\breve{T}_0 = kT_s) = a_k \left[c \cdot f_{\widehat{T}_0}(\widehat{T}_0 = kT_s) - \sum_{m=1}^{k-1} \left[f_{\widehat{T}_0}(\widehat{T}_0 = mT_s) \left(m - k + \frac{2 - P_{fa}}{P_{fa}} \right) \right] + b_k \right], \tag{4.11}$$

where

$$a_k = \beta P_{fa}^2 (1 - P_{fa})^k, \tag{4.12a}$$

$$b_k = \frac{\mathbb{E}(\widehat{T}_0)}{T_s} + \frac{2 - P_{fa}}{P_{fa}} - k + \left(\frac{\mathbb{E}(\widehat{T}_1)}{T_s} - k - 1 \right) \frac{P_{md}^k (1 - P_{md})^2}{P_{fa}^2 (1 - P_{fa})^k}, \tag{4.12b}$$

$$c = \left(\frac{P_{fa}^2 - 2P_{fa} + 1}{P_{fa}^2} \right). \tag{4.12c}$$

The equation shown in (4.11) can then be solved to find $f_{\widehat{T}_0}(\widehat{T}_0 = kT_s)$ as follows:

$$f_{\widehat{T}_0}(\widehat{T}_0 = kT_s) = \frac{1}{c}\left[\frac{f_{\breve{T}_0}(\breve{T}_0 = kT_s)}{a_k} + \sum_{m=1}^{k-1}\left[f_{\widehat{T}_0}(\widehat{T}_0 = mT_s)\left(m - k + \frac{2 - P_{fa}}{P_{fa}}\right)\right] - b_k\right]. \tag{4.13}$$

Equation (4.13) can be used as a recursive formula (where $k \in \mathbb{N}^+$) whose initial value can be found for $k = 1$ and successive values can be found by iterating over k as shown in (4.14).

$$f_{\widehat{T}_0}(\widehat{T}_0 = kT_s) = \begin{cases} \dfrac{1}{c}\left[\dfrac{f_{\breve{T}_0}(\breve{T}_0 = kT_s)}{a_k} - b_k\right] & \text{for } k = 1 \\ \dfrac{1}{c}\left[\dfrac{f_{\breve{T}_0}(\breve{T}_0 = kT_s)}{a_k} + \displaystyle\sum_{m=1}^{k-1}\left[f_{\widehat{T}_0}(\widehat{T}_0 = mT_s)\left(m - k + \dfrac{2 - P_{fa}}{P_{fa}}\right)\right] - b_k\right] & \text{for } k > 1 \end{cases} \tag{4.14}$$

Note that the mean period $\mathbb{E}(\widehat{T}_i)$ in (4.12b) can be substituted with the corresponding mean estimator (4.7) proposed in Section 4.2.2. As a result, the expression in (4.14) represents the estimator for the PMF of the periods that would be observed under PSS as a function of the PMF and mean obtained under ISS as well as P_{fa}, P_{md}, and T_s. Notice that the resulting PMF estimated from (4.14) is still a discrete distribution. A continuous estimation of the original distribution can be obtained by interpolating through the middle points of each discrete step in this PMF $f_{\widehat{T}_0}(\widehat{T}_0 = kT_s)$, which is justified by the analytical result obtained in [49, eq. (38)].

In the same way, the accuracy of the novel estimator proposed in (4.14) can be examined for the estimation of the original distribution of the periods from the ISS observations. Therefore, by comparing the KS distance of the proposed estimator with the KS distance resulting from the direct estimation under ISS (without using any estimation method), the improvement of the proposed estimation approach can be assessed. As it can be appreciated in Fig. 4.4, the proposed estimator leads to a significantly improved accuracy in the estimation of the true distribution of the periods based on the ISS outcomes, providing a nearly perfect estimation ($D_{KS} \approx 0$) under low sensing error probability and a significantly more accurate estimation even under high sensing error probability (provided that the appropriate sensing period is selected). The estimation using this approach will also be compared with the performance of the

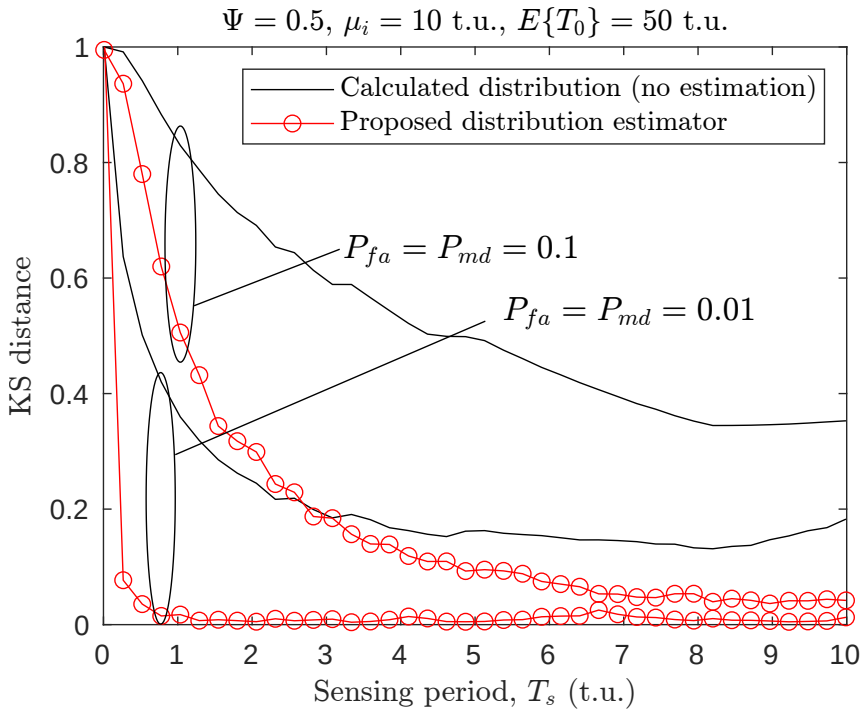


Figure 4.4: KS distance of the proposed distribution estimator.

other studied approaches later on in the the comparison section (Section 4.5). These results highlight the feasibility of obtaining a highly accurate estimation of the channel activity statistics from spectrum sensing observations, even in the presence of sensing errors, if the methods proposed in this section are employed.

4.3 Estimation Methods Based on Reconstruction Algorithms

A reconstruction technique was first introduced in [75], in which three different methods were presented to reconstruct the idle/busy periods estimated under ISS in order to improve the estimation accuracy of the statistical information of the primary channel. Then [77] developed three additional reconstruction algorithms which outperform the methods presented in [75]. However, all the above mentioned methods require perfect knowledge of the minimum period that the PU is active or in active within the primary channel. In a practical scenario such information may be unknown to DSA/CR systems, in which case the system is unable to obtain by itself an accurate estimation based on spectrum sensing observations as discussed in Section 3.4. In this context, this work

proposes a novel reconstruction algorithm that can reach the performance of the latest algorithm, which is presented in [77], but without requiring any additional knowledge about the primary channel and assuming that the DSA/CR system is blind to the PUs activity patterns.

In order to understand how the reconstruction technique can help obtain more accurate estimation for the primary channel statistics under ISS, the estimation of the statistical distribution of the primary channel periods is considered. The estimated distribution of these periods under ISS will be highly inaccurate compared with the original distribution as discussed in Chapter 3. A single false alarm error as shown in Fig. 3.1b, could corrupt the estimation of T_0 by producing three new shorter period durations, which are (idle) $\breve{T}_0$, (busy) $\breve{T}_1$, and (another idle) $\breve{T}_0$. In addition, the resulting $\breve{T}_1$ period from the false alarm itself has a duration equal to the sensing period T_s (where T_s should be shorter than the minimum period μ_i). Since the majority of the sensing errors could appear as individual short periods with a duration of T_s, these errors can easily be identified when the minimum period μ_i of the channel is known. This inspired [75] to propose three methods to reconstruct the sensing decisions affected by the errors in order to provide more accurate estimation for the statistical distribution of the idle/busy periods under ISS. Given a set $\{\breve{T}_{i,n}\}_{n=1}^{N_{iss}}$ of N_{iss} periods observed under ISS, the methods discussed below can be employed when the value of μ_i is perfectly known [75].

4.3.1 Method 1

This method (from [75]) simply assumes that any estimated period under ISS shorter than the minimum period μ_i is an error. Therefore, it discards any period $\breve{T}_{i,n} < \mu_i$ and does not include it in the computation of the distribution of the channel periods under ISS.

4.3.2 Method 2

This method (from [75]) not only discards the periods which are $\breve{T}_{i,n} < \mu_i$, but also discards the preceding ($\breve{T}_{i,n-1}$) and the subsequent ($\breve{T}_{i,n+1}$) periods since these periods could be the remaining fragments of the original period as illustrated in Fig. 3.1b.

4.3.3 Method 3

Instead of discarding the periods, this method (from [75]) suggests to reconstruct the incorrect periods by joining all the possible fragments of an original period together.

This can be performed by combining the periods which are $\breve{T}_{i,n} < \mu_i$ with the preceding and subsequent periods as $(\breve{T}_{i,n-1} + \breve{T}_{i,n} + \breve{T}_{i,n+1})$ and considering the resulting value as a single period of the opposite type to $\breve{T}_{i,n}$.

The explained methods of [75] can noticeably improve, to some extent, the accuracy of the estimated statistics of the primary channel under ISS. However, as shown in [75], Method 1 can perform better than the other methods. This is because, with the reconstruction technique in Method 3 the reconstructed periods could sometimes be incorrectly considered as the opposite type of the original type of the periods and this degrades the accuracy of statistics estimation. In this context, [77] developed three other reconstruction algorithms, which could outperform the methods in [75]. These reconstruction algorithms also assume perfect knowledge of the minimum period μ_i of the primary channel. In this work, the most significant reconstruction algorithm proposed by [77] is considered which is here referred to as Method 4.

4.3.4 Method 4

In this method (from [77]), a threshold $\beta T_s < \mu_i$ is defined (where $\beta \in \mathbb{N}^+$), which can be tuned as explained in [77]. Starting from an initial observed period $\breve{T}_{i,n}$ that has a duration less than the threshold (i.e., $\breve{T}_{i,n} < \beta T_s$), all the subsequent periods (idle and busy) are checked until a period of the opposite type with a duration greater than the threshold (i.e, $\breve{T}_{1-i,n+N} > \beta T_s$) is found. These periods are then reconstructed by summing $(\breve{T}_{i,n} + ... + \breve{T}_{i,n+N-1})$ to form a single reconstructed period of the same type as $\breve{T}_{i,n}$. The next period $\breve{T}_{1-i,n+N}$, which is the opposite type of the previously reconstructed period, is then taken as the new initial period for a new reconstruction. This process is repeated over the whole sequence of the observed periods in an attempt to reconstruct the whole set of fragments of the original idle/busy periods.

4.3.5 Method 5 (proposed method)

All the previous discussed methods, including the latest reconstruction algorithm in Section 4.3.4, assume a perfect knowledge of the minimum period μ_i of the PU occupancy patterns within a particular channel. In a practical scenario such information may be unknown to the DSA/CR system. Therefore, this section proposes a novel reconstruction method which can reach the performance achieved by [77], but without requiring any additional knowledge and assuming that the DSA/CR system is blind to the PUs activity and inactivity patterns [98]. In this method a new algorithm is developed that depends on another parameter, which is the mean of idle/busy periods

m_i, rather than the minimum period μ_i. This algorithm takes the advantage of the novel estimator proposed in (4.7) of Section 4.2.2 for estimating the mean of idle/busy periods accurately even under high probability of sensing error (i.e., no prior knowledge is required for this parameter since it can be estimated accurately based on the analysis achieved in 4.2.2).

As explained before, the estimated periods under ISS are divided into shorter fragments due to the sensing errors. As a result, the calculated mean of these periods will be much lower than its true value (i.e., the mean when there are no sensing errors). Based on this observation, the proposed algorithm reconstructs the periods in an iteration process and in each iteration the mean of the reconstructed periods will be calculated until it reaches the value of the mean obtained using the estimator proposed in (4.7) (the convergence performance of this process will also be evaluated in the next section). To explain this in more detail, let us consider the first iteration as an example. In this iteration the shortest periods, which are $\breve{T}_{i,n} = T_s$, will be reconstructed first as ($\breve{T}_{i,n-1} = \breve{T}_{i,n-1} + \breve{T}_{i,n} + \breve{T}_{i,n+1}$), then the mean of the reconstructed periods will be calculated and compared with the estimated mean using the estimator in (4.7). If the calculated mean is lower than the estimated one, the second iteration will take place where the second shortest periods, which are $\breve{T}_{i,n} = 2T_s$, will be reconstructed this time and in the same way as in the first iteration. Therefore, this process will be repeated until the calculated mean of the reconstructed periods reaches the estimated value of mean using the mean estimator proposed in (4.7). This proposed method relies on the fact that the estimator proposed in (4.7) can produce a highly accurate estimation of the original mean period, even in the presence of sensing errors, and thus can be exploited as an indicator of when periods are reconstructed correctly, without any prior knowledge of the PU activity pattern. The steps of this method can be further illustrated in Algorithm 1. Notice that the mean period is here used as a reference to determine when the periods are correctly reconstructed, however once the process is finished, other statistics (not only the mean) can also be estimated.

4.3.6 Performance Evaluation of Reconstruction Algorithms

To evaluate the performance of the proposed algorithm in Method 5 along with the previous methods, the system model is simulated using MATLAB, following a similar simulation procedure as in Section 3.8 with some slight changes. After a sequence of idle/busy periods T_0/T_1 is generated (with Generalised Pareto distribution) and observed under ISS as $\breve{T}_0/\breve{T}_1$ (based on predefined P_{fa} and P_{md}), the reconstruction

Algorithm 1: Proposed Method 5

Input: $(\breve{T}_i)$ The estimated periods under ISS
Output: $(\bar{T}_i)$ The reconstructed periods

Calculate the mean $(\breve{m}_i)$ of the periods under ISS
Estimate the mean $(\widetilde{m}_i)$ of the periods using the estimator in (4.7)
$k = 0$
$\bar{T}_i = \breve{T}_i$
while $\breve{m}_i < \widetilde{m}_i$ **do**
    $k = k + 1$
    for *each* $\breve{T}_{i,n} = kT_s$ **do**
        $\bar{T}_{i,n-1} = \breve{T}_{i,n-1} + \breve{T}_{i,n} + \breve{T}_{i,n+1}$
    end
    $\breve{m}_i = \mathbb{E}(\bar{T}_i)$ `// Calculate the mean of the reconstructed periods`
end
return $(\bar{T}_i)$

methods can then be applied to reconstruct the corrupted periods under ISS and referred to as $\bar{T}_0/\bar{T}_1$. The statistical parameters, e.g., Cumulative Distribution Function (CDF), can then be calculated for the reconstructed periods $\bar{T}_0/\bar{T}_1$ as well as for the unreconstructed periods $\breve{T}_0/\breve{T}_1$, and then compared with the CDF of the original periods T_0/T_1.

All discussed reconstruction methods in this work are implemented, and therefore the above simulation procedure is repeated five times, each time with a different reconstruction method. The accuracy of the reconstruction methods is evaluated by comparing the estimated CDF of periods (after and before reconstruction) with the CDF of the original periods. Since the original period durations (T_i) are continuous values, their CDF is continuous as well. In contrast, the periods $\breve{T}_i$ and $\bar{T}_i$ that are estimated under ISS before and after reconstruction, respectively, are discrete values, and their CDF therefore is discrete as well. Since it is impossible to compare between continuous and discrete distributions, the discrete distribution is interpolated in order to be comparable with the continuous one. This comparison can be performed using the KS distance (also explained in Section 3.8.3). For example, comparing the CDF of the original periods with the CDF of the ISS periods before reconstruction can be found as:

$$D_{KS} = \sup_{T_i} \mid F_{T_i}(T_i) - F_{\breve{T}_i}(T_i) \mid \tag{4.15}$$

where $F_{T_i}(T_i)$ represents the CDF of the original periods and $F_{\breve{T}_i}(T_i)$ is the interpo-

lated version of the discrete distribution $F_{\breve{T}_i}(\breve{T}_i)$, D_{KS} denotes the KS distance of the estimated distribution with respect to the original distribution.

Similarly, to compare the CDF of the original periods with the CDF of the periods after reconstruction, the following can be used:

$$D_{KS} = \sup_{T_i} \mid F_{T_i}(T_i) - F_{\bar{T}_i}(T_i) \mid \tag{4.16}$$

In this simulation the same settings are selected as in [75] and [77] in order to obtain a fair comparison. Therefore, the idle periods are drawn from a Generalised Pareto distribution using the following parameters: location (minimum period) $\mu_0 = 10$ t.u., scale $\lambda_0 = 30$ t.u., and shape $\alpha_0 = 0.25$. These parameters result in a mean period of $\mathbb{E}(T_0) = 50$ t.u. when using a duty cycle of $\Psi = 0.5$. However, for the proposed method these parameters are considered unknown to the DSA/CR system.

Proposed Method Operation

In order to understand how the periods are reconstructed in the time-domain using the proposed reconstruction algorithm of Method 5 and how the calculated mean converges to the true mean, a sequence of the estimated idle/busy periods under ISS selected from the simulation results is shown in Fig. 4.5 to illustrate the reconstruction process in each iteration and its corresponding calculated mean. First, an original sequence of the idle/busy periods is generated with a mean value equal to 50 t.u., then by employing a sensing period of $T_s = 1$ t.u. the estimated PSS sequence will be as shown in Fig. 4.5 (the first sequence). By applying sensing errors to this sequence with $P_{fa} = P_{md} = 0.1$ the ISS periods can be obtained as in the second sequence of the same figure. As it can be noticed, the calculated mean of the ISS sequence, about (5.1 t.u.), is much lower than its original value. The following sequences in Fig 4.5 represent the reconstruction process in each iteration. As it can be seen, the shortest periods, which are equal to $1T_s$, have been reconstructed in the first iteration and the calculated mean has become 26.2 t.u.. Similarly, the periods $2T_s, 3T_s$ and $4T_s$ have been reconstructed in the second, third and fourth iteration, respectively. In addition the calculated mean has converged gradually to the original value as $41.3, 44.1, 46.8$ t.u.. Since the original value of the mean should be considered unknown to the DSA/CR system, the achieved mean estimator in (4.7) is used here, which can provide a satisfactory accuracy for estimating the true mean. Therefore, in this case the estimated mean was $m = 44.4$ t.u., which was used as a threshold for the proposed algorithm to break the loop (i.e., stop the reconstruction process) whenever the calculated mean exceeds this value.

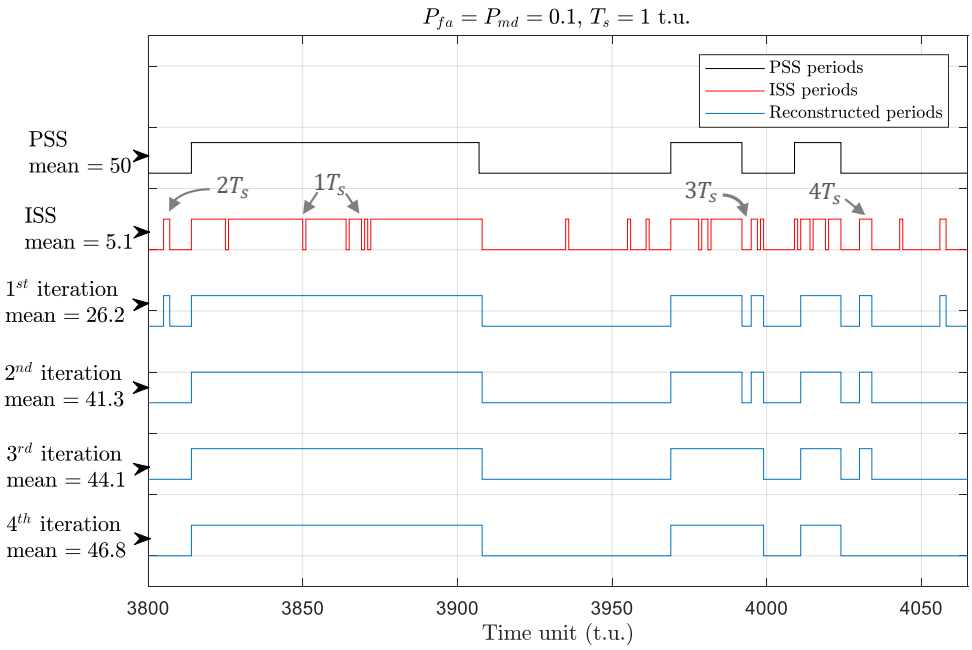


Figure 4.5: Reconstruction process of the idle/busy periods in time-domain using the proposed algorithm and comparing it with the periods under PSS and ISS.

In Fig. 4.6, on the other hand, the convergence performance of the proposed algorithm is demonstrated for different scenarios of probability of sensing error, in terms of the relative error with respect to the number of iterations required to achieve the accuracy given by the mean estimator in (4.7). As it can be seen, the accuracy of this method converges to zero (or close to zero) relative error as the number of iterations increases. In addition, more iterations are required to reconstruct the idle/busy periods and achieve the minimum estimation error when higher probability of sensing error is considered (e.g., 4 iterations for $P_{fa} = P_{md} = 0.1$, 2 iterations for $P_{fa} = P_{md} = 0.05$ and 1 iteration for $P_{fa} = P_{md} = 0.01$).

Performance Evaluation

Fig. 4.7 illustrates the accuracy of estimating the CDF under ISS using different methods, which is presented in terms of the KS distance versus the sensing period T_s. Regardless of which reconstruction method is used, the improvement in the CDF estimation can be clearly noticed as compared with the case when no reconstruction method is used. Knowing that as the KS distance approaches zero as the estimated CDF under ISS approaches the original CDF value.

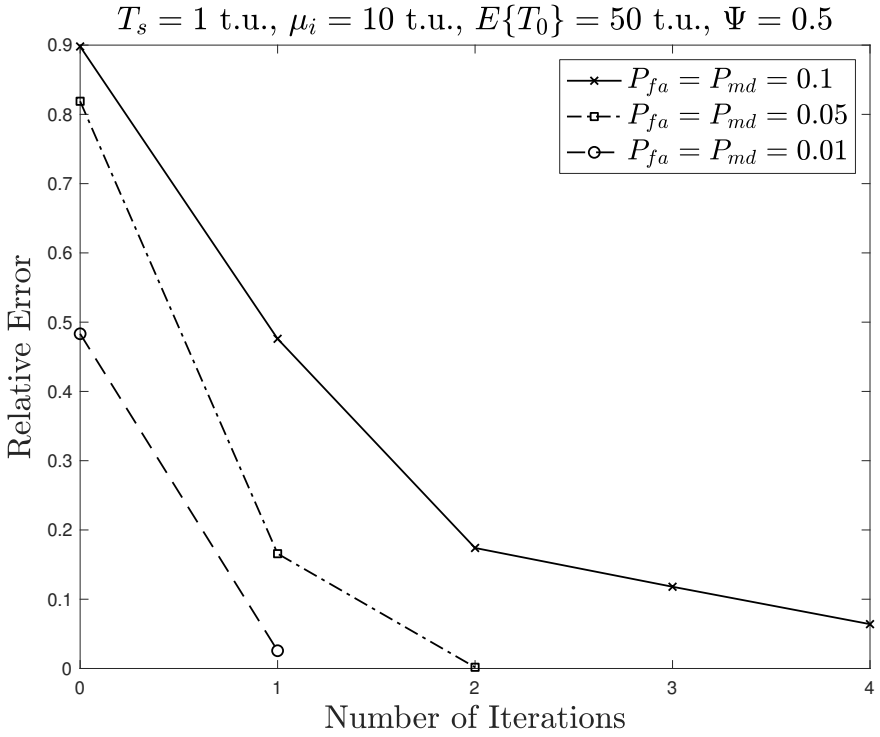


Figure 4.6: Convergence performance of the proposed algorithm in terms of the accuracy of estimating the mean period and the required number of iterations.

As it can be noticed, the methods proposed by [75] (Method 1,2 and 3) can improve, to some extent, the estimation of the CDF under ISS. However, Method 1 shows better performance than Method 2 and 3 as explained in Section 4.3.3. On the other hand, the reconstruction algorithm developed by [77] in Method 4 has further improved the accuracy of estimating the distribution under ISS, and it could outperform Method 1, 2 and 3 since its KS distance approaches closer to zero. The proposed algorithm, which is referred to as Method 5, has also been examined and compared with the previous methods. As shown in Fig. 4.7, Method 5 approaches the performance achieved by Method 4, while it outperforms Method 1, 2 and 3 as well. It can also be noticed that Method 5 outperforms Method 4 when the value of the sensing period T_s is high (note that T_s should not exceed μ_i), while it performs slightly worse when T_s is low. This is due to the fact that the accuracy of the mean estimator in (4.7) degrades as the employed sensing period T_s decreases, which in turn will degrade the performance of this method at the very low T_s. Meanwhile, the accuracy of the mean estimator in (4.7) increases as T_s increases (this is because there will be lower number of sensing

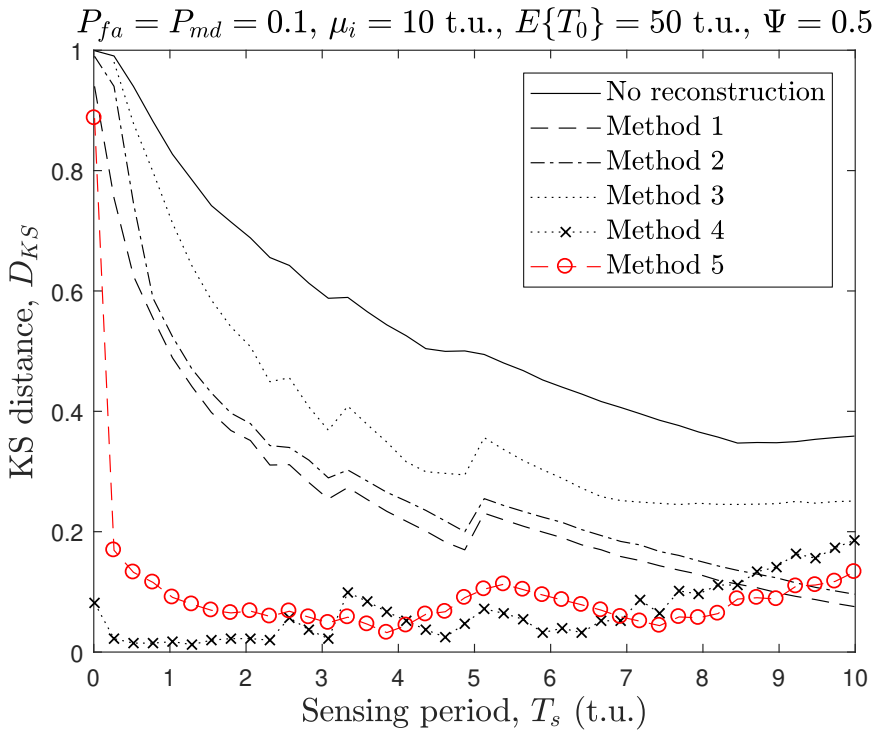


Figure 4.7: Performance of the reconstruction methods.

errors for higher T_s), thus Method 5 outperforms Method 4 for higher sensing period T_s. Overall, the proposed Method 5 can approximately achieve the same accuracy of the latest reconstruction method in the literature (i.e., Method 4) without requiring any knowledge about the activity patterns of the primary channel, opposite to all the previous methods (1-4) which always assume a perfect knowledge of the minimum period μ_i of the primary channel.

4.4 Estimation Methods Based on Deep Learning

The advancement of Deep Learning (DL) in computer vision, speech recognition and natural language processing domains has inspired a large community of experts in the communications field to exploit the potential of this technology for solving a wide range of problems in communication systems. Such problems are either difficult to represent with tractable mathematical models or impractical to be solved by following the classical methods and algorithms. In this context, there has been an increasing interest in exploiting DL in wireless communications, in particular, Spectrum Sharing

(SS) systems. This is due to the demonstrated improvements that DL has brought to several applications of SS such as spectrum management, spectrum sensing, spectrum prediction, network security and so on. These applications are crucial for the ongoing deployment of 5G technology, including but not limited to, 5G New Radio Unlicensed (NR-U) [99–101], unlicensed LTE [102, 103] and License Assisted Access (LAA) [104–106].

In recent works, the statistical information of the channel traffic has increasingly been exploited as input features to the neural network of DL models. These statistical features can make significant improvement in the performance of DL for solving particular problems in SS systems. For instance, in [107,108] traffic statistics (mean, variance and kurtosis) have been exploited as features for a neural network to recognise user-level applications such as YouTube$^{\text{TM}}$ and WhatsApp$^{\text{TM}}$. On the other hand, in [41] the accuracy of spectrum sensing in cognitive radio has remarkably been enhanced by exploiting channel activity statistics as input features to a DL model used to sense the spectrum. Moreover, [109, 110] has employed the historical samples of the channel traffic statistics to train a DL to predict the future channel occupancy ratio. Obtaining accurate statistical information of the channel traffic can also find a wide range of applications in enhancing the performance of cooperative spectrum sensing systems operating under realistic environmental conditions [111–113] as well as in cognitive radio for Vehicular Ad Hoc Network (VANET) [114].

From the above discussion, it is evident that traffic statistics play an effective role in the performance of various applications in SS systems which apply DL technique. The majority of these works, however, assume perfect estimation of these statistics, such that they can smoothly be exploited in DL models. In practice, however, these statistics can be corrupted due to signal detection errors as discussed in the previous sections. Inaccurate traffic statistics, consequently, can worsen the training process of a DL model and thus provide inaccurate results. Therefore, in order to exploit the channel traffic activity statistics in SS systems it is essential to estimate these statistics accurately especially under a realistic (i.e., ISS) scenario.

In the previous sections, the estimation of the channel activity statistics under ISS has been corrected through two approaches: i) closed-form expressions, where mathematical expressions are derived for the original statistics as a function of their corresponding statistics observed under ISS, probability of sensing error and sensing period. ii) reconstruction algorithms, where the observed idle/busy periods under ISS of the channel traffic are reconstructed to provide accurate statistics. Reconstruction methods can provide accuracy improvements but are typically based on heuristic al-

gorithms and therefore sub-optimal. Although closed-form expressions would be the most attractive solution to correct the estimation of traffic statistics under ISS, it is challenging sometimes to find these expressions for higher statistical moments such as variance, skewness and kurtosis under ISS. In addition, even though these expressions can provide accurate estimations, they may still show some considerable estimation errors when a short sensing period T_s is employed (e.g., see Fig. 4.2 (left) for the estimation of the mean period over short T_s values). In some cases, closed-form expressions are known or can be obtained but they are unable to lead to accurate estimations of the true traffic statistics under ISS as discussed in Section 4.2.1 for the estimation of the minimum period. Therefore, considering the aforementioned limitations of the previous approaches, this section proposes a novel approach, Traffic Learning (TL), as a DL approach to learn from the observations of the channel traffic activity under realistic ISS scenario in order to provide accurate statistical information of the channel traffic activity in SS systems [115].

4.4.1 Deep Learning Approach

The novelty of this section lies on the fact that it presents the first trial and investigation of exploiting DL in the estimation of the channel activity statistics under ISS. The consideration of deep learning here instead of the traditional machine learning methods is due to the difficulty sometimes to extract certain features by the traditional methods from the available dataset in order to provide accurate estimation results for the channel activity statistics. Learning from such unstructured dataset is possible via DL, however, this does not exclude the applicability of the traditional machine learning methods for the estimation of some of these statistical parameters when certain features can be obtained (e.g., mean or variance). In this context, this section proposes a novel approach for the estimation of the channel traffic statistics under ISS based on DL technique. The DL model in this section aims to provide an accurate estimation for the original statistical parameters of the channel traffic based on their corresponding (inaccurate) statistics observed under ISS. It is widely known that DL can solve various problems through formulating them as either classification or regression problems. The estimation of the statistical parameters, for example, mean, variance and minimum period is considered as a regression problem, while the estimation of the channel traffic distribution is solved by first classifying the type of the distribution, then finding its parameters. The estimation of these statistics can be solved using Multilayer Perceptron (MLP) fully-connected feedforward Neural Network (NN) [116].

An MLP with L (dense) layers maps the input layer x to the output layer y through one or more hidden layers in between. This mapping function can be written as $y = f(x;\theta)$, where θ denotes the NN parameters given by the weights W and biases b. Each layer of the NN consists of one or more neurons n, hence the output of the ℓ-th layer can be written as [117]:

$$f_\ell(x_{\ell-1};\theta_\ell) = \sigma_\ell\left(W_\ell x_{\ell-1} + b_\ell\right), \qquad \ell = 1, \cdots, L \tag{4.17}$$

where $W_\ell \in \mathbb{R}^{n_\ell \times n_{\ell-1}}$ is the weight matrix, $b_\ell \in \mathbb{R}^{n_\ell}$ is the bias vector (note that n_ℓ denotes the number of neurons at the ℓ-th layer), and $\sigma_\ell(\cdot)$ represents the non-linear activation function which can be given by, e.g., ReLU [118], sigmoid [119], softmax [120], etc. The output of the ℓ-th layer $f_\ell(x_{\ell-1};\theta_\ell)$ is based on the input $x_{\ell-1}$ from the previous layer and the parameter $\theta_\ell = \{W_\ell, b_\ell\}$ at the ℓ-th layer. In general, a Neural Network (NN) is trained based on a labelled training dataset, which is an input-output (x, y) vector pairs of data. In our scenario, the input vector is the observations of a statistical parameter $\breve{s}$ under ISS (e.g., mean, variance, etc.) and the output vector is the corresponding original statistical parameter s. Therefore, this input-output ($\breve{s}$, s) dataset is used to train a NN to find θ^* that minimises the loss function $\mathcal{L}(\breve{s}, s)$:

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, \mathcal{L}(\breve{s}, s) \tag{4.18}$$

This loss function can be given by a Mean Squared Error (MSE) as $\|s - f(\breve{s};\theta)\|^2$ to find θ that minimises the error. By selecting the appropriate hyper-parameters of the NN (e.g., number of layers, neurons, loss function) along with the useful input features, a DL model can be achieved to provide an accurate estimation for the statistical parameters of the channel traffic under ISS as it will be discussed next.

4.4.2 Mean, Variance and Minimum Estimation Based on DL

Let us first consider the estimation of the original mean m_i of the idle/busy periods (where i can be 0 referring to idle periods, or 1 referring to busy periods). A DL model using MLP NN is built to find the accurate estimation of the mean of the channel traffic from the corresponding mean observed under ISS. Therefore, the inaccurate means $\breve{m}_0$ and $\breve{m}_1$ of the idle/busy periods observed under ISS are used as inputs to the DL model to provide the accurate estimation of the mean period $\widetilde{m}_i$ (where $\widetilde{m}_i \approx m_i$). Since under ISS the presence of sensing errors corrupts the observation of the idle/busy periods as discussed in Chapter 3, the mean of these periods would be significantly

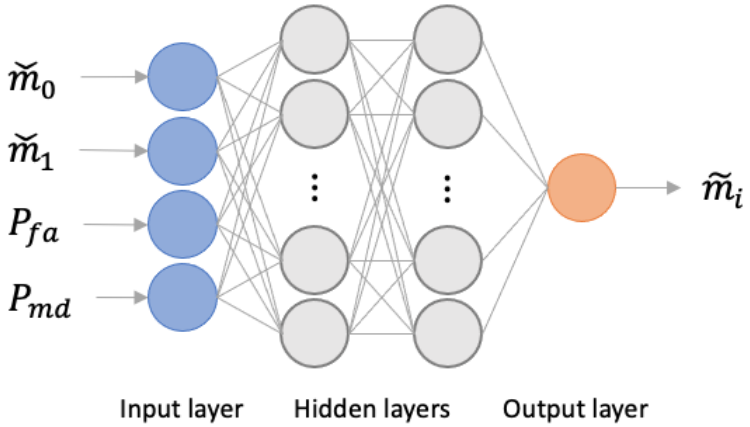


Figure 4.8: Deep Learning model for mean of periods estimation under ISS.

inaccurate depending on the probabilities of sensing error (i.e., P_{fa} and P_{md}). These probabilities can be pre-defined based on the employed sensing algorithm at the end terminal. Therefore, P_{fa} and P_{md} can also be exploited as input features to the DL model along with $\breve{m}_0$ and $\breve{m}_1$ observed under ISS. P_{fa} and P_{md} can assist a NN to learn from how these features affect the observation of $\breve{m}_0$ and $\breve{m}_1$ under ISS, which in turn will help predicting the actual mean value at the output as shown in Fig. 4.8. Note that when $P_{fa} = P_{md} = 0$, the observed mean will be equal to the original one [49]. A similar concept can also be applied to find a DL model for estimating higher statistical moments under ISS. Therefore, the second moment (variance v_i) of the idle/busy periods is here considered, which can similarly be found as shown in Fig. 4.9. As it can be noticed, the observed statistics of both idle and busy periods are always considered as input features because they are both affected by false alarms and missed detections, this effect can also be noticed in the closed-form expressions obtained in Section 4.2. Therefore, considering only the observed statistics for the same type of periods being estimated (idle or busy) would not provide complete input information.

On the other hand, the accurate estimation of the minimum period μ_i of the channel traffic under ISS is more challenging to find compared to the previous statistical parameters. This is because for any non-zero probability of sensing error ($P_{fa} > 0$ and $P_{md} > 0$) the observed minimum period $\breve{\mu}_i$ under ISS is always equal to the duration of a single sensing error, which is the same as the duration of the sensing period T_s (i.e., $\breve{\mu}_i \neq \mu_i$ and $\breve{\mu}_i = T_s$, $\forall P_{fa}, P_{md} > 0$) as it was explained in Section 4.2.1. Therefore, a NN can not learn anything from the observed minimum idle/busy periods $\breve{\mu}_0/\breve{\mu}_1$ under ISS (unlike the previous statistical parameters) since they are always equal to the

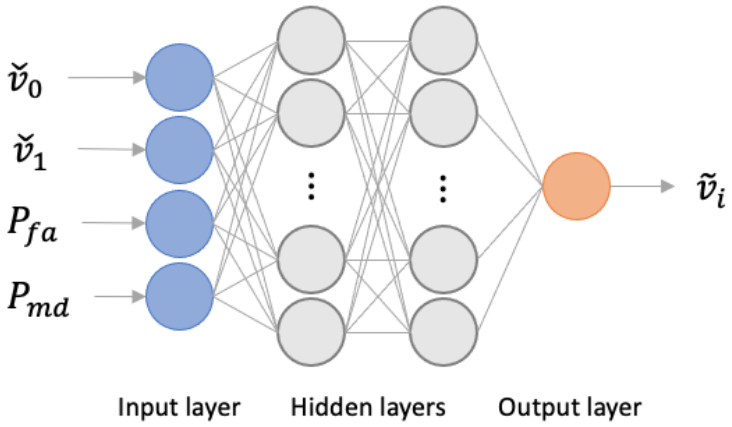


Figure 4.9: Deep Learning model for variance of periods estimation under ISS.

sensing period T_s, no matter how high or low the probability of sensing error is. In order to utilise a feature that can help a NN to predict the actual minimum period μ_i from the observations of the ISS, it is useful to look at the distribution of the observed periods under ISS. The observed periods under ISS have a discrete distribution with a bin size of T_s and starting at T_s as well. This distribution is distorted by the presence of sensing errors, however, it forms a distinguished pattern corresponding to a particular combination of probabilities of sensing error (P_{fa} and P_{md}). A NN can be trained to learn from these patterns of the observed distributions under ISS in order to locate the actual minimum period. As a result, it is found that by using the first h-th histogram bins of the observed periods under ISS along with the probabilities of sensing error (P_{fa} and P_{md}) it is possible to train a NN to provide an accurate estimation for the actual minimum period under ISS. The MLP NN in Fig. 4.10 shows an example of using 100 histogram bins of the observed periods under ISS as input features along with P_{fa} and P_{md}, where h_1 refers to the number of the observed periods under ISS within the first bin, while h_2 refers to the number of the observed periods under ISS within the second bin and so on. The number of bins was selected here after conducting several evaluations on the estimation accuracy of the minimum period under ISS while considering several scenarios of probabilities of sensing errors (P_{fa} and P_{md}), for which 100 bins were found to be sufficient to provide accurate results under any scenario of sensing errors. The output of this NN provides the accurate estimation $\widetilde{\mu}_i$ for the actual minimum period μ_i (where $\widetilde{\mu}_i \approx \mu_i$).

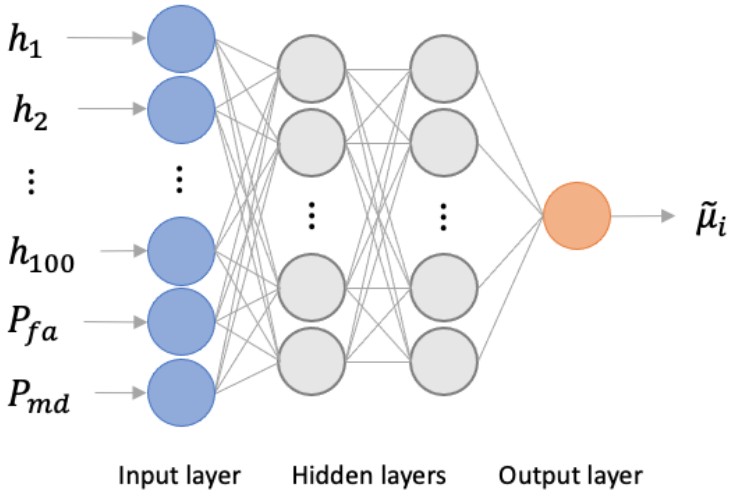


Figure 4.10: Deep Learning model for minimum of periods estimation under ISS.

Raw Dataset Construction and Preprocessing

In this approach, data are obtained and prepared in two stages, in the first stage raw datasets are generated using MATLAB, then in the second stage the generated datasets are preprocessed using Python to train, validate and test the proposed DL model. Dataset generation using MATLAB can be achieved as follows:

1. First, a channel traffic is modelled by generating a large sequence of idle/busy periods (T_0/T_1) in a frequency channel drawn from a particular distribution such as Generalised Pareto distribution.

2. Then spectrum sensing can be applied with periodic sensing period T_s, where T_s should be smaller than the minimum period of the channel idle/busy periods (i.e., $T_s < \mu_i$). In this work the use of short $T_s = 1$ t.u. is considered when the minimum period $\mu_i = 10$ t.u. (i.e., 10% of the minimum period). This is to show how the estimation methods perform under the worse scenario of using such short sensing period since higher sensing periods (e.g., 90%) can provide more accurate estimations for traffic statistics under ISS [75], which has been also noticed from the results of closed-form expression approach in Section 4.2.

3. Spectrum sensing is configured based on the selected probabilities of sensing error (i.e., P_{fa} and P_{md}), based on which a sensing threshold is adjusted to decide whether the channel is idle $\mathcal{H}_0$ or busy $\mathcal{H}_1$. Sensing decisions are then used to

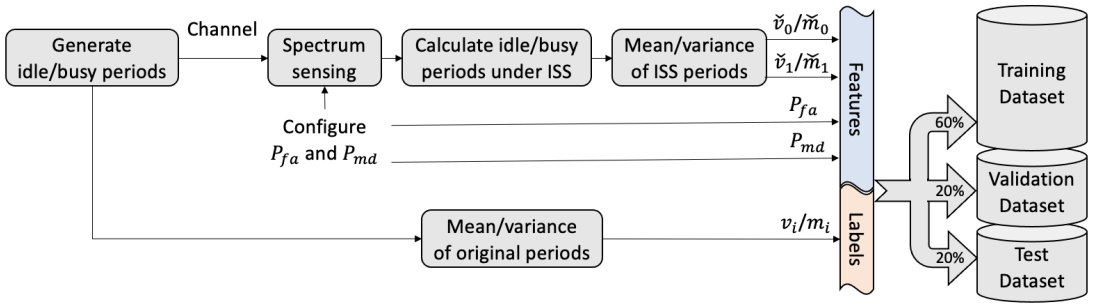


Figure 4.11: Datasets construction of DL for channel traffic mean and variance estimation.

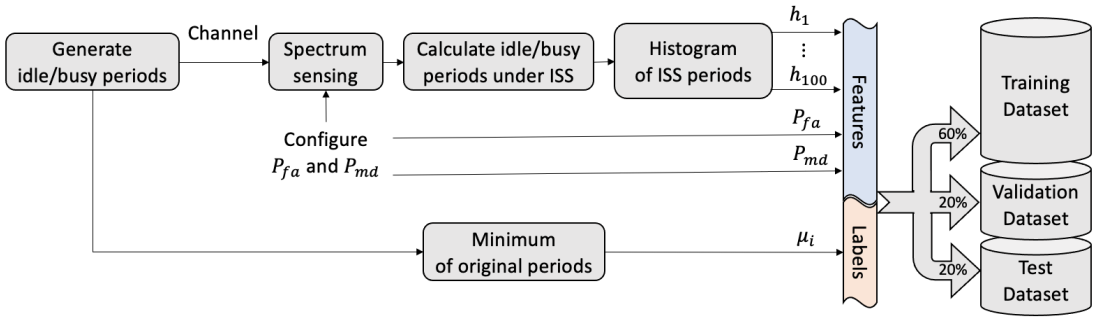


Figure 4.12: Datasets construction of DL for channel traffic minimum estimation.

calculate the duration of the idle/busy periods ($\breve{T}_0/\breve{T}_1$) observed under ISS.

4. The statistical parameters such as mean $\breve{m}_0/\breve{m}_1$, variance $\breve{v}_0/\breve{v}_1$ or histogram $\{h_1, \ldots, h_{100}\}$ can then be calculated from ($\breve{T}_0/\breve{T}_1$) periods observed under ISS in step 3. These statistics are saved into a `.mat` file along with the configured P_{fa} and P_{md} to represent the input vector (features). On the other hand, the corresponding original statistics m_0/m_1 for mean, v_0/v_1 for variance or μ_0/μ_1 for minimum of the idle/busy periods (T_0/T_1) generated in step 1 are also saved into the same `.mat` file to represent the output vector (labels).

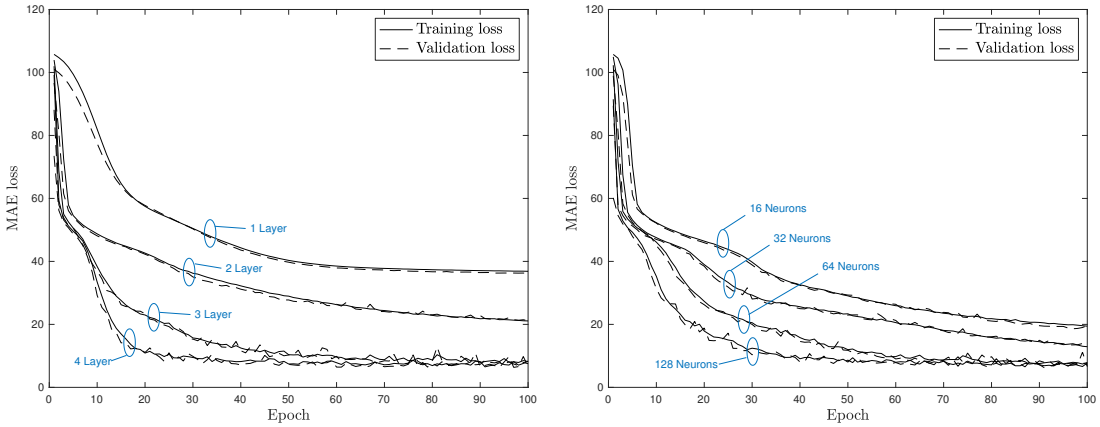
The obtained features and labels in `.mat` file are then used to construct the required dataset for DL, 60% of which is for training, 20% is for validation and the remaining 20% is for testing as shown in Figs. 4.11 and 4.12. It is a common practice to distribute the available dataset among training, validation and testing as above. However, this percentage can be adjusted based on the available size of the dataset while ensuring the selected portions of data from this dataset would be sufficient to train the DL models and then validate and test these models. These raw datasets require some pre-

processing before using them for DL training or testing. Python is used here, which offers numerous tools and advanced DL libraries (e.g., TensorFlow [121], Keras [122] and PyTorch [123]) that facilitate not only the preprocessing of the datasets, but also building, training and testing of the DL model. Therefore, the obtained dataset in `.mat` file is imported to Python for preprocessing, where the features and labels are extracted and stored into separate arrays. Since these data can hold any real values, it is a common practice to scale and normalise these values before learning from them. The `preprocessing.Normalization()` function from Keras library is used, which normalises its inputs into a distribution centred around zero with unit standard deviation. This is accomplished by applying the following normalisation relationship $(input - mean)/\sqrt{variance}$ to the input dataset.

Training, Validating and Testing the DL model

After preprocessing the datasets, they are ready to train, validate and test a DL model. An MLP NN has been examined using several hyper-parameter settings to build the required DL model for channel traffic statistics estimation under ISS. As shown in Fig. 4.13, different number of hidden layers {1, 2, 3 and 4} and neurons {16, 32, 64 and 128} are used to examine the accuracy of training based on Mean Absolute Error (MAE) loss function. It is found that a NN with 3 hidden layers can reach the same accuracy as a higher number of layers after 100 epochs of training. In the same way, 64 neurons per hidden layer can provide the same accuracy as a higher number of neurons after 100 epochs of training. As a result, the MLP NN shown in Table 4.1 is considered for this approach to provide the accurate estimation of the channel traffic statistics under ISS. The output of this model would be either the accurate estimation of the mean $\widetilde{m}_i$, variance $\widetilde{v}_i$ or minimum period $\widetilde{\mu}_i$ when the input is the corresponding ISS mean $\breve{m}_i$, variance $\breve{v}_i$, or histogram bins $\{h_1, \ldots, h_{100}\}$, respectively.

This MLP NN model is trained based on the 60% of the preprocessed features and labels, while 20% of which is used to validate the training process. This validation is important to make sure that the NN can generalise to new data and avoid the overfitting problem during training process. ReLU activation function is selected at each hidden layer, and Adam optimiser is used with learning rate 0.001. After training and validating the DL model, it can now be tested based on the remaining 20% of the dataset to evaluate its estimation performance. Although the testing dataset has both features and labels, only features are fed to the NN to predict the accurate channel traffic statistics, while labels are used to quantify the accuracy of the estimation pro-



(a) Using different number of hidden layers (64 neurons each).

(b) Using different number of neurons per layer (for 3 hidden layers).

Figure 4.13: Training and vlaidation accuracy of the MLP NN for the estimation of the mean, variance and minimum period.

Table 4.1: MLP NN model used for mean, variance and minimum period estimation under ISS.

Hyper-parameter	**Settings**
Number of hidden layers	3
Neurons per hidden layer	64
Activation function	ReLU
Optimiser	Adam
Learning rate	0.001
Loss function	MAE
Metric	Accuracy
Batch size	10
Epochs	100

vided by the NN, which will be shown in the results section 4.5 in comparison with the previous approaches.

4.4.3 Distribution Classification and Estimation Based on DL

Having an accurate estimation for the distribution of the idle/busy period durations completes the whole picture of learning about the channel traffic activity (i.e., Traffic Learning). In the literature, different distribution models have been considered for the

channel traffic. Exponential (E) distribution, for example, is one of the widely assumed models for channel traffic as in [73,74,124], which can simplify the mathematical analysis of the studies. However, field measurements in [78] have shown that the Generalised Pareto (GP) distribution is more realistic for channel traffic representation. In this section, however, the estimation of the channel traffic distribution under ISS is investigated using a DL approach without making any prior assumption about the original distribution type of the channel traffic, which then will be compared with the previous studied approaches. First, a DL model is used to classify the distribution type of the channel traffic based on the ISS observations. After classifying the distribution type, Method of Moments (MoM) inference technique [125] can then be used to estimate the distribution parameters (location μ , scale λ and shape α, if they all exist) from the sample moments obtained previously (i.e., mean, variance and minimum).

The classification problem can be solved using an MLP NN that selects a distribution class at the output based on the observations of the ISS for the channel traffic. Table 4.2 is considered for the list of the possible traffic distribution types that provides accurate representations for the empirical data [78], from which a NN can select the best match type for the channel traffic distribution. This list includes Exponential (E), Generalised Pareto (GP), Gamma (G) and Weibull (W) distributions (note that other distribution types can also be added to the list). Therefore, there is no particular type assumption for the channel traffic distribution (as often is assumed in the literature) since the list here can easily be extended to other distribution models. The input of the NN, as shown in Fig. 4.14, uses the first h-th histogram bins of the observed periods under ISS along with the probabilities of sensing error (P_{fa} and P_{md}) to predict the best classification for their distribution (the highest probability at the output). Note that the input of this NN is similar to the input of the NN used to find the minimum parameter μ in the previous section, however, the input here is used to solve a classification problem rather than a regression problem and as a result the NN has multiple outputs.

After classifying the distribution type of the channel traffic, MoM inference technique [125] is considered to estimate the distribution parameters (location μ_i , scale λ_i and shape α_i, if they all exist) from the sample moments obtained previously (i.e., mean, variance and minimum). The location parameter μ_i is the same as the minimum period estimated previously as $\widetilde{\mu}_i$ using DL approach, while the scale λ_i and shape α_i parameters can be found from the mean and variance of the selected distribution model. Since the moments (mean and variance) can also be estimated accurately using the DL approach as discussed before, the scale λ_i and shape α_i parameters of the selected dis-

Table 4.2: Considered probability distribution models for idle/busy period durations. Distribution names: E (Exponential), GP (Generalised Pareto), G (Gamma), and W (Weibull). Distribution parameters: μ_i (location), λ_i (scale), and α_i (shape). T_i represents the period length. $\mathbb{E}\{\cdot\}$ and $\mathbb{V}\{\cdot\}$ represent the mean and the variance of the distribution, respectively. $\gamma(\cdot,\cdot)$ is the lower incomplete Gamma function [126, 6.5.2] and $\Gamma(\cdot)$ is the (complete) Gamma function [126, 6.1.1]. (reproduced from [78]).

Distribution function	**Parameters**	**Moments**
$F_E(T_i;\mu_i,\lambda_i) = 1 - e^{-\lambda_i(T_i-\mu_i)}$	$T_i \geq \mu_i > 0$ $\lambda_i > 0$	$\mathbb{E}\{T_i\} = \mu_i + \frac{1}{\lambda_i}$ $\mathbb{V}\{T_i\} = \frac{1}{\lambda_i^2}$
$F_{GP}(T_i;\mu_i,\lambda_i,\alpha_i) = 1 - \left[1 + \frac{\alpha_i(T_i-\mu_i)}{\lambda_i}\right]^{-1/\alpha_i}$	$T_i \geq \mu_i$ $(\alpha_i \geq 0)$ $T_i \in [\mu_i, \mu_i - \frac{\lambda_i}{\alpha_i}]$ $(\alpha_i < 0)$ $\mu_i, \lambda_i > 0$, $\alpha_i < 1/2$	$\mathbb{E}\{T_i\} = \mu_i + \frac{\lambda_i}{1-\alpha_i}$ $\mathbb{V}\{T_i\} = \frac{\lambda_i^2}{(1-\alpha_i)^2(1-2\alpha_i)}$
$F_G(T_i;\mu_i,\lambda_i,\alpha_i) = \frac{\gamma\left(\alpha_i, \frac{T_i-\mu_i}{\lambda_i}\right)}{\Gamma(\alpha_i)}$	$T_i \geq \mu_i > 0$ $\lambda_i > 0$ $\alpha_i > 0$	$\mathbb{E}\{T_i\} = \mu_i + \lambda_i\alpha_i$ $\mathbb{V}\{T_i\} = \lambda_i^2\alpha_i$
$F_W(T_i;\mu_i,\lambda_i,\alpha_i) = 1 - exp\left[-\left(\frac{T_i-\mu_i}{\lambda_i}\right)^{\alpha_i}\right]$	$T_i \geq \mu_i > 0$ $\lambda_i > 0$ $\alpha_i > 0$	$\mathbb{E}\{T_i\} = \mu_i + \lambda_i\Gamma\left(1 + \frac{1}{\alpha_i}\right)$ $\mathbb{V}\{T_i\} = \lambda_i^2\left[\Gamma\left(1 + \frac{2}{\alpha_i}\right) - \Gamma^2\left(1 + \frac{1}{\alpha_i}\right)\right]$

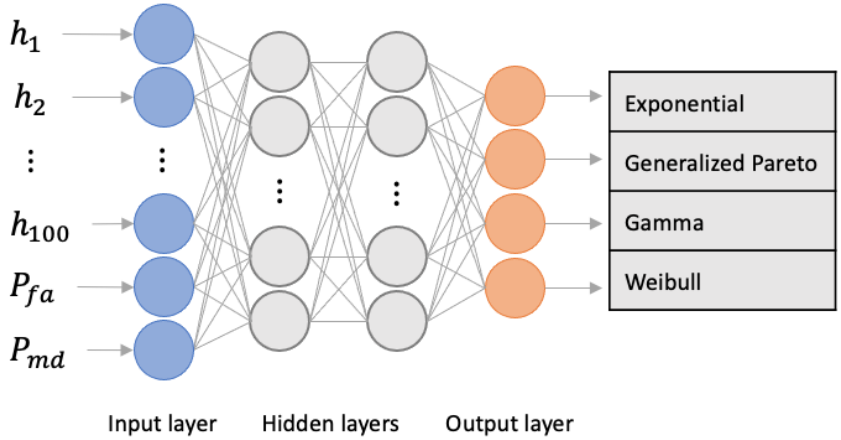


Figure 4.14: Deep Learning model for distribution type classification under ISS.

tribution can therefore be solved using MoM technique. For example, if the DL model shown in Fig. 4.14 classifies (with highest probability) the channel traffic observations as GP-distributed, their μ_i, λ_i and α_i parameters can then be found as [125, ch. 20]:

$$\mu_i \approx \widetilde{\mu}_i \tag{4.19a}$$

$$\lambda_i \approx \widetilde{\lambda}_i = \frac{1}{2}\left(1 + \frac{(\widetilde{m}_i - \widetilde{\mu}_i)^2}{\widetilde{v}_i}\right)(\widetilde{m}_i - \widetilde{\mu}_i) \tag{4.19b}$$

$$\alpha_i \approx \widetilde{\alpha}_i = \frac{1}{2}\left(1 - \frac{(\widetilde{m}_i - \widetilde{\mu}_i)^2}{\widetilde{v}_i}\right) \tag{4.19c}$$

where $\widetilde{\mu}_i$, $\widetilde{m}_i$ and $\widetilde{v}_i$ are the estimated minimum, mean and variance of the channel traffic using DL approach, respectively. Once the distribution parameters are found, the Cumulative Distribution Function (CDF) of the GP distribution F_{GP} can then be obtained from:

$$F_{GP} \approx \widetilde{F}_{GP}(T_i; \widetilde{\mu}_i, \widetilde{\lambda}_i, \widetilde{\alpha}_i) = 1 - \left[1 + \frac{\widetilde{\alpha}_i(T_i - \widetilde{\mu}_i)}{\widetilde{\lambda}_i}\right]^{-1/\widetilde{\alpha}_i} \tag{4.20}$$

In the same way the expressions for other channel traffic distributions can be found.

Raw Dataset Construction and Preprocessing

As discussed before, distribution estimation is achieved by first classifying the distribution type using DL model, then estimating the distribution function using MoM technique. To solve the classification problem, datasets are required to be obtained.

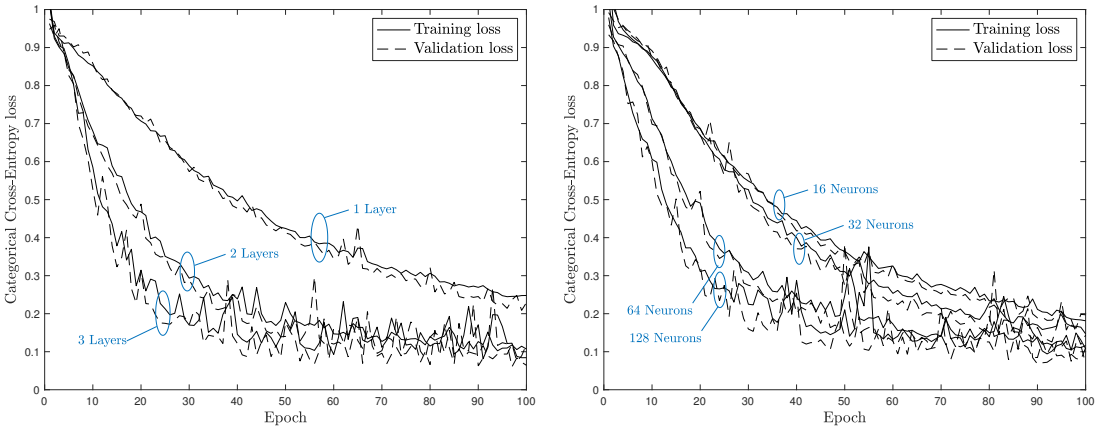
These datasets are constructed in the same way as step 1 to 4 in Section 4.4.2 with some slight differences. In step 1, channel traffic is modelled 4 times using (E, GP, G and W) distributions. Then spectrum sensing and probability of sensing error (P_{fa} and P_{md}) are configured in the same way as in step 2 and 3. In step 4, channel traffic statistics (histogram bins $\{h_1, \ldots, h_{100}\}$) are computed from the ISS observations. These observations along with the configured P_{fa} and P_{md} represent the input vector (features) of the DL model, whereas the output vector (labels) is given by the classes of the original distribution used to model the channel traffic in step 1. Since four distribution classes (E, GP, G and W) is used, they can be encoded as a *one-hot* vector $1_s \in \mathbb{R}^4$ (i.e., 4-dimensional vector, the s-th element of which is equal to one and zero otherwise [117]). These features and labels can then be saved into `.mat` file to be used later for training and testing.

However, preprocessing is required to be performed first on the produced dataset. Therefore, the obtained dataset in `.mat` file is imported to Python for preprocessing. Similar to section 4.4.2, `preprocessing.Normalization()` function from Keras library is used to normalise these datasets in order to be used for training and testing.

Training, Validating and Testing the DL model

After preprocessing the dataset, it can now be used to train, validate and test a DL model. An MLP NN with several settings has been examined to build the required DL model for classifying channel traffic distribution under ISS. As shown in Fig. 4.15, different number of hidden layers {1, 2, and 3} and neurons {16, 32, 64 and 128} are used to examine the accuracy of training based on Categorical Cross-Entropy loss function (which is a loss function used in multi-class classification tasks, where a decision can only belong to one out of many possible categories). It is found that a NN with 2 hidden layers can reach the same accuracy as a higher number of layers when 100 Epochs is used. In the same way, 64 neurons per hidden layer can provide the same accuracy as a higher number of neurons when 100 Epochs are used. As a result, an MLP NN shown in Table 4.3 is considered to provide accurate classification for the type of the channel traffic distribution under ISS. The output layer of this model has 4 neurons referring to the corresponding classes (E, GP, G and W). Therefore, by using Softmax activation function at this layer, the output of these 4 neurons will represent a probability of the corresponding distribution class. Hence, the output with the highest probability will indicate the best distribution class match for the observed channel traffic under ISS.

After preprocessing the features and labels in the `.mat` file, 60% of these data is used



(a) Using different number of hidden layers (64 neurons each).

(b) Using different number of neurons per layer (for 2 hidden layers).

Figure 4.15: Training and validation accuracy of the MLP NN for distribution classification.

Table 4.3: MLP NN model used for distribution classification under ISS

Hyper-parameter	**Settings**
Number of hidden layers	2
Neurons per hidden layer	64
Activation function	ReLU, Softmax (output)
Optimiser	Adam
Learning rate	0.001
Loss function	Categorical Cross-Entropy
Metric	Accuracy
Batch size	10
Epochs	100

to train this MLP NN model, while 20% is used to validate the training process. After training and validating the DL model, it can then be tested based on the remaining 20% of the dataset to evaluate its classification performance. Although testing dataset has both features and labels, only features are fed to the NN to classify channel traffic distribution, while labels are used to quantify the accuracy of the classification provided by the NN, which will be shown next in the results section in comparison with the previous approaches.

4.5 Comparison Results of Estimation Methods

In this section, the performance of the proposed deep learning approach in comparison with the previously investigated approaches in this chapter, which include the closed-form expression approach and the reconstruction algorithm approach, will be evaluated and compared altogether for the estimation of several statistical parameters of the channel activity under different scenarios of the ISS.

4.5.1 Mean, Variance and Minimum Period Estimation under ISS

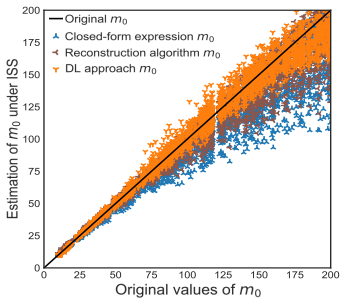
First, to evaluate the performance of the DL model proposed in Section 4.4.2 for the estimation of the mean, variance and minimum period of the channel traffic under ISS, a large dataset is produced to train the DL model such that it can generalise a problem, i.e., provide accurate estimation for the channel traffic statistics even when new data are observed under ISS. This can be achieved by repeating steps 1 to 4 in Section 4.4.2 several times to remodel the original channel traffic to cover a wide variety of traffic statistics, and for each traffic model spectrum sensing is applied and configured in step 3 using different combinations of P_{fa} and P_{md} ranging from low (0.01) to high (0.1) probability of sensing error. In the estimation of mean, for example, channel traffic in step 1 can be modelled repeatedly to have random mean values as $m_i \sim \mathcal{U}(10, 200)$ t.u., where $\mathcal{U}(a, b)$ denotes a uniform distribution between a and b. For each traffic mean, spectrum sensing is applied using several combinations of $P_{fa} \sim \mathcal{U}(0.01, 0.1)$ and $P_{md} \sim \mathcal{U}(0.01, 0.1)$ to observe the original mean under different scenarios of ISS. Similar procedures can also be followed to obtain the datasets for variance and minimum period statistics.

Then 60% and 20% of such datasets are used to train and validate the DL model, respectively, as discussed in Section 4.4.2, while the remaining 20% of the dataset is used to test the accuracy of the DL model. Fig. 4.16 shows the accuracy of estimating the mean of the channel traffic under ISS using different approaches of closed-form expression, reconstruction algorithm and DL. Each point in the figure represents the corrected estimation of the traffic mean observed under ISS for a particular P_{fa} and $P_{md} \sim \mathcal{U}(0.01, 0.1)$. As it can be noticed, DL approach outperforms the other approaches for providing accurate estimation, in which all the points are distributed closely around the straight line that corresponds to the original mean value. These results are obtained while using a short sensing period (in this case $T_s = 1$ t.u.) in order to consider the worst case scenario for the estimation of the channel activity statistics under ISS (as it has been also discussed and shown in the results of the closed-form

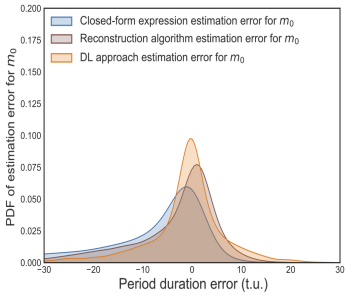
expression approach in Fig. 4.2 (left) for the estimation of the mean period over short T_s values). On the other hand, the reconstruction algorithm introduced in this work (in Section 4.3) performs better than the closed-from expression because the algorithm itself exploits the closed-form expression to improve the estimation of the mean. It can also be noticed that, as the mean value increases the estimation performance degrades for all approaches. This is due to the fact that the longer the periods the higher the number of sensing errors occur within those periods, thus less accurate estimation can be achieved. In Fig. 4.17 and 4.18, on the other hand, the DL approach also provides higher accuracy for the estimation of the variance and minimum period, respectively. Variance estimation in Fig. 4.17 is only provided for DL and reconstruction approaches since, to the best of the author's knowledge, no closed-form expression for such moment under ISS is available in the literature. In Fig. 4.18, on the other hand, even when a closed-form expression is provided for the estimation of the minimum period under ISS (which is simply given by $\breve{\mu} = T_s$ as discussed in Section 4.2.1), it does not lead to accurate estimation of the true minimum period. Similarly, the reconstruction method also fails to provide accurate estimation for the minimum period under ISS, this is because even after reconstructing the corrupted idle/busy periods under ISS there will be still some short periods which have not been reconstructed properly, thus providing incorrect minimum period estimation. The distribution of estimation error (found from the PDF of the relative error) for all approaches is also provided (in the middle plots), where it shows better performance for DL estimator as it is centred around zero with narrow standard deviation with respect to other approaches. This performance improvement can also be observed in the right hand side plots in terms of the Maximum Absolute Error (MAE) obtained within a 90% confidence interval. The performance shown in Figs. 4.16(a), 4.17(a) and 4.18(a) can also be presented in numerical form as shown in Table 4.4 by taking the average of the differences between the original values of these statistics and their estimations under ISS, for which it can be noticed that the proposed DL approach also, in average, provides less error in the estimation of the original statistics with respect to the other studied approaches.

4.5.2 Distribution Classification and Estimation under ISS

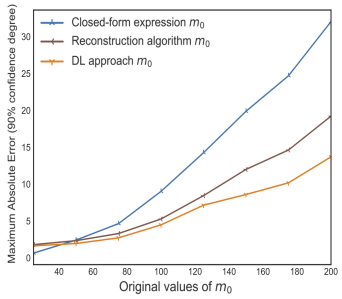
As discussed in Section 4.4.3, channel traffic distribution can be estimated using DL in two stages, first classifying the distribution type, second estimating the distribution parameters. To evaluate the performance of the DL model used to classify the distribution of the channel traffic, a large dataset of 4×10^5 histograms using 100 bins is produced



(a) Mean m_0 estimation under ISS.

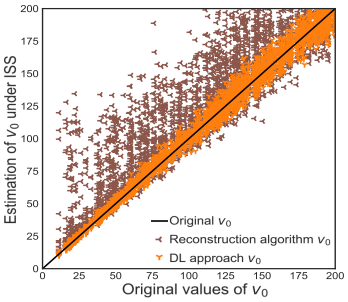


(b) PDF of estimation error for the mean m_0.

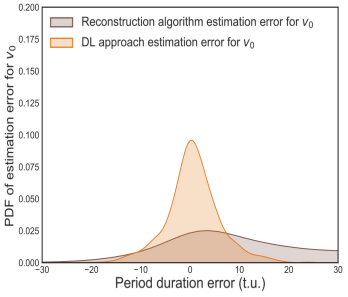


(c) Maximum Absolute Error for the mean m_0 estimation.

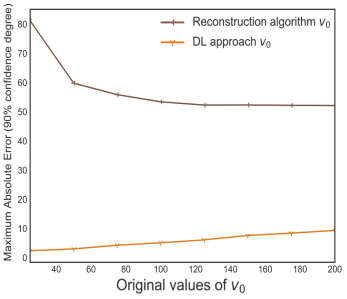
Figure 4.16: Simulation results for traffic mean estimation under ISS using different approaches, when P_{fa} and $P_{md} \sim \mathcal{U}(0.01, 0.1)$, $T_s = 1$ t.u..



(a) Variance v_0 estimation under ISS.

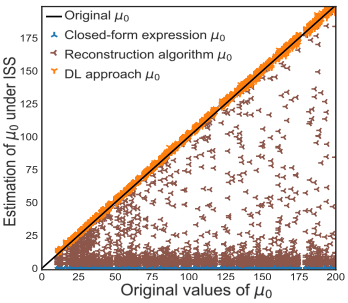


(b) PDF of estimation error for the variance v_0.

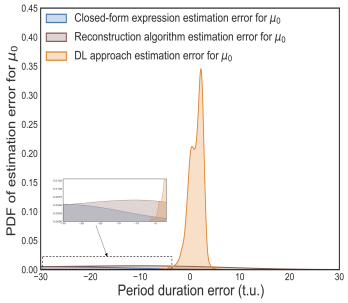


(c) Maximum Absolute Error for the variance v_0 estimation.

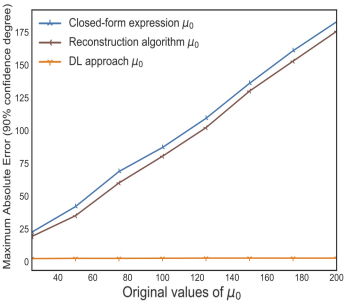
Figure 4.17: Simulation results for traffic variance estimation under ISS using different approaches, when P_{fa} and $P_{md} \sim \mathcal{U}(0.01, 0.1)$, $T_s = 1$ t.u..



(a) Minimum μ_0 estimation under ISS.



(b) PDF of estimation error for the minimum μ_0.



(c) Maximum Absolute Error for minimum μ_0 estimation.

Figure 4.18: Simulation results for traffic minimum estimation under ISS using different approaches, when P_{fa} and $P_{md} \sim \mathcal{U}(0.01, 0.1)$, $T_s = 1$ t.u..

Table 4.4: Average error for statistics estimation using different approaches.

Approach	**Mean (t.u.)**	**Variance (t.u.)^2**	**Minimum (t.u.)**
Closed-form expression	10.83	–	106.75
Reconstruction algorithm	7.95	60.56	86.41
Deep Learning	5.22	10.33	2.42

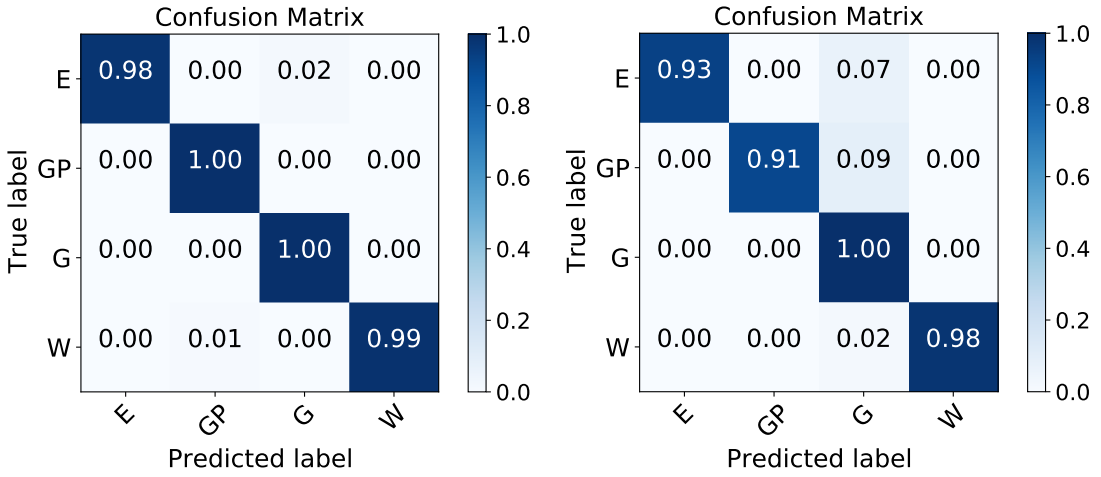


Figure 4.19: Channel traffic distribution classification under ISS when (a) $P_{fa} = 0.01$ and $P_{md} = 0.01$, (b) $P_{fa} = 0.1$ and $P_{md} = 0.1$.

by remodelling the channel traffic several times using (E, GP, G and W) distribution models. The corresponding observations of the channel traffic under ISS using random P_{fa} and $P_{md} \sim \mathcal{U}(0.01, 0.1)$ are obtained. Similar to the previous section, 60% and 20% of such dataset are used to train and validate the DL model, respectively, while the remaining 20% of the dataset is used to test the accuracy of classification. Fig. 4.19 shows the accuracy of classifying the distribution of the observed channel traffic under low (0.01) and high (0.1) probability of sensing error. Sensing errors can distort the shape of the observed traffic distribution. However, as it can be seen from the confusion matrix, even under high probability of sensing error the proposed DL model can still provide accurate classification for the observed channel traffic under ISS. To estimate the distribution parameters (μ_i, λ_i and α_i), MoM method can be applied according to the selected distribution type. Since the mean, variance and minimum period can be estimated accurately using DL approach as seen from the previous section results,

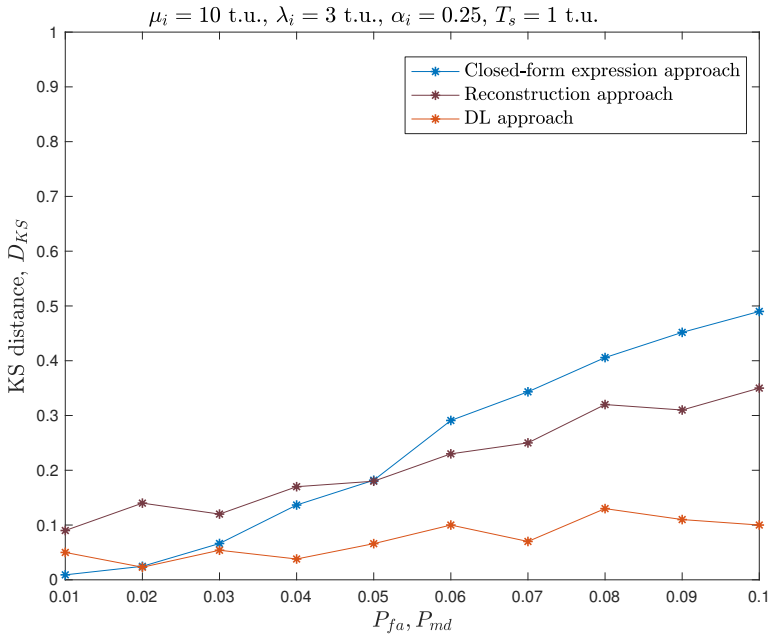


Figure 4.20: KS distance of the channel traffic CDF estimation under ISS.

accurate estimation can also be obtained for ($\widetilde{\mu}_i, \widetilde{\lambda}_i$ and $\widetilde{\alpha}_i$), based on which the CDF of the channel traffic $\widetilde{F}(T_i)$ can then be found as explained in Section 4.4.3. The accuracy of this estimation can be presented in terms of the Kolmogorov-Smirnov (KS) distance between the estimated CDF $\widetilde{F}(T_i)$ and the original CDF $F(T_i)$ of the channel traffic as:

$$D_{KS} = \sup_{T_i} |\widetilde{F}(T_i) - F(T_i)| \tag{4.21}$$

Therefore, based on (4.21), the accuracy of estimating the distribution of the channel traffic under ISS is shown in Fig. 4.20 using DL, reconstruction algorithm and closed-form expression approaches when the original traffic distribution is drawn from GP with $\mu_i = 10$ t.u., $\lambda_i = 3$ t.u. and $\alpha_i = 0.25$ parameters. As it can be appreciated, the proposed DL approach achieves lower KS distance (i.e., higher accuracy of estimation) than the previous approaches for different values of P_{fa} and P_{md}. Since the estimation of the traffic distribution using DL approach is dependent on the estimations of the mean, variance and minimum period, its accuracy changes according to the accuracy of estimating those moments, which are also obtained using DL approach for the given P_{fa} and P_{md}. Similar observations can be obtained as well for the estimation of other types of distributions, showing significant improvement in the distribution estimation through using the proposed DL approach.

4.5.3 Computational Complexity

The computational complexity of the different approaches used to estimate channel traffic statistics under ISS is an important aspect to investigate. Generally, closed-form expressions approach tends to be more attractive in terms of the complexity as it provides accurate estimations for the channel traffic statistics under ISS through using explicit mathematical equations. However, the accuracy of these equations tends to degrade as the sensing period T_s decreases. This is because decreasing the latter causes an increase in the number of the sensing events within an observed period, which in turn increases the occurrence of sensing errors as a result. In addition, regardless of being more attractive, closed-form expressions can be challenging sometimes to find for higher statistical parameters under ISS such as variance, skewness and kurtosis (where this can be noticed from the results of Fig. 4.17, the absence of the closed-from expression approach for variance estimation). The reconstruction algorithms approach, on the other hand, is less attractive in terms of the complexity as it performs heavily computational operations with several iterations in their algorithms in order to reconstruct the idle/busy periods corrupted by the sensing errors. In the reconstruction Algorithm 1, for example, each sensing error needs to be identified and then corrected using two arithmetic (addition) operations. These operations, therefore, increase significantly as the number of the sensing errors increases and they, even more, double for every iteration performed. In contrast, the complexity of the deep learning approach depends on the NN models used to perform estimation (i.e., number of layers, neurons, etc.). The computation requirements of this approach weighs more on the training process than on the prediction process of the DL models. However, this training operation does not take place often, in fact once a DL model is trained it can then be used to perform estimations for the channel traffic statistics.

Table 4.5 shows a comparison for the computational complexity of the considered approaches in this work in terms of the computation time taken to perform 100 samples of estimations for the channel traffic statistics under ISS. As it can be appreciated, the computational cost associated with the closed-form expressions approach is the most efficient one, while it is significantly higher for the reconstruction algorithm approach (using the proposed Method 5). The implications of the long time taken by the reconstruction algorithm to provide accurate statistical estimations, as shown in Table 4.5, could limit the applicability of this approach when fast statistics estimations (short-term decisions) for the channel activity is required (i.e., its applicability might be limited to long-term decisions). On the other hand, the deep learning approach is

Table 4.5: Computation time (in seconds) required by each approach to provide 100 estimations for different statistical metrics.

Approach	**Mean**	**Variance**	**Minimum**	**Distribution**
Closed-form expressions	0.037	–	0.001	0.52
Reconstruction algorithms	200.7	243	266.8	280
Deep Learning	0.4	0.42	0.47	1.7
+ training	50.8	51.4	50.74	162.1

considerably less complex than the algorithmic approach and reasonably more complex than the closed-form expression approach. It can also be noticed that the already trained DL models require significantly less computations than the resulting computations from the training process, however, as explained earlier, this training is not required to take place often to preform estimations for the channel traffic statistics. Therefore, considering the significant accuracy improvement with a reasonable increase in the complexity, the proposed DL approach can be considered an efficient solution for providing accurate estimation for the channel traffic statistics under ISS.

4.6 Summary

In the previous chapter, the observation of the channel activity statistics under ISS was analysed and closed-form expressions were provided for several statistical metrics to accurately represent these statistics mathematically as a function of their corresponding original statistics. The outcomes of these results have contributed, in this chapter, to find novel estimation methods for the channel activity statistics (mean, duty cycle and distribution) under ISS based on the obtained closed-form expressions. This closed-form expressions approach outperforms the conventional estimation methods calculated directly from the sensing decisions. Then, the reconstruction approach has also been investigated in this chapter for the estimation of the channel activity statistics under ISS, wherein a novel reconstruction algorithm has been proposed to be the first reconstruction algorithm that does not require any prior knowledge about the channel activity, and at the same time approaches the performance achieved by the latest reconstruction method available in the literature. This method utilises the closed-form expression achieved for the mean period as an indicator of when periods

are reconstructed correctly. Finally, the chapter introduces a new approach, Traffic Learning (TL), as a deep learning approach for obtaining accurate statistical information of the channel traffic in spectrum sharing systems. This approach learns from the imperfect observations of the traffic statistics in order to predict their accurate estimations. Several deep learning models have been validated for the estimation of the mean, variance, minimum and distribution of the channel traffic.

The performance and the complexity of all studied approaches have been analysed and compared altogether for the estimation of the channel activity statistics under ISS. It has been demonstrated that the deep learning approach outperforms the other approaches which are based on closed-form expressions and reconstruction algorithms, under different scenarios of sensing error probabilities and specially for the worst scenario of using a short sensing period. The closed-form expressions approach, on the other hand, can be more attractive in terms of the complexity, but it can be challenging to find for higher statistical moments (e.g., variance).

Chapter 5

Conclusions and Future Work

5.1 Conclusions

The spectrum sharing concept is one of the promising solutions to overcome the frequency scarcity problem and maximise spectrum utilisation efficiency. The harmonious coexistence of several wireless communication systems in a shared frequency spectrum is highly dependent on making effective decisions for the utilisation of such spectrum. These decisions are usually based on the users' activity within the channel and their traffic statistical information. Therefore, it is crucial for a spectrum sharing system to obtain accurate estimation of the channel activity statistics even under low SNR conditions (i.e., ISS). In this context, this thesis has addressed this challenging problem and conducted a detailed mathematical analysis on the estimation of the channel activity statistics under a realistic ISS scenario taking into account the main factors that would influence the estimation accuracy of the channel activity statistics. In addition, several approaches have been investigated to correct the estimation of these statistics under the realistic operational scenario of spectrum sensing. To this end, the main conclusions of this thesis are:

1. **Chapter 2:** Two research problems have been considered in this chapter. First, the problem of cooperative spectrum sensing in spectrum sharing systems has been investigated and analysed in terms of the collision ratio (i.e., interference) and missed-opportunity ratio (i.e., utilisation). A novel approach has been proposed, which can achieve minimum interference and maximum utilisation in spectrum sharing systems. It has been shown that the proposed cooperative approach outperforms the conventional approach by taking into account the impact of both sensing errors and sensing resolution. The second part of the chapter, on the

other hand, has introduced an experimental platform (USRP-based prototype) to support a wide range of experiments for the estimation of the channel activity statistics using spectrum sensing. The proposed prototype overcomes the limitations associated with the existing experimental platform in the literature (PECAS) for analysing the statistics of the channel activity in terms of its functionality and wide applicability. This prototype serves as a proof-of-concept for the conducted analyses throughout this thesis as well as similar research work conducted by other researchers and engineers in the related field.

2. **Chapter 3:** This chapter has addressed the significant problem of observing inaccurate channel activity statistics under the presence of sensing errors (i.e., ISS). Such problem has been poorly addressed in the literature without comprehensive mathematical analyses taking into account the factors that would influence the estimation accuracy of the channel activity statistics such as the probability of sensing error and the sensing period. Therefore, the impact of the sensing errors and sensing period on statistics estimation has been analysed and mathematical relationships (in closed-form expressions) between the observed channel activity statistics under ISS and their corresponding original statistics have been found. In addition, the impact of the sample size on the estimation of these statistics has also been analysed, and closed-form expressions for the required sample size under ISS to achieve a targeted level of accuracy have been obtained. The correctness of the achieved analytical results has been validated by means of simulation and experimental results, showing an excellent agreement for all cases.

3. **Chapter 4:** After addressing and analysing the problem of inaccurate estimation of channel activity statistics under ISS in the previous chapter, this chapter has investigated three approaches to correct the estimation of channel statistics corrupted by the presence of sensing errors. The first approach exploits the closed-form expressions obtained previously to find novel estimation methods for the channel activity statistics under ISS. These expressions provide accurate estimations for the channel statistics based on the inaccurate observations of their corresponding statistical parameters under ISS, outperforming the conventional estimation methods used to calculate these statistics directly from the sensing decisions. Second, another approach based on reconstruction algorithms has been studied and for which a new reconstruction algorithm has been proposed that does not require any prior knowledge on the channel activity to correct statistics estimation (opposite to all existing reconstruction methods in the literature) and

at the same time achieving the same performance achieved by the latest reconstruction method in the literature. Finally, the chapter has introduced a novel approach named Traffic Learning (TL) as a deep learning approach to provide accurate estimation of the channel activity statistics by learning from their imperfect observations under ISS. The studied mathematical, algorithmic and deep learning approaches have been evaluated and compared in terms of their performance and complexity. It has been concluded that the TL approach in general outperforms other approaches for providing accurate estimation of the channel activity statistics with some reasonable complexity for Neural Network (NN) training and prediction process. In contrast, the closed-form expressions approach is more attractive in terms of the complexity for the achieved estimation accuracy, but it can be limited to specific statistical parameters. The reconstruction algorithms approach, on the other hand, can be very costly in terms of the computational complexity required to correct the estimation of the channel activity statistics. In addition, it may require some prior information of the channel activity (as in the existing algorithms in the literature) or it may exploit a closed-form expression (as in the proposed algorithm).

In summary, the work carried out in this research has shown that it is possible to achieve an accurate estimation of the channel activity statistics in spectrum sharing systems under realistic operating conditions, including practical impairments and limitations of real systems.

5.2 Future Work

The conducted research throughout this thesis has broadened the author's horizon and opened further questions that require further research and investigations. These research topics can be considered in the future work as extensions to this thesis, which include but are not limited to the points discussed below:

1. As a part of the future work, the proposed Cooperative Spectrum Sensing (CSS) approach in Chapter 2 can further be developed to propose a hybrid approach, where both the proposed and the conventional CSS approaches can be combined in one system to achieve even better spectrum utilisation efficiency and minimum interference. In such approach, a number of SUs will be selected to operate under the conventional CSS approach to reduce the impact of the sensing error (i.e., ISS), while another number of SUs will be selected to operate under the proposed

CSS approach at the same time to reduce the impact of the sensing resolution (i.e., T_s). The selection procedure of the SUs will be required to be studied to find the best combinations of the SUs' groups that can provide accurate observations under each CSS approach based on their (i.e., SUs) individual operating conditions such as P_{fa}, P_{md} and T_s.

2. For the estimation of the channel activity statistics under ISS based on the machine learning approach, other types of neural network models can be considered and investigated for providing accurate statistical information of the channel activity. This can be studied more comprehensively through considering simpler models such as the traditional methods of machine learning, and more powerful models such as the Convolutional Neural Network (CNN) and/or Recurrent Neural Network (RNN). The selection of these models would be based on the difficulty for a specific statistical parameter to estimate from the ISS observations. In addition, the potential of using Multitask Learning in NN models can also be applied to solve several tasks through using a shared model. The complexity of these neural networks with respect to the ones considered in this work can then be evaluated and compared.

3. A useful extension of this work, furthermore, would be the exploitation of the proposed estimation methods in this thesis in various applications of spectrum sharing systems. This can be investigated by considering a realistic scenario (i.e., ISS) for a spectrum sharing system, and then apply the methods proposed in this work to help spectrum sharing systems take critical decisions regarding how to use the spectrum, for example, to select the optimum available channel, to predict the future behaviour of the spectrum, to optimise the radio resource management etc.. This would require a detailed study that is beyond the scope of this thesis and is suggested as future work.

Bibliography

[1] A. Goldsmith, *Wireless Communications.* Cambridge University Press, 2005.

[2] M. W. Akhtar, S. A. Hassan, R. Ghaffar, *et al.*, "The shift to 6G communications: vision and requirements," in *Human-centric Computing and Information Sciences*, vol. 10, no. 53, pp.1–27, 2020.

[3] P. Yang, Y. Xiao, M. Xiao and S. Li, "6G Wireless Communications: Vision and Potential Techniques," in *IEEE Network*, vol. 33, no. 4, pp. 70–75, 2019.

[4] M. Z. Chowdhury, M. Shahjalal, S. Ahmed and Y. M. Jang, "6G Wireless Communication Systems: Applications, Requirements, Technologies, Challenges, and Research Directions," in *IEEE Open Journal of the Communications Society*, vol. 1, pp. 957–975, 2020.

[5] A. Mourad, R. Yang, P. H. Lehne and A. de la Oliva, "Towards 6G: Evolution of Key Performance Indicators and Technology Trends," *2020 2nd 6G Wireless Summit (6G SUMMIT)*, 2020, pp. 1–5.

[6] ITU-R M.2370-0, *IMT traffic estimates for the years 2020 to 2030*, Jul. 2015.

[7] 3GPP TS 38.101-2 v15.3.0, *NR; User Equipment (UE) radio transmission and reception; Part 2: Range 2 Standalone (Release 15),* Oct. 2018.

[8] 3GPP TS 38.101-1 v15.3.0, *NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone (Release 15),* Oct. 2018.

[9] M. Marcus and B. Pattan, "Millimeter wave propagation: spectrum management implications," in *IEEE Microwave Magazine*, vol. 6, no. 2, pp. 54–62, June 2005.

[10] S. Haykin, "Cognitive radio: brain-empowered wireless communications," in *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[11] B. Wang and K. J. R. Liu, "Advances in cognitive radio networks: A survey," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 5–23, Feb. 2011.

[12] C. B. Papadias, T. Ratnarajah and D. T. M. Slock, *Spectrum Sharing: The Next Frontier in Wireless Networks.* John Wiley & Sons, 2020.

[13] S. Biswas, A. Bishnu, F. A. Khan and T. Ratnarajah, "In-Band Full-Duplex Dynamic Spectrum Sharing in Beyond 5G Networks," in *IEEE Communications Magazine*, vol. 59, no. 7, pp. 54–60, July 2021.

[14] X. Liu, K. -Y. Lam, F. Li, J. Zhao, L. Wang and T. S. Durrani, "Spectrum Sharing for 6G Integrated Satellite-Terrestrial Communication Networks Based on NOMA and CR," in *IEEE Network*, vol. 35, no. 4, pp. 28–34, July/August 2021.

[15] M. Ali, S. Qaisar, M. Naeem, W. Ejaz and N. Kvedaraite, "LTE-U WiFi HetNets: Enabling Spectrum Sharing for 5G/Beyond 5G Systems," in *IEEE Internet of Things Magazine*, vol. 3, no. 4, pp. 60–65, December 2020.

[16] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 79–89, May 2007.

[17] S. Wirsing and P. Reichl, "Dynamic spectrum access and the current spectrum management paradigm: On the challenges of dynamic licensing," *2015 13th International Conference on Telecommunications (ConTEL)*, 2015, pp. 1–8.

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

[19] Ingenious Consulting Network, *Authorised Shared Access (ASA). An evolutionary spectrum authorisation scheme for sustainable economic growth and consumer benefit*, ref. document FM(11)116, Germany, 2011.

[20] Radio Spectrum Policy Group RSPG13–538, *RSPG Opinion on Licensed Shared Access*, Nov. 2013.

[21] A. J. Morgado, "The Licensed Shared Access Approach," in *Spectrum Sharing: The Next Frontier in Wireless Networks*, IEEE, 2019, pp. 97–119.

[22] FCC document 15-47, *Report and Order and Second Further Notice of Proposed Rulemaking*, 2015.

[23] M. Mueck, M. Dolores (Lola) Pérez Guirao, R. Yallapragada and S. Srikanteswara, "Regulation and Standardization Activities Related to Spectrum Sharing," in *Spectrum Sharing: The Next Frontier in Wireless Networks*, IEEE, 2019, ch. 2, pp.17–33.

[24] Ł. Kułacz et al., "Coordinated Spectrum Allocation and Coexistence Management in CBRS-SAS Wireless Networks," in *IEEE Access*, vol. 7, pp. 139294–139316, 2019.

[25] C. B. Papadias, T. Ratnarajah and D. T. M. Slock, "Introduction," in *Spectrum Sharing: The Next Frontier in Wireless Networks*, IEEE, 2019, ch. 1, pp.1–15.

[26] FP7 Research Project ADEL (Advanced Dynamic spectrum 5G mobile networks Employing Licensed shared access), Grant no. 619647. [Online]. Available: `https://cordis.europa.eu/project/id/619647`.

[27] LTE-U Forum TR v1.0, *LTE-U Technical Report Coexistence Study for LTE-U SDL,* Feb. 2015.

[28] 3GPP TR 36.889 v13.0.0, *Study on Licensed-Assisted Access to Unlicensed Spectrum (Release 13),* Jun. 2015.

[29] 3GPP TR 38.889 v16.0.0, *Study on NR-based access to unlicensed spectrum (Release 16),* Dec. 2018.

[30] N. Patriciello, S. Lagén, B. Bojović and L. Giupponi, "NR-U and IEEE 802.11 Technologies Coexistence in Unlicensed mmWave Spectrum: Models and Evaluation," in *IEEE Access*, vol. 8, pp. 71254–71271, 2020.

[31] X. Lu *et al.*, "Integrated Use of Licensed- and Unlicensed-Band mmWave Radio Technology in 5G and Beyond," in *IEEE Access*, vol. 7, pp. 24376–24391, 2019.

[32] P. B. Oni and S. D. Blostein, "Optimized Carrier Sensing Thresholds for Dense mmWave Wireless Networks Coexistence," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[33] T. S. Dhope, *Cognitive radio Networks optimization with spectrum sensing Algorithms*, Vol. 44, River Publishers, 2015.

[34] H. Kokkinen, "TV White Space Spectrum Sharing Using Geolocation Databases," in *TV White Space Communications and Networks*, pp. 29–43. Woodhead Publishing, 2018.

[35] A. P. Hulbert, "Spectrum sharing through beacons," *2005 IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications*, 2005, pp. 989–993.

[36] M. López-Benítez, "Cognitive radio," in *Heterogeneous cellular networks: Theory, simulation and deployment.* Cambridge University Press, 2013, ch. 13.

[37] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proc. IEEE*, vol. 55, no. 4, pp. 523–531, Apr. 1967.

[38] A. Sharma and A. Chauhan, "Spectrum sensing based on multiple energy detector for cognitive radio systems under noise uncertainty," *2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPE-ICES)*, 2016, pp. 1–4.

[39] M. López-Benítez and F. Casadevall, "Improved energy detection spectrum sensing for cognitive radio," *IET Communications*, vol. 6, no. 8, pp. 785–796, 22 May 2012.

[40] M. Z. Alom, T. K. Godder, M. N. Morshed and A. Maali, "Enhanced spectrum sensing based on Energy detection in cognitive radio network using adaptive threshold," *2017 International Conference on Networking, Systems and Security (NSysS)*, Dhaka, Bangladesh, 2017, pp. 138–143.

[41] B. Soni, D. K. Patel and M. López-Benítez, "Long Short-Term Memory Based Spectrum Sensing Scheme for Cognitive Radio Using Primary Activity Statistics," in *IEEE Access*, vol. 8, pp. 97437–97451, 2020.

[42] F. Salahdine, H. E. Ghazi, N. Kaabouch, and W. F. Fihri, "Matched lter detection with dynamic threshold for cognitive radio networks," in *2015 International Conference on Wireless Networks and Mobile Communications (WINCOM)*, Oct 2015, pp. 1–6.

[43] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," in *Comput. Netw.*, vol. 50, no. 13, pp. 2127–2159, Sep. 2006.

[44] G. Huang and J. K. Tugnait, "On cyclostationarity based spectrum sensing under uncertain Gaussian noise," *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 2042–2054, April 2013.

[45] M. López-Benítez, "Sensing-based spectrum awareness in Cognitive Radio: Challenges and open research problems," *2014 9th International Symposium on Communication Systems, Networks & Digital Sign (CSNDSP)*, Manchester, UK, 2014, pp. 459–464.

[46] 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, May 2009.

[47] C. I. M. Althaf and S. C. Prema, "Covariance and eigenvalue based spectrum sensing using USRP in real environment," *2018 10th International Conference on Communication Systems & Networks (COMSNETS)*, 2018, pp. 414–417.

[48] A. Chen, Z. Shi and J. Xiong, "Generalized Real-Valued Weighted Covariance-Based Detection Methods for Cognitive Radio Networks With Correlated Multiple Antennas," in *IEEE Access*, vol. 7, pp. 34373–34382, 2019.

[49] M. López-Benítez, A. Al-Tahmeesschi, D. K. Patel, J. Lehtomäki and K. Umebayashi, "Estimation of Primary Channel Activity Statistics in Cognitive Radio Based on Periodic Spectrum Sensing Observations," in *IEEE Transactions on Wireless Communications*, vol. 18, no. 2, pp. 983–996, Feb. 2019.

[50] G. Ding, *et al.*, "Spectrum Inference in Cognitive Radio Networks: Algorithms and Applications," in *IEEE Communications Surveys & Tutorials*, vol. 20, no. 1, pp. 150–182, Firstquarter 2018.

[51] L. Yang, N. Lv, and Z. X. Xu, "Spectrum prediction for cognitive radio system based on optimally pruned extreme learning machine," *Appl. Mech. Mater.*, vols. 536–537, pp. 430–436, Apr. 2014.

[52] Y. Chen and H.-S. Oh, "Spectrum measurement modelling and prediction based on wavelets," *IET Communications*, vol. 10, no. 16, pp. 2192–2198, Oct 2016.

[53] X. Xing, T. Jing, W. Cheng, Y. Huo, and X. Cheng, "Spectrum prediction in cognitive radio networks," *IEEE Wireless Commun.*, vol. 20, no. 2, pp. 90–96, Apr. 2013.

[54] K. Umebayashi, K. Hayashi and J. J. Lehtomäki, "Threshold-Setting for Spectrum Sensing Based on Statistical Information," in *IEEE Communications Letters*, vol. 21, no. 7, pp. 1585–1588, July 2017.

[55] S. Sengottuvelan, J. Ansari, P. Mähönen, T. G. Venkatesh and M. Petrova, "Channel Selection Algorithm for Cognitive Radio Networks with Heavy-Tailed Idle Times," in *IEEE Trans. on Mobile Computing*, vol. 16, no. 5, pp. 1258–1271, 2017.

[56] M. B. Hosen, M. M. H. Mridha and M. A. Hamza, "Secondary User Channel Selection in Cognitive Radio Network Using Adaptive Method," *2018 3rd International Conference for Convergence in Technology (I2CT)*, Pune, 2018, pp. 1–6.

[57] X. Liu, B. Krishnamachari, and H. Liu, "Channel selection in multichannel opportunistic spectrum access networks with perfect sensing," in *Proc. 2010 IEEE Int'l. Symp. Dyn. Spect. Access Networks (DySPAN 2010)*, Apr. 2010, pp. 1–8.

[58] A. John and A. P. Mathew, "Channel selection in cognitive radio networks using exploration order and stopping rule," *2016 International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, 2016, pp. 1–5.

[59] G. C. Deepak, K. Navaie and Q. Ni, "Radio Resource Allocation in Collaborative Cognitive Radio Networks Based on Primary Sensing Profile," in *IEEE Access*, vol. 6, pp. 50344–50357, 2018.

[60] W. Zhang, C. Wang, X. Ge and Y. Chen, "Enhanced 5G Cognitive Radio Networks Based on Spectrum Sharing and Spectrum Aggregation," in *IEEE Trans. on Comms.*, vol. 66, no. 12, pp. 6304–6316, Dec. 2018.

[61] E. Jung and X. Liu, "Opportunistic spectrum access in multiple-primary-user environments under the packet collision constraint," *IEEE/ACM Trans. Networking*, vol. 20, no. 2, pp. 501–514, Apr. 2012.

[62] A. Al-Tahmeesschi, M. López-Benítez, K. Umebayashi and J. Lehtomäki, "Analytical study on the estimation of primary activity distribution based on spectrum sensing," *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Montreal, QC, 2017, pp. 1–5.

[63] A. Al-Tahmeesschi, M. López-Benítez, J. Lehtomäki and K. Umebayashi, "Accurate estimation of primary user traffic based on periodic spectrum sensing," *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, Barcelona, 2018, pp. 1–6.

[64] C. Liu, P. Pawelczak and D. Cabric, "Primary User Traffic Classification in Dynamic Spectrum Access Networks," in *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 11, pp. 2237–2251, November 2014.

[65] C. Liu, E. Rebeiz, P. Pawelczak and D. Cabric, "Primary user traffic classification in dynamic spectrum access networks," *2013 IEEE Global Communications Conference (GLOBECOM)*, Atlanta, GA, USA, 2013, pp. 1155–1160.

[66] M. López-Benítez, A. Al-Tahmeesschi and D. Patel, "Accurate Estimation of the Minimum Primary Channel Activity Time in Cognitive Radio Based on Periodic Spectrum Sensing Observations," *European Wireless 2018; 24th European Wireless Conference*, Catania, Italy, 2018, pp. 1–6.

[67] M. López-Benítez and J. Lehtomäki, "On the sensing sample size for the estimation of primary channel occupancy rate in cognitive radio," *2016 IEEE Wireless Communications and Networking Conference*, Doha, Qatar, 2016, pp. 1–6,

[68] A. Al-Tahmeesschi, M. López-Benítez, D. K. Patel, J. Lehtomäki and K. Umebayashi, "On the Sample Size for the Estimation of Primary Activity Statistics Based on Spectrum Sensing," in *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 1, pp. 59–72, March 2019.

[69] M. López-Benítez and J. Lehtomäki, "Energy detection based estimation of primary channel occupancy rate in cognitive radio," *2016 IEEE Wireless Communications and Networking Conference*, Doha, Qatar, 2016, pp. 1–6,

[70] J. J. Lehtomaki, M. López-Benítez, K. Umebayashi and M. Juntti, "Improved Channel Occupancy Rate Estimation," in *IEEE Transactions on Communications*, vol. 63, no. 3, pp. 643–654, March 2015.

[71] J. J. Lehtomaki, R. Vuohtoniemi and K. Umebayashi, "On the Measurement of Duty Cycle and Channel Occupancy Rate," in *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 11, pp. 2555–2565, November 2013.

[72] W. Gabran, P. Pawelczak, C. Liu and D. Cabric, "Blind estimation of primary user traffic parameters under sensing errors," *2013 IEEE International Conference on Communications (ICC)*, Budapest, 2013, pp. 2391–2396.

[73] W. Gabran, C. H. Liu, P. Pawelczak and D. Cabric, "Primary user traffic estimation for dynamic spectrum access," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 3, pp. 544–558, March 2013.

[74] Q. Liang, M. Liu and D. Yuan, "Channel Estimation for Opportunistic Spectrum Access: Uniform and Random Sensing," in *IEEE Transactions on Mobile Computing*, vol. 11, no. 8, pp. 1304–1316, Aug. 2012.

[75] M. López-Benítez, "Can primary activity statistics in cognitive radio be estimated under imperfect spectrum sensing?," *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, London, UK, 2013, pp. 750–755.

[76] A. Al-Tahmeesschi, M. López-Benítez, J. Lehtomäki and K. Umebayashi, "Investigating the Estimation of Primary Occupancy Patterns under Imperfect Spectrum Sensing," *2017 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, San Francisco, CA, 2017, pp. 1–6.

[77] A. Al-Tahmeesschi, M. López-Benítez, J. Lehtomäki and K. Umebayashi, "Improving primary statistics prediction under imperfect spectrum sensing," *2018 IEEE Wireless Comms. and Networking Conference (WCNC)*, Barcelona, Spain, 2018, pp. 1–6.

[78] M. López-Benítez and F. Casadevall, "Time-Dimension Models of Spectrum Usage for the Analysis, Design, and Simulation of Cognitive Radio Networks," in *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 2091–2104, Jun 2013.

[79] K. B. Letaief and W. Zhang, "Cooperative Communications for Cognitive Radio Networks," in *Proceedings of the IEEE*, vol. 97, no. 5, pp. 878–893, May 2009.

[80] I. F. Akyildiz, B. F. Lo and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical communication*, vol. 4, no. 1, pp. 40–62, 2011.

[81] W. Lee, M. Kim and D. Cho, "Deep Cooperative Sensing: Cooperative Spectrum Sensing Based on Convolutional Neural Networks," in *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 3005–3009, March 2019.

[82] W. Choi, M. Song, J. Ahn and G. Im, "Soft Combining for Cooperative Spectrum Sensing over Fast-Fading Channels," in *IEEE Communications Letters*, vol. 18, no. 2, pp. 193–196, February 2014.

[83] N. Salout, F. A. Awin, A. F. Alqawasmeh and E. Abdel-Raheem, "Hierarchical cluster-based cooperative spectrum sensing in cognitive radio employing soft-hard

combination," *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, Windsor, ON, 2017, pp. 1–4.

[84] W. Zhang, R. K. Mallik and K. B. Letaief, "Optimization of cooperative spectrum sensing with energy detection in cognitive radio networks," in *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 5761–5766, December 2009.

[85] O. H. Toma and M. López-Benítez, "Cooperative Spectrum Sensing: A New Approach for Minimum Interference and Maximum Utilisation," *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2021, pp. 1–6.

[86] M. López-Benítez, A. Al-Tahmeesschi, K. Umebayashi and J. Lehtomäki, "PECAS: A low-cost prototype for the estimation of channel activity statistics in cognitive radio," in *Proc. IEEE Wirel. Comms. and Net. Conf. (WCNC 2017)*, 2017, pp. 1–6.

[87] O. H. Toma and M. López-Benítez, "USRP-Based Prototype for Real-Time Estimation of Channel Activity Statistics in Spectrum Sharing," *2021 17th International Symposium on Wireless Communication Systems (ISWCS)*, 2021, pp. 1–6.

[88] O. H. Toma and M. López-Benítez (2020). *Source Code.* [Online]. Available: `https://github.com/ogeen-toma/USRP-prototype-for-channel-statistics-estimation`.

[89] Ettus Research, "USRP B200mini Series Data Sheet," Santa Clara, CA, USA. [Online]. Available: `https://www.ettus.com/wp-content/uploads/2019/01/USRP_B200mini_Data_Sheet.pdf`.

[90] L. Devroye, *Non-uniform random variate generation.* Springer-Verlag, 1986.

[91] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes: The art of scientific computing*, 3rd ed. Cambridge University Press, 2007.

[92] O. H. Toma, M. López-Benítez, D. K. Patel and K. Umebayashi, "Estimation of Primary Channel Activity Statistics in Cognitive Radio Based on Imperfect Spectrum Sensing," in *IEEE Transactions on Communications*, vol. 68, no. 4, pp. 2016–2031, April 2020.

[93] O. H. Toma, M. López-Benítez, D. K. Patel and K. Umebayashi, "Primary Channel Duty Cycle Estimation under Imperfect Spectrum Sensing Based on Mean Channel Periods," *2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, 2019, pp. 1–6.

[94] O. H. Toma, M. López-Benítez and D. K. Patel, "Analysis of the Sample Size Required for an Accurate Estimation of Primary Channel Activity Statistics under Imperfect Spectrum Sensing," *2020 IEEE 31st Annual Intl. Symp. Pers., Indoor and Mobile Radio Comms. (PIMRC)*, London, 2020, pp. 1–6.

[95] M. López-Benítez and F. Casadevall, "Methodological aspects of spectrum occupancy evaluation in the context of cognitive radio," *European Trans. Telecommun.*, vol. 21, no. 8, pp. 680–693, Dec. 2010.

[96] A. D. Spaulding and G. H. Hagn, "On the definition and estimation of spectrum occupancy," *IEEE Trans. Electromagn. Compat.*, vol. 19, no. 3, pp. 269–280, Aug. 1977.

[97] M. Wellens and P. Mähönen, "Lessons learned from an extensive spectrum occupancy measurement campaign and a stochastic duty cycle model," *Mobile Netw. Appl.*, vol. 15, no. 3, pp. 461–474, 2010.

[98] O. H. Toma, M. López-Benítez, D. K. Patel and K. Umebayashi, "Reconstruction Algorithm for Primary Channel Statistics Estimation Under Imperfect Spectrum Sensing," *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, South Korea, 2020, pp. 1–5.

[99] M. Hirzallah, M. Krunz, B. Kecicioglu and B. Hamzeh, "5G New Radio Unlicensed: Challenges and Evaluation," in *IEEE Transactions on Cognitive Communications and Networking*, 2020.

[100] R. Karaki, "5G New Radio in Unlicensed Spectrum," in *5G and Beyond: Fundamentals and Standards*, p. 401, 2021.

[101] J. Oh, Y. Kim, Y. Li, J. Bang and J. Lee, "Expanding 5G New Radio Technology to Unlicensed Spectrum," *2019 IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–6.

[102] R. Zhang, M. Wang, L. X. Cai, Z. Zheng, X. Shen and L. Xie, "LTE-unlicensed: the future of spectrum aggregation for cellular networks," in *IEEE Wireless Communications*, vol. 22, no. 3, pp. 150–159, June 2015.

[103] B. Bojovic, L. Giupponi, Z. Ali and M. Miozzo, "Evaluating Unlicensed LTE Technologies: LAA vs LTE-U," *IEEE Access*, vol. 7, pp. 89714–89751, 2019.

[104] H. Lee, H. Kim, H. J. Yang, J. T. Kim and S. Baek, "Performance Analysis of License Assisted Access LTE with Asymmetric Hidden Terminals," in *IEEE Transactions on Mobile Computing*, vol. 17, no. 9, pp. 2141–2154, 1 Sep. 2018.

[105] Y. Gao and S. Roy, "Achieving Proportional Fairness for LTE-LAA and Wi-Fi Coexistence in Unlicensed Spectrum," in *IEEE Transactions on Wireless Communications*, vol. 19, no. 5, pp. 3390–3404, May 2020.

[106] M. Han, S. Khairy, L. X. Cai, Y. Cheng and R. Zhang, "Reinforcement Learning for Efficient and Fair Coexistence Between LTE-LAA and Wi-Fi," in *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8764–8776, Aug. 2020.

[107] E. Testi, E. Favarelli and A. Giorgetti, "Machine Learning for User Traffic Classification in Wireless Systems," *2018 26th European Signal Processing Conference (EUSIPCO)*, Rome, 2018, pp. 2040–2044.

[108] E. Testi, L. Pucci, E. Favarelli and A. Giorgetti, "Blind Traffic Classification in Wireless Networks," *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1747–1751.

[109] A. Al-Tahmeesschi, K. Umebayashi, H. Iwata, M. López-Benítez and J. Lehtomäki, "Applying Deep Neural Networks for Duty Cycle Estimation," *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, Seoul, Korea (South), 2020, pp. 1–7.

[110] A. Al-Tahmeesschi, K. Umebayashi, H. Iwata, J. Lehtomäki and M. López-Benítez, "Feature-Based Deep Neural Networks for Short-Term Prediction of WiFi Channel Occupancy Rate," in *IEEE Access*, vol. 9, pp. 85645–85660, 2021.

[111] P. Salvo Rossi, D. Ciuonzo and G. Romano, "Orthogonality and Cooperation in Collaborative Spectrum Sensing through MIMO Decision Fusion," in *IEEE Transactions on Wireless Communications*, vol. 12, no. 11, pp. 5826–5836, November 2013.

[112] A. Patel, H. Ram, A. K. Jagannatham and P. K. Varshney, "Robust Cooperative Spectrum Sensing for MIMO Cognitive Radio Networks Under CSI Uncertainty," in *IEEE Transactions on Signal Processing*, vol. 66, no. 1, pp. 18–33, 1 Jan.1, 2018.

[113] I. Dey, D. Ciuonzo and P. S. Rossi, "Wideband Collaborative Spectrum Sensing Using Massive MIMO Decision Fusion," in *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5246–5260, Aug. 2020.

[114] M. A. Saleem et al., "Expansion of Cluster Head Stability Using Fuzzy in Cognitive Radio CR-VANET," in *IEEE Access*, vol. 7, pp. 173185–173195, 2019.

[115] O. H. Toma and M. López-Benítez, "Traffic Learning: A Deep Learning Approach for Obtaining Accurate Statistical Information of the Channel Traffic in Spectrum Sharing Systems," in *IEEE Access*, vol. 9, pp. 124324–124336, 2021.

[116] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[117] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," in *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, Dec. 2017.

[118] J. Schmidt-Hieber, "Nonparametric regression using deep neural networks with ReLU activation function," *The Annals of Statistics*, vol. 48, no. 4, pp. 1875–1897, August 2020.

[119] S. Langer, "Approximating smooth functions by deep neural networks with sigmoid activation function," *Journal of Multivariate Analysis*, vol. 182, 104696, 2021.

[120] S. Totaro, A. Hussain and S. Scardapane, "A non-parametric softmax for improving neural attention in time-series forecasting," *Neurocomputing*, vol. 381, pp.177–185, 2020.

[121] M. Abadi *et al.*, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems," 2015. [Online]. Available: http://tensorflow.org/.

[122] F. Chollet. *et al.*, "Keras," 2015. [Online]. Available: https://github.com/fchollet/keras.

[123] Paszke, Adam *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in neural information processing systems*, pp. 8026–8037, 2019.

[124] P. Tehrani, L. Tong, and Q. Zhao, "Asymptotically efficient multichannel estimation for opportunistic spectrum access," *IEEE Trans. Signal Processing*, vol. 60, no. 10, pp. 5347–5360, Oct. 2012.

[125] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous univariate distributions*, 2nd ed. Wiley, Nov. 1994, vol. 1.

[126] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, 10th ed. New York: Dover, 1972.