

**QUANTIFYING SECONDARY SPECTRUM AND OPTIMAL
TIMESCALES IN MOBILE WIRELESS COMMUNICATIONS**

by

Pierce Matthew Rixon



MACQUARIE
University
SYDNEY · AUSTRALIA

Dissertation submitted in fulfilment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

Department of Engineering
Faculty of Science
Macquarie University
Sydney, Australia

25th of June 2019

ABSTRACT

The sub-optimal methods currently used to allocate and manage electromagnetic spectrum results in a significant waste of spectrum resources. This wasted ‘secondary’ spectrum, presents a significant engineering opportunity for developing novel methods and systems to ‘capture’ and utilise. Limited techniques are available for determining, to a high degree, the underlying structure of secondary spectrum and the implications that timescales of observation and operation present to this structure. Additionally, no formal method exists for quantifying and evaluating these structures. This lack of understanding is instead substituted for randomness, making it difficult to develop effective secondary spectrum access systems.

To address this lack of insight, power and latency implications for future secondary spectrum access systems with respect to timescale is identified, followed by the construction of a low cost high resolution measurement system to provide real world context. Leveraging these findings and capabilities, a novel ‘Whitespace Opportunity Distribution’ algorithm is developed, providing an unprecedented view into the structure of secondary spectrum opportunities in both time and frequency. These insights are captured in the collaborative development of several novel secondary spectrum access and spectrum management technologies.

This thesis presents several inventions, novel analysis techniques and evaluation methods to further the understanding of the impact of timescales on secondary spectrum utilisation.

STATEMENT OF ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree to any other university or institution other than Macquarie University.

I also certify that the thesis is an original piece of research and it has been written by me.

To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

Signed: _____

Pierce Matthew Rixon

Date: _____

ACKNOWLEDGMENTS

First and foremost I would like to thank my supervisor Prof. Michael Heimplich. You are a brilliant teacher, co-inventor and colleague. Your secret challenges and constant flow of ideas have significantly evolved the way I think and perceive the world. Thank you for selecting me as your student and persevering (putting up) with me.

The Intel Project team, for the fantastic collaboration and invention.

Prof. Eryk Dutkiewicz, for your guidance and leadership during the Project.

Dr. Gengfa Fang, for the travels in China and abroad.

Dr. Markus Mueck, your knowledge and ruthless work ethic during the entire Project and beyond is exceptional and inspirational. I look forward to the next collaboration!

Dr. Beeshanga Abewardana Jayawickrama and Dr. Ying He, the banter, the adventure, the innovation, the friendship.

Calla, for your constant support, affection and insight. It wasn't easy, but finally, success!

Finally, my family, for constantly reminding me during the entire writing process that I must embody the Nike slogan. Thank you for providing me with the resolve to never give up and to always keep charging forward.

To My Family

“It takes less time to do a good job, than it does to do a bad one.”

– Frank Pusswald (1930 - 2019)

Contents

Table of Contents	xi
List of Figures	xvii
Listings	xxi
Glossary	xxv
1 Introduction	1
1.1 Summary	3
1.1.1 Quantifying Secondary Spectrum and Optimal Timescales	4
1.1.2 Intel Project	5
1.2 Power, Latency and Cost of Secondary Spectrum Access	6
1.3 A High Resolution Spectrum Measurement System	7
1.4 Quantifying Secondary Spectrum Resources	8
1.5 Licensed Shared Access Technology	10
1.5.1 Signal Buffering for Licensed Shared Access (LSA) Technology	11
1.5.2 Evolved Node-B, Local Controller and Method for Allocation of Spectrum for Secondary Usage	11
1.5.3 Methods for Performing Mobile Communications Between Mobile Terminal Devices, Base Stations and Network Control Devices	12

1.5.4	Methods and Devices for Shared Spectrum Allocation	13
1.6	Scope	14
2	Background and Related Work	17
2.1	Introduction	17
2.2	Electromagnetic Spectrum Properties	18
2.2.1	Electromagnetic Spectrum Properties	19
2.2.2	Dimensions and Constraints of Spectrum Usage	21
2.2.3	Time and Frequency	22
2.2.4	Space	23
2.3	Electromagnetic Spectrum Management	27
2.3.1	Spectrum Allocation	29
2.3.2	Spectrum Assignment	30
2.3.3	Mobile Communications Spectrum	32
2.3.4	Spectrum Management Policy	35
2.4	Radio Resource Management	37
2.4.1	MAC and PHY Layers	37
2.4.2	Orthogonal Frequency Division Multiplexing	39
2.4.3	Channel Duplexing and Aggregation	45
2.5	Secondary Spectrum	50
2.5.1	Managed and Unmanaged Secondary Spectrum	53
2.5.2	Television Band Whitespace (TVWS)	54
2.5.3	LSA and SAS	56
2.6	Software Defined Radios	62
2.6.1	Cognitive Radios	66
2.6.2	Sensing	66

3	Power, Latency and Cost of Secondary Spectrum Access	73
3.1	Introduction	73
3.2	Factors Impacting Power and Latency	75
3.2.1	Power	76
3.2.2	Latency	79
3.3	OFDM Receiver Chain Model	81
3.3.1	Latency-bound Energy Conservation	85
3.4	Energy Efficient Secondary Spectrum Access	86
3.4.1	Simple Whitespace Opportunity Model	86
3.4.2	Cost of Spectrum Access	88
3.4.3	Time Risk of Spectrum Access	92
3.5	Summary	94
4	A High Resolution Spectrum Measurement System	97
4.1	Introduction	97
4.2	Existing Spectrum Measurement Studies	99
4.3	Spectrum Analyser Processing Stages	102
4.3.1	Sample Normalisation	102
4.3.2	Windowing	102
4.3.3	Overlapping	105
4.3.4	Fast Fourier Transform	105
4.3.5	Power Estimation	107
4.4	Energy Detection Based Sensing	108
4.4.1	Energy Detector Performance	109
4.4.2	Receiver Operating Characteristic	113
4.4.3	Averaging Parameter	114
4.4.4	Energy Detection Performance	114

4.5	Custom Measurement System Design & Implementation	116
4.5.1	Requirements	116
4.5.2	Commercially-Available Hardware Platform	118
4.5.3	Custom Measurement System Architecture	122
4.5.4	Software Platforms	124
4.5.5	USRP Hardware Drivers	126
4.5.6	USRP Firmware	127
4.5.7	Software Architecture and Development	128
4.5.8	Calibration and Verification	136
4.6	Filtering and Threshold Selection	137
4.6.1	Spectrum Noise Image Filter Performance	138
4.6.2	Noise Image Filter Evaluation	143
4.7	Summary	145
5	Quantifying Secondary Spectrum Resources	147
5.1	Introduction	147
5.2	Secondary Spectrum Use-Cases	149
5.3	Measurement of Candidate Bands	151
5.4	Spectrum Analysis	153
5.4.1	Existing Spectrum Analysis Methods	154
5.4.2	Spectrum Analysis Tools	155
5.4.3	Timescale Bound Detectability of Whitespace	157
5.4.4	Whitespace Resource Blocks	160
5.4.5	FFT Depth Selection	161
5.5	Measuring Functional Whitespace	169
5.5.1	Whitespace Slices	170
5.6	Whitespace Resource Partitioning	172

5.6.1	Problem Identification	172
5.6.2	Brute Force Approach	174
5.6.3	Procedural Generation Approach	174
5.7	Whitespace Opportunity Distributions	180
5.7.1	WOD Evaluation on 465MHz and 880MHz Whitespace Structure .	181
5.8	Summary	186
6	Intel Project Patent Submissions	189
6.1	Introduction	189
6.2	Methods and Devices for Shared Spectrum Allocation	191
6.2.1	Agreement Matching and Negotiation	191
6.2.2	Time and Availability Dependent Sharing Agreements	193
6.2.3	Summary	194
6.3	Evolved Node-B, Local Controller and Method for Allocation of Spectrum for Secondary Usage	196
6.3.1	Control Domains	197
6.3.2	Interfaces	199
6.3.3	Control Groups	201
6.3.4	Summary	203
6.4	Signal Buffering for Licensed Shared Access (LSA) technology	204
6.4.1	B2TF Protocol and System Architecture	205
6.4.2	Summary	208
6.5	Methods for Performing Mobile Communications Between Mobile Terminal Devices, Base Stations and Network Control Devices	209
6.5.1	Duplex Balancing Schemes	210
6.5.2	Applications for Dynamic Carriers	214
6.5.3	Summary	215

7	Conclusions and Future Work	217
7.1	Conclusions	217
7.2	Future Work	219
	Bibliography	223

List of Figures

2.1	Example spectrum occupancy diagram.	23
2.2	Example of geographical reuse of the same frequency band.	24
2.3	Depiction of geographic multiplexing by taking advantage of physical entities to separate networks	25
2.4	Depiction of four spatial streams between two MIMO antenna arrays	26
2.5	ITU Spectrum Regions	28
2.6	Australian Radiofrequency Spectrum allocations chart	31
2.7	Open Systems Interconnection (OSI) architecture	38
2.8	Depiction of subcarriers in the frequency domain	41
2.9	LTE Resource grid structure, detailing Resource Blocks and Resource Elements.	42
2.10	Time and frequency resource allocation of an LTE downlink channel	44
2.11	Comparison between FDD and TDD spectrum duplexing and access schemes	47
2.12	Carrier Aggregation types	49
2.13	Block Diagram of the LSA Architecture	59
2.14	Block Diagram of the SAS Architecture	62
2.15	Ideal SDR Architecture	64
2.16	Conventional SDR Receiver	65

2.17 Comparison of Spectrum Sensing Methods. Complexity of implementing and designing the sensor versus the accuracy of the sensing method	68
2.18 Illustration of the hidden primary user problem	70
3.1 Received Signal Chain Block Diagram	81
3.2 Receiver Latency and Power Consumption for User Plane Operation	83
3.3 Whitespace Window and Receive Frame Model	87
3.4 Setup versus Data Receive Energy Efficiency Curve	90
4.1 Spectrum Analyser Processing Stages	102
4.2 Amplitude Flatness (AF), Power Flatness (PF) and Overlap Correlation (OC) of a 4 term Blackmann-Harris Window	104
4.3 Architecture of Spectrum Analyser and Energy Detector Processing Stages	109
4.4 Receiver Operating Characteristic Curves	113
4.5 Example 25MHz Bandwidth PSD Plot with Filter Roll-off Regions and Noise Estimation	115
4.6 Diagram of measurement system high level physical architecture	122
4.7 Image of spectrum measurement and analysis system set up.	124
4.8 Flow Diagram of Spectrum Sampling Program	130
4.9 Output from the Spectrum GUI Program	133
4.10 CUDA Module Flow Diagram.	134
4.11 Spectrum GUI Flow Diagram	135
4.12 Calibration and Verification Setup	136
4.13 Thresholded power estimation at -93dBm and 5 by 5 filter kernel applied, $F_L = 16$	140
4.14 Thresholded power estimation at -95dBm and 5 by 5 filter kernel applied, $F_L = 16$	140

4.15	Thresholded power estimation at -96dBm and 5 by 5 filter kernel applied, $F_L = 16$ and 13 respectively	141
4.16	Thresholded power estimation at -97dBm and 5 by 5 filter kernel applied, $F_L = 16$ and 13 respectively	142
4.17	Power Spectrum Density of a candidate frequency band, highlighting the noise floor of the observation and threshold settings investigated for filter performance.	144
5.1	PSD of 12.5/25kHz Mobile/Fixed Band, 452.5 to 477.5 MHz Bandwidth, 190Hz resolution.	152
5.2	PSD of 3G/4G Mobile Band, from 867.5 to 892.5 MHz Bandwidth, 190Hz resolution.	153
5.3	Example Duty Cycle Plot	155
5.4	Process Samples package and Analysis Tools flow diagrams	156
5.5	Timescale Bound Detectability of Whitespace	158
5.6	Resulting spectrum with FFT depth set to 2048 bits (2kb)	162
5.7	Resulting spectrum with FFT depth set to 4096 bits (4kb)	163
5.8	Resulting spectrum with FFT depth set to 8192 bits (8kb)	163
5.9	Resulting spectrum with FFT depth set to 16384 bits (16kb)	164
5.10	Resulting spectrum with FFT depth set to 32768 bits (32kb)	164
5.11	Resulting spectrum with FFT depth set to 65536 bits (64kb)	165
5.12	Resulting spectrum with FFT depth set to 131072 bits (128kb)	165
5.13	FFT Depth vs Discovered Whitespace	166
5.14	FFT Bandwidth Resolution Loss	168
5.15	Single Dimension Analysis (Slicing) of \mathbf{X}	170
5.16	MWIS Partitioning of Spectrum Occupancy Grid \mathbf{X}	173
5.17	Flowchart of Partitioning Algorithm Main Functions	175

5.18	Recording Structure of WSRB Cluster	176
5.19	A Detailed Look at Packing Algorithm Performance	178
5.20	Spectrum Occupancy for 3G/4G Mobile Band, 880MHz centre frequency, 20MHz bandwidth (effective), 5 minute observation period	182
5.21	Whitespace Opportunity Distribution for 3G/4G Mobile Band, 880MHz centre frequency, 20MHz bandwidth (effective), 5 minute observation period	182
5.22	Spectrum Occupancy for Land Mobile/Fixed 12.5/25kHz Band, 465MHz centre frequency, 20MHz bandwidth (effective), 5 minute observation period	183
5.23	Whitespace Opportunity Distribution for Land Mobile/Fixed 12.5/25kHz Band, 465MHz centre frequency, 20MHz bandwidth (effective), 5 minute observation period	183
5.24	Whitespace Opportunity Distribution for Land Mobile/Fixed 12.5/25kHz Band, 465MHz centre frequency, 20MHz bandwidth (effective), 1 hour ob- servation period	185
6.1	Modified Block Diagram of the LSA Architecture from Figure 2.13	192
6.2	Modified Block Diagram of the SAS Architecture from Figure 2.14	192
6.3	System Domain Diagram for the ‘Regional Controller’, ‘Local Controller’ and Interfaces	198
6.4	System Diagram Detailing Control Groups and Regional Service Areas . . .	202
6.5	Diagram of the B2TF User and Control Plane Signaling	206
6.6	Diagram of the B2TF Protocol Analog Processing and Decode Stages	207
6.7	Carrier Aggregation with Switched Balancing	211
6.8	Carrier Aggregation with Simultaneous Balancing	211
6.9	Single Channel Duplex Optimisation	212
6.10	Spatial Optimisation	212
6.11	Phased Array SCs within Base Station Cell	213

Listings

4.1	Spectrum Measurement Configuration File	131
5.1	Partitioning Algorithm Output File	177

Acronyms

3GPP The 3rd Generation Partnership Project.	CN Core Network.
ACK Acknowledged.	CP Cyclic Prefix.
ACMA Australian Communications and Media Authority.	CPU Central Processing Unit.
ADC Analog to Digital Converter.	CR Cognitive Radio.
AF Amplitude Flatness.	CSMA Carrier Sense Multiple Access.
AFE Analog Front End.	D2D Device to Device.
AP Access Point.	DAC Digital to Analog Converter.
API Application Programming Interface.	DBP Duration Bandwidth Product.
AWGN Additive White Gaussian Noise.	DC Direct Current.
B2TF Back To The Future (Protocol).	DDR Double Data Rate.
BHW Blackmann-Harris Window.	DFE Digital Front End.
BS Base Station.	DFT Discrete Fourier Transform.
BW Bandwidth.	DL Downlink.
CA Carrier Aggregation.	DLL Data Link Layer.
CAGR Compound Annual Growth Rate.	DSP Digital Signal Processing.
CBRS Citizens Broadband Radio Service.	DVB Digital Video Broadcasting.
CBSD Citizens Broadband Radio Service Device.	ECC Electronic Communications Committee.
CDF Cumulative Density Function.	ED Energy Detection.
CEPT European Conference of Postal and Telecommunications Administrations.	eNB Evolved Node-B.
CMOS Composite Metal Oxide Semiconductor.	ETSI European Telecommunications Standards Institute.
	FCC Federal Communications Commission.

FDD Frequency Division Duplex.	LC Local Controller.
FEM Front End Module.	LCCG Local Controller Control Group.
FFT Fast Fourier Transform.	LLC Logical Link Control.
FPGA Field Programmable Gate Array.	LNA Low Noise Amplifier.
FPR False Positive Rate.	LO Local Oscillator.
FSDR Full Scale Dynamic Range.	LSA Licensed Shared Access.
GAA General Authorised Access.	LTE Long Term Evolution.
GB Gigabyte.	M2M Machine to Machine.
GPS Global Positioning System.	MAC Medium Access Control (Layer).
GPU Graphics Processing Unit.	MAN Metropolitan Area Network.
GUI Graphical User Interface.	MB Megabyte.
HDD Hard Disk Drive.	MDS Minimum Detectable Signal.
HSDR High Definition Software Defined Radio.	MEMS Micro Electro-Mechanical Systems.
HSPA High Speed Packet Access.	MIFR Master International Frequency Register.
IC Integrated Circuit.	MIMO Multiple Input Multiple Output.
ICT Information and Communication Technologies.	MNO Mobile Network Operator.
IEEE Institute of Electrical and Electronics Engineers.	MOSFET Metal Oxide Semiconductor Field-Effect Transistor.
IF Immediate Frequency.	MS Megasamples.
IFFT Inverse Fast Fourier Transform.	MWIS Maximum Weighted Independent Set.
IMT International Mobile Telecommunications.	NAK Not Acknowledged.
IoT Internet of Things.	NF Noise Figure.
IQ In-Phase and Quadrature.	OA&M Operations, Administration and Management.
ISI Inter-Symbol Interference.	OC Overlap Correlation.
ISM Industrial Scientific and Medical.	OFCOM the Office of Communications.
ITU International Telecommunications Union.	OFDM Orthogonal Frequency Division Multiplexing.
IWCMC International Wireless Communications and Mobile Computing Conference.	OSI Open Systems Interconnect.
LAN Local Area Network.	OWD One Way Delay.

PAL Priority Access License.	SMA SubMiniature version A.
PCIe Peripheral Component Interconnect Express.	SNR Signal to Noise Ratio.
PF Power Flatness.	SoC System on a Chip.
PHY Physical (Layer).	SSD Solid State Drive.
PPB Ping Pong Buffer.	SU Secondary User.
PSD Power Spectrum Density.	TB Terrabyte.
PSK Phase Shift Keying.	TDD Time Division Duplex.
PU Primary User.	TPR True Positive Rate.
QAM Quadrature Amplitude Multiplexing.	TS Timescale.
QPSK Quadrature Phase Shift Keying.	TSG Technical Specification Groups.
RAID Redundant Array of Inexpensive Disks.	TSMC Taiwan Semiconductor Manufacturing Company.
RB Resource Block.	TV Television.
RBW Resolution Bandwidth.	TVWS Television WhiteSpace.
RC Regional Controller.	UE User Equipment.
RCSA Regional Controller Service Area.	UHD USRP Hardware Driver.
RE Resource Elements.	UL Uplink.
REM Radio Environment Map.	USB Universal Serial Bus.
RF Radio Frequency.	USRP Universal Software Radio Peripheral.
ROC Receiver Operating Characteristic.	WCDMA Wideband Code Division Multiple Access.
RRC Radio Resource Controller.	WG Working Groups.
RRM Radio Resource Management.	WOD Whitespace Opportunity Distribution.
RSPG Radio Spectrum Policy Group.	WPMC Wireless Personal Mobile Communications Conference.
SAB Services Ancillary to Broadcasting.	WRAN Wireless Regional Area Network.
SAP Services Ancillary to Programme.	WS WhiteSpace.
SAS Spectrum Access System.	WSRB WhiteSpace Resource Block.
SC Sub-Cell.	
SDR Software Defined Radio.	
SLME Service Level Management Entity.	

Chapter 1

Introduction

The increased consumer appetite for always-on wireless internet connectivity, to access image and video heavy social media, high definition video streaming and real time applications have had a profound impact on global societies and economies. Adding to this demand, Internet of Things (IoT) devices and the growing number of data collection services fueling artificial intelligence algorithms have also been growing at an aggressive rate. Other applications dependent on high speed, wireless, always-on internet connectivity include self driving vehicles, Machine to Machine (M2M) communications, security systems, environmental sensing and measurement, and financial transaction systems. These numerous applications are increasing strain on conventional wireless communications services every day and by association, the limited available electromagnetic spectrum allocated to support these communications, which has given rise to significant commercial competition for access and control of this ‘finite’ strategic resource.

There are good reasons for exclusively owned frequency allocations of spectrum for commercial, industrial, government and research purposes. However, much of this exclusively assigned spectrum is not leveraged to its fullest potential, and in particular, the spectrum below 6GHz that is significant interest and utility for consumer mobile devices

and Mobile Network Operators (MNOs) [1, 2]. The legacy models and technological approaches toward spectrum management has lead to inefficiencies resulting in potential over-allocations of spectrum in some bands, with heavy congestion in other bands [3–5].

The recent development of reconfigurable and highly flexible software defined radios and so called cognitive radios, have opened up a new way of approaching spectrum management. Rather than taking the broad sweeping inefficient (in some cases) and arguably ‘anti-competitive’ approach of exclusively allocating chunks of spectrum to an individual user or company, a more dynamic approach to spectrum could be taken. This approach would involve a network of radios that can sense their radio environment, make intelligent decisions when accessing spectrum and coordinate with local datastores for spectrum usage data, to ensure that their activity will not negatively impact any other ‘less sophisticated’ users of the spectrum. This approach could enable a new generation of mobile devices to capture all the possible unused spectrum within its local area, compensating for existing spectrum assignments while significantly increasing the overall utilisation of the spectrum.

Such a utopian ideal for spectrum management is not yet possible, however significant international efforts are being made in this area to make it a reality. As part of these efforts, this body of work develops numerous novel solutions that were awarded international patents, providing enhancements to existing and proposed secondary spectrum management systems for wireless communications.

During the development of these solutions it was determined that there exists a fundamental gap within the communication of secondary spectrum investigations that has been largely ignored by academia and industry. This gap is the lack of a formal and intuitive classification method for assessing the viability of a candidate spectrum band and, in particular, the secondary spectrum whitespace that it contains.

This work addresses the issue of a formal classification and evaluation method by

introducing the notion of a Whitespace Opportunity Distribution (WOD). This evaluation tool leverages a novel spectrum partitioning algorithm that quantifies secondary spectrum whitespace opportunities as a function of the timescales and bandwidths of the opportunities in a band of interest. This information is then aggregated in an intuitive yet powerful manner for assessing the viability of a candidate spectrum band. This tool presents a significant leap in how secondary spectrum whitespace can be communicated, evaluated and thus, exploited.

The end objective of this work is to better inform device ¹ and secondary spectrum system requirements for operation in this new spectrum paradigm, enabled by a greater understanding of the secondary spectrum environment.

1.1 Summary

The research conducted over the course of this thesis integrates two key motivations. The first motivation pertaining to the quantification of secondary spectrum and optimal timescales ² for spectrum sharing. The second motivation pertaining to exploiting the insight gained through the investigation of secondary spectrum and optimal timescales through a research partnership with the Intel Corporation focused on developing secondary spectrum focused 5th generation mobile communications technologies.

¹Unless otherwise defined, a ‘device’ or ‘devices’ are either typical User Equipment (UE) such as: Mobile Handsets, Fixed Wireless Systems (i.e. desktop computer, sensors etc), Mobile Wireless Systems (i.e. laptop, USB cellular modem, vehicle etc) or supporting radio infrastructure components including Access Points and Base Station (BS), operating within a secondary spectrum context.

²Unless otherwise specified, any reference to ‘timescales’ within this document, is explicitly referring to the duration of a, or sequence of, secondary spectrum opportunities.

1.1.1 Quantifying Secondary Spectrum and Optimal Timescales

The primary investigation evolved into a series of investigations as familiarity with the topic increased and gaps in the literature or current state of the art were discovered. These investigations are captured by the three core chapters in this thesis.

The first investigation in Chapter 3 investigates the concept of an ‘optimal timescale’ for communications within a secondary spectrum context. The focus of this investigation is to explore the relationship between latency and power consumption within mobile handsets. This relationship is used as a baseline for determining the impacts and constraints on device performance under a hypothetical opportunistic secondary spectrum access scenario.

Upon conclusion of the findings from the initial investigation it is identified that in order to determine the optimality of timescales, it is fundamental to identify what spectrum ‘looked like’ from a ‘resource’ perspective to identify what timescales could be supported by the spectrum in question. Specifically, the ‘resources’ we were interested in were the time and frequency of allocated spectrum that were occupied or unoccupied by incumbent activity.

To find these resources, the development of a custom spectrum measurement and analysis system is conducted in Chapter 4. This system is designed to capture high resolution spectrum measurements in realtime, lasting days but measured in microseconds. The development of this system is necessitated by the lack of a cost-effective spectrum measurement solution with the resolutions, observation periods, and required flexibility.

The completion of the measurement system generated large datasets and a method of processing and identifying secondary spectrum opportunities in a computationally efficient manner is devised and explored in Chapter 5. This information was then represented in an innovative and informative method by use of a partitioning method to group chunks of contiguous time and frequency located secondary spectrum whitespace, and then display-

ing the data as a set of cumulative distributions and a two dimensional visual clustering of the opportunities. This analysis technique is termed a ‘Whitespace Opportunity Distribution’ enables rapid interpretation and understanding of the distribution of secondary spectrum resources.

The whitespace opportunity distribution analysis technique is designed to inform device requirements for future secondary spectrum access systems and to enable superior modeling of secondary spectrum opportunities from both a time and frequency dimension. This method of displaying information has the added impact of providing an ‘apples to apples’ evaluation of spectrum resources, enabling the rapid assessment and comparison of spectrum quality. Finally, this system and analysis technique enabled a new understanding of what an optimal timescale for spectrum sharing is, and how this is of benefit to the design of secondary spectrum systems.

1.1.2 Intel Project

This work was additionally motivated by a research partnership formed with the Intel Corporation and termed ‘The Intel Project’. This collaboration was conducted over the first two and a half years of the investigation. The research interests of the Intel Project were focused upon the development of fifth generation (5G) mobile communications technologies. Specifically, the project interests were within the areas of managed spectrum sharing systems: Licensed Shared Access (LSA) standardisation activities in the EU and the Spectrum Access System (SAS) standardisation activities in the USA. This research in tandem with the research objective of the overarching thesis topic saw the development and submission of numerous international patent filings in the area of managed secondary spectrum systems, of which most have been awarded and are discussed in length in Chapter 6.

1.2 Power, Latency and Cost of Secondary Spectrum Access

The study of timescales lead into analysing the current limit of state of the art communication protocols and consumer hardware to determine the best possible case, for the shortest timescale that could feasibly be achieved for secondary spectrum communications. As will be presented in chapter 3, it is found that timescales are particularly limited to the latency penalties incurred by attempting to access spectrum. Access latency is found to be dependent on the protocols and specific hardware implementations used. There is a connection between latency and power/energy consumption as well. A high-power low-latency system can have a performance benefit over a low-power high-latency system, with respect the amount of energy expended per unit of information (bit), this relationship is further emphasised under a highly dynamic secondary spectrum environment.

In Chapter 3 the relationship of power consumption and latency is used to form a limit when determining the optimal timescale for gaining access to spectrum, and hence providing insight to the feasibility from an energy consumption perspective of spectrum sharing in highly dynamic environments for mobile devices. These findings were published and presented at the 2014 WPMC Conference in Sydney, Australia, titled: “Power and Latency Limitations in Secondary Spectrum Reuse for Mobile and Home Wireless Systems” [6]. This paper studied the impact of OFDM receiver architectures and the latency of operation incurred at each fundamental signal processing stage with the associated amount of required power to decode more information from the signals it receives. The latency vs power metric is then tied to short timescale secondary spectrum (whitespace) opportunities to formulate an energy efficient decision for accessing spectrum. This decision aims to capitalise on detected whitespace opportunities that given the cost of accessing the spectrum and the duration of the whitespace opportunity, a return on energy

required to access the opportunity can be determined.

1.3 A High Resolution Spectrum Measurement System

To determine and better define an optimal timescale, or many optimal timescales for secondary spectrum access, observations of real spectrum environments would need to be performed under different spectrum use cases. The timescales at which secondary spectrum 'exists' within given bands needed to be identified to determine how fast a secondary spectrum device would need to react to a secondary spectrum opportunity without any a-priori knowledge in order to capitalise upon it. Observation of spectrum is also required to determine if any particular timescales that may result in an abundance of available secondary spectrum exist as well as identification of other interesting relationships and models that could be derived from the data.

To perform spectrum observations, a spectrum analyser is required to capture, store and subsequently interrogate spectrum measurements to determine how secondary spectrum evolved over time. The spectrum measurements needed to be at a sufficiently high resolution, with good dynamic range and sensitivity, to identify any potential incumbent activity within the measured bands of interest. The television whitespace (TVWS) standards were used as a baseline for the required sensitivities for identifying incumbent activity. The system developed in a cost constrained manner, and is able to perform measurements reliably for several days at a time.

The platform is constructed with solution specific software combined with commercial off the shelf software defined radios for measurement, a custom built storage and analysis computer and the use of a graphics processing unit to perform the eventual analysis. The decision to develop the measurement system is the combination of cost, capabilities

and interest. This sub-project represented a reasonable amount of time dedicated to development, optimisation and validation. The details of this sub-project are captured within Chapter 4.

The platform enabled the necessary measurement resolution and sensitivity metrics for spectrum observation with the added convenience of high computational efficiency. This platform provided rapid iteration and learning from the captured data, enabling a measurement campaign to be conducted that captured hundreds of terrabytes of spectrum measurement data, across various frequency bands of interest and times of the day. These datasets were subsequently analysed and formed the basis of the research conducted in Chapter 5.

1.4 Quantifying Secondary Spectrum Resources

To take advantage of a highly dynamic secondary spectrum environment with low coordination, very short timescales and very high resolution bandwidths would be required to detect and subsequently maximise the utilisation of all secondary spectrum opportunities. It stands to reason that a time and bandwidth metric would be beneficial for evaluating secondary spectrum opportunities, providing an ‘apples-to-apples’ comparison of different bands and use-cases/scenarios, by purely representing the range of opportunities that exist. These metrics will also reduce evaluation time of candidate bands and assist in informing design requirements for agile radios and secondary spectrum devices.

High resolution time and frequency measurements generate enormous amounts of multidimensional data, posing a significant challenge in interpreting the dataset in a tractable and informative manner. This has led to several attempts at displaying measurement data in an informative manner, however the only metrics that are typically conveyed are relatively uninformative duty cycle, or spectrum occupancy plots [7]. Arguably, these

simple spectrum data representations are due to complexities in both capturing and displaying large datasets. Unfortunately, these methods are insufficient to represent the timescales of secondary spectrum resources available within an observation window and are simply ‘ballpark measures’, that are not useful in informing device requirements.

To address the lack of comprehensive and generally accepted secondary spectrum resource analysis methods for evaluating short timescale whitespace opportunities, we developed a novel method for quantifying the availability of spectrum resources within an observed secondary spectrum environment. This method we refer to as a ‘Whitespace Opportunity Distribution’ (WOD) explained in depth in Chapter 5. A WOD captures an accurate approximation of discrete contiguous secondary spectrum opportunities in both time and frequency and displays these opportunities across a ‘hexbin density plot’, to adequately present the concentration and size of opportunities observed.

To achieve this, the dimensionality of spectrum measurement data is reduced by applying a known NP-Hard partitioning problem to successfully capture discrete whitespace opportunities at millisecond timescales and sub kilohertz resolutions, enabling the available spectrum resources to be quantified. This method was presented and applied to two different spectrum use case measurements, to contrast whitespace availability based on incumbent spectrum usage, at the 2017 IWCMC Conference in Valencia, Spain, titled: “Quantifying Secondary Spectrum Distributions” [8].

The capture of secondary spectrum opportunities under this method opened up several new opportunities for modeling spectrum occupancy and analysis of other spectrum occupancy features such as the fractal relationship between total secondary spectrum availability and timescale over different spectrum use cases.

1.5 Licensed Shared Access Technology

The collaboration with the Intel Corporation ran over a two and a half year period. The focus of this collaboration is within the context of the 5th generation mobile communications standardisation efforts, and in particular contributions to the development of the ‘Licenced Shared Access’ (LSA) standard. The objective of the LSA research and standardisation efforts is to achieve a complete shared spectrum solution. The objectives of this partnership coupled well with the research focus of this thesis on whitespace analysis and mobile device performance in a secondary spectrum context.

These interactions fostered important academic and professional outputs for both the Intel Corporation and Macquarie University. The project sought to bring together expertise from both industrial and research backgrounds from numerous universities across the world to shape the future technology that would become part of the 5G standardisation activities. The project saw the successful delivery of many patent filings and research papers within the secondary spectrum management space as contributions toward the greater 5G standardisation effort.

Patents of note and with connection to the greater thesis topic that were developed with the Intel Corporation have been included within this document. Each of the patents are presented with a high level overview and then explained further in depth throughout the body of the document under relevant sections. This structure is used as the individual patent submissions require the relevant background and context associated with their development to be appreciated for validation of their contribution. The works are presented here chronologically, however they do not appear in this order within the text due to the relevant background and context. While the patents themselves do not directly assist in answering the core thesis topic, they were substantial drivers toward the motivation of the topic and the final direction taken.

1.5.1 Signal Buffering for Licensed Shared Access (LSA) Technology

Drawing on the power and latency relationships studied in the first publication and combining this approach with managing secondary spectrum in highly dynamic environments to maximise the capture of secondary spectrum, a concept for a device is developed to receive data across available secondary spectrum before that device is made aware via a chain of trust that the data received is intended for it [9]. The primary function of the invention is to maximise the exploitation of temporal whitespace opportunities, reducing the overall required time of exposure to the whitespace channel for a given payload.

As a general rule, the less time that a whitespace opportunity is required to be utilised, the less risk that the particular communication is exposed to where an incumbent could retake the spectrum. Thus there is motivation to develop a system that simultaneously transmits a payload of interest during control plane signaling within a device to device communications context, to offset the latency 'penalty' incurred by the control signaling. The necessity of avoiding this penalty and the benefits of doing so are covered in Section 6.4.

1.5.2 Evolved Node-B, Local Controller and Method for Allocation of Spectrum for Secondary Usage

The research into the LSA and later SAS system revealed some important issues that the existing standardisation efforts had not addressed sufficiently. The first issue is how to handle LSA spectrum allocation to a multitude of MNO's including the case of shared Base Stations. LSA spectrum is only assigned to a single MNO through a binary sharing agreement, without considering potential multi-party scenarios. The second issue is how to anonymise LSA spectrum allocations through distributed spectrum management if the

multi-party scenario existed. This is important to address as MNO's and Base-Station (BS) operators may not want to expose specifics of their network activity to competitors or third parties.

The solution devised, split the spectrum management domains between the LSA operator and the MNO's network components attached to the LSA network [10]. This is to be achieved through anonymised data passing interfaces for sharing agreement data, with the core decision making for spectrum allocation (i.e. who is the successful operator at a point in time to be granted LSA spectrum) being completely hidden from the LSA network. The core decision making entity (termed the Local Controller (LC)) is responsible for enforcing sharing agreements within its region of operation, but would frequently check with a centralised policy entity to ensure that the decisions being made are correctly conforming to the policies pertinent to that LCs region of operation. Section 6.3 delves further into this concept and how anonymity is assured for the MNO's from each other and the greater LSA network, while still facilitating efficient secondary spectrum management for all interested parties.

1.5.3 Methods for Performing Mobile Communications Between Mobile Terminal Devices, Base Stations and Network Control Devices

LSA and SAS systems seek to facilitate the availability of additional electromagnetic spectrum, which enable MNO's to perform Long Term Evolution (LTE) (i.e 4G and beyond) based Carrier Aggregation (CA) between spectrum owned by the MNO and the secondary spectrum to provide greater amounts of bandwidth for data communications. In Australia, Europe and the USA, the majority of LTE channels used are based on Frequency Division Duplex (FDD), with handsets designed for the Chinese market being

predominantly based on Time Division Duplex (TDD) duplexing methods. The bands that LSA and SAS will be operating within fall directly within bands specified under the LTE standard as TDD bands, this would require CA of both TDD and FDD bands, which is a challenge for both device manufacturers and network operators. To exploit this diversity in duplex schemes, balancing and optimising user data payloads across the two duplex methods (and hence carrier frequencies) depending on particular user contexts [11]. The solution explored in Section 6.5 proposes a mapping of user payload types to the different duplex methods for optimising network throughput with the unique use-case posed by LSA and SAS, and improve user quality of service by managing system latency, power consumption, operational range and throughput metrics.

1.5.4 Methods and Devices for Shared Spectrum Allocation

This work addressed another deficiency within the LSA framework, where sharing agreements are expected to be ‘private discussions’ between incumbents and MNO’s. Thus it would be likely that a precise process would become dependent on a preferred proprietary solution. It would also be likely that these sharing agreements could shift overtime/change dynamically and to facilitate these dynamic changes, the LSA architecture could be modified and enhanced by the solution.

The solution introduces a ‘Sharing Level Management Entity’ (SLME) [12] for the LSA architecture and an adaptation for the SAS architecture that could also benefit from such a capability. This entity is responsible for automated negotiation and dynamic updating of sharing agreements between incumbents and licensees, with the method examined further in Section 6.2. This solution also proposes a negotiation mechanism for both the incumbents and licensees for access to secondary spectrum, as well as establishing a framework for outlining the requirements of the sharing agreements and an announcement mechanism to inform the greater network of updates to particular sharing agreements.

1.6 Scope

Chapter 1 provides an outline of the core structure of the document acting as a roadmap for the contributions presented and how the contributions relate within the document.

Chapter 2 is a comprehensive literature collection, detailing electromagnetic spectrum and its properties, how electromagnetic spectrum is managed and utilised for modern communications systems and the methods required to perform that management effectively. The second part of the literature review covers the topic of secondary spectrum in depth, including current and future developments for managing and utilising secondary spectrum effectively.

The four primary sections in Chapter 1 map directly to the remaining four Chapters under the same name. The following summaries detail original contributions that were not part of the Intel Project:

1. Section 1.2 provides an overview of Chapter 3 with Primary contributions:
 - (a) A survey of the relationship between power consumption and latency for conventional wireless receivers
 - (b) Introducing a simple whitespace model for evaluating timescale of device operation
 - (c) Providing a measure to calculate energy efficient secondary spectrum access
 - (d) A thought experiment, of how energy consumption is impacted as a function of a secondary spectrum environment and the inherent risks of operating in that environment.

2. Section 1.3 provides an overview of Chapter 4 with Primary contributions:
 - (a) Design, development and implementation of a low cost and high resolution

spectrum measurement system

- (b) Implementation of a novel 2D image noise filter for improving energy based detection of incumbent signals

3. Section 1.4 provides an overview of Chapter 5 with Primary Contributions:

- (a) Measurement of several candidate spectrum bands
- (b) Introduction of the ‘Functional Whitespace’ concept, for evaluation of discrete whitespace opportunities
- (c) Defining a unit of whitespace (the WhiteSpace Resource Block (WSRB))
- (d) Clustering of WSRBs to increase the ‘Functional Value’ of the discrete blocks
- (e) Defining the optimal clustering of WSRBs to maximise Functional Value as an NP-Hard problem
- (f) A novel partitioning algorithm to generate clusters of WSRBs in real time from an observed spectrum
- (g) Using the partitioning result to generate a novel Whitespace Opportunity Distribution (WOD)
- (h) Demonstration of WODs as a superior method for representing and evaluating secondary spectrum whitespace over conventional means of communicating secondary spectrum whitespace

4. Section 1.5 introduces the topics of LSA and SAS, covered in depth in Chapter 2. The four Subsections in Section 1.5 provide a brief of the contributions achieved and map directly to Chapter 6 under the Sections of the same name. The primary contributions in this chapter were achieved as part of the Intel Project collaboration:

- (a) Section 6.2 details extensions to the LSA and SAS systems are to facilitate a method of dynamic renegotiation of secondary spectrum agreements and how those agreements are defined.

- (b) Section 6.3 details a further extension to the LSA and SAS systems by providing a mechanism and framework to separate the control domains of the LSA and SAS system to encourage both sharing of physical infrastructure and to prevent private network configuration information from being exposed to the greater system.

- (c) Section 6.4 develops a novel concept of capturing secondary spectrum opportunities before a formal control and connection set up sequence has to be performed, capitalising on secondary spectrum opportunities that would otherwise have been unable to be captured.

- (d) Section 6.5 examines the merits of Time Division and Frequency Division Duplex schemes for mobile communications and details a method for determining the optimal selection of either of these duplex schemes given a user context. Duplex optimisation forms an important component of the success of secondary spectrum management systems such as LSA and SAS.

Chapter 2

Background and Related Work

2.1 Introduction

The electromagnetic spectrum or simply, ‘spectrum’, presents significant scientific, social, defence and economic benefits for modern society [13]. With the constant evolution of technological development from smartphones and wireless multimedia systems, the finite spectrum resources which were once considered abundant are now heavily stressed in the bands designated for these services [3–5]. This increased demand for the constrained spectrum resources calls for efficient management processes to be designed and implemented, particularly with regards to the primary occupants and services within a given spectrum band. All aspects of spectrum management need to be explored to achieve this goal. In particular, it is critical to optimise the spectral efficiency of primary spectrum occupants and subsequently, develop efficient strategies to harness unused or underutilised spectrum.

This chapter seeks to build a fundamental understanding into the broad topic of electromagnetic spectrum utilisation and management, current driving forces for spectrum strain, the limitations that exist with currently implemented management strategies, and forward facing solutions to further increase the efficiency of spectrum management.

Section 2.2 provides a background into electromagnetic spectrum properties, specifically the dimensions in which spectrum can be utilised. This section also introduces the concept of electromagnetic spectrum as a ‘finite’ resource, providing the motivation for harnessing secondary spectrum. Section 2.3 investigates the governing bodies responsible for spectrum management and spectrum policy and key standards bodies responsible for developing communication device requirements both domestically and globally.

Section 2.4 provides a foundation into the architecture of wireless devices, and methods used to access electromagnetic spectrum for communication. This section primarily provides a base for the power consumption and latency investigation in Chapter 3 and several contributions in Chapter 6.

Section 2.5 opens up the topic of secondary spectrum, how it can be harnessed, and the current state of the art to address congestion and inefficient spectrum usage. The Section formulates key motivations for much of the research and patent development in this document.

Section 2.6 introduces Software Defined Radios (SDR) and the impact they have had on developing solutions for harnessing secondary spectrum. The capabilities and core functions of SDRs are presented as well as sensing techniques used for the observation and classification of spectrum, forming an important basis for Chapter 4 and 6.

2.2 Electromagnetic Spectrum Properties

The concept of spectrum being ‘finite’ requires a general understanding of how communication systems access and utilise spectrum as well as the intrinsic physical properties of electromagnetic spectrum. Electromagnetic spectrum is temporally renewable; if at one instance in time a ‘piece’ of spectrum is ‘consumed’, every instance after that the spectrum will still be available to be consumed again, whether by the same user or some

other user or users.

Application performance is highly dependent on the intrinsic physical properties of spectrum, as the signal propagation characteristics of spectrum varies significantly depending on frequency. Due to the way in which certain systems can and are engineered, there is a finite supply of ‘suitable’ spectrum for these applications. The desirable propagation characteristics combined with finite supply of frequencies, gives rise to certain frequencies being highly prized and protected.

Thus the dichotomy of spectrum being infinite and simultaneously finite exists, where the hard limit of spectrum ‘capacity’ is due to how effectively the spectrum is managed and the technology that harnesses it.

2.2.1 Electromagnetic Spectrum Properties

In communications theory, the the Shannon-Hartley theorem describes the maximum rate at which information can be transmitted over a given bandwidth with a given signal to noise ratio of the received signal. (C) is given as the channel capacity in bits/Hz, (B) is the bandwidth of a channel, (S) is the signal power in Watts and (N_0) is the channel noise in Watts. The theorem to determine capacity is given by:

$$C = B \log_2 \left(1 + \frac{S}{N_0} \right) \quad (2.1)$$

Electromagnetic radiation models used in communications are subject to a phenomena known as path-loss [14]. Path-loss is defined as the amount of power that a signal attenuates by (L) over a given distance (d) between a transmitter and receiver at a given wavelength (λ) (given by $\lambda = \frac{v}{f}$ where (f) is frequency and (v) is the speed of the electromagnetic wave, in the free space example this is the speed of light (c)). To highlight the frequency dependency and for ‘ease of use’ the free space path-loss formulation represented in ‘log domain units’ (dBmW) of loss from the transmitted power (P_{Tx}) to received

power (P_{Rx}) for two ideal isotropic antennas in free space [15]:

$$\frac{P_{Rx}}{P_{Tx}} = 20 \log_{10} \left(\frac{4 \pi d f}{c} \right) \quad (2.2)$$

Adjusting the formula to units of kilometres (km) instead of m and frequency in megahertz (MHz) instead of Hz, this can be further be simplified to [15]:

$$P_{Rx} = P_{Tx} - (20 \log_{10}(d) + 20 \log_{10}(f) + 32.45) \quad (2.3)$$

This relationship between signal frequency and path-loss yields that lower frequency signals are less susceptible to path-loss over extended distances and are thus preferable for applications that require the transmitter and receiver to be considerable distances apart from each other.

Typical practical path-loss models will not simply assume free space for propagation and will introduce additional source of loss or ‘fade’ to the model. These models introduce a stochastic random variable to account for additional loss within the wireless channel. The most common fading ‘log-distance fading’ is used to account for other properties and randomness within a given channel.

In the ‘log-distance’ model [15], the path-loss at a reference distance is measured and then the path-loss at the desired distance is modeled based on that reference distance, a fixed quantity γ (the path-loss exponent) is accounted for with the distance portion of the calculation and a zero mean Gaussian random variable with zero mean (X_σ) is used to model the ‘flat fading’ (frequency independent constant signal loss), within the channel.

$$P_{Rx} = P_{Tx} - (10 \gamma \log_{10}(d) + 20 \log_{10}(f) + 32.45 + X_\sigma) \quad (2.4)$$

It can be seen from equations 2.1 and 2.2 that there is a relationship between the distance that a transmitter and receiver can be in space relative to each other for a

given transmission power with respect to the data rate that can be expected between the two devices. However, this is not the complete picture for the frequency dependency of propagation characteristics. The environments and other use cases such as movement in which the vast majority of consumer mobile devices operate is quite hostile for signal propagation, and is unaccounted by simple fading models.

‘Rayleigh Fading’ is a statistical model that aims to encompass the effects of an environment (urban/rural) and the velocity at which a transmitter is traveling on the propagation of radio signals. This model takes into account ‘scattering’ and the resulting ‘multipath’ caused by objects in an environment blocking the ‘line of sight’ between a transmitter and receiver. Many other fading models exist to account for additional sources of loss and randomness due to a devices operating environment and use case, these are thoroughly investigated in [15].

2.2.2 Dimensions and Constraints of Spectrum Usage

Equations 2.1 and 2.2 shows that there are three primary physical dimensions for which spectrum can be managed across, these are: time, frequency (or bandwidth) and space. Thus it is useful to gain an appreciation for how each of these dimensions in isolation can be managed or exploited to maximise achievable spectrum capacity.

Here we refer to both diversity and multiplexing as ways of explaining how spectrum can be better exploited for radio communications. Diversity and multiplexing are somewhat the inverse of each other, where a diversity scheme aims to trade off capacity to increase reliability, and a multiplexing scheme aims to maximise capacity with less regard to reliability of signal reception [7, 14].

A diversity scheme is a method employed to improve the reliability of a communications signal by routing that signal over two or more channels with different characteristics, such as fading, geography, frequency and interference. The same signal is essentially sent

multiple times in different ways with the aim of ensuring the successful reception of the signal by the intended receiver. These methods are typically employed in hostile spectrum environments where the capacity needs to be traded off for reliability. For example, multipath (the same signal taking many different spatial routes) can be exploited to improve the gain of that signal at a receiver.

Similarly, a multiplexing scheme seeks to exploit channel characteristics to improve the amount of data that can be simultaneously transmitted. For example, a transmitter may have knowledge of two channels that are the same in frequency but are separated in space, the transmitter could transmit a separate data stream in each physical direction, using a directional antenna to effectively double the information capacity of the spectrum at that frequency. As the data streams are transmitted in different directions, there is little chance that they will interfere with each other thus successfully multiplexing the spectrum.

2.2.3 Time and Frequency

It is useful for spectrum management to deal with time and frequency in discretised quantities [16]. An example of the discretisation of time and frequency is illustrated in fig 2.1, where certain users of the spectrum are constrained to different frequency bands and either access the spectrum for very long durations, or intermittent periods. In the case where usage of frequency bands are intermittent, there presents an opportunity for utilising that unused spectrum for additional purposes.

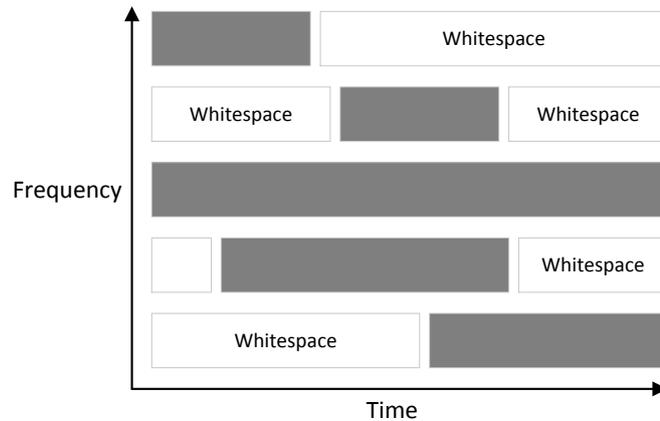


Figure 2.1: Example spectrum occupancy diagram. Incumbent activity depicted by grey bars and white bars depicting ‘Whitespace’ or unused allocated spectrum.

2.2.4 Space

The space dimension for spectrum allocation is becoming a point of increasing importance with respect to the current development of 5th generation mobile communication technologies (5G). To provide a brief and somewhat frivolous example: if every user had a discrete unit of space allocated just to themselves, then the entirety of the electromagnetic spectrum would be available to them, provided they did not interfere with neighbouring units of space. Due to the propagation of electromagnetic waves, the difficulty in controlling these propagations, the overwhelmingly complex infrastructural and cost requirements to service each unit of space adequately, and the simple use case of a user moving around to other units of space, spectrum must be shared with over areas of space and managed accordingly for multiple users [16].

‘Geographic multiplexing’ is the process of reusing spectrum in sufficiently separated geographic locations, where there is little to no interference from communications between region A and region B. These regions for example could be as between separate rooms within a building (provided that the interference between the two rooms is managed and

signal powers are limited), different sides of a large structure, or segments of a highway. Geographic multiplexing enables users to move around to different areas while still being connected to a mobile network and additionally enables more users to be served by the same allocation of spectrum.

Figure 2.2 illustrates a basic example of two separate mobile networks (network A and B), that are using the same frequency at the same time. The dotted area represents the effective range of the basestation. Transmitter A will not interfere with transmitter B and can thus access the same frequency and time dimension resources as transmitter B. This effectively doubles the capacity of the entire network when compared to a network with just a single much farther reaching transmitter and this concept can be scaled down much further to smaller network sizes and even take advantage of the physical environment to increase overall spectrum utilisation.

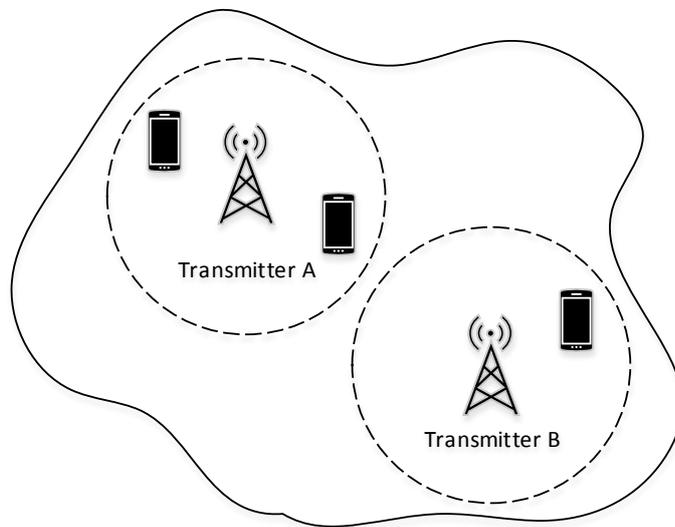


Figure 2.2: Example of geographical reuse of the same frequency band. Transmitter A can communicate simultaneously at the same frequency as Transmitter B without interfering with each other due to geographical separation

Recent developments in small cell technology and heterogeneous networks expand on the the concept of allocating a unique space to each user to maximise available spectrum resources to each user. By decreasing the size of a region, and increasing the density of basestation deployments, effectively, many more users can be served at higher datarates than under a single basestation serving all of those regions. This concept known as ‘densification’ is one of the key enablers for 5G systems.

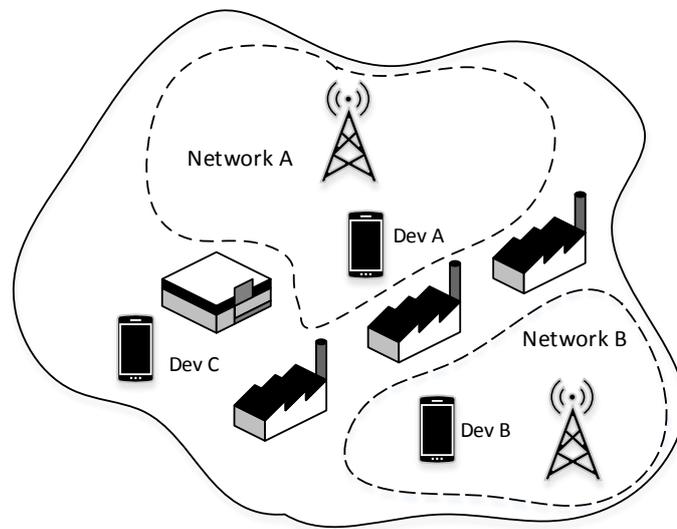


Figure 2.3: Depiction of geographic multiplexing by taking advantage of physical entities to separate networks

Figure 2.3 illustrates how the physical environment can impact upon the spatial reuse or ‘geographic multiplexing’ of spectrum. The structures form a barrier for the electromagnetic waves, preventing them from reaching further due to high signal attenuation. While this can be inconvenient in some cases, with respect to Device C, as it is unable to communicate with either Network. The structures however do present an advantage when the intent of the network is to maximise the total amount of throughput. As Networks A and B have been deployed strategically to maximise the geographic multiplexing offered

by the environment, thus eliminating the need for Device A and B to simultaneously share spectrum resources, which would negatively impact on both their power consumption and network throughput. The subject of mobile device power consumption will be explored in depth in Chapter 3.

In addition to geographic multiplexing, the concept of ‘spatial multiplexing’ also exists. Specifically this concept is used when referring to a technology called MIMO (Multiple Input Multiple Output) [17, 18]. This technology utilises an antenna array to perform multiplexing on a relatively small physical scale, and by taking advantage of the ‘multipath’ within the environment. Several separate ‘spatial streams’ can be created due to environmental diversity between individual transmitter and receiver antennas, and thus separate channels for transmitting data can be established.

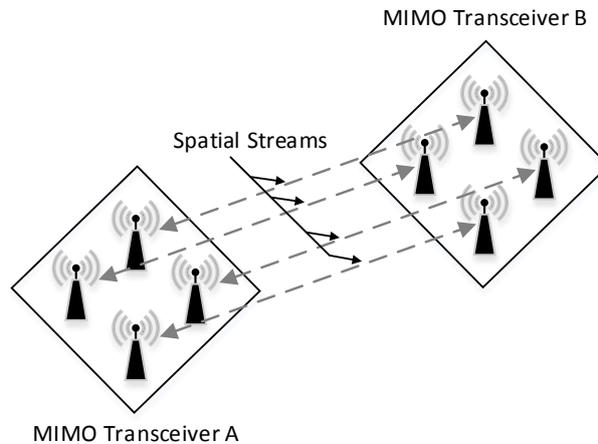


Figure 2.4: Depiction of four spatial streams between two MIMO antenna arrays

‘Multipath’ describes the multiple paths that a given or multiple electromagnetic waves can ‘travel’ from a transmitter to a receiver. These different ‘paths’ are created by the different angles and surfaces each that stream bounces off before it eventually arrives at the receiver. Often multipath becomes a source of interference, both destructive and

constructive for a given signal [19]. In environments with low diversity, typically only a low number of MIMO channels can only be supported.

2.3 Electromagnetic Spectrum Management

The Electromagnetic Spectrum is used in a large number of consumer, commercial, industrial and federal applications for observation and communications. Some particular applications for electromagnetic spectrum are as follows: mobile networks, financial trading systems, military, satellite communications, medical imaging, terrestrial and sea navigation (GPS (Global Positioning System), GLONASS, Galileo and BeiDou), exploration (Space: stars, blackholes, exoplanets and, Earth: gas, oil, water, minerals). Often these services and applications compete for spectrum due to the aforementioned physical properties of spectrum which necessitates effective management of it, to meet all the demands placed upon it.

The Electromagnetic Spectrum is treated as a finite national strategic resource that is administered and managed by the federal body of the relevant nation as it should see fit. The international standards for spectrum regulation are outlined by the International Telecommunications Union (ITU) [20], which is a specialised agency for information and communication technologies under the United Nations. The ITU is responsible for developing technical standards that ensure the interoperability of communications networks and technologies.

Globally, each (UN Member) country allocates spectrum as they best see fit with various domestic and international allocation standards and policies within the framework and standards outlined by the ITU.

The ITU has developed international radio regulations to manage the allocation of spectrum and satellite orbits, providing guidelines for the preferred services to be allocated

within particular bands of the spectrum. There are 3 distinct regions that exist under these regulations: Region 1 spans Europe, Africa, the Middle East west of the Persian gulf, the former USSR and Mongolia. Region 2 spans the Americas, Greenland, and some of the Eastern Pacific Islands. Region 3 covers the remaining areas of Asia, Iran and most of Oceania. These regulations additionally detail the mandatory operating parameters of transmitters for given frequencies, and the requirements of governments to update the ‘Master International Frequency Register’ (MIFR) [21] for spectrum assignments made in their respective jurisdictions.

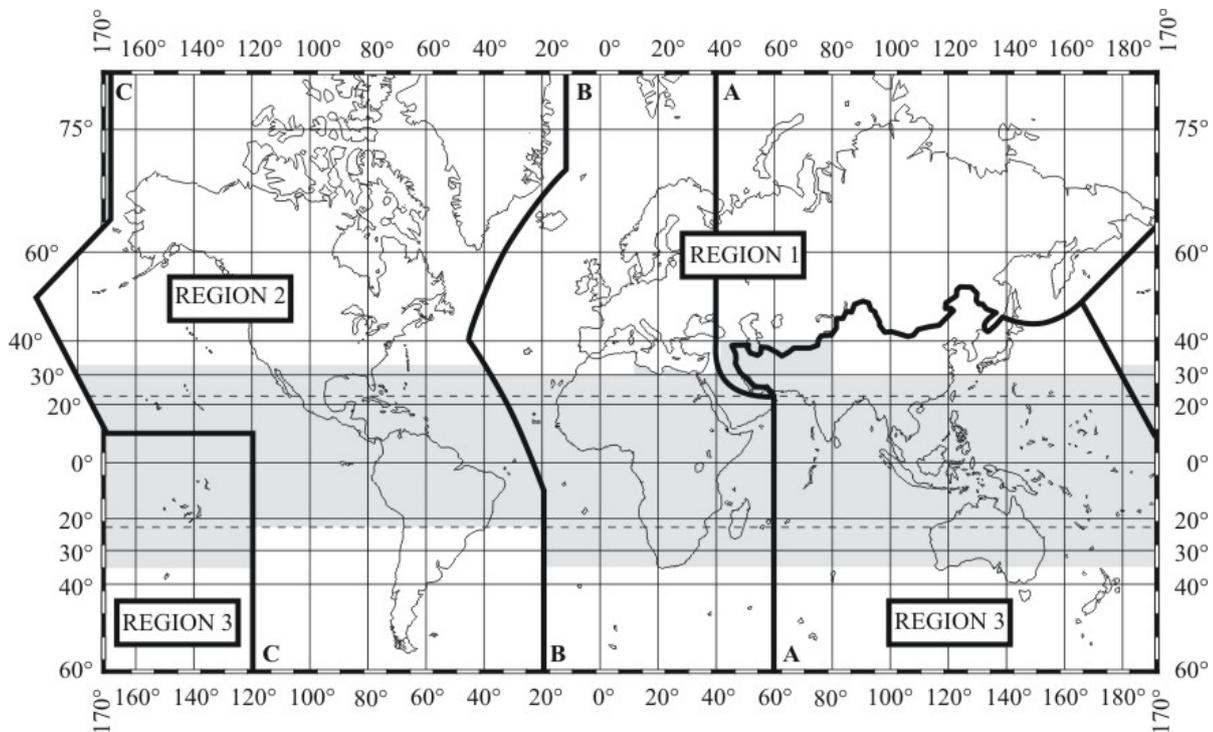


Figure 2.5: ITU Spectrum Regions [22]

Despite the ITU creating guidelines for spectrum allocation, the responsibility of discrete spectrum allocation falls to the government of a given country. Typically this is performed at the federal level by a communications regulator, for example, in the USA the FCC (Federal Communications Commission) [23], in the UK, OFCOM (the Office

of Communications) [24], in the EU significant work is being done to harmonise spectrum allocations for each member state under the EU's Spectrum Policy Framework [25], and in Australia the ACMA (Australian Communications and Media Authority) [26] is responsible for communications regulation. These bodies are responsible for administering spectrum for their respective domestic military, federal, commercial, industrial and general consumer applications.

2.3.1 Spectrum Allocation

Under the ITU Radio Regulations [27] there are four categories of spectrum assignments Primary, Secondary, Exclusive and Shared.

Primary spectrum is directly licensed to a user, users or service typically referred to as the Primary User (PU). The PU has a right to access the given spectrum and are provided with a guarantee that the expected interference within that spectrum will be at or below a certain level, as well as outlining particular services that may be used within that band and transmission requirements.

Secondary spectrum assignments are provided to enable greater spectrum usage in the absence of a primary service. Secondary assignees have no guarantee for interference and hence quality of service. They are additionally required to not interfere with the primary service of that same band and may not claim protection from the primary service, they may however claim protection from transmitters of secondary nature. Exclusive allocations are reserved for a single communications service, such as: distress, safety and emergency services, maritime navigational and meteorological warnings, space communications, aeronautical radionavigation systems and medical transports.

Shared assignments are used where a particular band can be used for two or more communication services, provided they adhere to certain interference minimisation policies.

2.3.2 Spectrum Assignment

Spectrum that is not exclusive or utilised for certain scientific, military and government purposes, is typically assigned via auctions for commercial and industrial uses [1, 2]. As it stands there is no uniform agreement for spectrum pricing, however large contiguous chunks of spectrum that are of high value (due to their functional efficiency) for mobile network operators (MNO) have commanded very large prices in recent times [28, 29].

Previous approaches to the problem of spectrum allocation was either via first come first serve over the counter licences, ‘spectrum lotteries’ where interested parties would apply and then be randomly drawn from a pool, where interested parties were evaluated for policy, societal and economic benefit. These approaches were often met with poor outcomes for either the governments, interested parties or overall spectrum productivity. Thus a more economically viable solution was converged to via spectrum auctions [28, 29].

The value of a given spectrum is dependent on the contiguous allocated frequency, the geographical area or volume that the spectrum licence is valid for, intrinsic propagation characteristics of the given frequency, the number of people that would be encompassed, the number of competitors/demand in the area and associated administrative costs. These auctions have been found to provide an effective method for a government to allocate the limited spectrum resources, with greater productivity of the spectrum being achieved and simultaneously generating revenue for governments [30].

The auction model of spectrum allocation, in Australia was adopted by the ACMA in 1992 under section 60 of the Radiocommunications Act 1992 [31].

Figure 2.6 illustrates at a high level view how a federal body depicts the assignment of spectrum resources. The allocations at these resolutions are typically for designating types of services that are able to access the various bands. Once these initial high level designations are agreed upon, the designated bands are then sliced up into chunks in bandwidth and geography or regions, to then be licensed out to services that purchase

2.3.3 Mobile Communications Spectrum

For the specific interest of mobile communications, the primary consumers of this spectrum are users connecting to the internet via devices such as mobile handsets, mobile modems, or static/fixed systems using the wireless network. These communications networks are operated and maintained by entities known as Mobile Network Operators (MNOs). These entities apply for spectrum licences from the federal government of the country in which they operate to gain exclusive access to large contiguous or several large discontinuous bands (1.4MHz to 20MHz per band) for which to provide internet connectivity and data transmission services to their customers via 4G and future 5G communications technologies.

As projected by the Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update [33] over the 2016 to 2021 horizon, there is expected that over 50 percent of global connected devices will be comprised of smartphones making up 86 percent of all mobile traffic. 20 percent of global internet protocol traffic will be created solely by mobile devices, and monthly global mobile data traffic will be approximately 49 exabytes per month by 2020. They also expect that mobile traffic will grow at a compound annual growth rate (CAGR) of 47 percent over the 2016 to 2021 horizon. These projections are also agree quite similarly to the Ericsson Mobility Report (Nov 2017) [34], with a CAGR of 42% toward to 2023 horizon. An expected monthly data traffic on the same order as the Cisco report by 2020, with an expected 110 exabytes per month by 2023.

MNOs require access to large contiguous bands of spectrum to provide their services due to the large data consumption requirements of their customers. It is foreseeable that these consumer requirements will continue to grow into the future. In fact the development of higher bandwidth and lower latency communications systems has considerably increased the amount of traffic generated by users of that technology. High bandwidth and low latency offers considerable quality of service improvements to the user, making the user

more intent on using that service. For example: waiting time for files to transfer are lower so larger files can be transferred, high definition content can be viewed in realtime and time sensitive applications such as stock trading, high definition voice communications and other financial transactions also become a possibility.

With the clear advantages for user based applications that these systems offer, there will only be a greater uptake over time of these services which will further strain existing spectrum capacity and infrastructure (fronthaul and backhaul) supporting the network. The typical fronthaul network or ‘access network’ comprises of numerous basestations (BS) organised into ‘cells’ that provide regions of connectivity for mobile users authorised to connect to that network. An access network is interested in providing the greatest amount of capacity to the users of that network to secure good quality of service for the users. Capacity can be increased in the following ways: increase power (or reduce noise), decrease the service area (to gain greater geographical efficiency) or increase the available amount of spectrum available. To address the growing user appetite for data consumption, there too must be a growth in spectral capacity to meet this demand.

A range of frequencies with good physical object penetration, and low attenuation, in the interests of minimising power consumption, are highly desirable for all mobile communication systems. It is also apparent that the range of frequencies that fit these requirements primarily exist below 6GHz. These spectral properties are also highly desirable for numerous other applications, not just mobile data networks and systems, resulting in a significant strain on frequencies below 6GHz with respect to allocation.

A key limitation of the growth in general connection speeds, is the suitability of spectrum for general consumer applications. Thus a significant interest in spectrum above 6GHz for future 5G mobile systems, as a prospective solution to meet this every growing demand has been generated [35, 36]. Due to the nature of mobile communications not requiring a user to be physically tethered to a location, a user will typically move around

buildings and the outdoor environment while using the device. This mobility aspect presents a challenge for communications engineers, especially at above 6GHz frequencies, as a device that constantly moves around its environment, presents constant variation in the physical channel between the device and basestation. Sources of variance are due to objects such as structures, bodies, cars or foliage blocking the direct line of sight (LoS) between a transmitter and a receiver. Thus a frequency that will propagate well and penetrate these objects with little loss of signal integrity and no impact to the physical object is important for the success of a mobile device, or, a means to overcome the penetration issue will need to be developed for a successful above 6GHz system [37].

Power consumption is another point of interest for communications in general, but especially with respect to mobile communications and even more so at higher frequencies due to the challenging signal environment. As mobile devices are primarily battery powered it is critical that the energy consumed to perform communications is kept to an absolute minimum, without negatively impacting on user experience. It is also of interest to keep transmission power low due to the close proximity to people that communication systems are conventionally used. These factors make high transmission powers unacceptable for mobile communication systems, so below 6GHz communication frequencies that experience low attenuation are already at a significant advantage when compared to above 6GHz systems. One of the primary advantages of above 6GHz systems, is access to very high allocations of contiguous bandwidth due to both the limited number of other users from 28GHz to 100GHz [36] and the significantly greater amount of spectrum at those frequencies. Thus, in addition to the study of above 6GHz 5G systems, investigation of efficient spectrum utilisation in sub 6GHz bands for 5G systems is also of high interest.

2.3.4 Spectrum Management Policy

A high level understanding of the interactions between the 3GPP (The 3rd Generation Partnership Project), ETSI (European Telecommunications Standards Institute) the ITU-R (ITU Radiocommunication Sector - a subgroup of the ITU), the IEEE (Institute of Electrical and Electronics Engineers), and CEPT (European Conference of Postal and Telecommunications Administrations) ECC (Electronic Communications Committee) is needed to appreciate the interactions and structure behind spectrum management and management policy within the European Union and globally. The purpose of this explanation is to demystify each bodies' responsibility and role in the context of management of spectrum and communications with respect to 4G, 5G and secondary spectrum access technologies. These governing bodies have particular importance in positioning the contributions in Chapter 6 and understanding the general policy and global framework interactions that mobile communications systems are designed toward.

The ITU-R tasks itself with: “Ensuring the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including those using satellite orbits, and to carry out studies and approve recommendations on radiocommunication matters.” They are responsible for (but not exclusively) approving the standards presented to them in accordance with existing radio regulations and regional agreements set out by the ITU-R. ITU-R also ensures that the necessary performance and quality of proposed radio systems are maintained and conservation of spectrum and more flexible methods of accessing spectrum are considered for addressing forward facing spectrum needs. The ITU-R is also responsible for issuing requirements for the IMT (International Mobile Telecommunications) standards: IMT-2000 (3G); IMT-Advanced (4G); IMT-2020 (5G) [38].

The CEPT, specifically the ECC business committee of the CEPT operates at a regional level within the EU for certain policy, coordination and harmonisation activi-

ties [39]. In addition to the ECC there are independent spectrum authorities responsible for the administration of spectrum activities in every EU member country on a per country basis. These local administrations perform a similar function in EU member countries as the ACMA in Australia, the FCC in the United States of America, and OFCOM in the United Kingdom. The responsibilities of these entities are to develop and administrate spectrum and spectrum policy, legislation and regulations for their relevant jurisdiction.

ETSI is a not for profit standardisation organisation that produces globally applicable standards to Information and Communication Technologies (ICT), for fixed, mobile, and internet technologies. Out of ETSI was created a technical working group for the 3G series of standards, which formed into the 3GPP, of which now ETSI is a partner organisation [40].

The 3rd Generation Partnership Project is a collaborative agreement between standards development organisations that develop specifications from contribution-driven member companies that are organised into Working Groups (WG) within three different Technical Specification Groups (TSG). The name of the partnership is a legacy name from when the 3G technical specifications and reports were under development. Their activities are now concerned with technical specifications and reports for 4G (LTE, LTE-A, LTE-A Pro) and 5G (5G New Radio) specifications [41].

The IEEE is a global professional engineering and standardisation organisation that tasks itself with numerous activities in the fields of engineering, computing and technology information, such as professional conferences, technology standards, highly cited publications and education. They are responsible for developing and maintaining the 802 family of standards (among many others), which in particular deal with LANs (Local Area Networks) and MANs (Metropolitan Area Networks). The particular standards of interest are IEEE 802.11 (WiFi) and IEEE 802.22 (TVWS) [42].

In summary, the ITU-R establishes what should and can be done with spectrum (i.e.

types of services, device categories and harmonisation efforts), the CEPT (FCC, ACMA, OFCOM etc) establishes what will be done with the spectrum at a federal level (i.e. explicit use cases and discrete band allocations/auctions) and ETSI, the 3GPP and the IEEE outline the systems and methods of accessing spectrum (i.e. specifications for the hardware to perform communications).

2.4 Radio Resource Management

A thorough understanding of Radio Resource Management (RRM) reveals that it is foundational to power consumption, access latency and secondary spectrum use because it provides both a logical and physical architecture for efficiently utilising spectrum as well as defining key metrics for evaluating the performance of these topics.

The field of RRM is focused on the design of systems to ubiquitously manage, access and navigate the spectrum and physical environments to perform successful communications between two or more devices, in the shortest possible time, with the lowest possible energy [43]. The objective of RRM is to maximize total system spectrum efficiency in bits per second per unit area, via PHY and MAC layer systems and algorithms to overcome and/or manage the physical environment, fading, shadowing, multipath, Doppler shift, interference, noise; and to maximise the efficiency of handover, modulation, error coding and scheduling [44, 45].

2.4.1 MAC and PHY Layers

With respect to the standardisation and development of mobile communication devices, T-REC-X.200 (1994) Section 6.1 presents the reference model for the Open Systems Interconnection (OSI) architecture [46]. It defines the seven layers of an ‘open system’, which provides a common basis for the coordination of standards development for the purpose

of systems interconnection. Additionally it allows existing standards to be placed into perspective within the overall Reference Model. The layers of this model are captured in Figure 2.7.

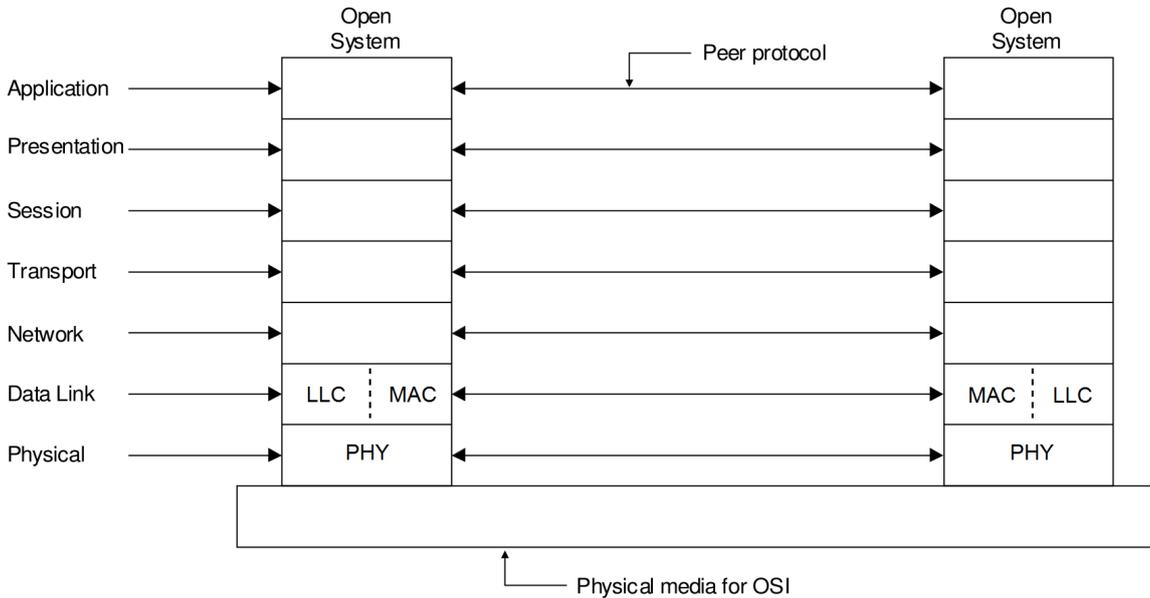


Figure 2.7: Open Systems Interconnection (OSI) architecture [46]

Under the ITU-T definition: “The Physical Layer (PHY) provides the mechanical, electrical, functional and procedural means to activate, maintain and de-activate physical connections for bit transmission between data-link-entities (MAC layers). A physical-connection may involve intermediate open systems, each relaying bit transmission within the Physical Layer. Physical Layer entities are interconnected by means of a physical medium” [46]. The PHY layer encompasses what is known as the ‘air interface’, the radios, multiplexing, quality of service measurements and calibrations, antennas and the analog/digital frontends within mobile devices. Effectively the PHY layer is responsible for the successful transmission and reception of bits over the air interface.

Layer 2, the Data Link Layer (DLL), encompasses the Medium Access Control (MAC) and Logical Link Control (LLC) sub-layers. By the ITU-T definition: “The Data Link

Layer provides functional and procedural means for connectionless-mode among network-entities, and for connection-mode establishment, maintenance, and release data-link-connections among network-entities and for the transfer of data-link-service-data-units. A data-link-connection is built upon one or several physical-connections” [46]. Fundamentally, the MAC sub-layer is responsible for error correction such as requesting re-transmission of information, setting and adjusting power amplification for transmission and reception, performing radio resource scheduling tasks in coordination with other connected entities and determining where chunks of information received from the PHY layer should be sent to or acted upon.

2.4.2 Orthogonal Frequency Division Multiplexing

Orthogonal Frequency Division Multiplexing (OFDM) is a technique used by modern transceivers for transmitting and receiving digital information by encoding that information over multiple ‘subcarrier’ frequencies.

Due to the tight packing of the subcarriers, the sidebands of each carrier signal will overlap with neighbouring subcarriers. This is circumvented due to each of the subcarriers being orthogonal to each other, and is achieved by setting the spacing of each subcarrier to the inverse of the symbol rate [47].

The symbol rate, or more commonly known as the ‘baud rate’ (f_B), is the rate at which communication signals are generated over a carrier. Each of these symbols has a bit representation, determined by the number of distinct levels (M) the symbol can represent. For example, these may be represented by M unique voltage levels, which can encode $\log_2(M)$ bits. The ‘bit rate’ (R) of a system is equal to its baud rate, times the number of bits encoded within each symbol, expressed in bits per second.

$$R = f_B \log_2(M) \tag{2.5}$$

Each of the subcarriers are modulated using either Quadrature Amplitude Multiplexing (QAM) or Phase Shift Keying (PSK) to encode the digital bits. By utilising N subcarriers within a given band rather than a single carrier spanning the entire band. The robustness of the transmitted data is significantly higher with respect to spurious channel noise, due to only a subset of the carriers being impacted, rather than the entire signal.

Intersymbol Interference

Figure 2.8 demonstrates the Nyquist Inter-Symbol Interference (ISI) criterion, with channel response ($h(t)$) on the y axis and time (t) on the x axis. Under the ISI criterion, if the symbol period (T_s) is equal to the inverse of the bandwidth of each subcarrier, then the raised cosine response of each (n^{th}) subcarrier with centre frequency (n/T_s at baseband), when sampled at the nyquist rate will ideally only contain the signal power for that respective subcarrier [14]. Under these conditions the impulse response of every other subcarrier is 0, with the response of the intended subcarrier being equal to 1, thus eliminating ISI $\forall n \in \mathbb{Z}$:

$$h(nT_s) = \begin{cases} 1; & n = 0 \\ 0; & n \neq 0 \end{cases} \quad (2.6)$$

For the Fourier Transform $H(f)$ of $h(t)$, to satisfy the ISI criterion [14]:

$$\frac{1}{T_s} \sum_{k=-\infty}^{+\infty} H\left(f - \frac{k}{T_s}\right) = 1 \quad (2.7)$$

Figure 2.8 provides a graphical representation of the conditions satisfying Eq 2.6 and Eq 2.7.

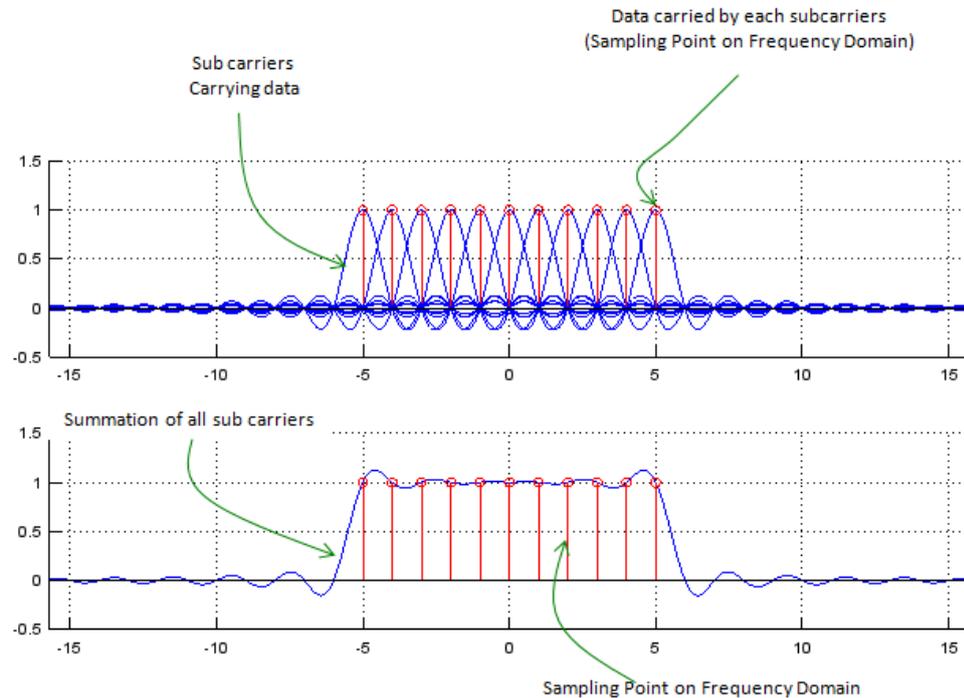


Figure 2.8: Depiction of subcarriers in the frequency domain, with identified subcarrier centre frequencies forming the sample points/bins for a FFT to capture each symbol [48]

The trade-off with this approach is that the symbol holding times must be significantly longer than if the signal was permitted to occupy the entire band. Thus there is an increase in transmission delay before a symbol can be successfully received and decoded as the integration of the signal must be performed for as long as the holding time. Due to this extended holding time, the signal integrity becomes more susceptible to deviations in frequency. This can occur either due mismatched transmitter and receiver oscillators or Doppler shift from moving at speed [49].

Resource Blocks

OFDM symbols are grouped into ‘Resource Blocks’ (RB). These resource blocks are comprised of multiple ‘Resource Elements’ (RE), with each resource element occupying a

discrete period in time over a particular subcarrier frequency. These resource elements form the smallest spectrum units that an OFDM system, such as LTE utilises. The limits on symbol bandwidth and holding duration are inversely related by $\frac{1}{f}$. A resource element represents a discrete chunk of time and frequency within the band of interest within an LTE system, that can be allocated for data communications.

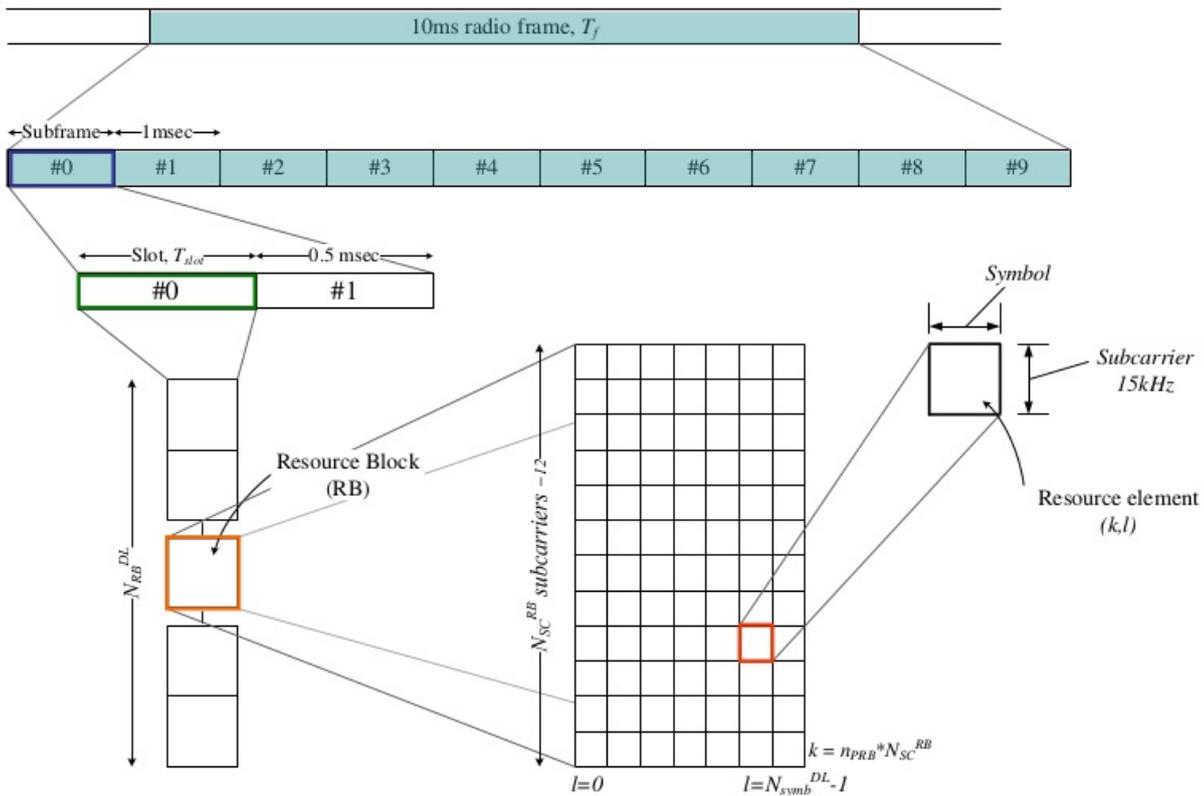


Figure 2.9: LTE Resource grid structure, detailing Resource Blocks and Resource Elements. The grid, bottom right, Y axis represents frequency (number of subcarriers and bandwidth), and X axis represents time (symbol holding time) [50]

Each of the resource elements comprises a single OFDM symbol. In the case of QPSK (Quadrature Phase Shift Keying), a single RE comprises 2 bits of information, 16QAM: 4 bits, 64QAM: 6 bits. In the LTE case, a resource grid can look like the following, Resource Blocks (180kHz each - 12 subcarriers of 15kHz) and subframes (1ms) being used to break

up and schedule spectrum resources at a high level, before further breaking the resources down into their final Resource Elements, detailed in Figure 2.10.

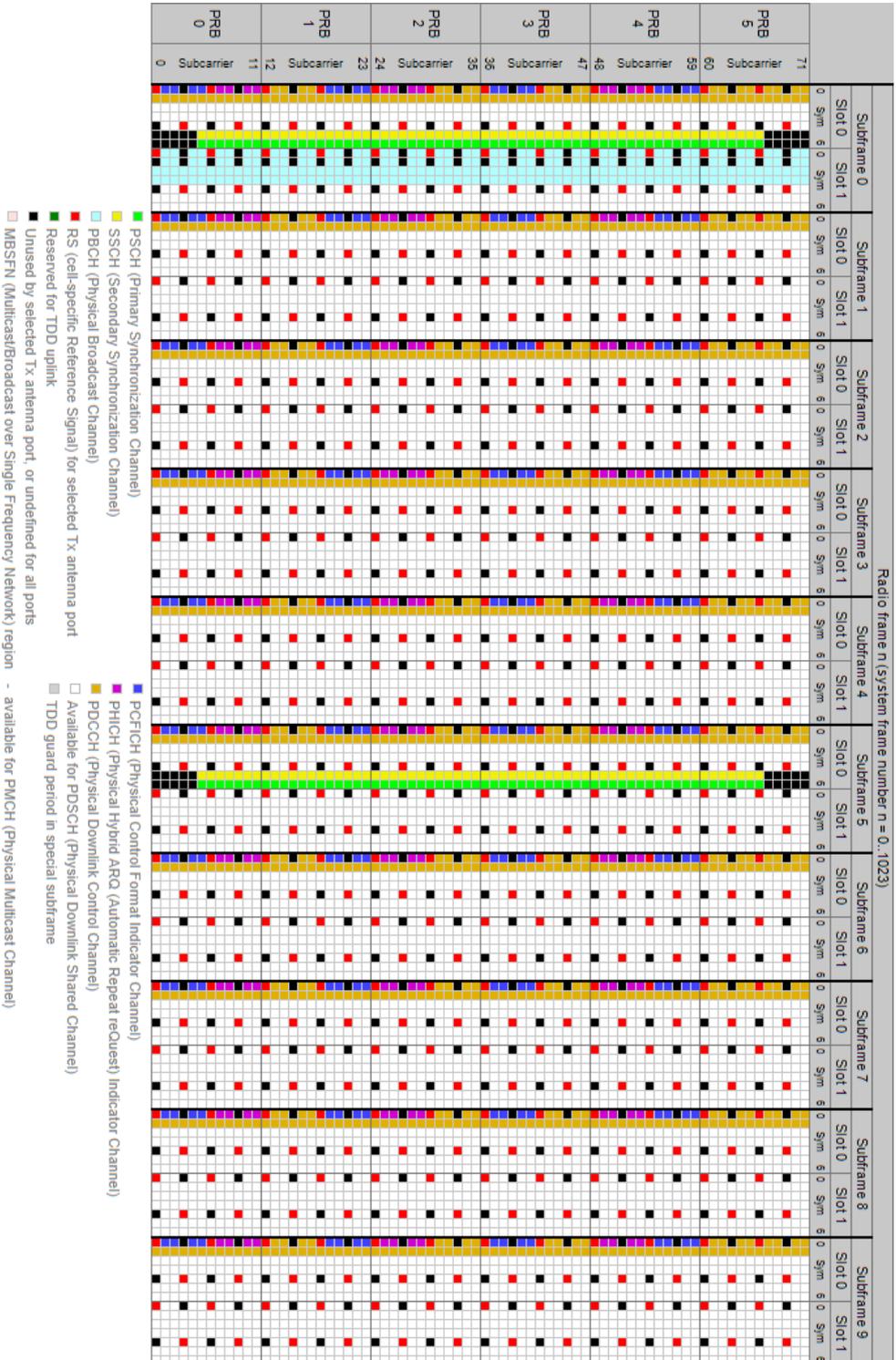


Figure 2.10: Time and frequency resource allocation of an LTE downlink channel for 6 RBs (1.4MHz total) over 10 subframes (10ms). Resource elements are identified on the grid, within their slot and frame structure [51]

Pilot Signals

To account for time variant interference and fading within an OFDM communications channel, channel estimation is physically integrated directly into the operation of OFDM. By selecting a few equidistantly spaced subcarriers that span the band of interest, a known ‘pilot’ signal can be broadcast over these subcarriers [52]. These ‘pilot signals’ or ‘pilot tones’ (depicted in red in Figure 2.10) are transmitted (by the basestation) with a given power level, on a particular set of subcarriers and with a particular cyclic prefix (or training data sequence) for accurate channel estimation. Typically these tones are scheduled for specific resource blocks under a specific pattern to assist with estimation functions. This is also performed on the uplink, but in a somewhat different manner.

When the (UE) receiver measures the power and accompanying training dataset of the pilot tones [52], the receiver can model the time variability within the channel, and correct for it. This leads to a reduction in the bit error rate in the overall channel. Rather than performing channel estimation on every subcarrier across the entire channel, it is more efficient and preferable to select a few, equally spaced carriers to perform pilot signal estimation.

2.4.3 Channel Duplexing and Aggregation

Frequency Division Duplex (FDD) and Time Division Duplex (TDD) are communication schemes that enable bi-directional (full duplex) communications between two parties simultaneously [53]. While both these methods provide simultaneous bi-directional communications, they are quite different in carrier frequency management and possess different properties that make them more or less suitable in certain conditions.

Carrier Aggregation (CA) [54, 55] (particularly in the context of LTE) is a method for enabling the collection (‘aggregation’) of one or more separate carrier frequencies, to

increase the amount of available bandwidth to a device for uplink or downlink communications. The carriers do not have to be contiguous in frequency (under certain CA schemes), a key benefit for maximising the utilisation of fragmented spectrum.

TDD and FDD

A FDD scheme, uses separate bands for the uplink and for the downlink, these are known as ‘paired’ bands [53]. The paired frequencies (with necessary guard bands between them) are used exclusively for the uplink or the downlink and are in fact two paired simplex communication channels, functioning simultaneously. Typically these bands are equal in bandwidth, which results in reduced spectral efficiency under asymmetrical traffic. Due to each carrier being dedicated to a single direction, delay between transmission and reception is less critical than under a TDD scheme, increasing the effective range of an FDD deployment.

A TDD scheme only requires a single frequency band for operation, but slices the band into units of time, which are allocated to both uplink and downlink [53]. This form of duplex, at a small timescale is in fact a simplex scheme (achieving only one way communication at any point in time), however the duration between uplink and downlink slots are sufficiently short. This short duration between slots results in practical application, emulating simultaneous bi-directional communications with delays (guard periods) in between communication frames to minimise interference.

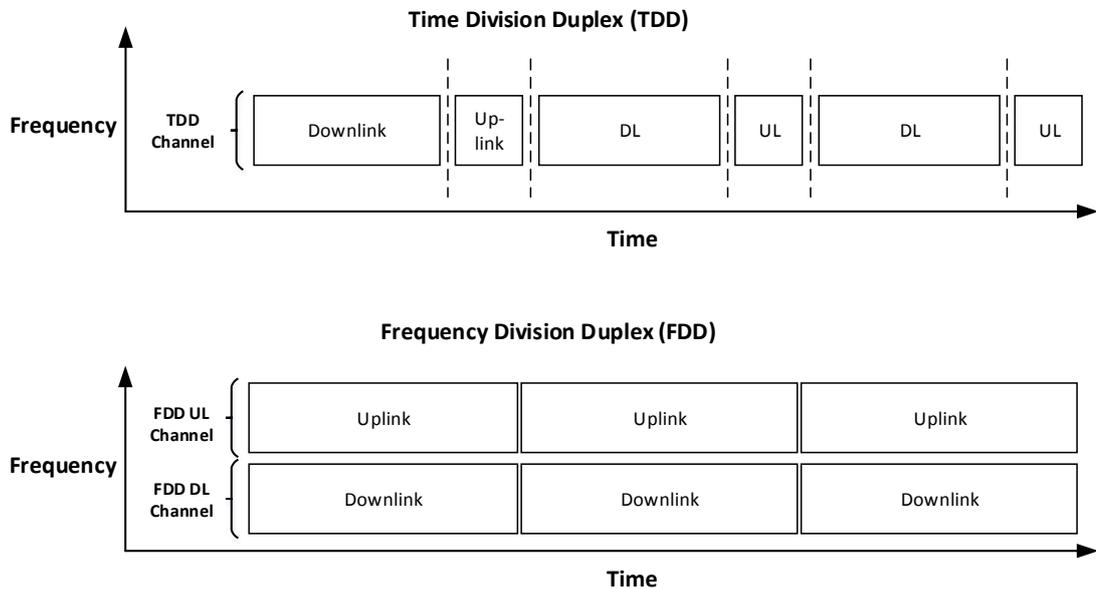


Figure 2.11: Comparison between FDD and TDD spectrum duplexing and access schemes. Note the stringent time spacing between uplink and downlink periods in the TDD case.

Figure 2.11, illustrates the key differences in time and frequency behind an implementation of both TDD and FDD duplex schemes.

The primary advantage of a TDD system over an FDD system is that the TDD system does not require paired spectrum in order to provide full duplex communications, and can also achieve greater spectral utilisation (in some scenarios). The time slot allocation between the size of uplink and downlink slots can be dynamically varied if more downlink traffic is required over the uplink or vice versa. Which is a key benefit for spectrum efficiency as traffic is commonly asymmetrical [53].

Disadvantages of TDD include: the increased time delay between slots to ensure that uplink and downlink traffic do not collide; and, higher cost due to the requirement of more complex radios and infrastructure. The more stringent timing requirements when compared to an FDD scheme also results in reduced ability for the scheme to handle long

distance communications due to the delay spread (or reduced throughput by extending the guard period), requiring more dense deployments of TDD basestations to serve the same area as a single FDD basestation. A further TDD disadvantage is the notion of cross slot interference, for example: two neighbouring BS's with different up/down allocations, creating conflicting slots and substantial interference at the cell edge.

TDD based asynchronous traffic has the added knock on effect of causing increased communication delays and energy consumption between the UE and BS (due to a mismatch in spectrum access requirements between a given UE and BS) that are not present in FDD scenarios. If a UE has different access requirements (such as more upload traffic than the BS is allocating as the BS may be optimising for heavy download traffic, i.e. significantly more down frames are allocated than up frames) the UE's transmitter will have to be active for an extended period, consuming more energy and experiencing significantly more delay in communication when compared to uploading the same payload under an FDD strategy. This problem can be further exacerbated if an up/down frame is allocated by the BS and only a small portion of the traffic actually occupies the spectrum resources, traffic in the opposite direction incurs a substantial delay penalty until the next slot, as it cannot be transmitted on the unoccupied resources during that frame.

Conversely, the inverse of the problem exists in FDD systems if there is substantially more traffic than either channel (up or down) can support, and if one channel has a very low amount of traffic, similar latency and energy consumption penalties will be incurred by the UE's competing for resources on the congested channel.

TDD access schemes provide more options with spectrum utilisation however can cause unintended and somewhat unpredictable delay and energy consumption penalties due to conflicting channel direction allocations, while FDD schemes provide a more predictable communication environment, but are more susceptible to congestion in a single direction which can also incur delay and energy consumption penalties for UEs.

Carrier Aggregation

Carrier Aggregation is a method of achieving higher available bandwidth to a device that is limited to only communicating on a single allocated carrier frequency, by simultaneously communicating over multiple allocated carriers. The advantage of this method is the increase in available bandwidth offered to a device by using the available hardware on the device (assuming the device is at least 3GPP Release 8 compliant, and the MNO has supporting basestations and spectrum).

Carrier aggregation is performed in three distinct methods, highlighted in Figure 2.12. Each of these aggregation types pose unique challenges for protocol and hardware design [55]. Aggregated channels do not have to be contiguous in frequency, and can be made up of several discontinuous carrier frequencies [54]. This method of harnessing multiple discontinuous bands, addresses a significant challenge for spectrum regulators and licensing agreements with respect to allocating large contiguous blocks of spectrum.

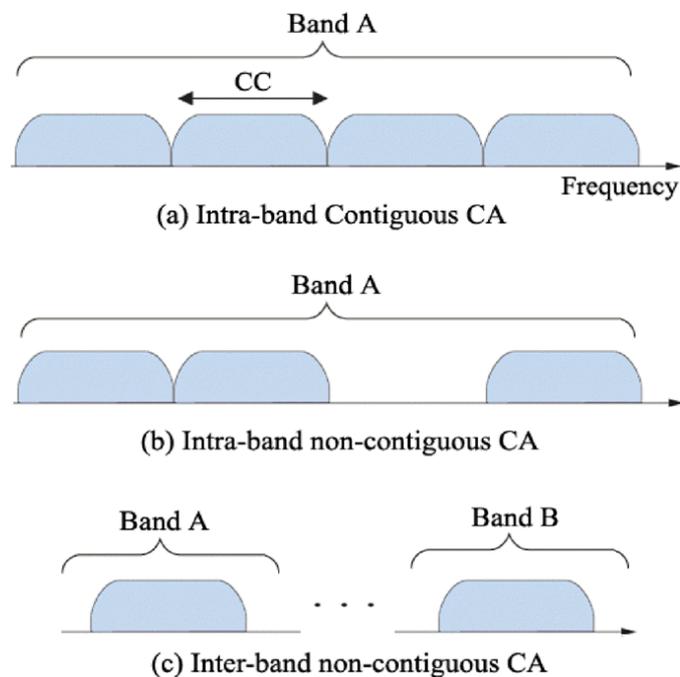


Figure 2.12: Carrier Aggregation types [55]

Carrier aggregation is also a key enabler for maximising the utility of secondary spectrum access systems, and is fundamental to their commercial success. Section 2.5.3 includes the numerous ways CA can be exploited to achieve greater available bandwidth to communications devices.

2.5 Secondary Spectrum

Building upon Section 2.3.3, it follows, that by the large number of spectrum-consuming applications, the economic benefit and quality of life improvements that spectrum enables, that the finite amount of electromagnetic spectrum (on a frequency basis) would already be heavily constrained and thus used very efficiently. This unfortunately is not the case.

It is known that a scarcity of spectrum exists for frequencies below 6GHz as the spectrum below 6GHz are completely allocated for various services or leased to long term incumbents [5]. Recent studies, however, have found that the utilisation of the spectrum below 6GHz is very low on temporal, bandwidth and geographical [3,4] dimensions.

Limitations of licensing structures, spectrum assignments and spectrum access methods to address the difficult task of managing interference from various sources, results in much of the spectrum being wasted. It must be expressed, however, that many of the existing licensing and spectrum assignment structures are created under the assumption that neighbouring services are uncoordinated and have no interest in coordination, likely due to cost and complexity.

The inefficient management of spectrum to address the problem of interference and uncoordination, results in a substantial loss in ‘functional efficiency’ of the spectrum which can be attributed to several factors:

- Physical factors: large geographic regions or over sized buffer zones than necessary being assigned to isolate the incumbent from an uncoordinated neighbour, or over

allocation of bandwidth incorporating larger than necessary guard bands to avoid interference from an uncoordinated neighbour.

- Operational factors: services that typically communicate with a less than 100% uptime, have intermittent usage, are spurious in operation, or employ older technologies with low spectral efficiency.

The combination of the aforementioned spectrum usage factors, lack of coordination and interference management leads to substantial inefficiencies in the available spectrum that can be exploited, and hence, the low ‘functional efficiency’ of spectrum. This issue of low functional efficiency however, presents an opportunity. As much of the spectrum is unused, there is presented a substantial opportunity for a system to detect the ‘gaps’ in spectrum use and exploit those gaps.

The notion of secondary spectrum assignments was introduced in Section 2.3.1 as an assigned secondary application or user of an allocated spectrum band, with no guarantee for quality of service. Secondary spectrum going forward is defined here as: any electromagnetic spectrum that is currently unutilised on a temporal, geographical or frequency dimension that is designated for a service or use, due to the absence of the primary user of that spectrum.

The primary user typically possesses a license for a given carrier frequency, bandwidth and area of operations which grants them exclusive use of that spectrum over the regions defined by the license. A primary user often has to pay for the privilege of having ‘exclusive access’ to that spectrum. Due to the desirability of ownership of spectrum for numerous applications the majority is already allocated to a primary service or user. However, if that primary service or user is not operating within a particular geographical area/space, time or frequency, at the highest possible efficiency (i.e. saturating the channel capacity) then that particular spectrum can be used for a secondary service. Partitioning the spectrum

into useful-sized blocks on the part of the primary user (in space, time, frequency or any other measurable quantity), provides an opportunity for improving the functional efficiency of that spectrum as well as effectively enabling opportunities for secondary users (presumably who have run out of their own dedicated spectrum).

The granularity at which spectrum needs to be managed to combat the aforementioned inefficiencies is a complex issue to address. The abundance of unused spectrum, or, ‘whitespace’ that exists needs to be leveraged to address the functional efficiency of spectrum. This has been identified by numerous global organisations and as a result significant efforts have recently been put into designing and defining systems for secondary spectrum reuse in an effort to reclaim a portion of the under-utilised spectrum within a particular frequency band.

One of the main motivations for a primary user (or ‘incumbent’ with respect to a secondary spectrum service) to purchase a spectrum licence, other than having exclusive access to that strategic resource, is to protect the primary users’ communications or activities within that spectrum from external man-made sources of noise. As the capacity of spectrum is noise limited (see equation 2.1) it is within a primary users’ best interest to minimise sources of noise, to therefore maximise the capacity of a given channel. As such, the exclusive licence grants a regulatory and legally enforceable contract to the primary user with the effective guarantee that their signals will not be interfered with by a third party.

This becomes the main point of contention for a secondary spectrum service. As the service will be operating at the same frequency as the primary service there are significant concerns that the secondary service may degrade the quality of the primary incumbents signals leading to potentially negative performance and financial impacts upon the incumbent. Thus it is of utmost importance that for a secondary spectrum service to be designed with all aspects of interference with respect to the primary users’ operation.

As a fundamental principle of managing secondary spectrum, secondary services must cede control of the spectrum to their primary counterparts whenever the primary service begins operating within that space, time or frequency, so that a Secondary User (SU) will not interfere with a PU. The coordination between primary and secondary users gives rise to unique engineering challenges which are being addressed through two combined government and private enterprise initiatives: the Licensed Shared Access (LSA) [56, 57] system under development in Europe and the Spectrum Access System (SAS) as part of the Citizens Broadband Radio Service (CBRS) [58] within the United States of America. Managing and identifying secondary spectrum is an ongoing area of study and is the core focus of much of our research.

2.5.1 Managed and Unmanaged Secondary Spectrum

An ideal secondary spectrum service will have to be aware of and opportunistically consume any available spectrum that the primary incumbent does not make use of, with no impact toward the primary user/incumbent. It is obvious that a ‘quasi-omnipotent’ system required to fulfill this requirement will be quite challenging to design and implement. To this end, typically secondary services are only deployed in locations where it is highly likely that there will be no presence of a primary user at all (such as the case with TVWS devices) [59].

The construction of an aforementioned ‘quasi-omnipotent’ system (or a crude approximation), is the objective of several technologies with several different schools of thought with respect to how much control or management should be needed or required to make these systems a possibility. Broadly, there are two umbrellas for secondary spectrum access systems, managed and unmanaged systems. Managed systems such as LSA and SAS [60, 61] (Section 2.5.3) are infrastructure and legal agreement heavy with stringent restrictions on spectrum access. What these systems do provide however, is a guaran-

tee or assurance that secondary users accessing spectrum resources with this system will have zero impact upon primary user activity situated within that band, with methods for recourse if the secondary user does in fact break that agreement.

Unmanaged systems, in contrast, such as Cognitive Radios (CR) seek to address the avoiding of interference with a primary user issue by: having the radio completely aware of the spectrum around it, locally and regionally; designing capabilities for identification of spectrum access patterns or signals/triggers from the incumbent services within the band of interest; and finally, making decisions for when to access the spectrum using gathered information so as to not impact the primary user but still have access to the band of interest. The topic of Cognitive Radios will be explored in more detail in Section 2.6.1.

2.5.2 Television Band Whitespace (TVWS)

A system to make use of Television Whitespace (TVWS) was conceived in the mid 2000's as a possible method of harnessing unused secondary spectrum resources. The first efforts were undertaken by the IEEE 802.22 working group set with the task of developing the architecture and standards for a Wireless Regional Area Network (WRAN) that can operate within the Digital Video Broadcasting - Terrestrial (DVB-T) bands (470MHz to 790MHz), to access any unused channels on a geographical and temporal basis [62, 63].

Specifically, TVWS refers to a technology that implements the regulatory requirements for unlicensed sharing of unused space in licensed, local terrestrial TV (Television) channel coverage areas. The basic principle relates to the idea of allowing unlicensed, secondary devices to access spectrum at specific geographic locations and/or during specific time intervals, in geographically-limited spectrum where it would not interfere with terrestrial TV transmission or reception.

This is achieved through the use of a centralised geo-tagged spectrum database, that

holds records for all operating television stations across any regions of interest. A TVWS device is required to obtain authorization before they can transmit, and requires those devices to cease operation if they come into close range or are located inside protected areas [64]. These protected areas are created at incumbent boundaries or within areas where there are no available channels to ensure that the incumbents are protected from secondary devices.

Within the U.S. [65] and the UK [66], TVWS regulatory frameworks have been developed and introduced. There have also been efforts within the WiFi standard to introduce TVWS capabilities [67]. However, due to the key areas where additional spectrum would have the greatest commercial impact (i.e. within densely populated urban areas), the majority of television bands are already owned or occupied by an incumbent, resulting in a significant lack of exploitable whitespace. In addition to the lack of availability, there also exists no mechanism within the TVWS system to ensure quality of service (QoS) for secondary systems communications, which reduces its desirability for time sensitive and high throughput applications.

To further complicate, and challenge the viability of TVWS systems, there is a growing interest from national regulators to repackage existing spectrum previously assigned for television (analog and now primarily digital), and relicense it, rather than to leave unused channels idle for access on a secondary spectrum basis. For example within Australia in 2015, spectrum from 694MHz to 820MHz (which included a number of analog and digital television channels), was consolidated and repurposed under the “Digital Dividend” program. This spectrum was subsequently auctioned to MNOs for dedicated mobile networks [68]. In a similar vein, the FCC auctioned off reclaimed spectrum from 698MHz to 806MHz [69], to various MNOs and interested parties under auctions 73 and 92. These factors globally have resulted in the limited success of TVWS systems.

2.5.3 LSA and SAS

Proceeding the development, deployment and somewhat limited success of TVWS systems, an improved generation of managed secondary spectrum access systems has been under development. These systems would be able to access a somewhat smaller amount of spectrum (approx 150MHz, compared with TVWS estimated 320MHz), but with the possibility of achieving greater secondary spectrum utilisation due to less constrained scenarios. These systems have been trialed and developed both in Europe (LSA) and the United States (SAS), and while similar in intent are rather different in implementation, initial frequency bands of interest, incumbents present and types of secondary users [59].

LSA

LSA is a secondary spectrum sharing standard currently set for deployment within the European Union, first proposed in 2011 by the Radio Spectrum Policy Group (RSPG) [70]. It is poised to be a significant improvement on existing regulatory frameworks enabled by building intelligent management and licensing schemes to more effectively manage complex spectrum use cases. The premise of ‘licensed shared access’ is to develop a system that enables a PU (or owner of spectrum) to enter into a bilateral (or multilateral) sharing agreement with one or more SUs to enable the SU to make use of the primary spectrum without negatively impacting the PU’s mission.

LSA as per its initial specifications completed in 2017 [71], opens up the 2.3GHz to 2.4GHz band in Europe to MNO’s with LTE Band 40 (TDD) networks, enabling a form of coexistence with the existing incumbents through the use of binary sharing agreements. The primary incumbents within this band (in particular within Germany) are: SAP/SAB links (Services Ancillary to Programme/Services Ancillary to Broadcasting) i.e. wireless high definition television cameras, portable video links and mobile video links (for news related purposes or sporting/special events coverage); and various airborne telemetry and

radiolocation services [72]. Most of these applications have a limited range resulting in a small geographical incumbent footprint at any one point in time. These particular incumbents are also only active during times of particular interest that are not necessarily known ahead of time or to be planned for (other than the sports and events coverage scenario). It is foreseeable that these services will not be on 24/7 in a concentrated location due to the opportunistic nature of these services, yielding a considerable opportunity for secondary spectrum utilisation.

Sharing agreements are set up as a form of temporary lease agreement between the PU and SU. The SU leases spectrum over defined time ranges, regions, bandwidths sensing requirements and other parameters such as transmission powers, incumbent reclaim conditions and allowable interference levels, outlined within the sharing agreement. The advantage in LSA sharing agreements, is the provision of a guarantee to the PU that any SU activity will not negatively interfere with the PU. The agreement outlines conditions and penalties if interference events do occur and specifies the procedure for an event occurs where the PU must retake control of the spectrum from an active SU. These agreements are enforced by the physical LSA control system and sets of protocols for the rapid transition of ownership of the spectrum from the PU to SU and vice versa.

Several advantages of the LSA system are:

- LSA can be deployed while maintaining existing incumbent services.
- QoS is guaranteed for the incumbent and some level of QoS can also be guaranteed for the SU, which is not possible under either ‘opportunistic spectrum access’ approaches or conventional ‘secondary use’ allocations.
- LSA is a complimentary spectrum management tool that fits under ‘individual licensing regimes’ [56].

As an example for the use-case of an LSA system, an explicit agreement can be set

up where a PU has agreed to vacate the spectrum for a period of time, or, in the event when the PU is not using the spectrum at a given point in time for whatever reason, the secondary user may have a method to access the spectrum at that point in time. The objective of this system is to enable parties with ownership of spectrum or services that are explicitly allocated to a portion of spectrum, be able to coexist with intelligent secondary services that seamlessly ‘work around’ the comparatively unsophisticated primary users.

At a high level, this work-around is achieved by sensing the presence and activities of the PU, then building a model and conforming to a particular set of rules around that activity, so that the spectrum is only accessed when the SU is certain that they will not interfere with PU activities. Effectively, the system guarantees the QoS for the incumbent and provides minimal QoS guarantees for the secondary user trying to squeeze value from inefficient use of the spectrum.

Without access to *a-priori* spectrum allocation information, real-time sensing information, and a system to provide the decision making or administration of the spectrum, it would be impossible to ensure the coexistence of primary and secondary services while maintaining QoS for all interested parties. The LSA architecture in Figure 2.13 is designed with these constraints in mind.

As per Figure 2.13, the system architecture requires the following system blocks [60]: an LSA Repository, an LSA Controller, an OA&M (Operations, Administration and Management, or, a Network Operator) entity, and an interface for regulators to update agreements and assess performance of the system

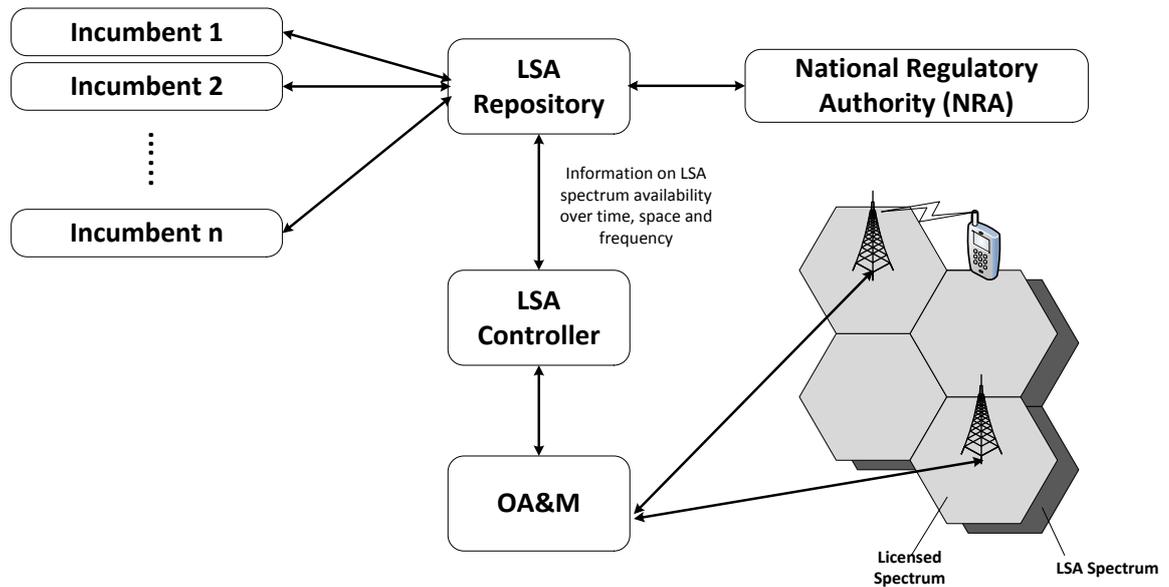


Figure 2.13: Block Diagram of the LSA Architecture

The LSA repository maintains a database of geographically tagged spectrum measurements, spectrum access conditions, sharing agreements and incumbent spectrum usage information. Incumbents are required to provide *a-priori* usage information to this database for their particular spectrum usage patterns in space, time and frequency. This data is then used to create primary user access models enabling greater accuracy for determining whether a secondary service can be used within a given region at a particular point in time [73]. Additionally, allowable transmission power ranges and interference limits can be determined from these datasets in combination with general regional radiation propagation models and incumbent data. There have been several pilot tests of this system in recent times [74].

The LSA controller is responsible for handling OA&M requests for accessing the secondary spectrum. The controller is also responsible for interrogating the LSA repository to identify possible conflicting services within the region of spectrum requested by the OA&M service. Once a request is received it is combined with regional information (such

as incumbent services present, expected regions that interference could be caused etc) and the particular sharing agreement that the secondary user on the OA&M network has with the relevant incumbents. A decision is then made by the controller to approve or deny the spectrum access request to ensure that any incumbents will not be affected by the secondary system accessing the spectrum.

SAS

The Spectrum Access System (SAS) is a USA FCC initiative to improve the functional efficiency of the 3.55 - 3.7GHz band in the USA mainland, by devising a three tier model for primary and secondary spectrum reuse within that band. The 3.55 - 3.7GHz frequency band is termed the “innovation band” and while its objective too is to enable managed secondary spectrum reuse of the given band, this approach takes a different approach to LSA. Within this frequency band the secondary users are encompassed by the CBRS, and the access technology proposed to access the spectrum is a variant of LTE called LTE-U (LTE unlicensed) that uses the LTE TDD Band 42 and 43, or WiFi type opportunistic spectrum access with modified collision avoidance systems [75].

The key difference with SAS when compared to LSA, is that certain incumbents are unable or not required to inform the central database of their activity, and thus the secondary users must adapt to this use case. The incumbents that fall under this category are military services such as naval radar systems, who do not want to broadcast their position or operational area to any central broker or entity. Thus their positional information must be obfuscated from the SAS network at large, and particular coastal areas within the mainland United States are exclusion/protection zones that SAS devices cannot use.

Under SAS there are three classifications of users (compared to LSA’s two), the primary user or incumbent which has the highest level of protection and QoS guarantee

against all CBRS devices; A priority secondary user or PAL (Priority Access License) that has some guarantee of QoS and protection against the last class of devices; and the third class, an unprioritised secondary user that has no guarantee of QoS and can only access the spectrum ‘opportunistically’ (by sensing spectrum and reporting its measurements and location to the SAS to gain access) called the GAA (General Authorised Access) [61].

A general model of the SAS architecture is present in Figure 2.14. The SAS entities are responsible for providing secondary spectrum access licenses to interested Citizens Broadband Radio Service Devices (CBSD), and ensuring that the CBSDs do not interfere with incumbent services. Similar to the LSA architecture, an interface between all SAS entities, spectrum databases and federal incumbents present within given bands are provided to define the parameters for spectrum sharing. The CBSD is representative of a node, access point or base station that provides the access network for users over the secured secondary spectrum access.

The PALs (priority CBSDs) can engage in a bidding process to secure access to the secondary spectrum for a given geographical region. These regions defined by the FCC notice are split up according to USA federal census tracts, where a licensee that has secured spectrum will have access to that spectrum over the entire census tract (with limited transmission powers allowed around the borders of the tract). As per FCC 15-47 section 1 point 4 [75] this PAL license grants a 10MHz band of secondary spectrum over that tract for a period of 3 years, however only a maximum of 7 licences (70MHz total) can exist within a given census tract within the 3550 - 3650MHz portion of the band. This limitation on the number of licences ensures that there will always be at least some spectrum available for GAA purposes. GAAs are able to move freely within the entirety of the 150MHz band (excluding PAL owned spectrum or active incumbents), however they are afforded no QoS protection from other GAAs or PALs.

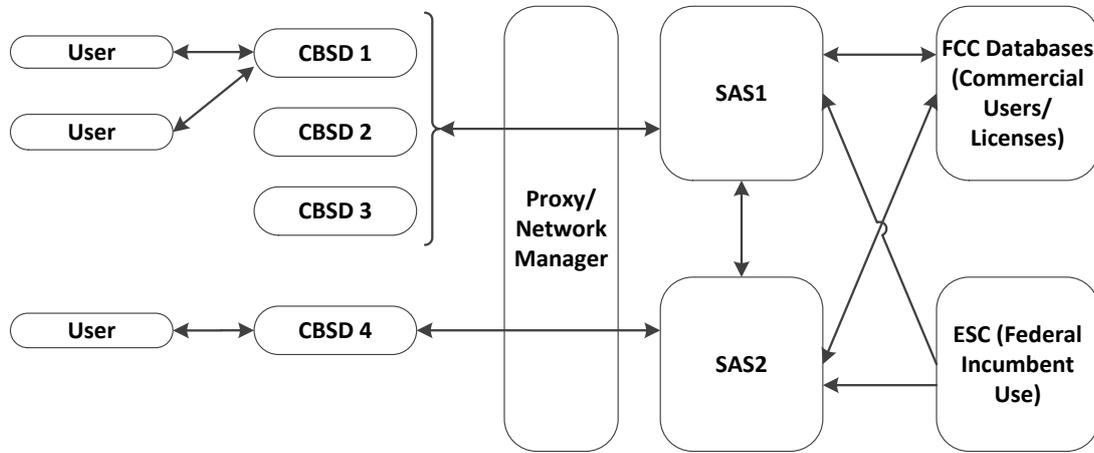


Figure 2.14: Block Diagram of the SAS Architecture

2.6 Software Defined Radios

The concept of a Software Defined Radio (SDR) has been in existence since the mid 1990's under a United States Department of Defence program under the name "SPEAKeasy" [76]. This radio was designed to address an issue faced by the USA military, in that it had over 15 existing radio platforms, RF bands and waveform modulations for which all of defence command and control operations were conducted over. The Speakeasy program was designed to emulate the entire suite of radios under operation on a single device. To achieve this, the radio needed to be reconfigurable and modular to suit all the various waveforms and signal processing tasks in order to successfully communicate with all the existing radios in the field.

The SPEAKeasy produced a functional prototype combining software programmable baseband processing and wideband interchangeable RF modules to enable communication over a wide range of RF bands and waveform modulations. This was a revolution in

modern radio architecture as no longer did a single system have to be designed for every niche frequency band and waveform modulation, now existed a general purpose device that could address all of these requirements [7, 77].

Since then the concept has been further refined, commercialised and deployed broadly. The system that used to occupy the space of a large truck can now occupy seldom more space than a credit card and for only hundreds of dollars [78]. More powerful consumer level SDRs such as Ettus Research USRPs (Universal Software Radio Peripheral), is a core component of many wireless technology development and research labs across the world [79].

The capabilities of an ideal SDR (Figure 2.15) are enabled by a series of important processing blocks:

- Reconfigurable front-end (antennas, amplifiers, mixers and variable frequency oscillators) to change the frequency band of operation and signal amplification
- Digital Upconversion and Downconversion to perform frequency conversion from and to baseband.
- High sample rate Analog to Digital Converter (ADC) and Digital to Analog Converter (DAC) to convert the analog information to and from digitised information respectively.
- A suitable baseband processor for digital signal processing, decoding, radio control and protocol enforcement.
- Fast Fourier Transform (FFT) capabilities to translate the captured waveform to the time domain for analysis.
- Inverse Fast Fourier Transform (IFFT) capabilities to convert a digitally encoded time domain signal to the frequency for transmission.

Digitised waveforms are typically captured by a direct-conversion receiver (also known as a zero-IF receiver), that comprises a demodulation stage for capturing samples in the complex domain with separate in-phase (I) and quadrature (Q) signal paths [80]. These samples are then manipulated to extract the information encoded within the captured waveform, enabling the successful reception of any waveform at any frequency band within the range of operation of the radio [81]. The specifics of capturing and manipulating an analog signal with a SDR is studied in depth in Chapter 4. An ideal SDR receiver (Figure 2.15, utilises direct sampling of the carrier frequency, high resolution ADC sampling at at least twice the frequency of the carrier and digital downconversion to extract the desired signal.

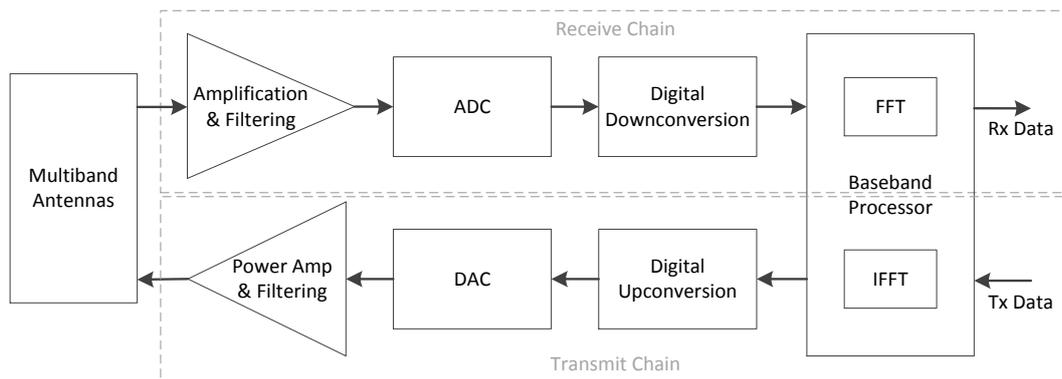


Figure 2.15: Ideal SDR Architecture

From a practical and conventional implementation however, Figure 2.16 highlights the core components of a reconfigurable ‘Direct Conversion’ or ‘Zero IF’ radio receiver. A reconfigurable channel select filter is used to pick out the desired carrier while minimising other signals present in the spectrum. Downconversion is used to mix the radio frequency (RF) down to a baseband frequency, rather than direct sampling of the RF. To achieve this, a frequency synthesiser is used for the LO input, to match the carrier frequency of the

desired signal. The carrier frequency is situated at the centre frequency of the waveform, and since the LO frequency matches the centre frequency (creating a ‘Zero Intermediate Frequency’), in-phase and quadrature (IQ) signal paths are used to accurately capture the signal. Downconversion to a baseband requires a much lower sampling frequency to capture the desired waveform than directly sampling a multiple GHz carrier. ADCs that sample at slower rates are also offered superior resolution which enables more precise power estimation and signal reconstruction for the baseband processor to decode [82].

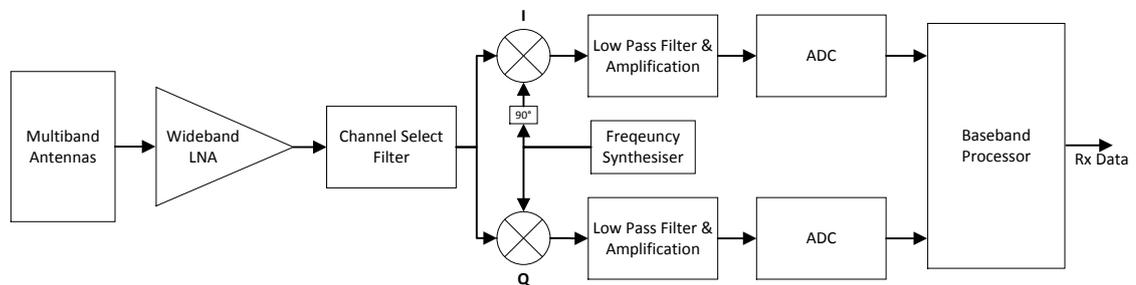


Figure 2.16: Conventional SDR Receiver

On the transmit side, any digitised waveform intended to be transmitted is encoded as digital information in the time domain, translated to the frequency domain, digitally upconverted and transformed to the analog domain via the DACs, propagated through the analog front end, amplified, and finally transmitted via the antenna to an intended recipient.

Due to the large range of capabilities offered by a modern SDR platform, the power consumption and physical footprint required to implement the computationally intensive signal processing and reconfigurable analog front-end components still make highly reconfigurable SDR capabilities out of reach for modern low power mobile handsets.

2.6.1 Cognitive Radios

A Cognitive Radio is a communications device proposed by J. Mitola in 1998-1999 [83] as a way to expand available bandwidth for digital telecommunications by utilising spectrum more efficiently. The CR was proposed as an extension of SDRs to better exploit the general purpose and reconfigurability offered by an SDR platform, with the intent of developing a device that capitalises on the ‘ideal’ functional use case for SDRs.

A CR builds upon the reconfigurable air interfaces, waveform generation and reprogrammability offered by a SDR platform and leverages its fullest potential by integrating spectrum sensing, continuous monitoring of system and network performance, combined with machine learning for spectrum prediction, modeling and decision making. The integration of these capabilities in a single device or network of devices should enable greater electromagnetic spectrum utilisation and thus significantly improving the aforementioned functional efficiency of spectrum, by detecting and exploiting inefficient uses of spectrum in realtime.

It is the primary function of a CR to coexist with existing primary users, without the CR impacting negatively in any way upon that primary user. CR’s ideally manage secondary spectrum across a decentralised network of nodes (i.e. across all the CR’s in the network), whereas systems like LSA and SAS (see section 2.5.3) perform secondary spectrum management centrally, with inputs from decentralised nodes. In practice however, this is difficult to achieve.

2.6.2 Sensing

The first and most prominent difference that a CR encapsulates over a conventional SDR is spectrum sensing and decision making or cognition. Spectrum sensing is fundamental for a secondary spectrum system to make decisions to gain access to interference free

spectrum and at the same time, to not interfere with other users.

There are many spectrum sensing algorithms that are currently in use or are areas of study today. These include: energy detection, cyclostationary filtering, radio identification, matched filtering and waveform based sensing [16, 84].

Energy detection is the simplest and most common form of spectrum sensing due to its low computational and implementation complexity. It is also a very good method of observing a channel without *a-priori* knowledge of the users within that channel. It is a powerful tool for generalised incumbent detection. Incumbents are detected by their relative power level to the noise floor of the detector. Thresholding or minimum power levels above the noise floor can then be used to broadly determine if an incumbent is present or not. The tradeoff with this method versus other sensing methods is its relatively low accuracy and susceptibility to noise within the sensing environment, which can lead to false detections or misdetections. Sensing performance and evaluation parameters for energy detectors are covered further in depth in Section 4.4.

Cyclostationary feature detection exploits the ‘cyclostationarity’ of features within received signals [85]. These features exploit the so called ‘spectral redundancy’ that is present within man made signals and are caused by the periodicity of the statistical variance in the signal. When a quadratic non-linear transform is applied to the variance, ‘spectral lines’ are generated as a form of fingerprint of the signal. This method of detection is substantially more robust to noise than energy detection, it does however require knowledge of the various ‘fingerprints’ that map to relevant incumbents present.

Waveform based sensing is generally used to assist with synchronisation of a receiver to a given transmission signal, or a pilot signal for channel sounding. The sensor aims to exploit these synchronisation or training indicators for detecting the incumbent, but however requires *a-priori* knowledge of those particular signal patterns to make an accurate detection.

Matched filtering is considered the optimal method (minimal misdetection error) for detecting incumbents when the transmitted signal is known. It requires that the sensor has perfect knowledge of all incumbent signaling features and has the capability to demodulate all of the incumbent signals detected. While this method achieves the highest rate of incumbent detection, it is only practical if only one or several types of incumbents exist within a given band of interest, as sensor complexity will grow considerably as more incumbents are accounted for.

Radio identification applies in a similar vein to matched filtering but is less ‘all encompassing’. The sensor is designed to pick out particular properties of a transmission technology in order to determine the access technology used, such as WiFi or LTE. Once the access technology is determined, it will calibrate its sensing functions accordingly to be more sensitive to that particular type of incumbent to minimise false detections.

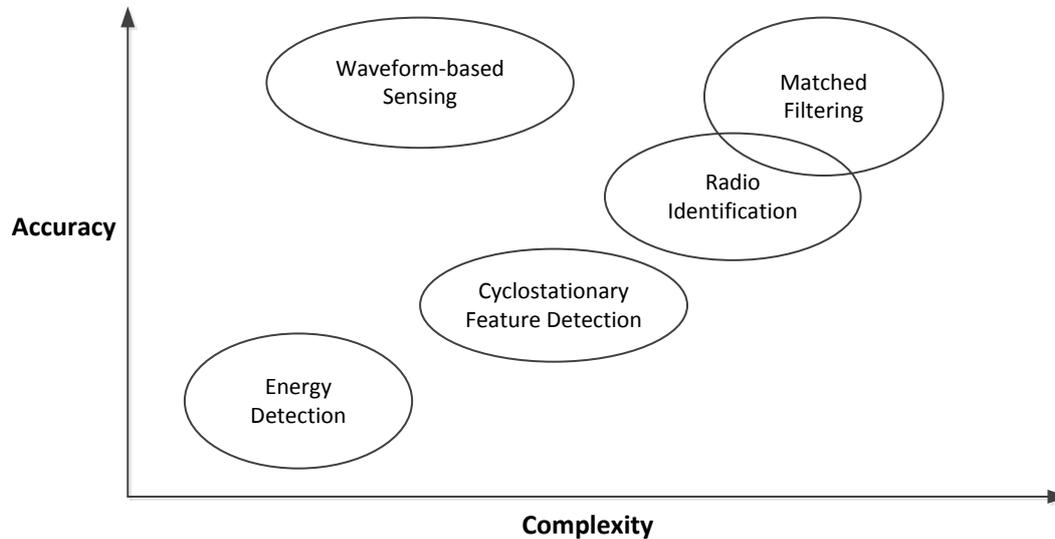


Figure 2.17: Comparison of Spectrum Sensing Methods. Complexity of implementing and designing the sensor versus the accuracy of the sensing method [16]

Sensing duration also forms a critical part of a CR's success, as longer observation windows will result in more accurate sensing results. However as a CR's function is to opportunistically access spectrum the duration for which observations can be performed is a direct loss in transmission opportunity if the spectrum is clear of incumbents without the CR's knowledge. So an adequate balance between enough sensing to ensure no incumbents are present versus transmitting the intended payload needs to be sought.

Another problem that persists with the performance of individual CR's, is the 'hidden primary user problem'. This is a common issue, prevalent in wireless communications that is similar to the 'hidden node' issue in Carrier Sense Multiple Access (CSMA) networks such as WiFi. This issue is particularly important for CRs with imperfect knowledge of the spectrum environment (possibly due to multipath or fading effects) and typically occurs near the edge of an incumbents transmission range [86]. For example: if an incumbent basestation is transmitting to an edge user, but the CR is out of range of the incumbent signal, then the CR will determine the channel to be clear. The CR's transmission range however can overlap with that same edge user, causing interference with the user (not the basestation), even though the CR correctly determined the channel to be clear. Figure 2.18 illustrates the hidden primary user problem:

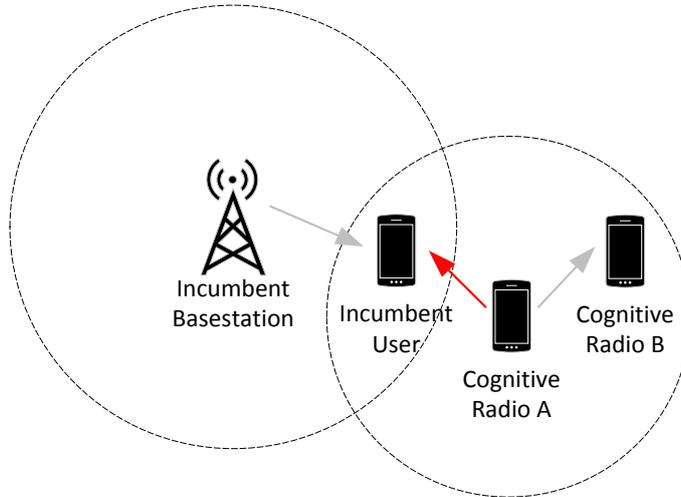


Figure 2.18: Illustration of the hidden primary user problem

There are several solutions to this problem such as cooperative sensing (CRs distributed widely, sensing together) or a geolocation database. A geolocation database can be used to keep a live record of all incumbents within an area of interest, the CR can then interrogate the database to ensure no hidden users are within the CR's range, or be used to inform maximum power output levels for the CR.

Cognition and Coordination

The primary objective of a cognitive radio and more generally, a secondary spectrum system, is to detect and subsequently exploit whitespaces within a spectrum of interest. However detecting an opportunity to exploit while good in theory, is still subject to randomness in channel useage by the incumbent or error in the detection method used. A more robust method to ensure that the secondary spectrum that has been found is indeed available for use, is through constant sampling and subsequent modeling of the spectrum to establish useage patterns that could be exploited by the CR.

Further, in the event that there are multiple CRs vying for the same spectrum resources, there needs to be a form of coordination between them to ensure that they do not interfere (collide) with each other and render the opportunity worthless [86,87]. Thus a level of cognition and coordination is required by these devices to maximise the efficiency of secondary spectrum capture.

Cognition or ‘machine learning’ is required to optimise a self organised network of CRs, in the absence of a centralised controller and to perform decision making for when to access spectrum. These decisions are formed in accordance with the protocols outlined by the network, a set of access constraints, currently available sensor information and other spectrum usage models (such as geolocation data or incumbent access modeling). Machine learning comes into play in assessing the performance of access constraints and whether they need to be tightened or relaxed by optimising for a particular cost function, such as wasted spectrum with respect to payload time to transmit/receive [88,89]. The decision making component can be as simple as ‘if-then’ logic, in combination with more complex methods for optimising access constraints.

Chapter 3

Power, Latency and Cost of Secondary Spectrum Access

3.1 Introduction

Power consumption and latency metrics of a modern handset are paramount to the commercial and functional success of the device, as these are key factors impacting the user experience: users are not fond of delays and dead batteries. Generally, modern handsets already implement quick connection set up times afforded by modern LTE and WiFi systems. These devices are also designed with the goal of maintaining low rates of energy consumption except in the most heavy usage cases. This design goal provides the user with an always-at-the-ready experience that will endure for the most part of a day before needing to be recharged again. Thus it is important to remain sensitive to the amount of energy that is consumed for every aspect of the devices operation and how quickly it performs those operations.

For a mobile handset to function within a secondary spectrum environment, the trade-off between power consumption and latency needs to be quantified with an eye to the impact

on user experience and overall device performance. It is necessary to identify real world scenarios for secondary spectrum reuse, and by association, the kinds of spectrum environments that are suitable for secondary communications. These are needed to establish the functional requirements for devices designed to operate within those environments and to inform design decisions made in implementing future devices.

Chapter 3 focuses on the relationship that power and latency exhibits with respect to the cost associated with accessing secondary spectrum. Section 3.2, provides background and an introduction into power and latency constraints on the function and performance of wireless devices and systems. Section 3.3 provides background on conventional OFDM receivers to provide context for the extended original work presented in: “Power and latency limitations in secondary spectrum reuse for mobile and home wireless systems” [6]. Subsection 3.3.1 presents original work that establishes the motivation for the patent submission in Section 6.4. Section 3.4 additionally comprises of original work, presenting a simple whitespace opportunity time and cost model, establishing a baseline for further whitespace evaluation. Subsection 3.4.3 presents a thought experiment on ‘time risk of spectrum access’ as a hurdle to achieving optimal energy efficiency in a secondary spectrum environment where a device can evaluate whether to access a secondary spectrum opportunity, or wait for a future opportunity under non deterministic opportunities.

Secondary spectrum, from the perspective and designs employed by conventional User Equipment (UE) and Base Stations (BS) operating within fixed, licensed spectrum, would view a secondary spectrum environment as highly dynamic, discontinuous and hostile, effectively writing off this spectrum as inaccessible for their purposes. This perspective can be attributed to the comparatively longer time frames that exclusive spectrum licensees enjoy for spectrum ownership, managed secondary spectrum systems are designed to somewhat accommodate ‘classical’ spectrum access strategies. The non-continuous and somewhat unpredictable nature of unmanaged secondary spectrum, in particular, presents

unique challenges to the mobile networks and devices required to access it. Both managed and unmanaged secondary spectrum are landscapes which alternate between primary (incumbent) and secondary user (BS/UE) utilisation.

Secondary spectrum whitespaces are artifacts resulting from poor or inefficient spectrum utilisation. These inefficiencies could be exploited by both secondary users who require access to spectrum and primary users in search of additional revenue with excess spectrum available. To take advantage of secondary spectrum opportunities however, updated wireless devices will need to be constructed that meet the stringent requirements of this new spectrum environment.

To adequately perform within a secondary spectrum environment, significant improvements on existing wireless electronics is required, such as: detection, modeling, sensing, decision making, access control and coordination. At the most fundamental level these additional operations come at a cost: energy consumption and latency.

To establish a baseline for consumer-targeted secondary spectrum systems, the relationships between power consumption and latency of a modern mobile device OFDM receiver are identified and modeled. This model is then used to assess suitability and limits of simple secondary spectrum whitespace opportunities. The suitability of these opportunities is then assessed with respect to how fast a device can react to an opportunity and how much energy it is willing to expend to take advantage of it.

3.2 Factors Impacting Power and Latency

The relationships between power consumption, latency of operation, time of operation, and energy are important to understand going forward. The simplest representation of Energy (E) in electronic devices is a product of Time (t) and Power (P). For a component or device operating with a constant power dissipation over a given duration the total

energy consumed:

$$E = Pt \tag{3.1}$$

Latency is the time a system takes to respond to a given input. For example, if two functions are queued, and must be completed sequentially, then the second function cannot be initiated until the first one is completed. Thus the latency for the second function to begin is at a minimum, equal to the duration that the first function takes to complete.

In the study of reducing energy consumption, either power consumed can be reduced, or the duration of consumption reduced. By reducing the time of operation, or speeding up the time of operation, it is often the case that latency is positively impacted. Thus a focus on reducing time of operation has the twofold benefit of minimising energy consumption (assuming that power consumed does not scale linearly or worse with the tighter time constraint), and improving the overall responsiveness of a system. Thus, to reduce energy consumption for a given function (assuming P and t are not interdependent), either P or t can be minimised to minimise E .

3.2.1 Power

One of the primary success factors for a mobile device is how long it can operate without requiring to be recharged. As the energy capacity of any given battery is finite, there is a trade-off that needs to be made between performance/user experience and power consumption to maximise the mean time to charge for a given device.

The primary energy consumers in a modern smartphone are the display, applications processor (running applications and operating system) and the baseband processor (responsible for basic communications calling/texting and data connections LTE/HSPA (High Speed Packet Access)/WiFi). Strategies for reducing the energy consumption

within these devices can be achieved through both hardware and software means.

Software can be implemented as part of the phones' firmware, applications, or overall management, to reduce power consumption. Several of these strategies include: increasing memory consumption (or precomputed values) to minimise redundant clock-cycles/repeated computations required to achieve some output. Less clock-cycles consumed leads to a direct energy consumption improvement, however memory in modern devices is finite and some computational operations cannot always be precomputed in an effort to reduce computational overhead. Oftentimes significant energy efficiencies are achieved through the maturation and development of the operating system, functional libraries, component drivers and specialised hardware.

Reduction in hardware-based power consumption is often achieved or attributed to silicon node improvements, by shrinking the process to pack more transistors over a unit area, and/or transistor designs that operate at lower energy consumptions [90]. Modern System-on-a-Chip (SoCs) take this a step further and combine the benefits of various semiconductor processes, not just silicon, to minimise power consumption in specific tasks. Each semiconductor process selected is typically optimised for the specific mission that the component fabricated from it is required to perform [91].

Doped Silicon CMOS (Composite Metal Oxide Semiconductor) processes such as the MOSFET (Metal Oxide Semiconductor Field-Effect Transistor) or the more recent FinFET (Fin Field-Effect Transistor) are used for applications processing, due to their low power and maturation of fabrication processes from companies such as Intel [92], TSMC [93] and Samsung [94]. SiGe (Silicon Germanium), GaN (Gallium Nitride) or the commonly preferred GaAs (Gallium Arsenide) and RF-CMOS processes used for power amplification, mixing and multiplexing. MEMS (Micro Electro-Mechanical Systems) are also utilised for applications such as filtering, employing materials such as LN (Lithium Niobium Oxide) or LT (Lithium Tantalum Oxide). These substrates are com-

monly employed in resonators, filters and oscillators from companies such as Qorvo [95], Skyworks [96] and Broadcom [97].

Further advancements in the area of stacked integrated circuit (IC) dies, for example stacking a memory chip directly ontop of the processor chip, are becoming more commonplace in an effort to further reduce latency of operation, improve performance and further the miniturisation of devices [98]. Other stacking technologies such as current 3D NAND technology for flash memory are also intense areas of research and development to further capacities and reduction of power consumption [99].

Hardware operation strategies exist to reduce energy consumption as well. For example, reducing how long a component stays unnecessarily on for otherwise known as ‘racing to idle’ or power gating/dark silicon, and dynamic frequency scaling based on the current computational workload [100,101]. These functionalities are commonly achieved through a combination of hardware, software and firmware components.

Wireless protocol implementations can also have significant impact upon how power is managed. Mobile handsets at a high level, have two modes or states of operation when initiating and performing communications, these are the ‘control plane’ and ‘user plane’. Control plane operation is reserved for when a connection is initiated, for the handset to make itself known to the network, perform network updates, such as location and channel state information to ensure acceptable quality of service (QoS) is achieved and then operation is handed over to the user plane [102]. The user plane is typically where the majority of communication time and energy are expended by transmitting and receiving data. User plane operations cover applications such as: email, social media, voice calls, images, and streamed video/audio content.

3.2.2 Latency

Pedantically, latency can be caused within a communications system from almost every element of the system, hardware, software, and the environment. Environmental latency is primarily caused from propagation delay as the speed of light is finite, there will be some period of time that must elapse before a particular signal is broadcast from a transmitter and then received at a given receiver. This is somewhat negligible for terrestrial communications however becomes an issue during operations at much greater distances such as in space. A good example of where this is the communications delay between earth and the moon (approx. 384,400,000 m) where at the speed of light in a vacuum (299,792,458 m/s) takes approximately 1.25 seconds [103].

In spite of environmentally bound latency, other more prominent sources of latency are caused from electronics and software implementations. From an electronics perspective every electrical component has some delay inherent within it between an input and the resulting output. For single discrete analog components this is somewhat negligible, however when many components are arranged together in a complex system, this can be substantial and is examined further in Figure 3.2. Digital electronics are an even greater source of latency due to smallest instance of time that a digital component can respond to an input is denoted by its clock speed, the slower that frequency the longer it will take for the digital system to respond to a given input. This further compounds when many operations are required to achieve the desired result, such as with CR.

Oftentimes, software is written with many human conveniences thrown in to make the process of writing the software more tractable or comfortable. However the more convenient that software is to write there is often a performance penalty that must be paid, in terms of latency and energy consumption. This can be due to many unnecessary operations being performed that wouldn't have to have been performed if the software was written in the most efficient way possible. The downside addressing this issue, is to

generate very efficient software it often becomes extremely difficult and unpleasant for an engineer/developer to write, develop and maintain. Unless the latency is of the utmost importance, oftentimes a more convenient route will be taken, with the drawback of adding to the latency and energy consumption of the system. Advances on the hardware side of the equation can and often do, alleviate these penalties, but are not a substitute for good software development and optimisation practices.

Latency in the perspective of user experience can make a device seem unresponsive to inputs or actions requested of the device, this is experienced by the user as a phenomenon referred to as ‘lag’. ‘Input lag’ is the time taken between the time for a system to respond to a given input. Lag is also used as a common colloquial term to explain the degraded or unresponsive user experience of an application due to the intermittent, ‘bursty’, or untimely reception of data. Lag poses a considerable QoS issue particularly plaguing latency-bound applications such as voice communications, audio/video streaming and online gaming. In each of these applications, the ability to buffer information cannot be used in an effort to smooth out intermittent connectivity, and data retransmission events to make up for dropped packets (or timed out packets that are dropped) can often cause undesirable performance within these applications. The inability to address latency in these scenarios often renders the functions of these applications unbearable for the user which is a considerable concern for the commercial success of a given device or application.

Latency within a system can lead to desynchronisations that can cause program instabilities, resulting in system crashes or in the communications context, wireless connections can be dropped. This is undesirable as when connections are dropped, control plane signaling must be repeated to re-establish the communications channel. This adds further latency to the communication and consumes additional energy to successively receive the desired information. Thus it is important to remove latency as cause of system instability, particularly within a mobile device.

Latency poses a similar issue in the secondary spectrum context. In aggressive or highly dynamic secondary spectrum environments where opportunities only persist for short durations, we aim to show that the ability for a secondary device to detect and then communicate during that opportunity is heavily latency-bound. If a system responds too slowly, then the total spectrum capture and exploitation would be heavily degraded or not existent. Thus, it is important that to establish a baseline of the latency limits that current access technologies have when accessing primary spectrum under a variety of scenarios. This established baseline can then be applied in an effort to assess the upper bound of secondary spectrum capture that could be feasible for these systems.

3.3 OFDM Receiver Chain Model

To further explore the relation between latency and energy consumption, the structure of an OFDM receiver is broken up into major functional blocks to identify contributors of energy consumption and latency within the receive chain. Analysing a receiver in this manner enables us to determine which elements may pose the greatest impact to the energy consumption and latency of operation. The analysis additionally sheds light on functional limits for similar receiver designs that could hypothetically operate within a secondary spectrum context.

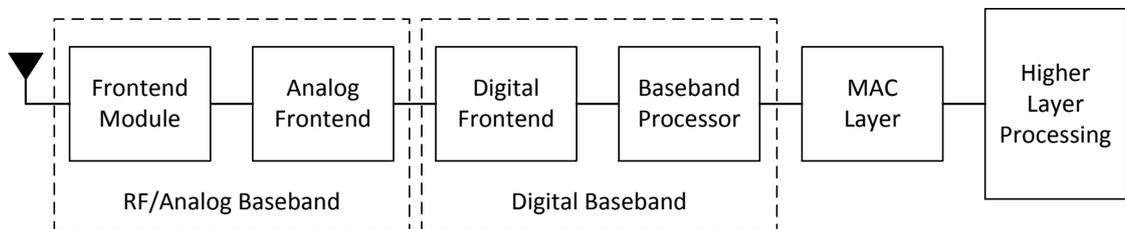


Figure 3.1: Received Signal Chain Block Diagram

An idealised OFDM receiver chain comprises of: a Front End Module (FEM), Analog Front End (AFE), Digital Front End (DFE), Baseband Processor and Medium Access Control (MAC). In typical implementations, the DFE is encompassed by the Baseband Processor. The AFE is also known as the ‘Analog Baseband’, and the DFE combined with Baseband Processor is also known as the ‘Digital Baseband’. Figure 3.1 provides a high level view of the functional processing stages:

In order to ascertain power and latency metrics for each of the subsystems in Figure 3.1, a literature search was performed to ascertain the state of the art for power and latency of mobile hardware. The time and power required to successfully receive data for each discrete processing stage is displayed within Figure 3.2. In reviewing the literature, it was found that there is little consistency in reporting of this data as different researchers were focusing on different aspects of the systems or the system as a whole rather than solely reporting power and latency according to this simplified system. The latency and power consumption values for the data points of the FEM and AFE sections of the graph were modeled off discrete components, rather than an integrated system. This is a reasonable approximation due to the similarities inherent within the front ends of each technology analysed [104–107]. Furthermore, and not surprising, the data from these sources demonstrate that the power consumption and latency become functions of the telecommunications technologies such that standards supporting higher data rates, or operating at higher data rates, required greater power consumptions than those supporting lesser rates.

To establish the baseline metrics of operation in Figure 3.2, the figures presented for MAC layer latency incorporates the time required for an ACK/NAK (Acknowledged/Not Acknowledged signal) to be transmitted, indicating successful reception of a packet and assumes no retransmission attempts (depending on technology standard) are made. The latency values shown are indicative of the latencies incurred by individual processing

stages and power consumption is presented as cumulative of previous stages. Thus to determine the discrete power consumption delta of the stage, subtract the previous stage from the final value. The latencies presented for the FEM, AFE, DFE and baseband processor are for receiving and successfully decoding a symbol. The identified receivers are often required to operate with very low amounts of available energy (within a mobile device context), requiring significantly limited periods in which the components were operating at 100% power draw.

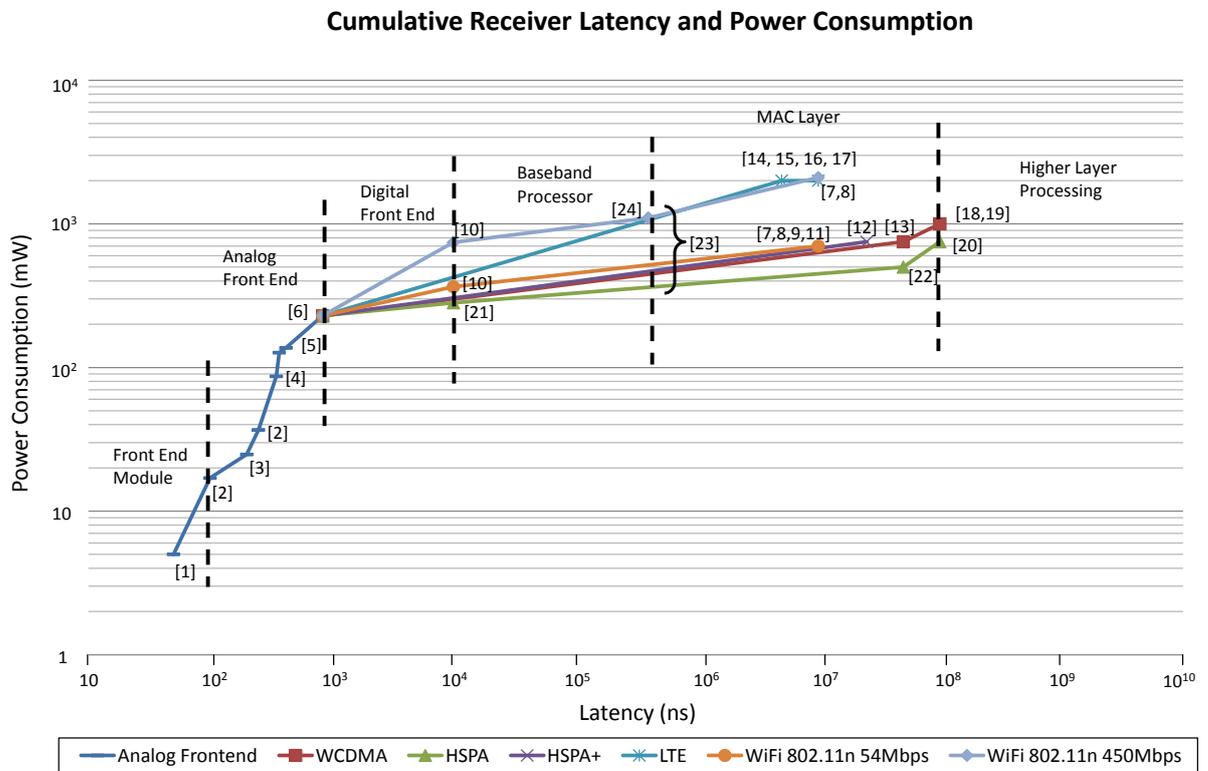


Figure 3.2: Receiver Latency and Power Consumption for User Plane Operation. Note that the references on the diagram are only correct for the original paper that the diagram was published within. See paper [6] for source inputs to the models.

An additional benefit of Figure 3.2 is the time and power required to extract successively more information from the received signal during user plane operation can be

visualised. As more information is ‘extracted’ from the underlying baseband signal there are higher amounts of energy and time required to ‘refine’ the signal into eventually, useful information. The series of operations and the ‘refinement’ of the information is as follows: within the FEM, the available information is an analog signal; within the AFE the available information are demodulated in-phase (I) and quadrature (Q) analog signals yielding constellation specific amplitude and phase information; in the DFE the analog signals are digitized and ‘soft bits’ are extracted (i.e. a rough classification of the mapped constellation point); the Baseband processor confirms the integrity of those bits (making them ‘hard bits’, reducing the error in the classification of the constellation point) and then maps the now hard bits to the appropriate physical channels; finally the MAC layer interprets useful information from the bits and can perform computational operations based on the bits.

The operation of the MAC layer when processing a raw signal to discernible information, is several orders of magnitude longer with respect to time, than any other processing stage. It can also be seen that from the FEM, AFE and DFE, over half of the power consumption of the receiver is accounted for within these stages, however these stages only take a 100^{th} to a $10,000^{th}$ of the time that the MAC layer requires to perform its functions. While the MAC layer consumes less power as a discrete stage than the front end components, the latency of its operation shows that significantly more energy will be consumed compared to the front end (PHY layer). This however only holds true if the front end components are only ‘on’ for the time at which they are required to complete their defined tasks, and are otherwise deactivated or ‘turned down’ when not in use. Deactivating or turning down components does present additional challenges for latency and power consumption particularly due to the ‘time till ready’ (i.e. operating correctly) and how the power consumption varies during this turning on period. Despite those challenges, strictly, if the front end components are active for the entire duration a device is active or

idle, then those components will likely consume more total energy than the MAC layer, due to the front ends' equal or higher power draw.

In summary, Figure 3.2 clearly identifies the baseband processing and MAC layer functions as interesting and viable targets for improving energy consumption and latency of operation.

3.3.1 Latency-bound Energy Conservation

Based on Figure 3.2, LTE has the lowest identified user plane latency (other than WiFi), and is also the technology of choice for 4th generation mobile telecommunications networks.

For LTE, the duration of a symbol is $66.7\mu\text{s}$ or, $6.67 \times 10^4\text{ns}$ [108]. When examining Figure 3.2, it can be seen that a symbol can propagate through the receive chain to the baseband processor within that symbol holding period. The signaling structure of LTE defines that a radio is to receive 2 slots worth of symbols, where a slot contains either 6 or 7 symbols. Each slot is 0.5ms, within a 1ms subframe. There are 10 subframes per each 10ms frame [109]. As a side note, the symbol duration is the same for both long and short cyclic prefix (CP) modes of transmission [108].

As a potential improvement and comment toward improving the operation of an LTE receiver: once the final symbol of the second slot for the scheduled subframe has been received and propagated to the baseband processor, the FEM, AFE and DFE (front end) could ideally be put into a low power state, until the next scheduled subframe. Scheduling information during data transmission is already communicated to the UE's Radio Resource Controller (RRC) as part of the BS's resource allocation algorithm [109].

Before the next scheduled subframe the UE would 'wake up' the front end prior to the first symbol being sent. Actively turning on and off the front end electronics could result in meaningful power savings due to the front ends high power dissipation and the need

for the front end to only be on for relatively short periods of time. Adopting this strategy in future front end control in combination with predictive scheduling, would significantly reduce the cumulative power dissipation of the receiver, reducing the net energy required to perform receive functionality.

3.4 Energy Efficient Secondary Spectrum Access

To determine the energy efficiency of wireless communications, a measure linking the total expended energy to the number of bits successfully transmitted and received during user plane operation is useful. This metric however, implies any energy expended during control plane operation as ‘wasted’ energy, as the energy expended does not translate directly to received or transmitted user data. Thus, it is desirable to limit or minimise the amount of time a radio spends in control plane operation, to maximise the energy spent in user plane operation. The added benefit of minimising time spent during control plane operation, is that the user plane operation can be commenced faster, reducing the overall latency of communication.

To apply this measure in the context of a secondary spectrum environment, where the opportunities for transmission are limited or heavily constrained in time, this measure can be used to determine the value or associated cost in accessing a discrete secondary spectrum ‘window’¹.

3.4.1 Simple Whitespace Opportunity Model

A model of a secondary spectrum whitespace opportunity or whitespace ‘window’ is presented in Figure 3.3, to evaluate the time and energy costs of accessing a secondary spectrum opportunity. LTE is used as the access technology for the purposes of the

¹A window within this context and going forward refers to a discrete secondary spectrum opportunity

model, and we assume modifications are made for operation within a secondary spectrum context.

This model assumes the downstream case (from the perspective of the UE), where either a UE has requested data to be sent to it from the BS or the BS is notifying the UE of an incoming payload. This model also assumes that every time a new whitespace opportunity is to be accessed, that the setup/control negotiation has to take place before data can be transmitted. The ideal whitespace window assumes that there is no sensing or channel migration time required by the UE so as soon as the window begins, the control period begins. This is possible if prior knowledge of the spectrum is available to both the UE and BS.

For a secondary spectrum window of length T_{ws} , a ‘receive frame’ exists between T_0 and T_{data} . The frame is broken down into three discrete periods:

1. Control plane (c-plane) latency, a fixed period between T_0 and $T_{c-plane}$ dependant on time to establish a connection between the BS and UE.
2. One Way Delay (OWD) [110] or user plane (u-plane) latency, a fixed period between $T_{c-plane}$ and T_{owd} . OWD is the latency incurred due to propagation delays in the downlink between the edge of the core network (CN) and UE.
3. Data/payload receive period, defined between T_{owd} and T_{data} .

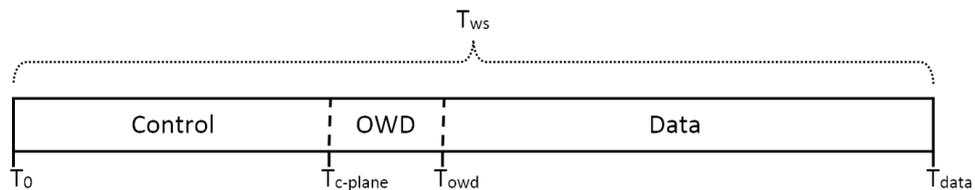


Figure 3.3: Whitespace Window and Receive Frame Model

The points on the model are as follows:

- T_0 : point in time where transmission begins in a pre-identified whitespace opportunity.
- $T_{c-plane}$: time for handset control plane functions to complete, switching the handset from idle to active. This point is typically a fixed duration from T_0 .
- T_{owd} : time taken for a packet to propagate from the edge of the CN to the UE.
- T_{data} : time taken to receive requested data.
- T_{ws} : length of whitespace window, can be arbitrary, however is finite and small for purposes of dynamic secondary spectrum.

OWD in this context does not account for other external latencies incurred, other than propagation delay in order to keep the model simple. Refinements of the model can include latencies incurred by variables external to the UE, BS and CN, provided that they do indeed add to the period of no whitespace window utilisation between the control plane and user plane transition. Ideally these sources need to be accounted for and quantified when determining appropriate whitespace windows to access. These refinements can be accommodated in the existing model by simply extending the OWD duration to encapsulate their impact.

The control period is dependent on the particular mobile technology (in this case LTE) in question to perform connection setup between the UE and BS to toggle the UE from the ‘idle’ state to the ‘connected’ state.

3.4.2 Cost of Spectrum Access

Utilising the model presented in Figure 3.3 the energy cost of accessing a whitespace window can be expressed as a function of the window duration, bounded by communication setup time (control plane latency). Note that the cost of access only accounts for energy

expended during control plane operation (i.e. ‘wasted’ energy, that must be amortised across the energy cost of transmitting the data payload) and does not include the energy cost of receiving user plane data.

If the length of the window is sufficient for all necessary data to be transmitted within that window, then only a single setup cost is incurred. If the window is smaller, then multiple setup costs will be incurred. If the window is smaller than the required setup time, then it is impossible to transmit any data. The following equations describe the energy cost of whitespace access.

$$E_{cost}(T_{ws}) = \begin{cases} T_{set}P_{set} & \text{if } T_{set} + T_{len} \leq T_{ws} \\ \lceil \frac{T_{len}}{T_{ws} - T_{set}} \rceil T_{set}P_{set} & \text{if } T_{set} < T_{ws} < T_{set} + T_{len} \\ \infty & \text{if } T_{set} \geq T_{ws} \end{cases} \quad (3.2)$$

Where:

- E_{cost} : total amount of energy required by the required setup period(s)
- T_{set} : setup time (c-plane latency and OWD)
- T_{len} : required time to successfully transmit all required data
- P_{set} : power dissipated during the setup

This cost assumes that subsequent whitespace windows of duration T_{ws} exist during case 2. It can be seen by inspection that as the duration of T_{ws} is shorter than 50% of T_{len} the cost of access grows rapidly as $(T_{ws} - T_{set})$ approaches 0. At $T_{set} \geq T_{ws}$ the window is too small to even allow for connection to be established, thus the cost of accessing a series of windows of that length (if not impractical), is infinite.

To offset the cost of connection set up (E_{cost}), a receiver must receive a certain amount of information or rather for a certain period of time, before the energy expended to begin the connection is justified. The simple cost model presented in Eq 3.2 is applied to several

control plane latency scenarios by varying T_{set} , to assess the range of spectrum window durations that would present a viable opportunity to access under each scenario. As a reference, 100ms is the c-plane latency requirement for LTE [111].

Receive efficiency within this context is the energy cost of setup plus the energy cost of receiving requested data, divided by the energy cost of receiving data. The energy cost of receiving data is calculated under the assumption that the receiver is operating at maximum power consumption during both control and user plane operations, with the consumption figures presented in [112].

In Figure 3.4 the power draw is kept constant for both control and user plane operation for simplicity. Thus the power draw during control and user plane operation can be canceled out. This yields the resulting expression to determine efficiency as: the ratio of time spent in control and latent operation, versus time spent during user plane operation for various window durations.

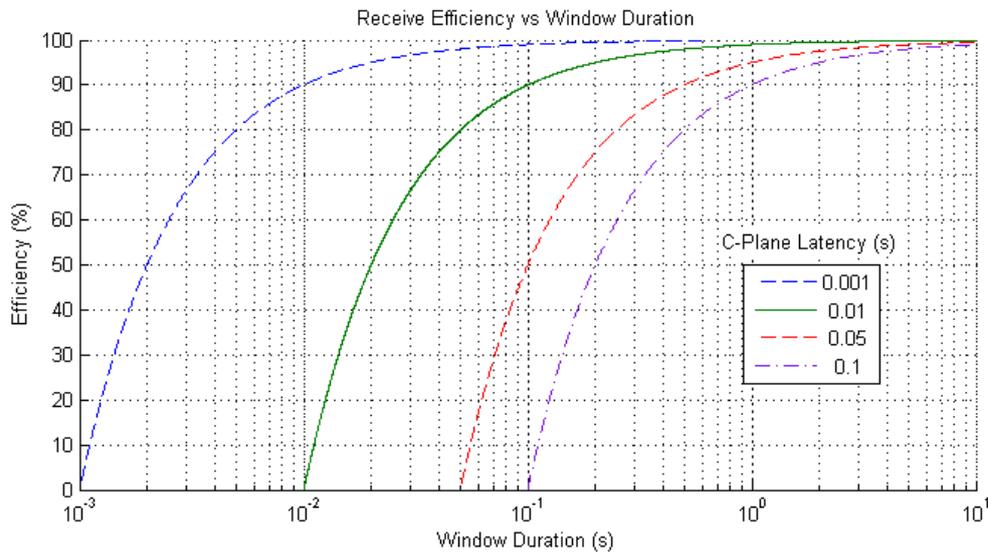


Figure 3.4: Setup versus Data Receive Energy Efficiency Curve

The results from Figure 3.4, show that the required window for efficient receiving of data, is tightly coupled to the control plane latency. This is due to the energy spent to

establish a connection, requires significantly more energy to be expended when receiving data, to reduce the connection overhead as a percentage of total energy. As energy is the product of time and power, a significantly longer data receive period is required, which increases the required duration of whitespace window to make the receive operation viable. Inefficient whitespace accesses will significantly degrade the operational time of the device, due to the batteries limited energy resources when receiving a defined amount of data.

The efficiency values presented in Figure 3.4 are shown as a simple example to detail energy efficiency with respect to window duration, under scenarios where the control plane operation expends energy at a much faster rate than the user plane operation, these values would vary. Under scenarios where the control plane consumes higher amounts of energy, this behaviour would significantly skew the efficiency curves to the right, requiring a longer duration opportunity in order to gain higher energy efficiencies over the cost of access.

The cost of access model assumes that every time a new whitespace window is to be accessed, that the access cost is incurred. This cost could be reduced if the UE and BS both had a priori knowledge of the next window to ‘jump’ to when the current window reaches its end (provided channel conditions allow it). The current active connection would be carried over to a different identified frequency in a similar method to how Bluetooth utilises frequency ‘hopping’ to combat interference without needing to renegotiate the connection with each frequency hop.

In a highly dynamic whitespace environment, significant increases in data throughput would be achieved versus a more latent receiver, particularly when the window is shorter than the c-plane latency of the more latent receiver. A receiver with low c-plane latency can efficiently access a larger number of potential whitespace windows due to it having to expend less energy to access additional windows to receive all requested data.

Extrapolating from Figure 3.4 and Eq 3.2 to a sequence of whitespace opportunities in a dynamic spectrum environment, the efficiency of spectrum access must also factor

in the real distribution of secondary spectrum opportunities and the duration of these opportunities. The distribution of opportunity durations can be used as a metric to determine the required control plane latency, that can be tolerated to operate in various secondary spectrum environments. This forms much of the focus of Chapters 4 and 5.

3.4.3 Time Risk of Spectrum Access

The time risk of access, is a thought experiment that extends Eq 3.2. In this experiment the priority of a payload is determined by its type and secondary spectrum opportunity duration and ordering is known to a degree. The ‘time risk’ is a measure of the tolerance of the system to delaying a communication in the interests of energy efficiency, with the expectation of capturing a longer contiguous whitespace opportunity in the future. The thought experiment is how to evaluate the time and/or energy penalty that is incurred by accessing a currently available whitespace opportunity and accepting the risk of that opportunity abruptly terminating or waiting in the interests of finding a longer duration whitespace opportunity, where, perhaps, greater energy efficiency can be achieved.

For example: if the secondary spectrum environment is very sparse, then there would be a high time risk associated with each opportunity, as missing an opportunity would incur a large time penalty which may be unacceptable for certain payloads. Conversely, in a whitespace abundant environment, there would be quite low time risk as there are an abundance of opportunities, however some opportunities may end abruptly or before the payload has completed, impacting efficiency.

In the absence of a well defined time risk of access scenario, either due to lack of information about a spectrum environment or the users occupying it, systems that are largely unaware of their spectrum environment can access the secondary spectrum in a couple ways, but with no objective of explicitly reducing energy consumption. One strategy is to use pre-determined frequency hopping patterns (such as in bluetooth) to avoid/reduce

collision with incumbents within a given spectrum, capturing some whitespace opportunities. In the second case, likely where energy consumption is a priority, random back-off and sensing strategies can be utilised, where sensing is required to be performed before communication can take place. If during a sensing period, an incumbent is detected the device will non deterministically wait for an opportunity to perform sensing again, in the hopes of detecting a whitespace opportunity at some future point for which to communicate over (such as in WiFi).

In a more well defined secondary spectrum scenario however, the risk to be calculated is the trade-off between the energy cost of accessing a current opportunity with a particular estimated duration versus waiting for a period to access a future opportunity with a greater duration. Payload efficiency in this case is dependent on the amount of data to be transmitted, the duration of the transmission period and the priority of the payload with respect to the actual whitespace opportunities captured. The performance of this approach is highly dependent on the data available and the predictive power of the whitespace and incumbent analysis to minimise interference while also minimising energy consumption.

Sensing is required by the device to determine when an incumbent is no longer present, and similarly to detect when an incumbent returns to the spectrum. Sensing however, is an expensive function and should be minimised to preserve energy. By reducing sensing intervals however, the delay between spectrum state updates is increased, which can either put the device at risk of interfering with an incumbent if it retakes the spectrum inbetween intervals. Or in the other case, where an incumbent releases the spectrum and a significant portion of the whitespace opportunity is wasted.

3.5 Summary

This chapter has presented general models for the power consumption and latency relationship of conventional radio receivers. The concept of whitespace windows within the context of secondary spectrum was outlined. Finally, the connection between the latency of a wireless system with respect to accessing electromagnetic spectrum and the performance implications for that device adapted to a secondary spectrum context were explored.

Section 3.3 details an OFDM receiver model and uses this model to perform an accompanying a survey of several OFDM receivers, providing a discretised breakdown of the receive chain with the power and latency incurred per stage. Figure 3.2 illustrates that of the technologies surveyed, technologies with higher power consumptions, have considerably lower latencies of operation, particularly with respect to the MAC layer latencies. Further, the technologies that have higher power consumptions, have higher theoretical throughputs over their lower powered counterparts.

The analysis of current OFDM technologies was fundamental to understanding hypothetical device performance with respect to spectrum reuse. The whitespace access cost model is then proposed to illustrate the cost of accessing whitespace with the knowledge of receiver power and latency constraints, and incorporating those performance metrics to make energy efficient spectrum access decisions.

Section 3.4 introduces a simple whitespace opportunity and energy cost model, that investigates the role of receiver latency with respect to the range of secondary spectrum opportunities that become viable as a function of that latency. The results of Figure 3.2 highlight that unless a receiver is constantly receiving information for every given point in time, there are substantial power savings and hence energy consumption reductions that can be realised due to latency of operation.

Finally a thought experiment in Subsection 3.4.3 is presented, connecting the timescale

at which a device operates with an underlying secondary spectrum model. The experiment posits how a device can make decisions on the secondary spectrum opportunities that it captures, via assessing the risk and reward of the decision. The energy required to seize an opportunity can thus be modeled as a function of the risk of the opportunity terminating, and the time risk between not taking advantage of the current opportunity and instead waiting for (an expected) next opportunity.

This Chapter has shown that future receiver and telecommunications standards will need to emphasise the reduction of UE receiver latency of operation as a means to reduce energy consumption. This is of particular importance for devices operating within dynamic whitespace environments, as decreasing control plane latency will result in higher device energy efficiencies and potential accessible ranges of whitespace opportunities.

Chapter 4

A High Resolution Spectrum Measurement System

4.1 Introduction

The study of short timescale secondary spectrum opportunities requires high resolution spectrum measurements and a suitable system to perform the measurement. High resolution spectrum measurement in this context, and going forward, is achieved with continuous direct sampling of the spectrum of interest, capturing the entire observation bandwidth continuously for an extended duration (days or more). High resolution spectrum measurements enable a discrete and progressive time analysis of the spectrum with relatively small time steps and small frequency bins to be performed. One can liken these small time steps and frequency bins to the size of pixels in an image, generally, the smaller the pixels the higher resolution the image, hence ‘high resolution’ spectrum measurement. Sparsely sampled spectrum would therefore yield ‘low resolution’ measurements of the spectrum.

To capture these measurements, a sufficiently well equipped radio and analysis plat-

form must be either built or obtained to measure, record, process and analyse the measurements. Simply performing simulations on arbitrary spectrum models is not sufficient to investigate the optimal timescale of operation for secondary systems in a real world context.

Section 4.2 provides a background into existing spectrum measurement studies and highlights several issues that arise with those existing sets of measurements and measurement techniques with respect to the investigation of optimal timescale operation within a secondary spectrum environment. Section 4.3 details the processing stages required to take raw samples from a radio receiver and compute those samples into its power spectrum and the resulting Power Spectrum Density (PSD) plot that can be used for energy detection and further spectrum analysis. Section 4.4 provides detailed background into the principles and performance of Energy Detection (ED) based sensing for spectrum measurement and classification of unknown signals, the impacts of threshold selection on the performance of the energy detector are then explained.

The primary contributions in this chapter step through the development and improvement of a low cost spectrum measurement system using energy detection based sensing for secondary spectrum observation. This system can be used for general spectrum observations and more particularly, is used for determining the optimal timescale of secondary spectrum operation.

Section 4.5 details the requirements, design and implementation of a low cost spectrum measurement system, including the detectors architecture, performance, hardware and software stack implemented for effective operation. Section 4.6 presents a novel image based noise filter as the second contribution for this chapter. Utilising the capabilities of the newly implemented spectrum measurement system, and then performing energy detection upon those measurements, the novel filtering approach significantly improves the accuracy of unknown signal detection and reduces the probability of misclassifying

low SNR events.

Chapter 5 uses this same measurement system to conduct secondary spectrum observation of multiple bands of interest and details the subsequent analysis of these observations.

4.2 Existing Spectrum Measurement Studies

Numerous studies have been performed across the world to collect whitespace measurements and provide the subsequent analysis of particular frequency bands [3–5, 64, 113]. The typical measurement strategy employed in these studies perform a frequency ‘sweep’ of the observable spectrum, rather than a continuous or ‘static’ sampling of a single band of interest.

A sweep is performed by progressively ‘scanning’ over a band of interest, using a spectrum analyser with a lower capture bandwidth (instantaneous bandwidth) than the spectrum band of interest being observed. Progressive snapshots are taken from the observable spectrum as it is scanned, equal to the spectrum analyser’s instantaneous bandwidth (or bandwidth ‘increment’). Once a ‘snapshot’ of the current bandwidth increment is captured, the next bandwidth increment is then captured until the entire band of interest is observed. Once the whole band of interest is captured, the spectrum analyser then steps back to original starting frequency and repeats the process continually until the observation is stopped. In some cases these sweeps can take several minutes per complete sweep due to the observed bandwidth being many times larger than the instantaneous bandwidth of the measurement systems used. Other reasons for the large time in between sweeps is also due to the settling time of the measurement hardware, and multiple observations are typically captured per increment for noise suppression.

The advantage that swept datasets have over static datasets (measurements taken where the spectrum analyser focuses on a single frequency range equal to its instantaneous

bandwidth), is that much more of the observable spectrum can be captured without requiring a spectrum analyser with an instantaneous bandwidth on the order of Gigahertz, which is useful when gaining a general understanding of how a given spectrum is utilised. Swept datasets are, however, ill suited for the study of optimal timescales due to their fragmented observation periods breaking the continuity of measurement of a single band.

Measurement timescales presented for bands such as TVWS [64] are on the order of days, weeks or years. Even though the measurements would have been instantaneously captured with sufficient time and frequency resolution, the measurements are not sustained or continuous. The discontinuous measurements are simply a series of snapshots stitched together to form the final dataset and this dataset is then used to evaluate the spectrum. As the study is not searching for smaller timescales less than the order of days, having a dataset with minute or hour spaced gaps is negligible. This data is however, ill suited for the investigation of optimal timescales.

To illustrate further: a measurement device may sample each spectrum increment for 5 seconds and takes a total of 5 minutes to complete a sweep. This dataset is then stitched together and used to determine spectrum occupancy. Measurements at this scale cannot be used to inform a UE of operational parameters such as latency of operation or power consumption requirements, due to these timescales being on the order of milliseconds. The dataset captured does not have sufficient fidelity in order to model the performance of those device parameters within that spectrum environment.

Additionally, the possibility of missing short timescale incumbents during sweeps becomes quite apparent due to the considerable amounts of short timescale data that is not actually observed:

- A dataset constructed using this sweeping method will be compiled with considerable information gaps within it (due to missed events in-between sweep periods), rendering it unsuitable for any investigations for incumbent events below the sweep

time of the system.

- As an added drawback to spectrum sweeps, when the spectrum is analysed in many of these studies [3, 4, 64, 113], small FFT lengths (or a small FFT depth ¹) are used, which provide a short timescale snapshot of the spectrum, however these observations are discontinuous as they are generated from stitched together sweep events, so the timescale that could be inferred from the observations is inaccurate.
- The second drawback of small length FFTs is the size of the resulting frequency bins. These large bins limit the prevalence of incumbent activity with bandwidths smaller than the bin widths, as the observed incumbent power is averaged across the entirety of the bin. This results in potentially significant errors in determining the availability of secondary spectrum, specifically with respect to the classification of contiguous chunks of whitespace (see Section 5.4).

Due to the shortcomings identified, most existing spectrum measurement datasets and databases suffer from these drawbacks, making them unsuitable for optimal timescale analysis. Additionally, independent measurements can be captured with sufficiently higher fidelity (sample rates and continuous observation periods) locally, compared to readily available online spectrum databases [114, 115]. This is also due to the very large amounts of data generated by capturing that fidelity. Thus, a system capable of performing high resolution spectrum capture, retrieval, and analysis, is necessary to investigate second, and sub-second secondary spectrum structure, to determine optimal timescales of operation.

¹An ‘FFT depth’ is the number of points used when calculating the FFT.

4.3 Spectrum Analyser Processing Stages

To compute the spectral power density from a series of spectrum measurements, several parameters must be taken into consideration to remove bias from the calculations and to ensure to overall correctness of power estimation. This section steps through the signal processing steps that were taken by the measurement system in performing power estimation. The processing stages of a spectrum analyser after spectrum samples are captured by a receiver are shown:

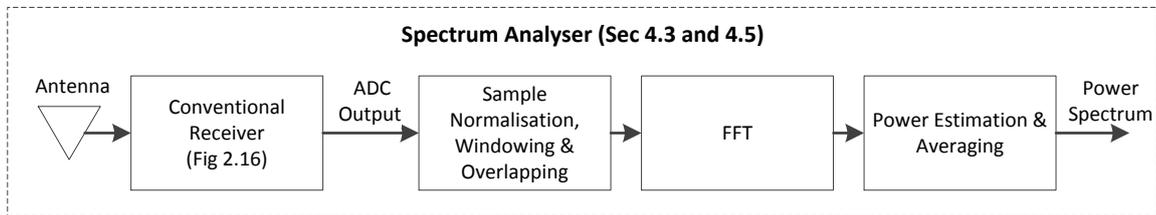


Figure 4.1: Spectrum Analyser Processing Stages

4.3.1 Sample Normalisation

Time series samples must be appropriately preprocessed before an FFT can be accurately computed. The complex sample $x[n] = a[n] + ib[n]$, is divided by the maximum output value of the ADC $= 2^r - 1$, where r is the number of bits of resolution the ADC supports. This result is then multiplied by an appropriate windowing coefficient.

4.3.2 Windowing

The samples must be windowed after normalisation to reduce the prominence of artifacts and power leakage due to the time series being abruptly truncated prior to the samples being fed into the FFT algorithm. The window of choice in this context is the four term Blackmann-Harris Window (BHW) [116] due to its superior frequency selectivity

characteristics and side lobe suppression. The BHW coefficients are computed as follows where N is the number number of samples (or FFT depth):

$$w_{BH}[n] = a_0 - a_1 \cos\left(\frac{2\pi n}{N}\right) + a_2 \cos\left(\frac{4\pi n}{N}\right) - a_3 \cos\left(\frac{6\pi n}{N}\right) \quad (4.1)$$

$$a_0 = 0.35875; a_1 = 0.48829; a_2 = 0.14128; a_3 = 0.01168$$

The power contribution of the window can be determined and removed from the final power calculation by:

$$P_{w_{BH}} = \frac{1}{N} \sum_{n=0}^N |w_{BH}[n]|^2 \quad (4.2)$$

It is of note however, that the BHW requires a 66.1% sample overlap for the best trade-off between amplitude flatness and overlap correlation of the windowed signal as shown in Figure 4.2. At this overlap percentage however, the power flatness is approximately 0.7 which is not ideal for power estimation. A significantly higher overlapping will result in a little to no improvement in amplitude flatness while considerably increasing the power flatness up to an 80% overlap where ideal power flatness is achieved. Conversely, the correlation between windowed samples also increases to a 0.6 correlation between samples at 80% overlapping, resulting in wasted computation. An overlapping factor of approximately 66% is a useful minimum level for the best performance vs accuracy trade-off for computation of successive FFTs.

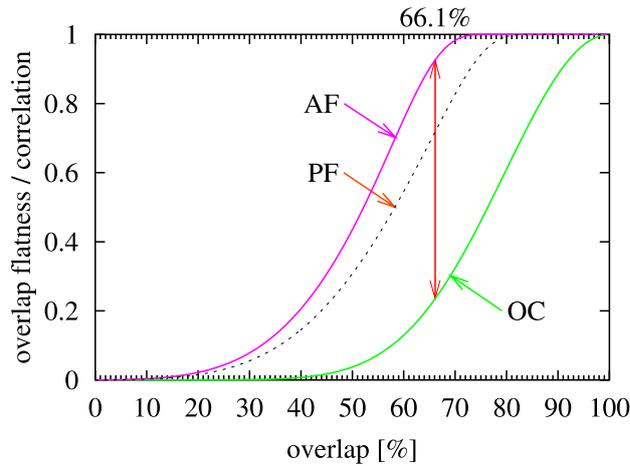


Figure 4.2: Amplitude Flatness (AF), Power Flatness (PF) and Overlap Correlation (OC) of a 4 term Blackmann-Harris Window [116]

Amplitude Flatness (AF) is the ratio of the minimal total weight that may be applied to any data point, with respect to the maximal total weight. It evaluates the disparity in weighting of samples due to a windowing function being applied. As the samples being taken are expected to have the same validity, the curve of summed window values should be as close to 1 as possible. Typically, by increasing the overlap of samples between windows, the curve of summed windows tends to flatten out (i.e. approach 1 or optimal flatness) removing any bias applied to the samples due to the windowing function.

Power Flatness (PF) is similar to amplitude flatness, however the square of the windowing biases are averaged, to take into account incoherent signals (i.e. non sinusoidal signals, such as noise) [117]. By increasing the overlap of the samples, the biases introduced from the windowing function on the power calculation can be minimised, resulting in greater accuracy of measured instantaneous power.

Overlap Correlation (OC) is a measure of how correlated subsequent spectrum estimates (spectrum frames) are upon evaluation. If the overlap of a signal becomes too great, spectrum estimates becomes strongly correlated, resulting in wasted computational effort. Thus it is also necessary to minimise OC to eliminate unnecessary processing of

highly correlated samples.

Overall, sample overlapping depends on the stationarity of data. If the data is in fact stationary, then there is no need to perform overlapping [116]. As the data series we are dealing with is non-stationary, overlapping is essential to remove bias from the measurements. Flatness is thus chosen to be as high as possible, while minimising correlation for efficient and bias minimised spectrum evaluation.

4.3.3 Overlapping

The effect that overlapping has on the samples is that the total number of output frames will substantially increase. If $k \geq 1$ is the total number of FFT output frames in the dataset, then the total number of frames created J from the original number of discrete frames k due to overlapping is:

$$J = 1 + \frac{k - 1}{1 - \%overlap} \text{ frames} \quad (4.3)$$

For our purposes an 80% overlap is selected to achieve optimal power flatness while keeping the computational requirement minimised. This aggressive overlap has a relatively high correlation between discrete frames of 0.6. This level of overlap and correlation (or wasted computation) is reasonable for our configuration as the computational power afforded by the measurement system can handle the increased load sufficiently, with the benefit of increased power flatness, increasing estimation accuracy.

4.3.4 Fast Fourier Transform

Once normalisation, windowing and overlapping is completed, the FFT can then be computed on the samples.

The Fast Fourier Transform (FFT) is a computationally efficient method of calculating the Discrete Fourier Transform (DFT) of a sequence of values (discrete time based measurements) into a set of characteristic frequencies and magnitudes present within the sequence. The FFT will compute the exact same result as evaluating a DFT according to its definition where $\{x_0, \dots, x_{N-1} \mid x \in \mathbb{C}\}$:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N} \quad k = 0, \dots, N-1. \quad (4.4)$$

The evaluation of the DFT directly, yields an upper bound of $\mathcal{O}(N^2)$, whereas it is well known that the FFT can correctly compute the same result in $\Theta(N \log N)$ [118]. To maintain this efficiency N must be a multiple of 2^m , where m is a positive integer.

Applying this to spectrum measurement, enables the frequency components of a series of time samples (measurements) to be extracted, providing insight to the occupancy of a given spectrum with respect to frequency. In spectrum measurement, the domain of the output spectrum from the FFT is represented from $[-N/2, N/2 - 1]$. This causes a flipping of the right hand and left hand sides of the DFT, placing the DC (Direct Current) component (0 frequency in the present case of direct down conversion) at the centre of the domain.

As this transform is from the time domain to the frequency domain, there is an evident trade-off between time and frequency resolution. The number of complex samples N , requires an amount of time τ to collect based on the sampling frequency f_S . In this work it is also referred to as the ‘timescale’ of measurement, as a smaller time increment cannot be resolved without modifying N . Thus, when an N is chosen:

$$\tau = N/f_S \text{ seconds per frame} \quad (4.5)$$

The chosen value for N will also define how many ‘bins’ or ‘frequency slices’ the resulting spectrum is broken up into, this is also referred to as the ‘Resolution Bandwidth’ (RBW). The advantage of increasing N yields a higher spectral resolution (number of slices) given by:

$$\beta = f_s/N \text{ Hz per bin} \quad (4.6)$$

4.3.5 Power Estimation

The power for each FFT frame is calculated by taking the magnitude of each complex bin result within that frame. This is then converted to fullscale power in dB and normalised by dividing the total number of bins that are present in the frame (equal to the number of samples N):

$$P_{n_{FS}} = \frac{10 \log_{10} \sqrt{Re_n^2 + Im_n^2}}{N} \quad (4.7)$$

The final power figure P_n for each bin, has a final ‘calibration’ value added to it, to correctly position the fullscale measurement with respect to a power value in dBW. As dBFS is unitless, the fullscale value can be added to the maximum detectable signal of the ADCs; the additional power bleed due to the window function can be accounted for; and the front end gain can be subtracted. The maximum detectable signal (P_{SAT}) as previously mentioned, forms the upper bound for the detectable power value. Once all these are accounted for, the resulting value is the correct power value in dBW for the bin in question.

$$P_{Cal}(dBW) = P_{SAT} - G_{FE} + P_{WBH} \quad (4.8)$$

The final power value is achieved as follows:

$$P_n = P_{n_{FS}} + P_{Cal} \quad (4.9)$$

These values can then be plotted alongside their appropriate frequency bin, as a power spectrum. Sequential power spectra can be averaged over an observation period to create a power spectrum density plot (PSD), similar to the one shown in Figure 4.5.

4.4 Energy Detection Based Sensing

Even though energy detection is somewhat ill suited for highly accurate spectrum measurement (see Figure 2.17), the complexity of implementing an energy detector is a good test of the systems basic functionality and does provide a robust capability for performing (low precision) spectrum measurement.

An energy detector performs incumbent detection by comparing the spectral power density of an observation (typically captured by a spectrum analyser) with a predetermined (or otherwise calculated) power threshold. Energy detection based sensing systems have lower levels of accuracy when compared to other sensing methods, but are however largely preferred due to their simplicity in implementation and function [16, 84].

The primary challenge with energy based detection within a secondary spectrum observation context is that the measurements are being performed under uncertainty. That is, that there are no deterministic signals such as training sequences available for the energy detector to train on in order to achieve accurate noise estimation. Thus, this uncertainty in the contributing effect of noise on the power estimation becomes a key limiting factor on detector performance and threshold selection [119–122].

Figure 4.3 provides an overview of the processing stages required for spectrum observation and subsequent energy detection. The spectrum analyser comprises of a conventional receiver (Figure 2.16) and the necessary processing (Section 4.3) to perform power estimation. Once the power spectrum has been estimated from the observations, these are then input to the energy detector to perform classification (Sections 4.4.1 and 4.6) of the observations.

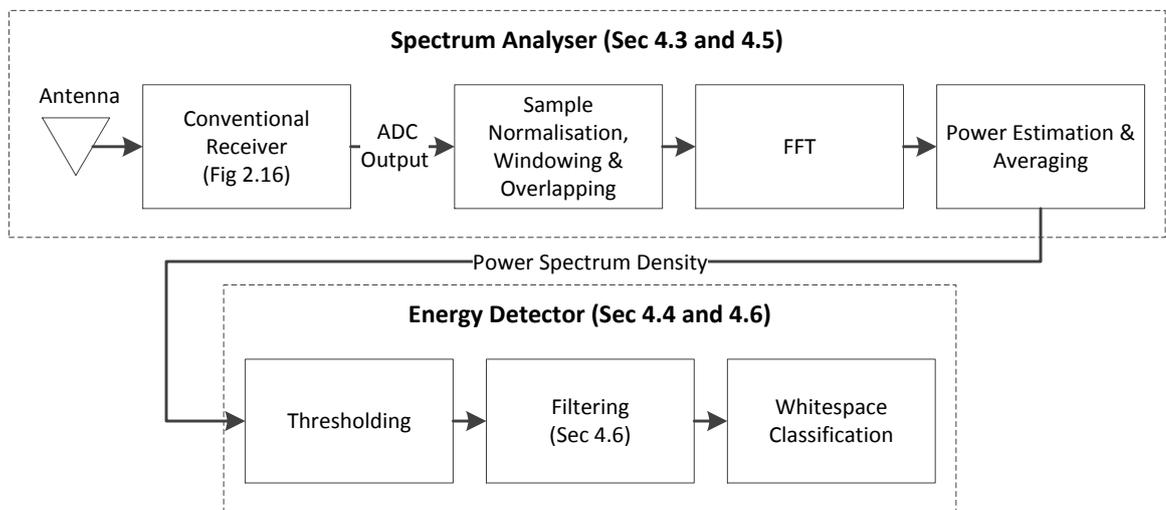


Figure 4.3: Architecture of Spectrum Analyser and Energy Detector Processing Stages

4.4.1 Energy Detector Performance

An energy detector used for the purpose of determining the difference between an incumbent signal and just noise, is an implementation of a binary classifier. A binary classifier is designed to separate observations by means of a particular threshold selected to discriminate between an observation being either true, or false. For example, a single power measurement and using that recorded power value to determine whether the power in that observation is the result of an incumbent being present and transmitting at that point of observation, or simply noise [119–121].

Ideally, the true positive rate for a detection system would be 1 and the false positive rate would be 0, however in practice this is rarely the case. The threshold must be selected such that, for a given sample set or set of observations Y , the false positive and true negative rates are minimised.

To define the performance of an energy detector more concretely, we evaluate the performance of the detector in the frequency domain after an FFT is performed on the time series observations. For $\{y[n] \in Y \mid n \in [0, \dots, N - 1]\}$ where N is the number of samples, the energy detectors' objective is to discriminate between the two hypotheses:

$$y[n] = \begin{cases} w[n] & : \mathcal{H}_0 \\ s[n] + w[n] & : \mathcal{H}_1 \end{cases} \quad (4.10)$$

To model the performance of the energy detector, it is assumed the channel noise $w[n]$ is Additive White Gaussian Noise (AWGN), distributed by ²: $w[n] \sim \mathcal{CN}(0, 2\sigma^2)$, with mean equal to 0 and identical independently distributed (i.i.d.) Real and Imaginary parts, each with variance σ^2 .

As the data collected from spectrum measurement is unlabeled i.e. it is unknown or unknowable whether an incumbent event occurred during a given observation, there is no feasible concrete validation method to determine whether an incumbent detection event is a true positive or a false positive. Thus to attempt to determine whether an incumbent event may have occurred at a given point, it can be concluded that any observed energy somewhere above the noise floor of the receiver and channel noise is likely due to an incumbent event/signal being present. Thus is necessary to select a classification threshold or specific energy level (perhaps conservatively) to confidently discriminate between an unknown incumbent event/signal and channel/receiver noise.

In the greater context of Cognitive Radios, the primary goal of setting a detection

² $X \sim \mathcal{CN}(\mu, \sigma^2)$ denotes a parameter X distributed by a Complex Gaussian Random Variable.

threshold is to ensure that an incumbent event is not missed, as the cost of missing an incumbent event and interfering with them (i.e. a false negative or misdetection) is undesirable. To select an optimal threshold we want to be certain that an incumbent event is a true incumbent event and not just noise (i.e a false positive, or false detection) by minimising the probability of a false detection (\Pr_{FD}) under \mathcal{H}_0 . Additionally, the probability of misdetection (\Pr_{MD}) under \mathcal{H}_1 is also to be minimised, by maximising the probability of detection (\Pr_D).

$$\Pr_{MD} = 1 - \Pr_D \quad (4.11)$$

We define the false detection and misdetection probabilities under both hypotheses, where λ is the detection threshold and $\Pr[\cdot]$ is an event probability:

$$\begin{aligned} \Pr_{FD} &= \Pr[\Lambda > \lambda \mid \mathcal{H}_0] \\ \Pr_{MD} &= \Pr[\Lambda < \lambda \mid \mathcal{H}_1] \\ \Pr_D &= \Pr[\Lambda \geq \lambda \mid \mathcal{H}_1] \end{aligned} \quad (4.12)$$

For both hypotheses, it is to be noted that the value for λ is equal and thus needs to be selected appropriately to minimise \Pr_{FD} and \Pr_{MD} .

As we are interested in the binned frequency location of an incumbent signal, energy detection is performed using the following metrics on a sequence of FFT frames Y . The discrete FFT bin $y[k, f]$ is the result from a frequency ascending n point complex FFT where k is the sequential bin number $\{k \in \mathbb{Z} \mid k = 0, \dots, n\}$ and f is a discrete time step (FFT frame).

The test statistic Λ represents the average power captured by the detector, for each FFT bin, calculated where $\{f - m \mid m \geq 1\}$ is a previous FFT frame (or timestep) from the current frame f , ($m = 0$) and A is the averaging depth (or number of frame lags averaged

over) [119–121]:

$$\Lambda = \frac{1}{A} \sum_{a=0}^{A-1} |y[k, f - a]|^2 \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\geq}} \lambda \quad (4.13)$$

Under \mathcal{H}_0 , Λ can be modeled as a chi-squared distribution with $2A$ degrees of freedom [121]. Thus P_{FD} is the Cumulative Density Function (CDF) of the chi-squared distribution:

$$\Pr_{FD} = \frac{\Gamma(A, \frac{\lambda}{2\sigma_w^2})}{\Gamma(A)} \quad (4.14)$$

Where $\Gamma(z, x)$ is the lower incomplete gamma function.

Since Λ is a power function, comprising a sum of $2A$ squares of independent and non-identically distributed Gaussian Random Variables with non-zero mean, Λ follows a non-central chi-squared distribution with non-centrality parameter $\gamma = \frac{\Lambda}{2\sigma^2}$ [119–122]³.

$$\Pr_{MD} = 1 - Q_A \left(\sqrt{2A\gamma}, \frac{\sqrt{\lambda}}{\sigma_w} \right) \quad (4.15)$$

$$\therefore \Pr_D = Q_A \left(\sqrt{2A\gamma}, \frac{\sqrt{\lambda}}{\sigma_w} \right)$$

The Marcum Q function ($Q_m(z, x)$) [119] is defined as per [123] Eq (2-1-123) where $m = A/2$.

³The non-centrality parameter is the SNR (γ) of the received signal. As the noise power of a double sided AWGN is equal to twice the variance of the noise, i.e. $P_w = 2\sigma_w^2$, and Λ is the received signal power.

4.4.2 Receiver Operating Characteristic

To evaluate the performance of a binary classification system (in this case an energy detector), a Receiver Operating Characteristic (ROC) [124] is generated to display the accuracy or performance of the classifier. This performance characteristic is denoted by the level of confidence between the ‘true positive rate’ (TPR) and ‘false positive rate’ (FPR) of the classifier.

The TPR, in the context of energy detection, is determined by the probability of ‘correctly detecting’ the incumbent (P_D) under a threshold λ (Eq 4.15), and the FPR is determined by the probability of a ‘false detection’ (P_{FD}) under the same threshold λ (Eq 4.14) i.e. the incorrect classification of an event as an incumbent being present, where only noise is in fact present. The joint false alarm and detection probability for several threshold levels are displayed by the ROC curves:

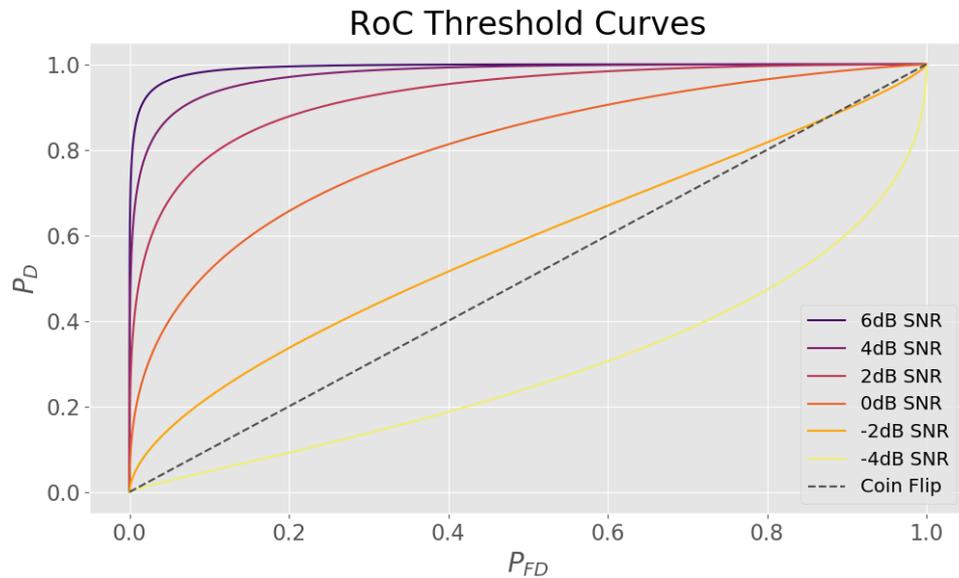


Figure 4.4: Receiver Operating Characteristic Curves for energy detector at various threshold levels, with $A = 10$ frames.

4.4.3 Averaging Parameter

These curves, under a low averaging parameter ($A = 10$), show that at a SNR of approximately 6dB is required to confidently discriminate between an incumbent event and noise. This low averaging parameter is used as it preserves a reasonable timescale for discerning incumbent activity over time without ‘smearing’ the power of those events in time, as a high power instantaneous event will bleed over into subsequent time steps while the averaging window decays. This will result in a less than ideal reconstruction of actual spectrum events when a simple moving average is used for noise suppression. Unfortunately under this regime, the high required SNR, while increasing the accuracy of the detector, causes numerous potential incumbent events to be missed (i.e. the probability of a misdetection increases) that are close to the noise floor of the detector (below the threshold).

4.4.4 Energy Detection Performance

From a practical perspective however, the probability of a misdetection is somewhat difficult to determine (in the absence of Eq 4.15), as the actual incumbent signals are unknown and to a degree, unknowable. While Eq 4.15 provides a reasonable estimate of unknown signal activity, it is somewhat general. An aggressive threshold should thus be used to minimise the probability of missed unknown signals, however as Figure 4.4 clearly shows, a low SNR threshold will introduce a significant proportion of false detections.

To accurately compute the SNR, absolute noise power (N_0) must first be estimated at the receiver, in addition to the noise variance σ_w^2 and total received signal strength, averaged across all bins (N) and over all observed frames (A). The noise power is estimated by:

$$N_0 = \frac{1}{2NA} \sum_{a=0}^{A-1} \sum_{n=0}^{N-1} |y[n, a]|^2 \quad (4.16)$$

Note, that the range of the sum and the observed bins for energy detection, in a practical energy detector should be truncated to regions of the measurement not susceptible to roll-off from the filter, to ensure that the performance of the filter within the energy detectors electronics has no impact on the resulting observation classifications with respect to a given threshold. This artifact is highlighted in the following figure, which portrays a sample output from the energy detector detailed and implemented in Section 4.5):

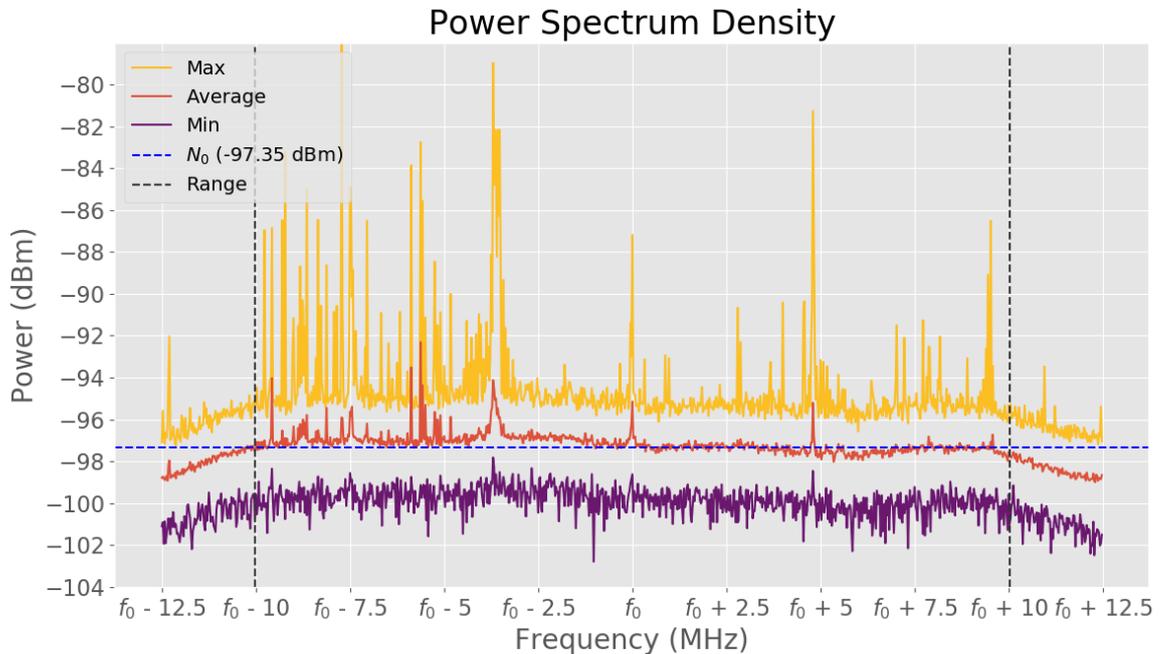


Figure 4.5: Example 25MHz Bandwidth PSD Plot of Incumbent Activity, Filter Roll-off Regions (frequency range bounded by black dotted lines) and Noise Power Estimation (blue dotted line)

As a way to improve the SNR conditions of the measurements and hence detection accuracy, it is to be noted that the resolution of the measurements performed by the mea-

surement system detailed in Section 4.5, yields underlying structure of incumbent events which can be seen in Figs 4.13, 4.14, 4.15 and 4.16. This structure is exploited under a more robust sensing and/or filtering approach in Section 4.6, which details the development of a novel image based filtering technique to exploit this structure and improve threshold selection for the energy detector.

4.5 Custom Measurement System Design & Implementation

As it was found that no completely contained, readily available, low cost spectrum measurement solution was available to meet the demands of the intended spectrum measurement scenario, a custom measurement solution was designed and implemented. Further, as highlighted in Section 4.2, there are no readily available datasets with sufficient fidelity in order to investigate second and sub-second secondary spectrum structure, thus those measurements would have to be performed directly.

4.5.1 Requirements

To determine optimal timescales for operation, the spectrum observations would have to be at timescales at least an order of magnitude smaller than the operational latencies of current state of the art communication systems. The control plane latency of an LTE-A signal is 50ms for a cold start and 5ms for a warm start [125], thus as a baseline, the measurements would require a time resolution of 500us (to yield precision at 5ms timescales).

Wide frequency diversity is required to provide the greatest flexibility of the system, if some candidate band is found to not be useful for secondary spectrum investigation, then the same system should be able to switch to measure a completely different band with

minimal reconfiguration. The range from 100MHz to 6GHz are frequencies that most mobile communication terrestrial links are focused [4,5] due to the superior transmission properties of this spectrum. Thus, this entire frequency range needs to be accessible by the system for measurement.

An instantaneous bandwidth of 20MHz or greater provides a useful slice of spectrum for observation as LTE and WiFi both use 20MHz as a typical operating band, as well as 20MHz being a sufficiently large instantaneous bandwidth for observing spectrum activity of many smaller bandwidth services. High frequency resolution (i.e. large FFT depths) are also important to have access to, to accurately determine spectrum occupancy across frequency and time for various wireless services.

It is critical that direct access to the raw spectrum samples before any preprocessing or DSP (Digital Signal Processing) stages are performed, to preserve the integrity of the measurements for independent analysis of the data. In tandem with direct access to the samples a sufficient bandwidth interface be present between the measurement device and the host computer for real time access and recording of the observed samples.

To meet the aforementioned challenges several key requirements for the measurement system were developed:

- (R1) - Wide frequency diversity to observe multiple candidate frequency bands from 100MHz to 6GHz
- (R2) - Instantaneous bandwidth of at least 20MHz
- (R3) - Measurements with time resolution of at least 5ms
- (R4) - High frequency resolution to assess spectral congestion or utilisation
- (R5) - Up to several days of error free continuous measurement

- (R6) - Expandability to gang multiple radios together to expand observation bandwidth or multiple simultaneous spatial measurements of the same band
- (R7) - Direct access to measurement data stream
- (R8) - Sufficient network bandwidth for real-time visualisation of measurements and storage for post processing/analysis
- (R9) - Sufficient computational resources for storage and processing tasks
- (R10) - Analysis software to evaluate measurements

4.5.2 Commercially-Available Hardware Platform

The radio hardware immediately available for development was a USRP n210 from Ettus Research [79]. This software defined radio platform was previously used for WiFi-based research thus has sufficient performance to meet the requirements of the measurement system, notably: 20MHz instantaneous bandwidth and sufficient frequency diversity (depending on the attached daughterboard) to access the frequency range of interest.

The USRP n210 provides on board networking, signal processing and ADCs and DACs with 100MS/s and 400MS/s sample rates respectively. Both of these converters are connected to a respective signal conditioning chain that provides an interface to a serial connector specified by Ettus, to interface with a daughterboard that is mounted over the top of the USRP motherboard. These daughterboards provide a set of different frequency oscillators, mixers and amplifiers to broaden the frequency range that the motherboard can access. This modular approach to a hardware platform is one of the key reasons for the USRPs success and reconfigurability (Req. 1).

Another advantage of using the USRP platform, is the significant amounts of documentation available for the Application Programming Interface (API) developed by Ettus

Research to facilitate the integration of USRPs within custom applications. The USRP is a favored platform for researchers and enthusiasts due to the platform's flexibility. In addition to this flexibility, the platform has an active forum/ mailing list where many development questions, bug reports, and workarounds are easy to find and assisted in the development process.

USRP Constraints

The measurement of samples poses a three-part problem: The rate at which data can be captured by the USRP's ADCs, the throughput of the physical interconnect between the USRP and the host machine (in this case 1GbE), and the capacity for the host machine to capture the sustained 1Gbps data rate of the network interface.

To achieve the 20MHz bandwidth requirement, for a practicable system, approximately 25MS/s of complex samples (i.e. 50MS/s effective sample rate, exceeding the Nyquist criterion for a 20MHz bandwidth) (Req. 2, 3, 4) is required to completely capture the 20MHz bandwidth with the 2.5MHz of either side of the captured spectrum cut off due to the bandpass filter roll-off. This phenomenon can be seen on both the right and left hand side of the power spectrum in Figure 4.9.

The FPGA (Field Programmable Gate Array) onboard the USRP provides an option to strip the resolution to 8b from 14b and with the decimation offered by the ADC, the sampling stream is reduced from 50MS/s to 25MS/s to meet the throughput constraint of the networking interface. A 1GbE networking connection can support a maximum of 125MB/s transfer rate, thus with a resolution of 14b (rounded to 16b in the protocol) to provide sufficient dynamic range, only a 25MS/s complex (I/Q) can be captured. The

total required transfer rate R_T is as follows where ζ is the network overhead:

$$\begin{aligned}
 R_T &= 16 \times 2 \times 25 \times 10^6 \div 8 \text{ b/s} \\
 R_T &= 100 \text{ MB/s} \\
 R_T + \zeta &\leq 125 \text{ MB/s}
 \end{aligned}
 \tag{4.17}$$

It is quite reasonable that ζ accounts for less than 20% of the total transfer rate, thus the required datarate is effectively supported by the 1GbE interface (Req. 8). At this rate, the amount of data captured over a 24 hour period is approximately 8.6TB.

$$24(h) \times 60(min) \times 60(sec) \times 100MB/s = 8.64TB \tag{4.18}$$

Keeping 8TB+ of samples in a contiguous file is an unrealistic and highly cumbersome prospect, as most file systems would not handle such a large file well and random reads would be very expensive. Thus the captured samples are split into timestamped 4GB chunks. This selection of chunk size also simplifies the retrieval of the data for analysis. Each sample set is arranged in a separate flat file structure (of multiple 4GB chunks). The use of a real-time database was deemed unnecessary as the raw samples do not have any additional information other than the time-stamp and their position in the sequence. The samples are streamed once and processed multiple times with only sequential reads used in analysis.

Analog to Digital Conversion

The ADCs (Texas Instruments ADS62P4X) [126] on-board the USRP are capable of capturing 14bits of resolution at 100MS/s with 88dB of Full Scale Dynamic Range (FSDR). In practice the dynamic range (G_{FS}) is calculated based on the number of bits of resolution

n. The practical FSDR is limited, $G_{FS} \approx 84dB$ given by:

$$\text{dBFS} = 20 \log(2^n) \quad (4.19)$$

The peak to peak voltage of the ADCs is 2V [126] and the ADC resolution is 14bits, where $V_{p-p} = 2V$ and reference impedance $R = 50\Omega$:

$$\begin{aligned} V_P &= V_{P-P}/2 \\ V_{RMS} &= V_P/\sqrt{2} \\ (V_{RMS})^2/R &= 0.01V \end{aligned} \quad (4.20)$$

Finally, to compute the input signal saturation power:

$$\begin{aligned} P_{SAT} &= 10 \log(0.01) + 30dB \\ P_{SAT} &= 10dBm \end{aligned} \quad (4.21)$$

The frontend of the WBX daughterboard provides up to 30dB of gain (G_{FE}) for the receive channel with a 6dB noise figure (NF) and 3dBi antenna (G_A) [127]. Accounting for saturation power, fullscale gain, frontend gain and antenna gain, yields a minimum detectable signal of $\approx -107dBm$.

$$MDS = P_{SAT} - G_{FS} - G_{FE} - G_A \approx -107dBm \quad (4.22)$$

The minimum detectable signal is primarily due to the quantisation error experienced due to the resolution of the ADC and amplification stages. The theoretical Noise Floor in this case would be $MDS + NF \approx -101dBm$.

Note that thermal noise power $P_{TN} = 10 \log_{10}(k T B) + 30$ where k is the Boltzmann constant $1.38064852 \times 10^{-23} J K^{-1}$, temperature T of the receiver in Kelvin and noise bandwidth (or resolution bandwidth) B in Hz, is commonly used as the minimum detectable signal (MDS). P_{TN} yields a thermal noise power of approx. $-160dBm + 30dB = -130dBm$

at room temperature with an RBW of 24.4kHz (25MHz/1024 samples - far lower resolution than the intended scenario). Thus the minimum detectable signal in Eq 4.22 is entirely dependent on the measurement hardware and noise figure of the system, and not limited by the selected resolution bandwidth and resulting thermal noise power.

4.5.3 Custom Measurement System Architecture

Figure 4.6 shows the basic system model for the measurement system. The USRP n210 with a mounted WBX daughterboard and attached antenna is connected to the analysis workstation via a 1GbE connection. After several data collection and processing iterations, the Central Processing Unit (CPU) onboard the workstation is determined to be insufficient to keep up with the computational load due to the numerous number of FFT operations. Thus an NVIDIA co-processor is recruited to offload the FFT computation, increasing the rate of analysis that could be performed.

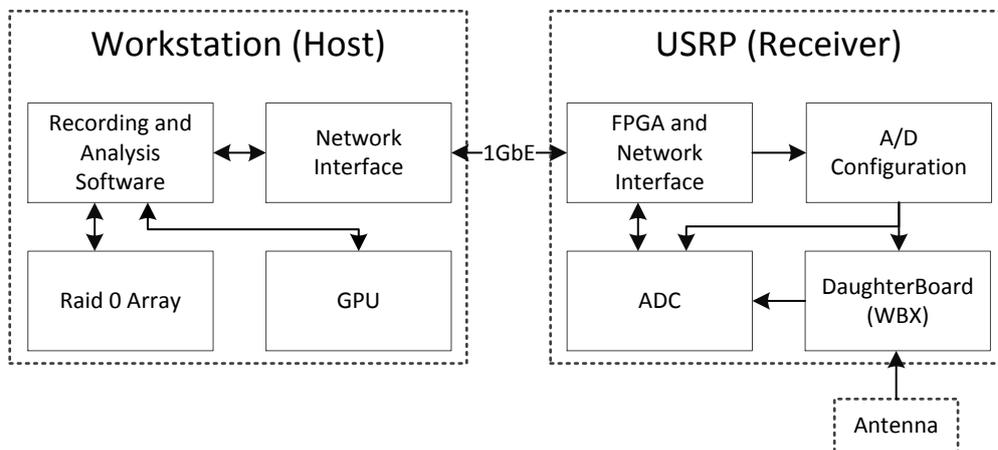


Figure 4.6: Diagram of measurement system high level physical architecture

The workstation comprises of a Xeon E5 1650, 32GB DDR3, 4 x 6TB HDDs in RAID 0, a NVIDIA Tesla K40 and, the developed measurement, storage and analysis packages

detailed in Section 4.5.4 (Req. 9). The particular RAID configuration provides sufficient storage capacity for approximately 2.7 days (Req. 5) ($24\text{TB}/8.6\text{TB Day}^{-1} = 2.7 \text{ Days}$, Eq 4.18) of continuous measurements from the USRP and supports a Read/Write of over 600MB/s. This high throughput is required to both handle the measurement data rate from the n210 comfortably and with the added benefit of speeding up data retrieval. This configuration is preferred due to cost with respect to storage capacity versus throughput. M.2 and other PCIe based solid state drives would support much higher data rates, however at the time of acquisition, high capacity SSD's (2TB+) were not available for the consumer market nor would they be cost effective.

The only physical bottleneck from this system architecture that cannot be addressed through faster computer hardware and more efficient software, is the physical interface between the USRP and host i.e. the Ethernet connection. In the X300 series of USRPs this physical throughput limitation is overcome by the inclusion of a 10GbE networking interface, or an onboard 4x PCIe connection that can be used to connect the USRP to the motherboard of a host device with an appropriate cable, should larger bandwidth datasets be required.

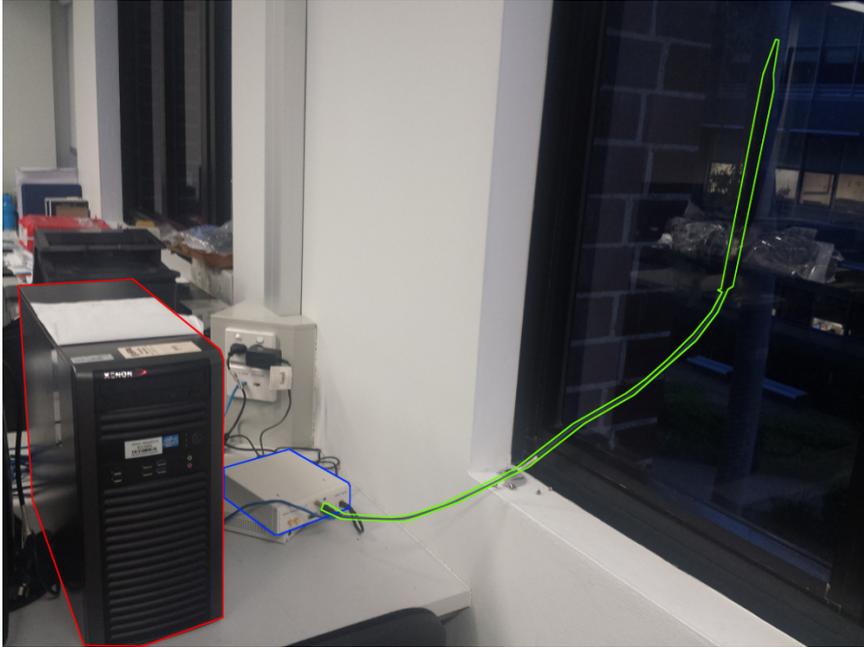


Figure 4.7: Image of spectrum measurement and analysis system set up. Workstation (red), USRP N210 (blue) and SMA (SubMiniature version A) cable with attached VERT 900 antenna (green) mounted to a window for measurements. Post-photographic outlines used for emphasis and identification.

Figure 4.7 shows the physical setup configuration for most of the spectrum measurement activities. A static configuration is determined to be ideal for rapid iteration and improvement of the system. The static setup is also preferred to eliminate environmental variance across numerous measurement over time, enabling the datasets to be compared equally from a physical environment perspective. The antenna is affixed to the inside of the window, as mounting it outside is not possible.

4.5.4 Software Platforms

Several candidate software platforms were investigated to determine their suitability for spectrum measurement activities. Matlab, HDSDR and GNU Radio were all freeware or

available at no additional cost (Matlab) as potential software candidates. Each of the candidates were assessed for their suitability in the high performance application that is required of the measurement system.

Matlab

Matlab [128] is a ubiquitous standard for engineering computing, especially within the communications field. MathWorks, the developer of Matlab, provides a toolbox for connecting a Matlab script to a connected USRP. However, due to limitations in the implementation of the associated toolbox for USRPs and the cumbersome nature of Matlab programs due to user code being interpreted and not compiled, the throughput of the attached USRP is severely impacted, resulting in very low throughput for realtime data capture and processing.

HSDSR

HSDSR (High Definition Software Defined Radio) is a freeware SDR program [129] that enables the user to quickly connect to a USRP and have some direct control of the USRPs functions. For the purposes of spectrum measurement however, due to the way in which HSDSR was written, there is a severe limitation on the instantaneous bandwidth that the program can capture. It is found that only 8MHz of instantaneous bandwidth are able to be captured by the program before significant lagging and dropping of samples are incurred, likely due to program inefficiencies in how the data stream is processed.

The additional disadvantage of this program is that the samples that were retrieved were unable to be dumped directly to a storage media for later playback and/or analysis, eliminating its applicability for the desired use-case.

GNU Radio

GNU Radio [130] is an open source graphical programming based software platform that offers deep control over the processing and data capture chain for many SDRs and in particular the USRP ecosystem. GNU Radio uses a GUI (Graphical User Interface) based programming approach, to enable the programmer to quickly ‘connect’ a signal processing chain to a source device or dataset and perform signal processing on the retrieved data. It can also interface to a USRP in realtime to pull samples directly from the device. This provides an easy to use and prototype package to get projects working quickly.

GNU Radio suffers from a similar problem to Matlab, in that its core functions are written in Python. Python is a language interpreted at runtime and is (typically) not pre-compiled (bar certain wrapped C/C++ library functions). In light of this, overall program performance is limited, especially in the context of processing large numbers of repetitive operations. Thus Python (and by virtue Matlab/GNU Radio) is unsuitable for very high throughput measurements and processing, but still remains useful for analysis.

It is determined that GNU Radio is also unsuitable for the application and a higher performance alternative needs to be selected.

4.5.5 USRP Hardware Drivers

After investigation of existing software solutions, the three alternatives were all deemed to be unsuitable for the required task, the only remaining option was to interface directly with the USRP hardware via a custom software layer and the associated UHD drivers supplied by Ettus Research. Ettus Research provides a well documented driver API to interface with all USRP platforms (Req. 6,7), making this a viable alternative, with unprecedented control over the USRP when compared with the alternatives.

The UHD Package offers a direct C++ API to interface with an attached USRP. This

API is used by both Matlab and GNU Radio implementations to communicate with a USRP (and supposedly HSDR as well). Matlab interfaces with the API via wrapping the relevant C++ functions and then referencing those, while GNU Radio takes a similar approach, by wrapping the C++ code with Python wrapper functions to integrate the USRP with the existing GNU Radio codebase that is written in Python.

The API provides all the necessary function calls to correctly configure the USRP for a range of wireless communication applications, providing fine grained reconfigurable control over: carrier frequency, bandwidth, sampling rate and amplification levels. Direct access to the USRP via the API enables the best possible performance that a host can achieve when communicating with a USRP, provided that the code is written appropriately to realise the required performance.

As none of the existing spectrum measurement solutions that existed were suitable for the use case required of this research. A custom solution directly accessing the UHD drivers to control and extract the raw information we required, had to be designed to fulfill our specific data and interface requirements.

4.5.6 USRP Firmware

The firmware for each USRP device is provided by Ettus Research as part of the USRP Hardware Driver (UHD) driver package (Section 4.5.5). The UHD package offers the user the ability to modify the firmware and the functions of the underlying FPGA onboard the USRP. The firmware can be easily flashed to a USRP via the included flashing utility as part of the UHD package using (in the case of the USRP n210) the ethernet connection to a host computer.

Should a USRP become bricked from a bad flash due to power or networking interruptions, the USRP has an onboard ‘flashback’ function triggered via an onboard push button. Upon activation the internal flash memory backup will flash a basic networking

firmware to the working memory set enabling the USRP to communicate with a host for a complete reflash to be performed.

4.5.7 Software Architecture and Development

To meet Req. 3, 4, 5, 6, 7, 8, and 10 the software platform had to be designed with an eye toward efficiency, performance, flexibility and reliability. Hence C/C++ is the chosen language for high throughput processing and interfacing operations, with Python selected as the language for the data analysis stages.

The initial objective for the software platform, is to capture and process the measured data in real time and then draw the necessary analysis from the data directly after the end of the measurement phase. By doing this, the amount of data generated from the measurement phase could be significantly reduced as there is no need to keep the raw samples and only the processed result of the samples are stored.

This rationale however is deemed to be insufficient: if there were an improvement or change in how the data is analysed, the measurements would have to be re-run to reflect this change. Each new run of measurements would thus result in an entirely different dataset, eliminating the possibility of accurate comparisons between older and newer processing methods.

To address this issue the samples would need to be stored and maintained appropriately. The associated hardware and software solution is developed to address the necessity of storing measurements as they are captured by the USRP on a high speed disk array for post processing.

To process the samples into useful information, large numbers of high resolution FFTs would have to be performed alongside various DSP and analysis algorithms. As most of the code would have to be low level as previously identified, C/C++ is the preferred language for a good balance of performance and flexibility, with Python being used as the

preferred language for manipulating the large data sets after they were processed.

FFTW [131] is a high performance FFT package for C/C++ and is used for initial development and within the recording program to implement a lightweight low resolution spectrum analyser. For the purpose of high FFT depth processing, it is found that even using the built in parallelisation options within FFTW and an 8 core Xeon platform, the throughput is insufficient to analyse the dataset at the rate that the RAID array could provide data, resulting in a processing bottleneck, giving rise to the need for GPU (Graphics Processing Unit) accelerated FFT processing. However, some FFTW code is preserved for lower throughput operations and as a verification step for the results from the GPU accelerated output.

To reduce development time to hit the necessary throughput targets, the boost for C++ libraries were used to improve ease of implementation multi-threading within the software stack (as well as being a requirement for compiling the UHD driver code). The NVIDIA CUDA libraries (that are natively written in C) were also used to develop GPU code to enable high throughput DSP and FFT functions.

The software development is broken up into four separate packages with a common DSP CUDA accelerated module used by several of the packages. These four packages are primarily written in C/C++ and consist of:

- ‘Samples To File’ a program to correctly configure the USRP to perform new spectrum measurements and save the continuous sample stream into discrete manageable data files with a built in realtime spectrum analyser to ensure correct function.
- ‘Spectrum GUI’ a program to view and ‘navigate’ through the saved samples as a power spectrum.
- ‘Process Samples’ a program to iterate over the saved samples to perform DSP on the spectrum for energy detection and subsequent opportunity detection, covered

in Chapter 5.

- ‘Analysis Tools’ an analysis package written in Python to perform deeper analysis and data manipulation of the processed spectrum, also covered in Chapter 5 (Req. 10).

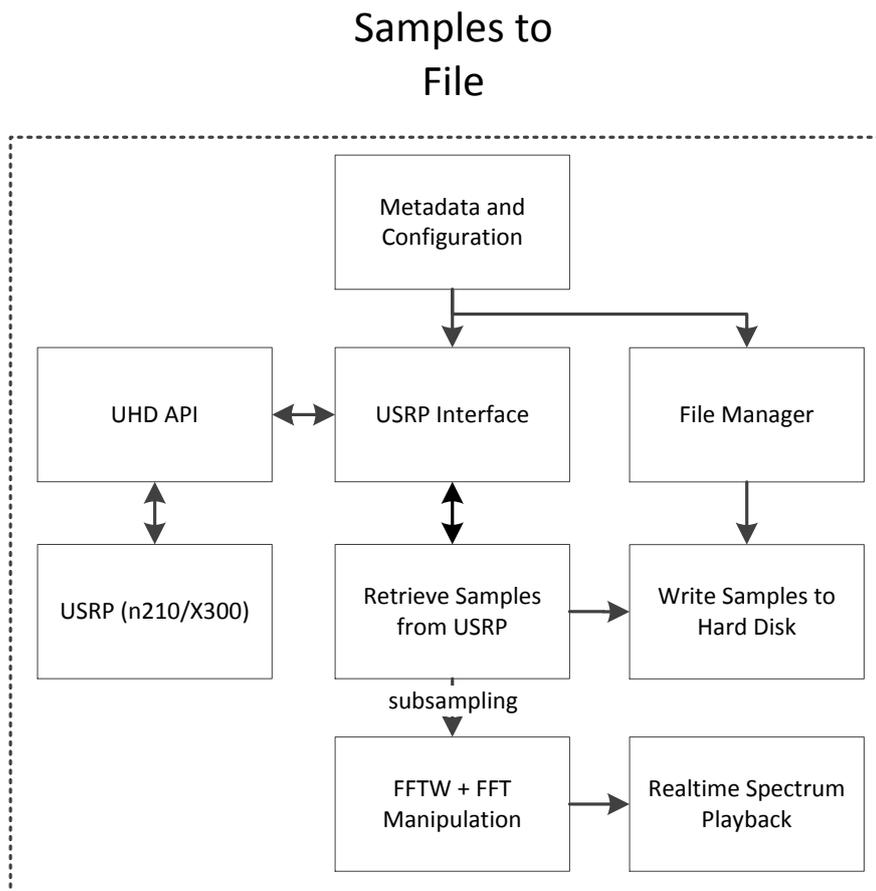


Figure 4.8: Flow Diagram of Spectrum Sampling Program, highlighting the main software functions and data flows to perform spectrum measurement.

Figure 4.8 describes the architecture of the spectrum sampling program, including how the USRP, drivers and software layer interact, and how raw spectrum samples are propagated through the system for verifying the system is functioning correctly and recorded

for later retrieval for analysis and processing.

The core measurement program, takes advantage of the UHD API to perform configuration and data retrieval from a connected USRP device. The configuration information is either input by the user, or a default configuration is used. This configuration information is saved alongside other measurement metadata to identify the measurement parameters for later retrieval. An example configuration file snippet is as follows:

Listing 4.1: Spectrum Measurement Configuration File

```
// 20160317_1834_465MHz.cfg
dev: N210r4
f: 4.65e+008
bw: 27.5e+006
s: 25e+006
g: 30
```

The configuration file preserves the start date and time in minutes of the measurement, the device and version, centre frequency, channel select filter bandwidth (found to have no effect), sampling rate and front end amplification stage gain.

A thread that is responsible for managing the USRP is then created, which feeds requests to the USRP. This thread is also responsible for retrieving the samples from the USRP and managing the storage buffer. This storage buffer is implemented as a ‘ping pong buffer’ (PPB). A PPB is a memory management technique to avoid race conditions/memory contention by implementing two separate buffers A and B, shared and coordinated between two separate routines (pointers): a storage pointer (S) and a retrieval pointer (R). This is also known as a ‘producer and consumer’ sharing mechanism.

Initially the R pointer is set to null, the S pointer is set to A. Buffer A is filled with

data from the S pointer, then S is updated to point to buffer B and proceeds to fill it. Once S is updated to point to B, R is then updated to point to A. S then consumes the data in A and waits until it is next updated to point to B. This update is triggered by R finishing filling B and then pointing back to A.

This process continues by alternating each of the pointers between the buffers, with S ‘producing’ data and R ‘consuming data’. As the retrieval (the writing samples to a file) function is faster than the storage function (getting samples from the USRP), the storage function is always assigned priority and the retrieval function will wait until the storage function is complete. This effectively enabled realtime streaming of data from the USRP as it is generated, as the data could be retrieved much faster than the USRP is able to generate it. Checks were also put in place to ensure that S and R could not desynchronise, which may result in measurements being overwritten.

For debugging purposes, the data stream is subsampled by a separate routine to enable live viewing of the power spectrum of the samples captured. The data stream is fed into a FFTW function running on the CPU to perform the FFT, normalisation and power estimation is done within the same routine, and then finally sent to a display routine which plotted the power spectrum for the user to view.

A separate mode is also implemented that did not save the samples to a file, and just enabled the user to view the live power spectrum captured by the USRP, which is used during the calibration and validation activities to ensure the estimated power spectrum is accurate. This functionality can be seen in-situ in Figure 4.7.

Various sample batch sizes were tested to determine software and architectural limitations on processing throughput. Most batch sizes tested ($n \times$ size of an FFT frame) resulted in similar throughput figures, with significant deterioration in throughput as n is less than 100. This is likely due to the small amount of data transferred in each read cycle causing the mechanical drives to constantly park, resulting in much greater numbers of

read cycles, and the FFT routine being constantly stopped and started.

Smaller batch sizes also had negative implications for the throughput of DSP functions, requiring many additional concatenation and buffering cycles to satisfy overlapping and averaging parameters.

A more efficient buffering solution such as a PPB could have been used at the time to speed up the CPU throughput, however when the initial tests were conducted the data retrieval and processing is implemented serially. A PPB solution is subsequently developed for the GPU code (and the aforementioned Samples to File program), to pipeline the constant reading and writing from GPU memory to main memory.

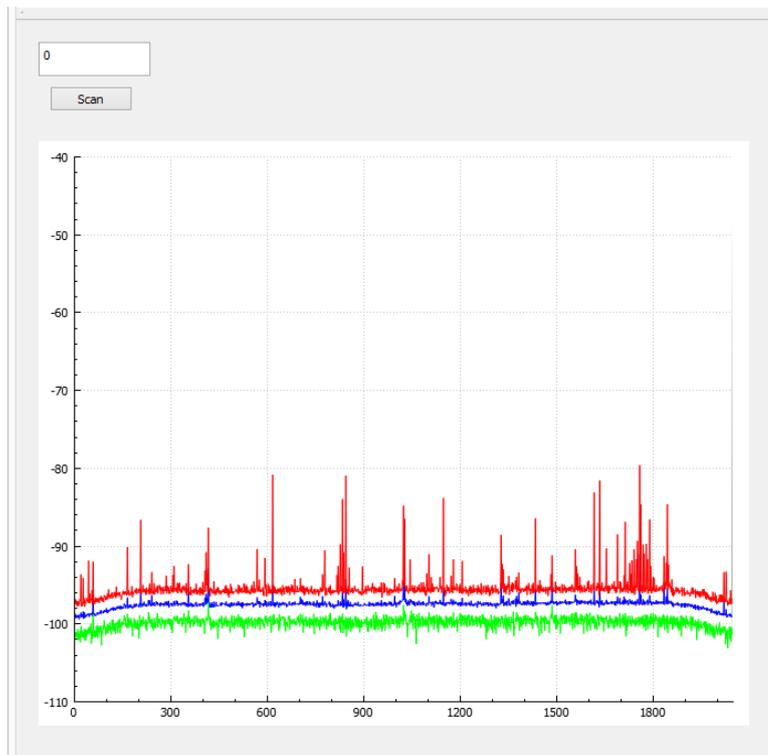


Figure 4.9: Output from the Spectrum GUI Program. Note that the y axis displays observed power, and x axis displays clusters of bins, which can be scaled off a known bandwidth and centre frequency

To alleviate the aforementioned processing bottleneck, a GPU solution is developed in CUDA using the cuFFT library and a NVIDIA Tesla K40. This setup is more than sufficient at performing the high depth FFT calculations and additional DSP functions on the datastream in realtime as it is retrieved from the RAID array. The total throughput increase is approximately 40x over the CPU alone, benchmarked using the rate of batched data retrieval and the time taken to complete the processing of each batch.

The functions that the CUDA solution provided were (Figure 4.10): sample normalisation to correctly weight the observations with respect to the maximum observable value; sample overlapping for optimal power flatness (see Section 4.3.2); windowing of the raw samples, using a Blackmann-Harris window; computation of the FFT on a subset of samples (a ‘frame’); computing power of the FFT frame, and normalising the power by the number of FFT bins; averaging of several FFT frames to reduce noise in measurement; thresholding in accordance with the energy detection threshold (to discriminate between an incumbent signal or background noise); an optional 2D ‘image’ filtering function (see Section 4.6) to remove spurious transmissions from the power spectrum; and finally the resulting power spectrum is written to main memory.

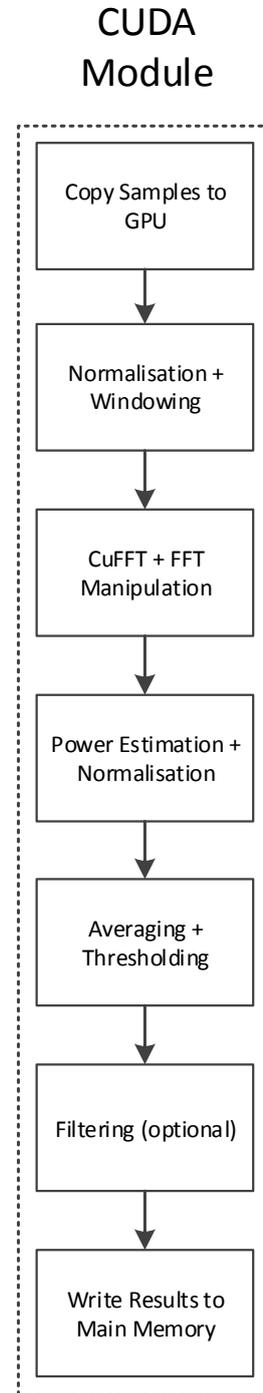


Figure 4.10: CUDA Module Flow Diagram.

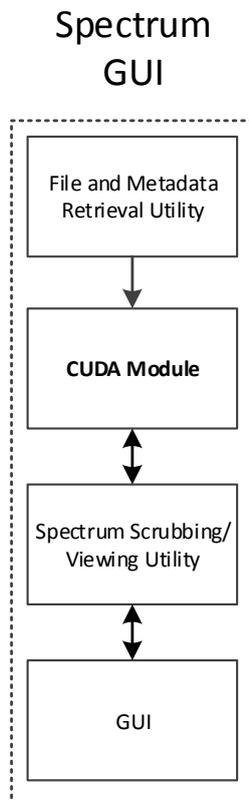


Figure 4.11: Spectrum GUI Flow Diagram, a utility designed visual inspection of captured spectrum samples.

The increase in computational throughput by using the GPU to offload the FFT and DSP functions, resulted in the new bottleneck due to the RAID array read speed. Being bound by only the storage media is ideal in this configuration.

To complete the spectrum measurement capabilities, a simple GUI (Figure 4.11) is written in C++ using the Qt GUI library (Figure 4.11), enabling playback of the Power Spectrum Density (PSD) of the observed spectrum samples by a user. The utility provides functionality to scan throughout a dataset to look for particular times and signals of interest as well as providing a graphical view of the power spectrum for the user shown in Figure 4.9. The user is able to either step through the spectrum frame by frame with the arrow keys, or can immediately jump to a desired frame through the manual input on the top left and pressing the scan button. The arrow keys can be held down for spectrum playback at about 30 frames per second.

As the amount of data contained within each frame is very high, each pixel represents numerous FFT bins and scales with FFT depth. For example, an FFT with depth 32k, will display 16 bins per pixel (2048 pixels horizontal on the chart). To broadly display this data, the red line is the maximum observed power within each bin cluster, the blue line is the average power and the green line is the minimum power observed in each cluster.

4.5.8 Calibration and Verification

To ensure correct function of the software, and that the hardware onboard the USRP is correctly calibrated for measurement, the USRP is connected to a signal generator with a known transmission power and frequency of operation.

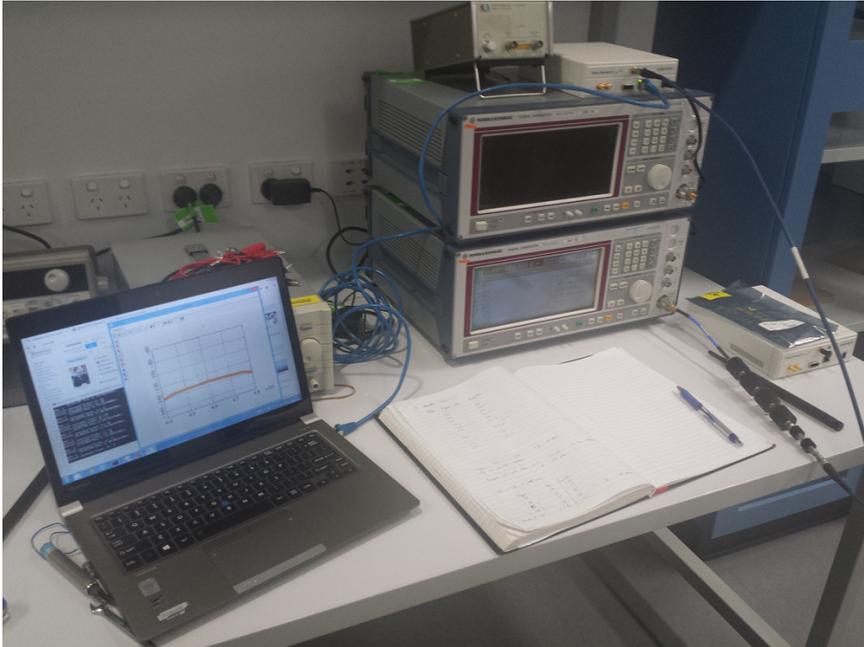


Figure 4.12: Calibration and verification setup: USRP is connected to a Rohde and Schwartz Signal Generator with a range of known output powers and frequencies. Laptop is used to display in realtime the measured signal strength.

The receive (Rx) port of the USRP is connected to a Rohde and Schwartz function generator via a SMA cable with 30dB of in-line attenuation to limit the possible power transmitted to the USRP. The USRP is connected to a portable host computer running the measurement software. Figure 4.12 displays the calibration and verification setup, with an earlier version of the Samples to File program running and collecting measurements. The setup is useful in debugging certain algorithmic errors in the power estimation calculations, as well as providing confidence in the detection thresholds selected for in-

cumbent identification, with respect to the minimum detectable signal (-107dBm Eq 4.22) and noise floor (-101dBm) of the receiver.

Ettus research also provides a loop back calibration function as part of its UHD driver package, which can also be used to perform closed loop calibration of the hardware frontend, in the absence of bench-top measurement hardware.

4.6 Filtering and Threshold Selection

From Section 4.4.2, it is evident that to reduce the number of missed incumbent events below the threshold, a method to either reduce the impact of noise on the classification of spectrum measurements is required to accurately discriminate incumbent events from noise.

At the final stages of the spectrum analysers function, after power estimation is completed, energy detection is performed. Functionally, this is achieved by comparing the estimated power of a bin with a Power Threshold P_L , where the power threshold is selected in accordance in Figure 4.4, for reliable detection. According to the threshold: if $P_n \geq P_L$ then the result is 0 (an incumbent event), else it is 1 (whitespace) ⁴.

To improve the SNR of the energy detector, a 2D image filter is implemented with a simple ‘filter kernel’, to perform robust noise filtering by assessing the clustering (or correlation) of incumbent events in time and frequency. A filter kernel is a sliding square frame that is applied discretely to every pixel in a 2D image. The kernel performs a function on the ‘target pixel’ (the central pixel) using the values of the surrounding pixels. The convolution of the underlying image and the kernel, results in the filtered image. A typical kernel size for image filtering is a u by u frame, where $\{u = 1 + 2n \mid n \in \mathbb{Z}\}$.

⁴While counterintuitive, as we are interested in whitespace, detection of whitespace is taken as the true/positive outcome i.e. 1

For our purposes, the kernel is used to remove noise from the estimated power spectrum, after thresholding is applied. This is able to be done, as, after thresholding, each pixel (FFT bin) in the image is reduced to either a 0 or 1 (whitespace), being above or below the threshold respectively.

When computing the convolution in this implementation, the number of 1's encompassed by the kernel are summed. If the number is equal to or greater than a Filtering Threshold F_L , then the output of the convolution will be a 1. If the number of 1's is less than the threshold, then the output will be a 0 (an incumbent event). Effectively the filter is looking for neighbouring whitespace to determine if the subject is whitespace or not.

This is a practical way of performing filtering as the resolution bandwidth of the observations is substantially finer than the smallest bandwidth used by a conventional telecommunications device, or conventional spectrum allocation, which is on the order of 10kHz - 12.5kHz [132, 133].

Using the correlation between neighbouring bins, that would also be occupied under a conventional communications scenario (as detected incumbent events at this resolution have to be correlated in frequency and time), to determine whether a point under investigation is spurious noise or not, is a rather logical approach. This approach is further preferred as the computation is taking place on a GPU, which is optimised for this type of (image based) operation.

4.6.1 Spectrum Noise Image Filter Performance

Initially a 3 by 3 kernel was investigated, however this was unable to cope with random clustering of noise and yielded virtually no improvement in incumbent identification when applied at a $P_L \leq -96dBm$ (noise clustering can be seen in Figure 4.15, Filter I). To better handle random clusters a larger kernel of 5 by 5 is selected with $F_L = 16$ for filter I, and

a more aggressive setting of $F_L = 13$ for filter II.

A larger kernel or more aggressive filtering thresholds could have been used to further eliminate the clustering observed in Figure 4.15 and 4.16. However, as these signals are unknown and (in a low SNR case) sparsely correlated, a more aggressive filter runs the risk of an increased misdetection rate at low SNR.

Each waterfall plot presented, represents a subset of the overall spectrum captured in order to intuitively display the information. The waterfall is generated by stacking discrete FFT frames sequentially in time (Y axis), with the X axis representing discrete frequency bins from the FFT. Each plot spans 384 frequency bins (190Hz per bin) on the X axis, and 384 time steps (5ms per step) on the Y axis such that each pixel portrays a discrete FFT bin at a particular point in time.

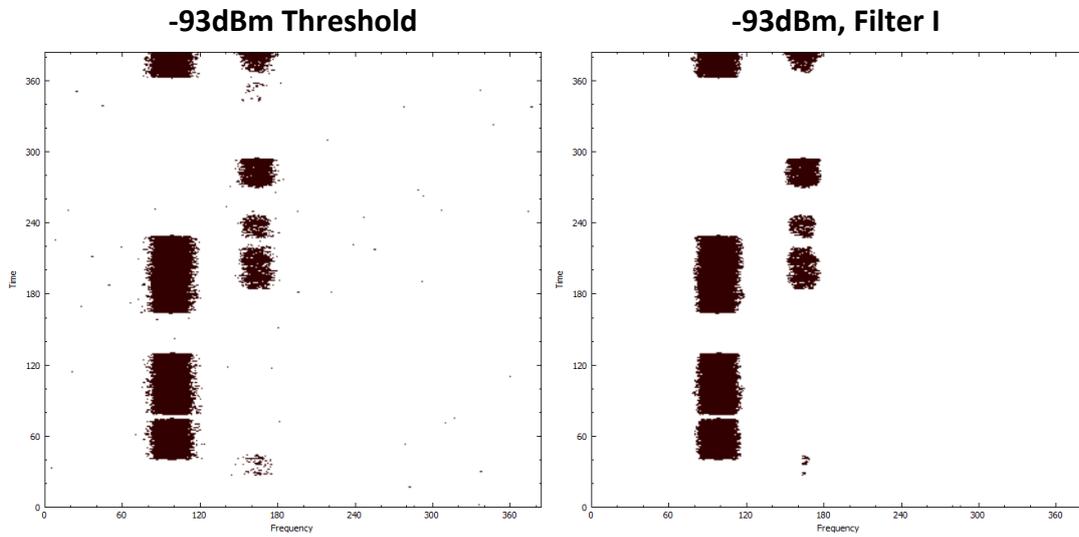


Figure 4.13: Thresholded power estimation at -93dBm and 5 by 5 filter kernel applied, $F_L = 16$

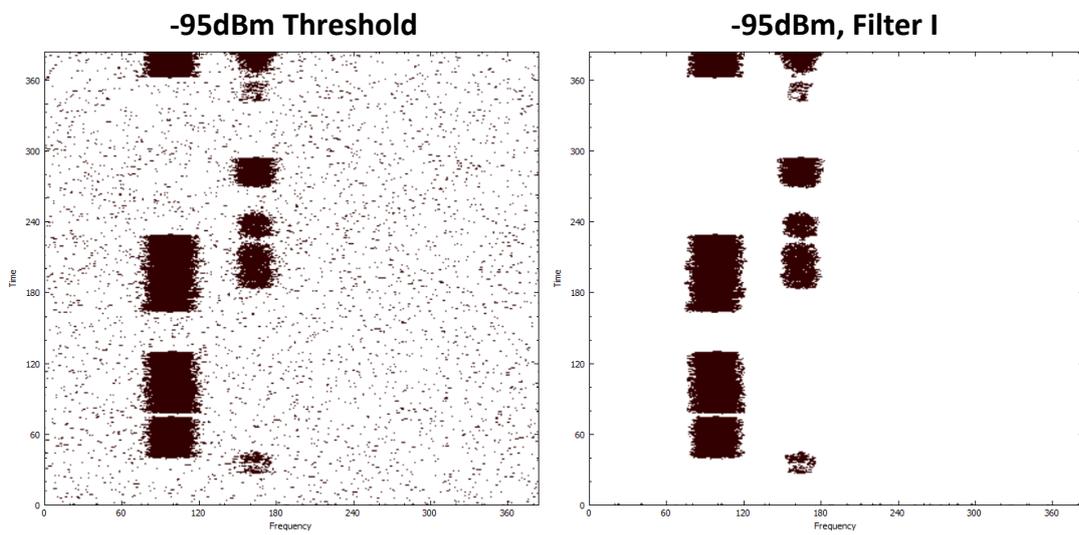


Figure 4.14: Thresholded power estimation at -95dBm and 5 by 5 filter kernel applied, $F_L = 16$

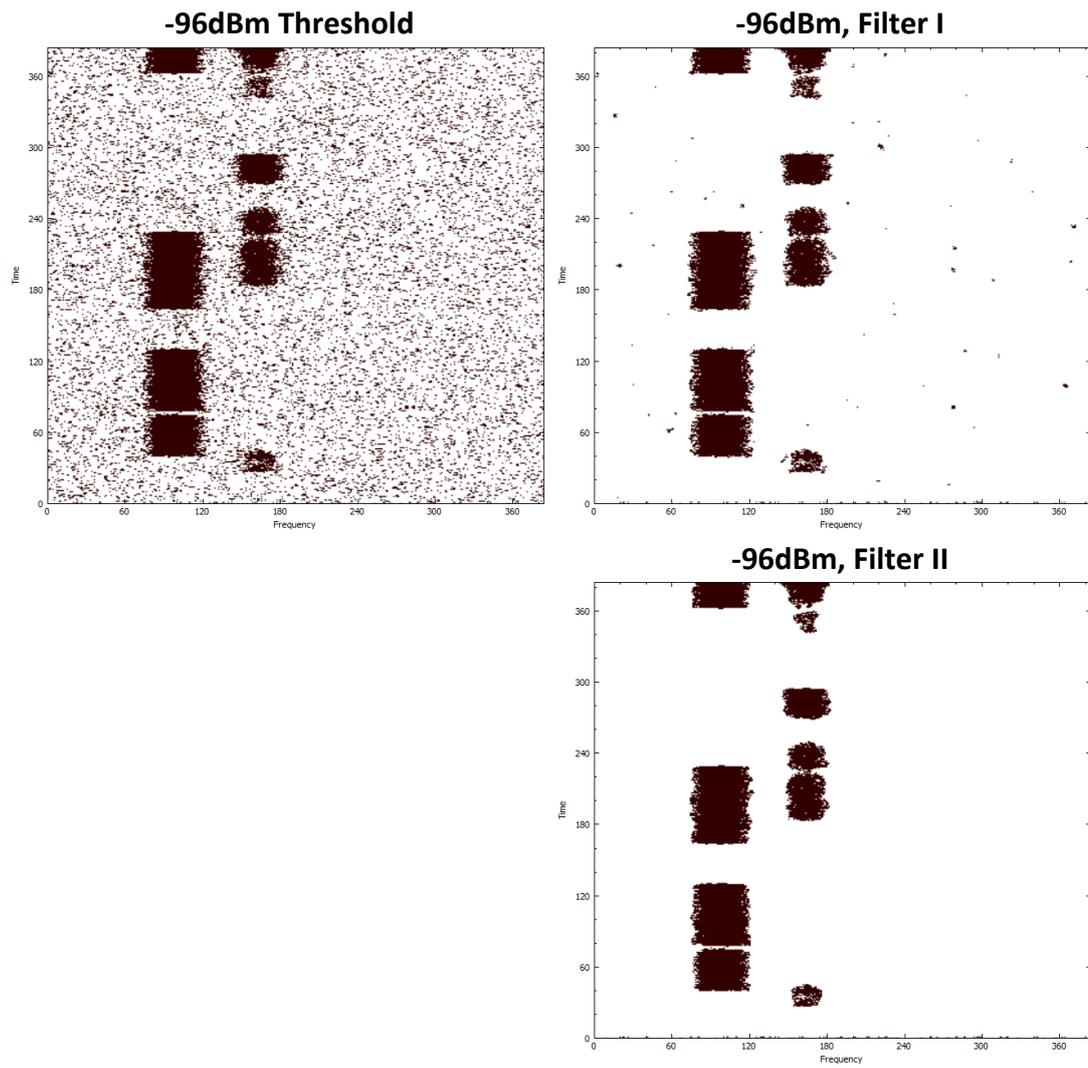


Figure 4.15: Thresholded power estimation at -96dBm and 5 by 5 filter kernel applied, $F_L = 16$ and 13 respectively

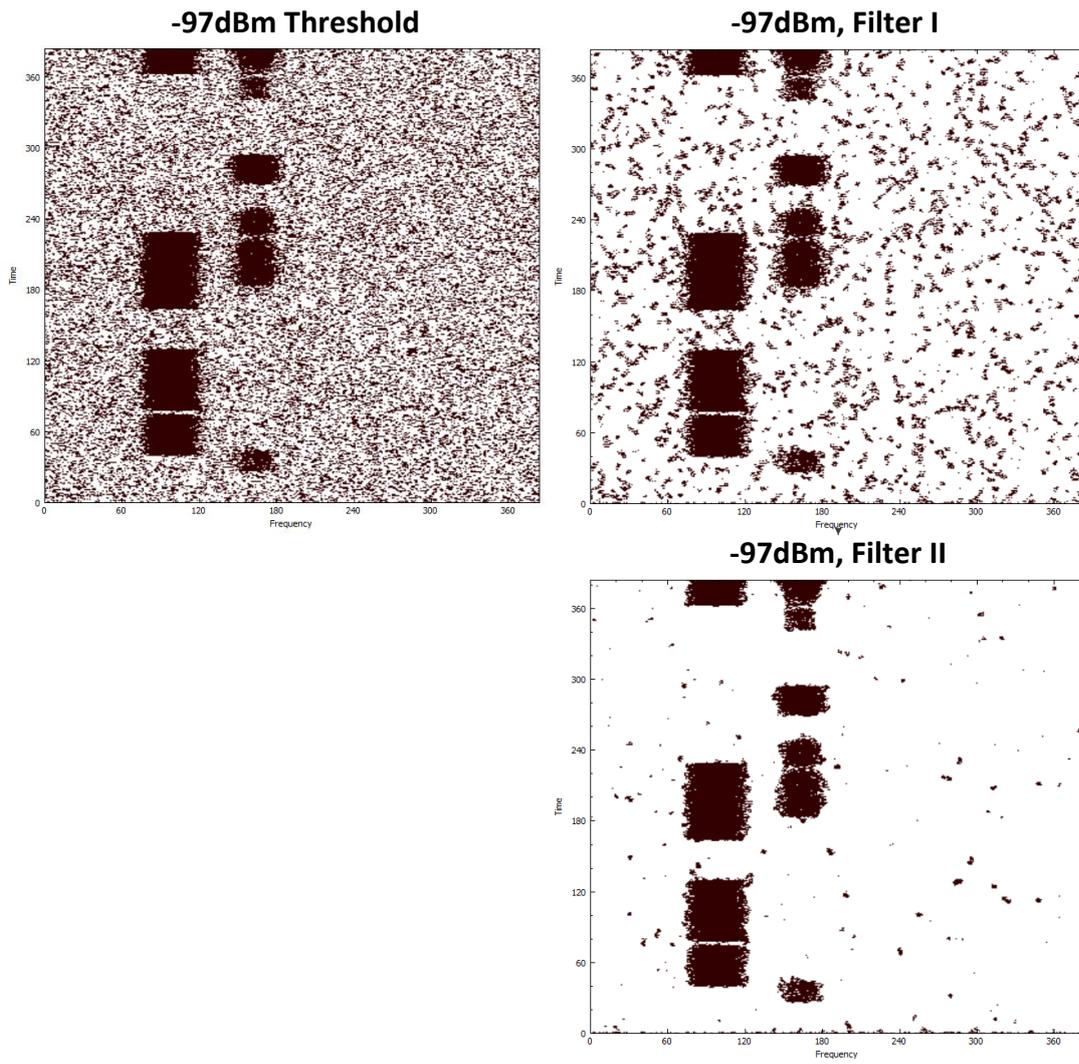


Figure 4.16: Thresholded power estimation at -97dBm and 5 by 5 filter kernel applied, $F_L = 16$ and 13 respectively

4.6.2 Noise Image Filter Evaluation

As the events being detected are from unknown incumbent sources or noise, it is hard to gauge the absolute improvement in performance within this context. One could assume that any small cluster of events that are not part of a larger cluster (like in Figure 4.16) is simply noise, however this naïve approach is invalid as it is unknown if the smaller clusters are in fact noise or an incumbent, thus all events must be treated as incumbent events to avoid misclassifying a true incumbent event. The downside of this, as is explored in Section 5.6 can result in highly fragmented secondary spectrum, which can render it unsuitable. Thus a filtering threshold that makes a reasonable attempt at detecting probable incumbent clusters, with minimal smaller fragmented clusters is desirable.

It is highlighted in the study [134], that noise variance plays a substantial role with the accuracy of energy detection of unknown signals. This can also be seen in the energy detector model presented in Eq 4.15, where the non centrality parameter is dependent on the noise variance. The more accurate that noise variance is able to be modeled, and subsequently removed, the greater the improvement in SNR (As is also intuitive with the Shannon-Hartley Theorem, Eq 2.1).

In regards to the conclusions in [134] the SNR improvement obtained by this filter is very close to the noise variance, due to the low averaging used, if higher averaging was used the SNR could have been further improved. Higher averaging however, negatively impacts timescale resolution (see Section 4.4.3) which is undesirable for precise spectrum measurement.

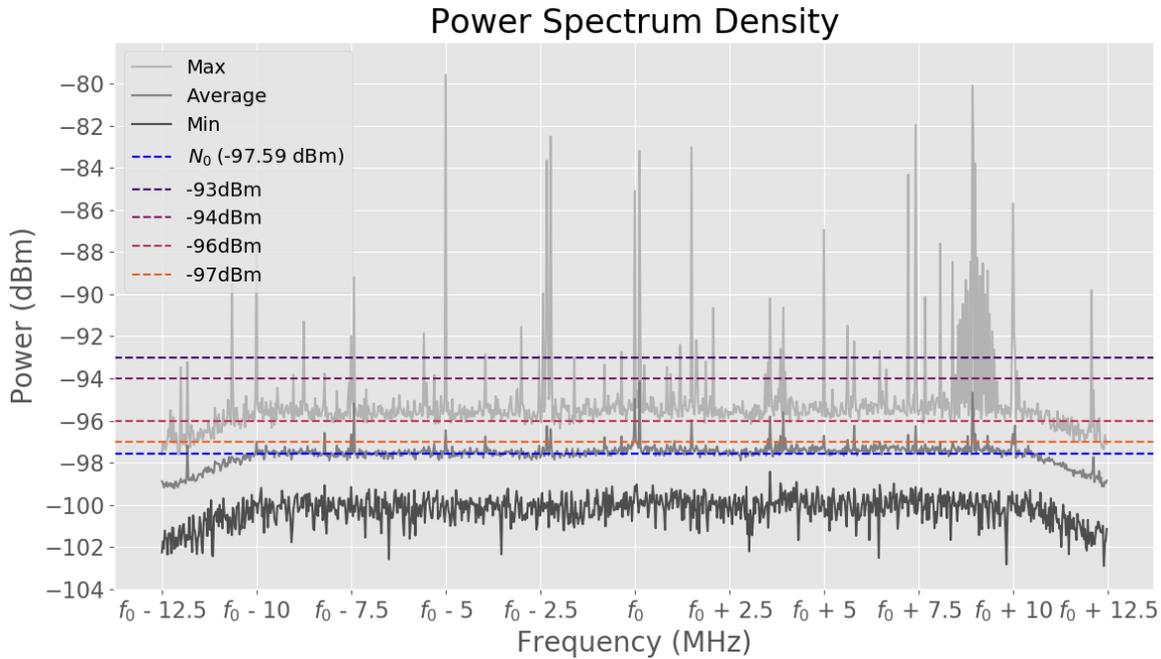


Figure 4.17: Power Spectrum Density of a candidate frequency band, highlighting the noise floor of the observation and threshold settings investigated for filter performance.

With respect to the noise floor measurements (N_0) in Figure 4.17 and the various threshold levels investigated, a threshold that has an SNR of approximately 1dB using the filtering method employed by Filter II, is used for whitespace classification and the subsequent analysis in Chapter 5. This threshold represents a 5dB SNR improvement, with a substantially improved Pr_{MD} at the equivalent 6dB threshold (in the classic energy detector case). The filter also provides greater access to low SNR events (between 6 and 1 dB SNR) that would have otherwise been misclassified by the classic energy detector, making it a highly preferable solution for improving the accuracy of observations going forward.

4.7 Summary

This Chapter focused on the design and development of a low cost, high performance spectrum measurement and analysis system. The method by which secondary spectrum is identified is through the use of conventional energy detection augmented with a novel image based filter to improve the SNR performance of the detector.

Section 4.5 detailed the complete design process for the spectrum analyser and energy detector required to perform secondary spectrum whitespace observation. The spectrum analyser created from this process performs exactly as intended, meeting all requirements identified, within acceptable tolerances. The high performance software platform also detailed, plays a critical role in enabling rapid analysis of the extremely large datasets captured by the spectrum analyser, significantly reducing the time required for measurement results to be processed and analysed, while offering unparalleled control of the spectrum analyser and the data stream it generated when compared to prior solutions.

The novel 2D image filtering approach for reducing noise from the classification of spectrum observations also proved to be a very useful addition to the overall system by substantially improving the SNR performance of the energy detector. This addition considerably increased the confidence that incumbent event misdetections will be reduced, and minimal noise will be incorrectly classified. This added advantage of this approach is the significant reduction in fragmentation of the resulting classified secondary spectrum, which proves to be a highly desirable attribute for the analysis and construction of Whitespace Opportunity Distributions in Chapter 5.

Chapter 5

Quantifying Secondary Spectrum

Resources

5.1 Introduction

This Chapter details several primary contributions made possible from the highly capable measurement system developed in Chapter 4. The new access to high resolution, long duration and large instantaneous bandwidth datasets, now enables the analysis and development of algorithms to determine both the distribution secondary spectrum resources and which timescales present optimal opportunities to exploit secondary spectrum resources.

Section 5.2 raises a distinction between simply observing incumbent activity in candidate bands simply on a frequency basis, versus determining how the spectrum is used (a use-case), and then mapping the incumbent activity to that use-case. By performing observations and then attaching those observations to a use-case, rather than a fixed range of frequencies, the observations can now be applied globally to other frequency bands where similar use-cases exist as comparison datasets.

Section 5.3 identifies two candidate frequency bands and their associated use-cases for which high resolution continuous spectrum observations are performed.

Section 5.4 first provides a background into existing spectrum analysis techniques, and how these techniques are quite limited in their application. This section then proposes a new approach to performing secondary spectrum analysis by selecting an appropriate timescale to measure discrete opportunities against and then details implications that the Fast Fourier Transform presents with respect to spectrum observation and analysis.

Section 5.5 introduces the notion of ‘functional whitespace’ and details a formal method for which to evaluate discrete whitespace opportunities based on the spectrum resources that it encapsulates. The notion of a WhiteSpace Resource Block (WSRB) is also introduced as a fundamental unit for which to perform evaluation of whitespace from.

Section 5.6 builds on the functional whitespace concept, by determining an optimal way to cluster groups of WSRBs such that the functional value that they present is maximised across an observed secondary spectrum. This original work was published under the title: “Quantifying secondary spectrum Whitespace Opportunity Distributions” [8]. The approximate optimal clustering is achieved through a novel procedural partitioning solution to an extreme type of ‘Maximum Wighted Independent Set’ (MWIS) problem, known to be NP-Hard [135, 136].

Section 5.7 presents a novel approach to displaying the partitioned secondary spectrum whitespace opportunities via means of a Whitespace Opportunity Distribution. This visual analysis tool enables rapid and accurate evaluation of the types of secondary spectrum opportunities that exist within a band of interest, providing a very powerful means of comparing candidate frequency bands and exposing vastly more information about the secondary spectrum opportunities than classical evaluation methods.

The lack of a formal classification method to evaluate or determine the quality of secondary spectrum whitespace is a fundamental issue when assessing the viability of

various candidate bands from the perspective of the whitespace resources that a potential candidate possesses in space, time and frequency. There are many potential measures for spectrum evaluation such as information capacity, bandwidth, intrinsic physical and spatial propagation properties, and geographic assignments. However no self contained evaluation method relating the time and frequency of secondary spectrum opportunity exists.

The focus of the analysis is in the time and frequency domains without considering information capacity, as the ‘raw components’ of spectrum capacity are time and bandwidth with respect to capacity. The analysis is performed for a discrete point in space, to make the initial problem more tractable, while still deriving useful information for the evaluation of the secondary resources contained within an observed band and geographic location.

Typical analysis methods for presenting whitespace availability are performed either via duty cycle or spectrum occupancy plots [7] highlighted in Subsection 5.4.1. Duty cycle and low resolution spectrum occupancy analysis of spectrum provides high level identification of candidate bands and snapshots into possible useage patterns over the course of an observation period. The resolution and timescales (typically minutes or hours) observed and subsequently presented, are insufficient to provide a complete analysis of all incumbent activity at the point of measurement within these bands.

5.2 Secondary Spectrum Use-Cases

With respect to the particular bands chosen for investigation, the overall utilisation of the band and the types of services present (and use-cases) are what are of interest, not the specific frequency ranges observed (other than within an Australian spectrum allocation context). This distinction is important as spectrum allocations and bands are not

harmonised globally and are thus not directly transferable from country to country, the analysis of the use of wireless devices in a specified use-case, however, is indeed transferable. Thus a use-case approach to performing spectrum observation and evaluation is useful to consider when modeling secondary spectrum environments.

The necessity for secondary spectrum environment models is directly connected to informing the requirements, design and thus impacting the overall performance of devices attempting to communicate within those environments. Suppose a device attempting to access secondary spectrum. For the device to function correctly and to ensure that its operation will not interfere with an incumbent, the device requires substantial information about the spectrum environment before the spectrum can be accessed. This information will need to be comprised of the types of whitespace opportunities that exist across the current bands of interest. In particular, the timescales of these opportunities and whether the device can select some optimal timescale where access to spectrum resources is maximised, with minimal wasted energy. As such, specific information regarding the duration, bandwidth and center frequency of the expected whitespace opportunities will also be of critical importance for communication planning and energy cost minimisation. Hence there is a significant need for robust identification of candidate bands and investigation into the structure of the whitespace resources and the timescales of those resources within each candidate.

Any sufficiently high performing secondary spectrum device will likely be designed for a target or range of bandwidths and transmission opportunity lengths (timescale), thus when designing this device a high degree of certainty for the type of secondary spectrum opportunities present from both a bandwidth and duration perspective needs to be known. These constraints will then be used to inform the optimal bandwidth to be used for communications and the time that a particular band or opportunity can be used for before having to hop to a new opportunity. These constraints will then factor in

to how quick the device needs to react to a secondary spectrum opportunity and inform models for the energy cost of accessing such opportunities.

5.3 Measurement of Candidate Bands

The band and dataset focused on in this Chapter is at 465MHz centre frequency with 25MHz instantaneous bandwidth. The band was observed over a 24 hour period generating over 8TB of data on the 17th/18th of March 2016 at Macquarie University, Sydney, Australia. The selection of the band was due to the devices and services allocated within this band in Australia. The band is divided completely into 12.5kHz and 25kHz allocations, which are allocated for services such as community broadcast radio, taxi cab radios, emergency services, rail corridor, and low data rate point to point links for industrial monitoring and communications [26].

The 465MHz band was a likely candidate to find significant amounts of whitespace with a high resolution analysis due to the spectrum use-cases within that band. Additionally, this study provides new insight into a lesser studied band as a potential candidate for secondary spectrum use. Most secondary spectrum measurement campaigns focus on TVWS bands (Section 2.5.2) or the 2.3GHz/3.5GHz bands for LSA and SAS (Section 2.5.3), however those particular use cases do not exist in Australia as of yet. The 465MHz band is an interesting candidate in this context as it is a band with numerous small channels and low total utilisation (similar to the 3.5GHz band in the US).

Due to the small bandwidths and expected spurious use cases in the 465MHz band, it was deemed ideal for testing the performance of the measurement system and development of whitespace evaluation techniques. The typical Power Spectrum Distribution (PSD) plot of this spectrum is below:

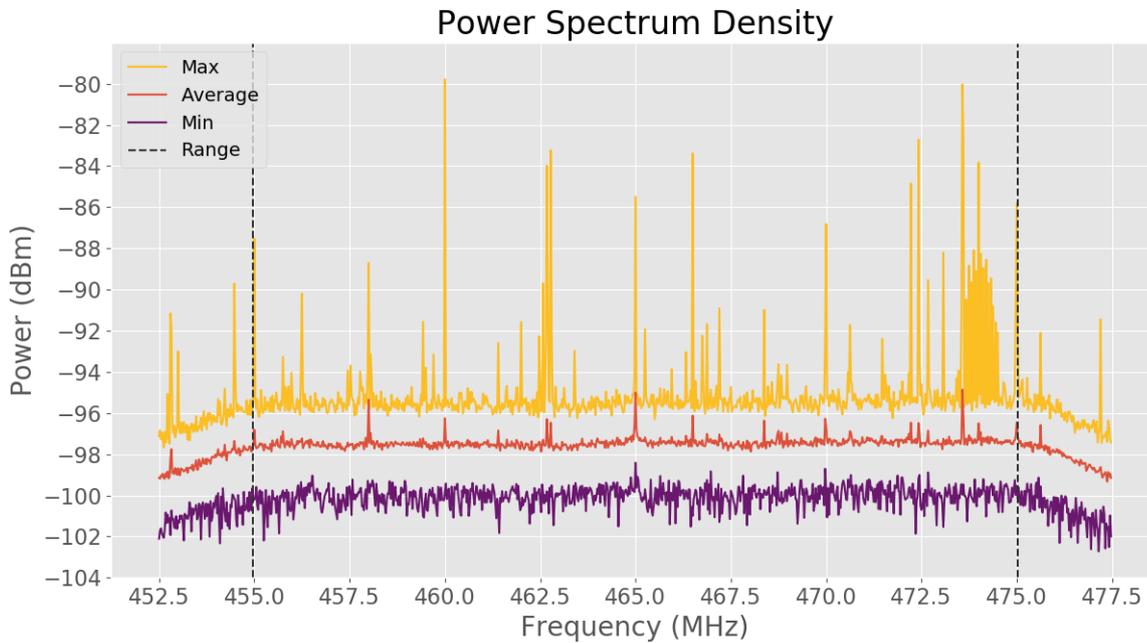


Figure 5.1: PSD of 12.5/25kHz Mobile/Fixed Band, 452.5 to 477.5 MHz Bandwidth, 190Hz resolution. Each min, max, avg point spans 64 bins.

Figure 5.1 shows a snapshot of the 465MHz band, the small spike at the centre of the band is due to the down conversion mixing from the receiver. There are numerous small bandwidth spikes across the band, which can be determined due to the low average and minimum power levels for the relevant frequency bin. Thus these spikes have bandwidths on the order of 12.5kHz ($190Hz/bin \times 64 bins = 12.16kHz$ spanned per pixel). The snapshot also covers a time period of 50ms, due to an averaging of 10 frames used (with approx 5ms timescale resolution per frame). An interesting cluster of incumbent activity can be seen at approx 474MHz, which is likely from a low bandwidth telemetry system.

To contrast with the expected sparse spectrum usage at 465MHz, a second dataset was collected at the 880MHz band. This band was selected as it comprises of two separate upper band mobile communications allocations: 870 — 880MHz owned by Vodafone Hutchinson Australia for 4G services, and 880 — 890MHz owned by Telstra Corporation for 3G services. This was captured to present the use case of a 3G/4G operator in contrast

to the 12.5/25kHz band. A PSD of the 3G/4G use case is presented below:

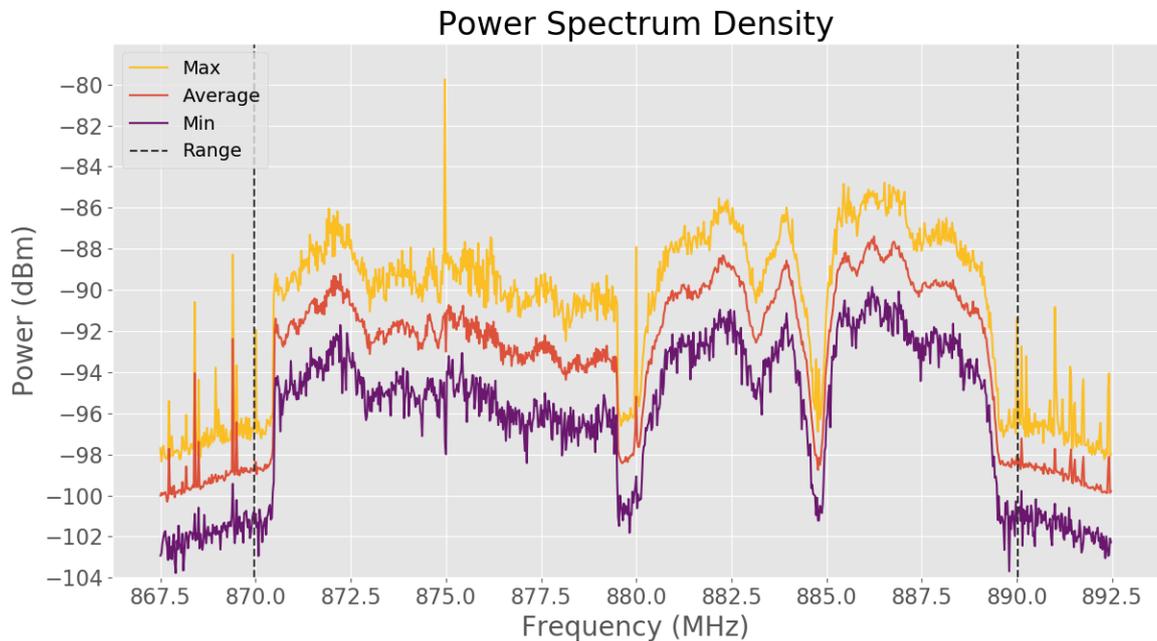


Figure 5.2: PSD of 3G/4G Mobile Band, from 867.5 to 892.5 MHz Bandwidth, 190Hz resolution. Each min, max, avg point spans 64 bins.

Figure 5.2 clearly shows the 4G and 3G services present in the observed band. At 875MHz the pilot signals can be clearly identified, and a significant decrease in signal power can be seen at approx 885MHz, which is the guard band in between two separate 3G signals (likely uplink and downlink). A guard band can also be identified at approx 879.5MHz to minimise interference between the 4G and 3G services

5.4 Spectrum Analysis

In order to design a device to perform dynamic secondary spectrum communications, there needs to be certain performance requirements established of the device that are informed by the spectrum resources that the device is attempting to access. As such, significant effort must be expended to observe to a high degree the ‘structure’ of the underlying

secondary spectrum resources within a given band of interest.

The structure of whitespace effects the ‘functional value’ of that whitespace opportunity. The concept of ‘functional efficiency’ of secondary spectrum was explored in Section 2.5, where spectrum allocations with long durations and bandwidths are preferred due to the large contiguous amounts of spectrum that can be optimally managed. ‘Functional value’ of whitespace is derived similarly, whereby a discrete whitespace opportunity that exists for several minutes with several megahertz of bandwidth, is of greater value to a secondary user, than an opportunity that exists for a couple seconds with the same bandwidth (however there may be more of these short timescale opportunities, which could be useful to exploit).

More concretely: an opportunity that persists for a very long period of time, presents a much more valuable proposition to take advantage of (from the perspective of a secondary user) than an opportunity that will expire rapidly. Further, maximising the bandwidth of an opportunity also increases its functional value as its instantaneous information capacity is greater. In essence, opportunities with very large bandwidths and long persistence times are highly valuable due to their greater utility, or functional value.

5.4.1 Existing Spectrum Analysis Methods

Existing spectrum analysis methods [7] such as duty cycle and occupancy plots only provide some (limited) pieces of the overall secondary spectrum picture. A duty cycle plot only indicates the proportion of the observation period that an incumbent event was detected. In some cases this is broken down into hour intervals yielding only limited additional information toward the structure of the secondary spectrum.

A duty cycle plot such as Figure 5.3, shows total incumbent activity for an observed frequency range as a fraction of the total observation period. It can effectively be used to gauge how active an incumbent is across the observation period.

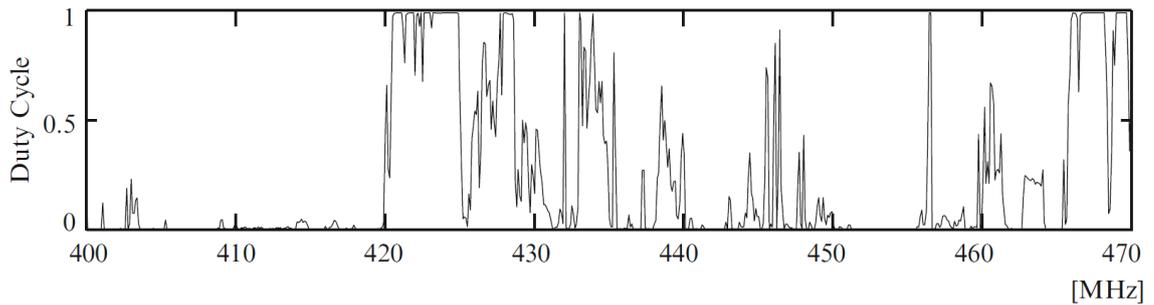


Figure 5.3: Example Duty Cycle Plot, ‘Overview of the UHF spectrum in The Netherlands: close-up of the 400 to 470 MHz band’ [7] pp.484

It is important to know the specific time periods that incumbent events are present in order to establish any kind of trend, pattern or underlying structure of the incumbent activity, rather than holding a running tally of time occupied vs time unoccupied. Typical spectrum measurement campaigns would preserve at least some of that information (possibly on minute or longer timescales).

An occupancy plot such as Figure 5.22 is useful for representing detected incumbent activity within a particular region over time, the activity or measured transmission power (given a known location) can be used to establish boundaries, exclusion zones and identifying high or low areas of incumbent activity. However this method too only presents information about the total spectrum occupancy over an observation duration and provides no further detail toward underlying secondary spectrum structure, especially when time intervals between observations are large. Occupancy plots taken from several locations over time are typically employed in constructing Radio Environment Maps (REMs).

5.4.2 Spectrum Analysis Tools

To meet the deficiencies identified with respect to existing commonly employed spectrum analysis and evaluation techniques, a set of tools were developed to perform deeper analysis of the datasets captured by the spectrum analyser developed in Chapter 4, to

investigate a more powerful method of representing secondary spectrum resources. The tools developed were separated into two packages: the first for processing the raw data into a useful output for further analysis, and the second for performing the various forms of analysis and evaluation.

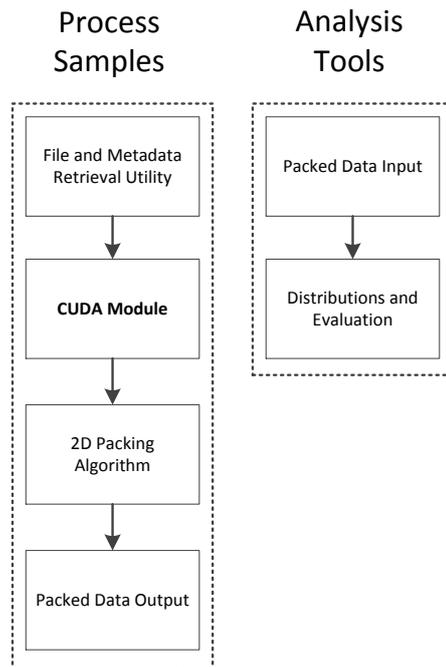


Figure 5.4: Process Samples package and Analysis Tools flow diagrams

Figure 5.4 provides a general overview of the main components that make up the two utilities for processing and analysing secondary spectrum opportunities. Process Samples is designed to ingest large spectrum measurement datasets and use the CUDA Module outlined in Figure 4.10, that performs the correct data conditioning and signal processing to calculate a time series power spectrum. Energy detection is then performed on the power spectrum by applying a threshold and filtering stage to classify spectrum events as an incumbent event or whitespace as outlined in Section 4.4. After classification, the now

classified data set is fed into the core feature of the Process Samples package, ‘the packing algorithm’ outlined in Section 5.6. The results of the packing/partitioning operation is output to a .csv file to be further analysed. The specifics of the analysis is explored in Section 5.7.

These tools were designed to enable a deeper investigation into how whitespace is structured and how that structure can be represented and evaluated at a glance, with the objective of yielding greater insight of the underlying secondary spectrum resources contained within a band.

5.4.3 Timescale Bound Detectability of Whitespace

As a thought experiment: if spectrometers were distributed over the entirety of a given geographic region with the objective of detecting spectrum utilisation, it would likely be found that significant periods of time exist where spectrum in this band is unused either geographically or temporally. For the purposes of the thought experiment, only the temporal case will be addressed.

There must also exist a particular timescale at which certain incumbent activities are performed, thus a smaller timescale has the potential for opening up secondary spectrum opportunities between incumbent activities that previously did not exist. Increasing the sampling rate of the spectrometers (or decreasing the averaging) further to reduce the observed timescale would increase the visibility of previously detected whitespace clusters, and also unveil new whitespace clusters. As the timescale is further reduced, more clusters would likely be detected between incumbent events and subsequently previously detected clusters would appear larger (or ‘stretched’, relative to the timescale). *For sufficiently small time intervals, a channel that once appeared to have no whitespace opportunities, appears more like free spectrum.*

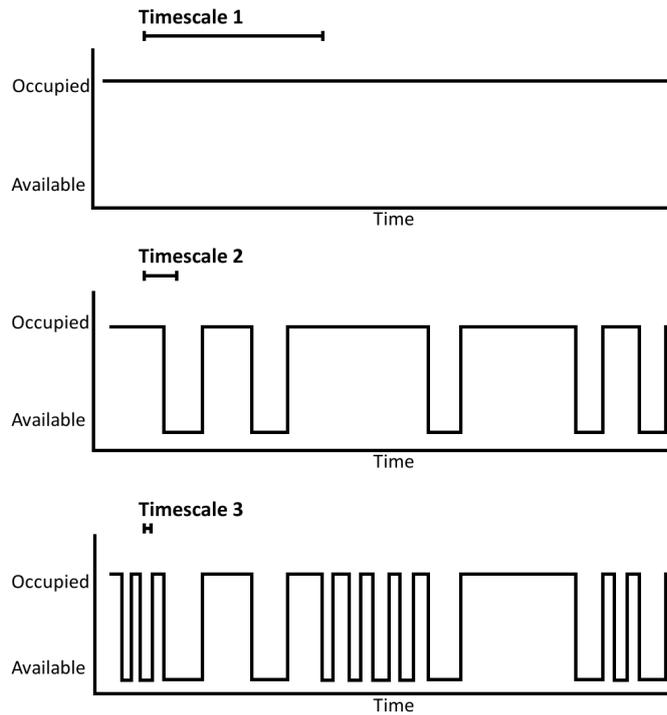


Figure 5.5: As the timescale at which secondary spectrum can be observed decreases, there is a higher likeliness that secondary spectrum opportunities will be detected, capturing a more accurate representation of the actual secondary spectrum resources present.

To illustrate this thought experiment and assuming a fixed location and spectrum modeled as a binary state either occupied (an active incumbent present) or available (no active incumbent). Figure 5.5 shows a primary spectrum owner reporting that the spectrum is ‘always’ unavailable for secondary users. For example, a military frequency band, TV channel, or commercial radio station may always be off-limits for secondary users (Timescale 1), especially for secondary users that require opportunities on the timescale of hours or longer. However, if we look more closely at that spectrum band with a timescale resolution on the order of minutes, we may find that there are indeed ‘minute intervals’ over which whitespace exists (Timescale 2). Increasing the timescale resolution further, toward seconds or sub-seconds, some of the previously unavailable portions of

timescale 2 may now present new opportunities that could be captured and subsequently exploited (Timescale 3). As this timescale is decreased further, subsequently more secondary spectrum can be detected.

This approach is of particular interest to spectrum environments that have a different set of use-cases pertinent to it than the ones previously mentioned. In particular, use cases that are completely allocated to a particular user, however that use case involves spurious use of the spectrum, such as: Emergency Services, Radiolocation, and Amateur Radio.

The challenge with harnessing this type of secondary spectrum (in contrast to the previously mentioned, coordinated secondary spectrum management methods presented in Chapter 2) is the absence of any centralised control or coordinated spectrum brokering to manage the secondary spectrum resources. To exploit the dynamic nature of this uncoordinated secondary spectrum, detection and spectrum analysis needs to be performed to a sufficiently high degree to capture the true nature or structure of the underlying secondary spectrum resources before any models and predictions can be developed.

General models for incumbent spectrum usage can be achieved with a sufficiently large spectrum observation data set, over time, space and frequency. However, minimum requirements for device performance, such as minimum set up time, and maximum continuous transmission length, can be established by determining the expected incumbent activity from spectrum observations.

The underlying structure of the secondary spectrum resources however is typically non-static and will vary over time, space and frequency. If a model is developed assuming that the secondary spectrum resources are static in any dimension, without supplemental realtime sensing at a minimum, there exists a significant risk of collision with a primary user, which is unacceptable. Therefore the following analyses and evaluations do not assume that the observed spectrum activities are static over time, only the underlying use

cases are static. Substantially more data is required before any kind of practical spectrum prediction/forecasting/trend analysis can be performed to any degree of certainty.

5.4.4 Whitespace Resource Blocks

To perform evaluation of whitespace, it must first be quantised in a way that is convenient for totaling and manipulating to enable calculations and comparisons between different whitespace opportunities. Quantising whitespace enables a connection to be made between the ‘structure’ of an opportunity and the impact of that structure on the overall functional value of the whitespace opportunity.

The structure of secondary spectrum is approached in the OFDM sense where time and frequency are represented discretely. Each ‘resource block’ represents a chunk of electromagnetic spectrum of a given duration and bandwidth, at a particular point in time and frequency.

A WhiteSpace Resource Block, or WSRB, represents available secondary spectrum analogous to how a resource block represents spectrum resources in LTE. In this context we introduce the WSRB as a fundamental quanta of whitespace for which we can evaluate a single unit of whitespace as a function of bandwidth β (Resolution Bandwidth, Section 4.6) and duration τ (Timescale, Section 4.5):

$$\beta = \frac{\text{Sample Rate}}{\text{FFT Depth}} Hz \quad (5.1)$$

$$\tau = \frac{\text{FFT Depth}}{\text{Sample Rate}} s \quad (5.2)$$

As can be seen, the fundamental whitespace unit is dependent on the selected FFT depth for which the observation is being performed, assuming a fixed sample rate.

5.4.5 FFT Depth Selection

As the FFT depth changes the time and bandwidth dimension of a WSRB, the impacts of particular FFT depths was investigated to determine the potential technical implications and challenges with respect to determining the structure of whitespace.

The computations were performed on the same dataset comprising of a 5 minute observation period with a sampling rate of 25MS/s (complex - 50MS/s effective, yielding 25MHz bandwidth). These measurements were then put through a classifier to reduce each observation point to either an incumbent detected event or whitespace opportunity.

These measurements were represented as a waterfall plot, which shows the spectrum activity over a contiguous period in discrete time steps.

The base case of these snapshots was performed at an FFT depth of 2^{17} points. Thus a single pixel represented in Figure 5.12 represents a WSRB with a bandwidth of approx 190Hz and 5ms duration, as per Eqns 5.1 and 5.2.

Each pixel in every image also represents the approximate 190Hz and 5ms time steps, which is the ‘native resolution’ for Figure 5.12. Native resolution in this context is analogous to the native resolution of a digital display, where when operating at its native resolution each pixel is correctly displayed without ‘smearing’ effects, a non-native resolution will result in pixels being incorrectly represented, resulting in a smearing effect where neighboring pixels blend together on the display. By representing each image with the same frequency bin width and time step duration per pixel, this a point by point comparison on both axes across all figures to visualise the impact that various FFT depths present.

As the time and frequency steps are constant across the images, ‘smearing’ of frequency and ‘compression’ in time is observed. Smearing and compression are related according to Eq 5.1 and Eq 5.2, where to reduce the smearing of power across the frequency, the FFT bin widths need to be smaller. Making these bins smaller requires the timescale to

increase proportionally. Similarly, if the timescale is reduced, to gain greater time step precision, the frequency bins will spread out causing the smearing effect. The increase in time steps requires ranges of time steps to be binned or ‘compressed’ into a single row of pixels.

When displaying these data points, this relationship is apparent as there are only a finite number of pixels available for the image to be displayed upon. This phenomena is due to the inherent uncertainty between the discrete frequency and time resolution of an FFT, otherwise known as the ‘Uncertainty Principle’ [137].

Thus, in Figure 5.6 each row of pixels, comprises of 64 discrete time steps, whereas in Figure 5.12 each row of pixels only comprises a single time step. The compression of time is calculated in each respective figure by $\frac{2^{17}}{\text{FFT Depth}}$ with the number of incumbent events represented per row of pixels, determined by its colour and the associated colour bar value for ‘whitespace density’. Similarly, the smearing in frequency can be calculated by the inverse as smaller FFT depths yield larger frequency bins.

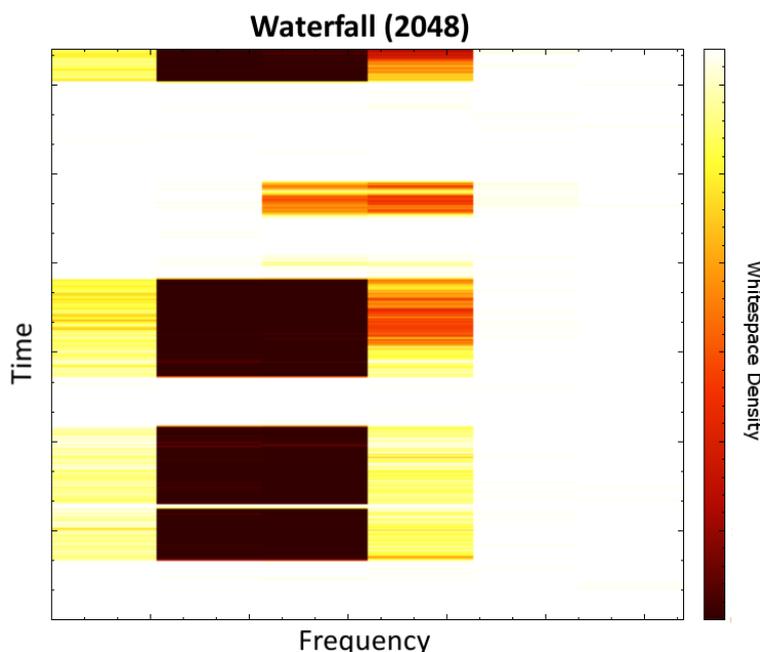


Figure 5.6: Resulting spectrum with FFT depth set to 2048 bits (2kb)

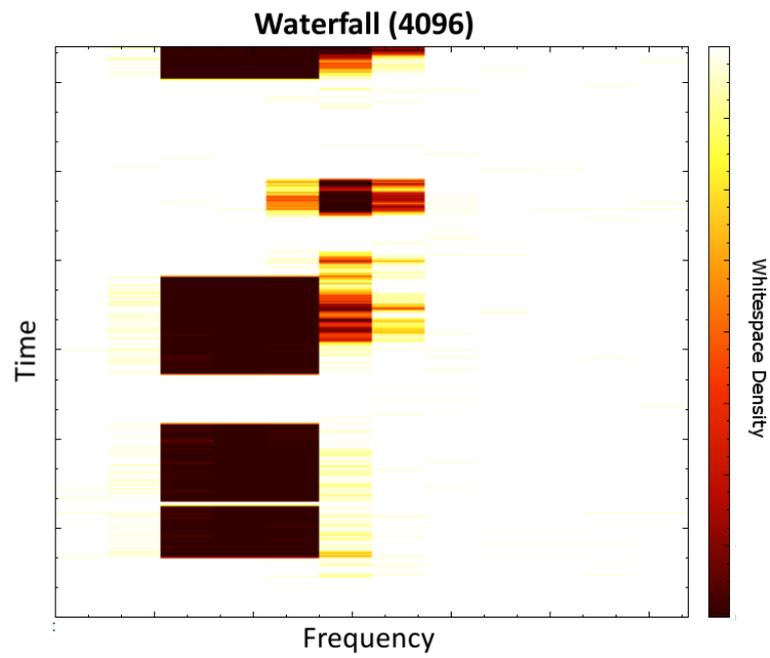


Figure 5.7: Resulting spectrum with FFT depth set to 4096 bits (4kb)

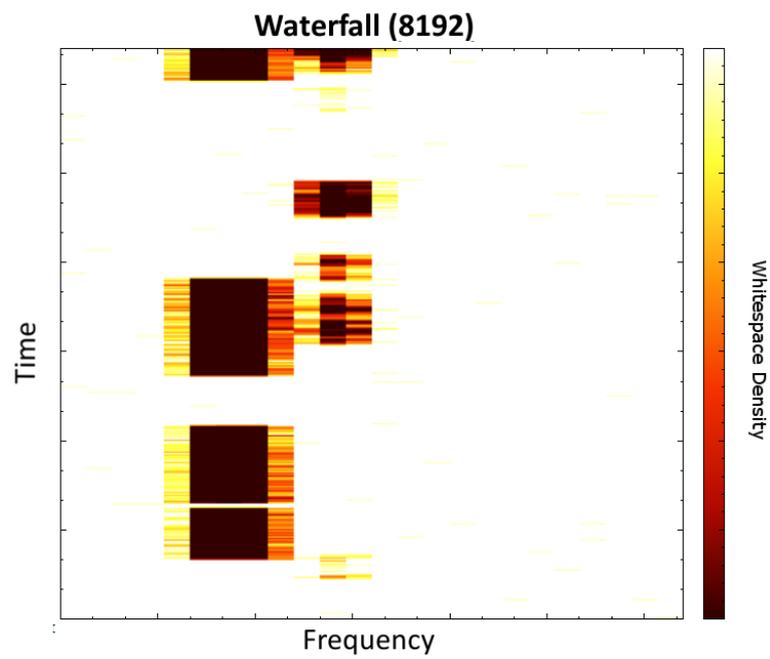


Figure 5.8: Resulting spectrum with FFT depth set to 8192 bits (8kb)

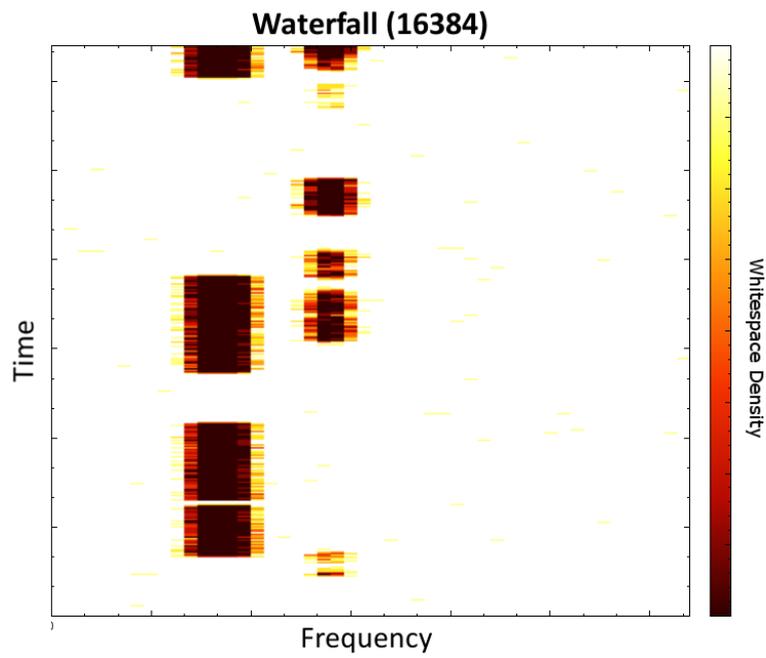


Figure 5.9: Resulting spectrum with FFT depth set to 16384 bits (16kb)

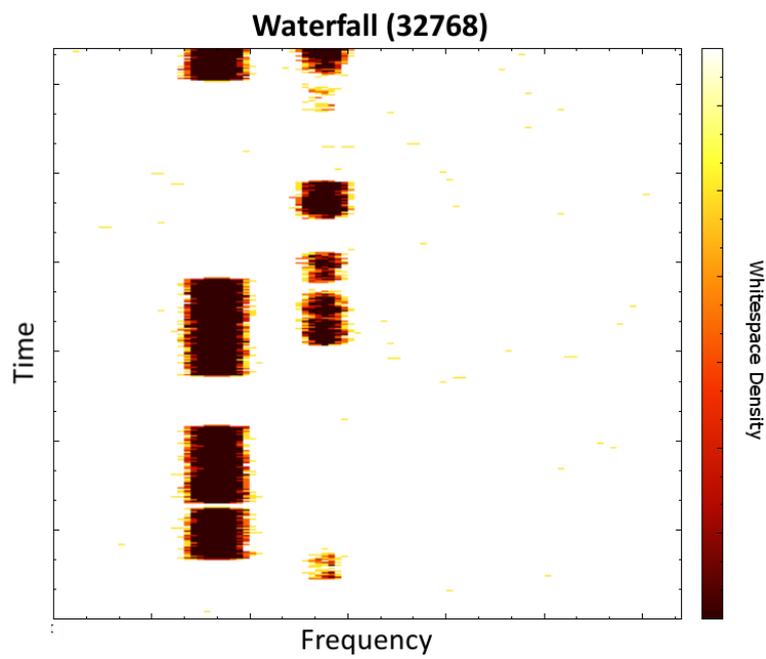


Figure 5.10: Resulting spectrum with FFT depth set to 32768 bits (32kb)

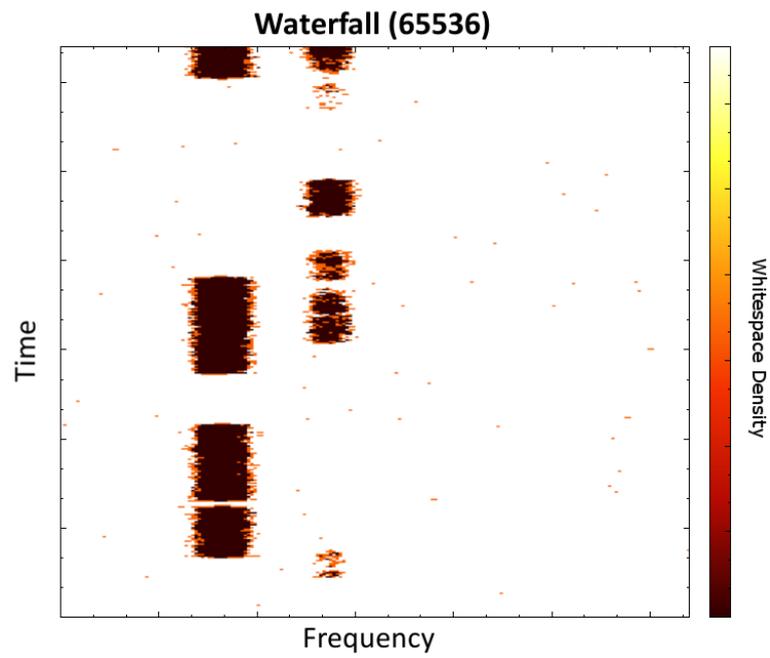


Figure 5.11: Resulting spectrum with FFT depth set to 65536 bits (64kb)

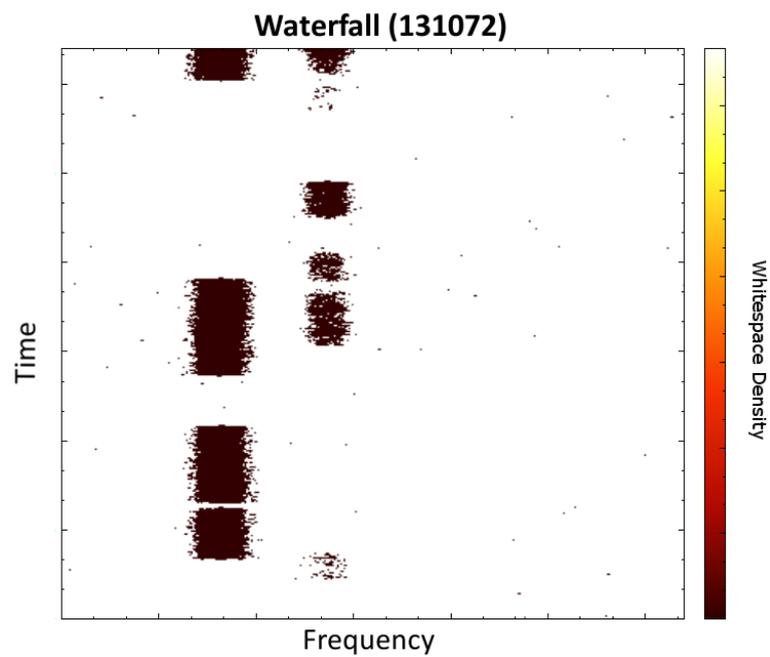


Figure 5.12: Resulting spectrum with FFT depth set to 131072 bits (128kb)

It can be clearly seen from these images that as the FFT depth increases, the certainty of the incumbent signal in frequency becomes much higher, while the time certainty decreases. Each WSRB is equivalent from a time and bandwidth perspective, no matter what FFT depth is used, the functional resources contained are unitary.

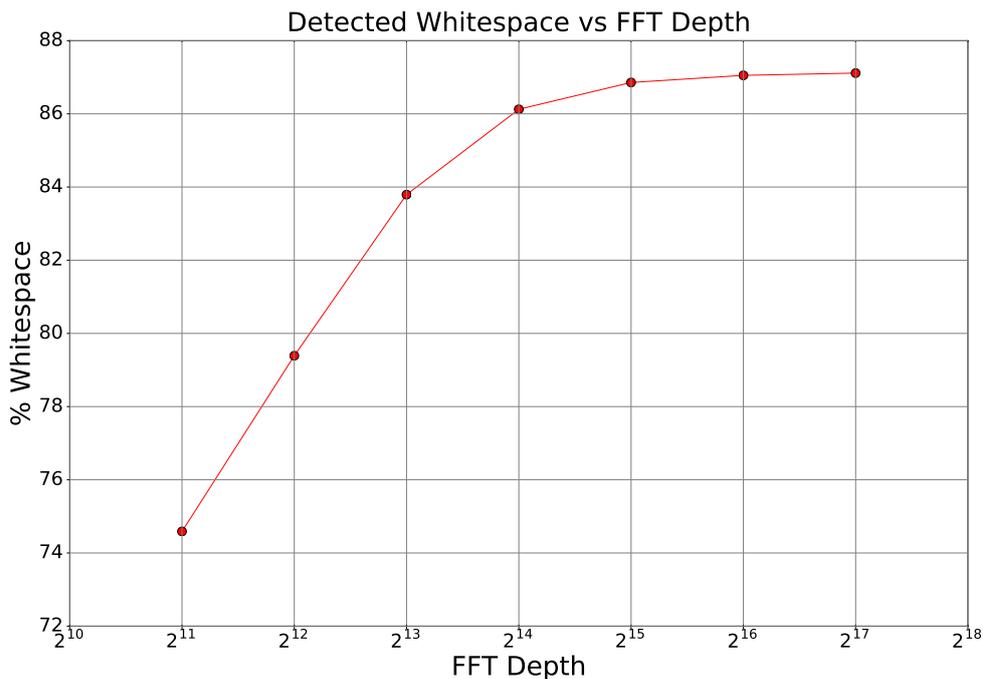


Figure 5.13: By increasing the FFT depth, higher frequency certainty is obtained for incumbent signals, which result in higher amounts of discovered whitespace due to the smearing of the incumbent signal power in the frequency domain being minimised

It is observed that at lower FFT depths, a significant amount of bandwidth smearing is observed, however the time certainty is higher. As WSRBs are unitary regardless of FFT depth, a comparison of all observed WSRBs is performed to calculate whether higher FFT depths should be used to detect whitespace with a higher frequency certainty, or with a higher time certainty. This comparison is performed across the entire 5 minute, 25MHz bandwidth sampling period where the computed incumbent activity is compared

against the entire spectrum resources. The results are contained in Figure 5.13, with the x axis noting FFT Depth and the y axis showing the percentage of total observed spectrum that is whitespace.

Higher RBWs result in a reduced smearing of the incumbent energy over the spectrum, which is desirable for maximising the detection of ‘total available whitespace’. The notion of total available whitespace fundamentally requires ‘perfect’ knowledge of the spectrum being observed in order to detect every single WSRB that exists. Perfect knowledge means to have an omnipotent view of all the spectrum events at all times to find exactly how much whitespace can and should exist, an ideal scenario that, perhaps due to the inherent randomness of spectrum, is highly improbable to achieve. Thus an upper bound for whitespace availability, will be used in place as a theoretical value that approximates the ‘perfect’ whitespace availability.

From Figure 5.13 higher RBWs provide a more accurate detection of available whitespace (at the expense of timescale resolution), approaching the upper bound for total available whitespace and likely the upper bound for functional whitespace (due to sub 5ms timescales being very hard to capitalise upon with modern devices). Perceivably there is an ideal FFT depth for which the best time resolution vs frequency resolution trade off exists, however due to computational efficiency a base 2 FFT depth will result in the closest approximation feasibly possible. The aforementioned ‘ideal resolution’ is the point whereby, increasing FFT resolution results in no additional secondary spectrum resources identified. It is also possible that an increase in resolution results in a reduction of observed secondary spectrum resources due to smearing effects. Efforts to detect ‘all’ discrete secondary spectrum resources is however practically infeasible and can really only be demonstrated in simulation where all WSRBs are known prior or are completely deterministic.

It should be noted that there is a significant fall-off in the amount of whitespace de-

tected as the FFT depth progresses past 2^{14} points, this is likely due to the RBW being significantly smaller than the bandwidth of the incumbent event bandwidths (12.5KHz), sufficiently minimising the frequency smearing. Figure 5.14 illustrates how incumbent energy is smeared across frequency due to the selection of the resolution bandwidth (RBW) due to FFT depth.

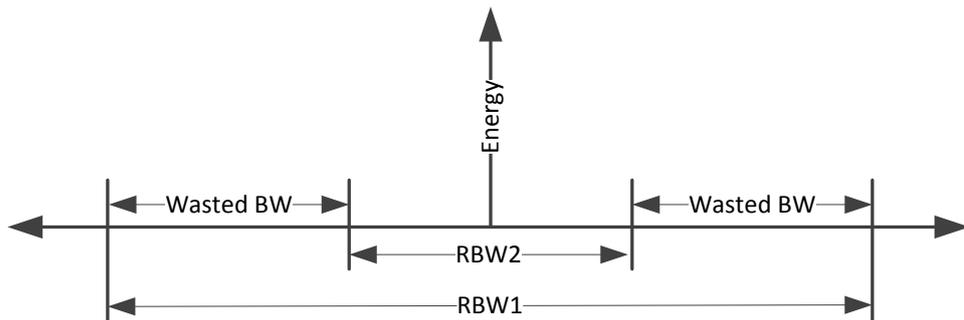


Figure 5.14: Diagram of the loss of Bandwidth resolution with respect to the energy of a spectrum event ‘smearing’ across different FFT bin sizes

These results provide an interesting insight. The impact that low FFT depth analysis of secondary spectrum paints a substantially pessimistic view of the actual secondary spectrum resources that are present within a band of interest. In fact, as incumbents are limited to specific frequency bands, it makes sense to increase the frequency precision of a measurement to correctly locate the incumbent signal energy in frequency, as neighbouring bands will have no incumbent activity, yielding more whitespace resources with higher value structuring. Further as an incumbent will likely never change its frequency of operation, slightly longer timescale steps will have limited impact on the overall detectability of whitespace. It is to be noted that no noise filtering is applied to the dataset, only thresholding is performed before the whitespace is tallied. The final reason for improved whitespace detection is that most incumbent events occur on timescales much greater

than 5ms. Increasing the timescale resolution beyond 5ms has little to no improvement in the detectability of whitespace in this particular dataset and band of interest, primarily due to the increased negative impacts of power smearing across frequency at shorter timescales and that the incumbents occupy relatively small channels.

5.5 Measuring Functional Whitespace

To evaluate whitespace efficiently there is limited value in taking note of every discrete whitespace block due to the bandwidth and time spanned by a single WSRB as tracking or creating a database of every discrete WSRB would generate a dataset an order of magnitude larger than the underlying sample set.

It is useful if not essential to group WSRBs into clusters. This approach has a twofold advantage, the amount of memory required to track the resulting whitespace opportunities is significantly less (than the raw samples), and grouping or clustering WSRBs is more useful for determining their structural value.

Fundamentally, the intrinsic value of a cluster is the information capacity contained by that cluster. This can be computed by the time and frequency spanned by the opportunity, times its information capacity. An accurate information capacity calculation for each cluster could require a channel model, fading model, transmitter and receiver locations, transmission powers, and communication technology to be known as well as bandwidth (and duration). Thus, we choose to simplify and only look at the Duration Bandwidth Product (DBP) $b \times d$ as a proxy for capacity for simplified evaluation of whitespace opportunities.

A whitespace cluster \mathbf{w} is a contiguous set of WSRBs aggregated over time and frequency to form a ‘useful’ slice of spectrum, represented as a two dimensional rectangular vector consisting of $(b \cdot \beta) \times (d \cdot \tau)$ WSRBs, of the form $\mathbf{w} = (b, d)$, where b is the number

of adjacent WSRBs in frequency and d is the number of sequential WSRBs in time.

To create a whitespace cluster, a 2 dimensional grid of sequential ‘spectrum frames’ \mathbf{X} is composed. A spectrum frame has time resolution τ and is the output of a single FFT computation after power estimation and thresholding is performed. \mathbf{X} is an m by n grid: the frame index $i = 0, \dots, m - 1$ where m is the maximum observation period in frames and frequency bin index $j = 0, \dots, n - 1$, where FFT depth is n . $x_{i,j} \in \{0, 1\}$, where 1 identifies a WSRB and 0 an incumbent event.

5.5.1 Whitespace Slices

To generate basic whitespace clusters, the most trivial case is to generate slices. These slices are generated by setting either the duration value to 1, $\mathbf{w}_k = (b, 1)$ or the bandwidth value to 1, $\mathbf{w}_k = (1, d)$, effectively becoming a single dimension analysis. These can be selected to either evaluate the possible bandwidths that the spectrum can support, or the expected duration of whitespace opportunities.

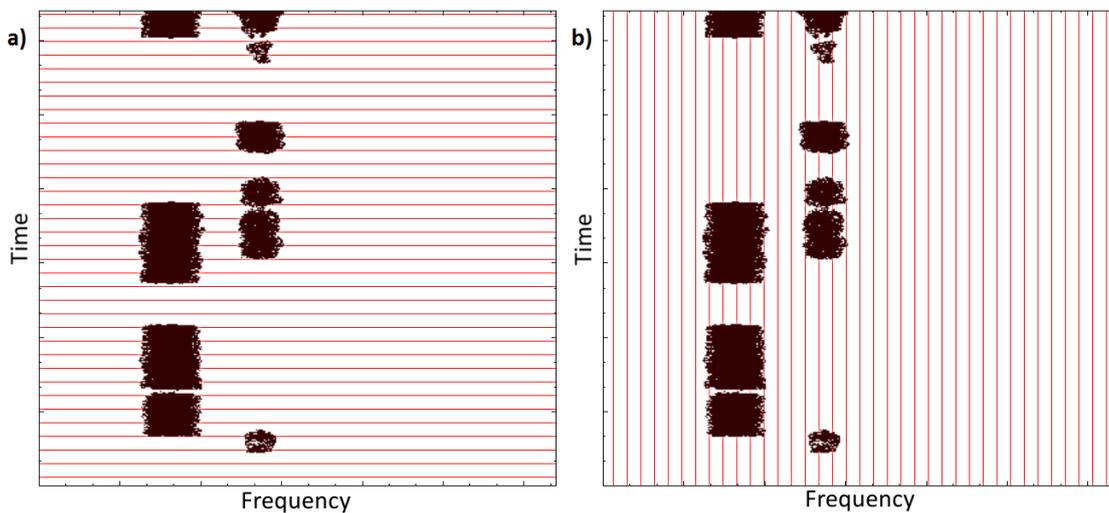


Figure 5.15: Single Dimension Analysis (Slicing) of \mathbf{X} (from Left to Right) a) Bandwidth Analysis Slices. b) Duration Analysis Slices

A bandwidth slice is obtained by setting the observed duration to 1. Each frame (row) in \mathbf{X} is iterated over and searched for continuous WSRBs in bandwidth. This yields the distribution of discrete bandwidths present within the spectrum over the observation period. This however, is effectively the same as stacking numerous separate data series on top of each other as there is no intrinsic information (or metadata) generated to link the rows in \mathbf{X} together as per Figure 5.15a.

A duration slice is obtained by setting the bandwidth to 1 and scanning each (column) of \mathbf{X} for contiguous WSRBs in time. This method is effectively the same as stacking numerous separate data series next to each other as there is no intrinsic information generated to link the columns together as per Figure 5.15b. The sum of each cluster duration per bin can be compared with the total observation period to determine the duty cycle.

These methods provide limited insight into the value of a spectrum and lead to misleading portrayals of whitespace. For example, if only the discrete bandwidths are known, there is no knowledge of how long each opportunity of a given bandwidth persists for and vice versa. It is possible that numerous large bandwidth opportunities are detected, however they may be scattered substantially across the observation period, rendering them useless for analysis and unusable for practical purposes.

Thus, a single dimension analysis is ineffective for evaluating whitespace, both the bandwidth and duration of an opportunity must be known simultaneously. To identify an opportunity as a function of bandwidth and duration, requires a 2D partitioning of \mathbf{X} . 2D partitioning will enable the DBP of an opportunity to be computed (and its capacity, channel models permitting) and hence enabling the evaluation of an opportunity.

5.6 Whitespace Resource Partitioning

By determining the absolute bandwidth and duration of each whitespace opportunity, a proper evaluation of secondary spectrum can be performed. The bandwidth and duration of a whitespace opportunity enable the DBP of the opportunity to be determined which is used to evaluate the raw resource that exists from the real bandwidth and duration measurements. This raw resource evaluation can then be used within channel simulations to test the performance of devices within the secondary spectrum environment.

5.6.1 Problem Identification

Once the grid \mathbf{X} is constructed, whitespace opportunities with contiguous blocks of WS-RBs in duration and frequency are then ‘carved out’ of \mathbf{X} until no regions within \mathbf{X} have WS-RBs that are not spanned by a carved out whitespace opportunity.

The approach of stacking spectrum frames and performing a search over the 2D dataset \mathbf{X} to find contiguous rectangular blocks of whitespace, is a subset of the classical rectangular partitioning problem [138]. More specifically, this problem falls under the category of problems known as ‘A Maximum Weighted Independent Rectangular Partitioning of a Rectilinear Polygon with Holes’, a known NP-Hard problem [135, 136].

To frame the problem: the observed spectrum forms a rectilinear container with frequency positioned along the X axis and continuous time along the Y axis. The ‘holes’ within the container are incumbent active events, as a secondary spectrum device is not permitted to use any resources occupied by an incumbent. The rectangles to be partitioned are chosen to form a Maximum Weighted Independent Set (MWIS), which are the whitespace clusters.

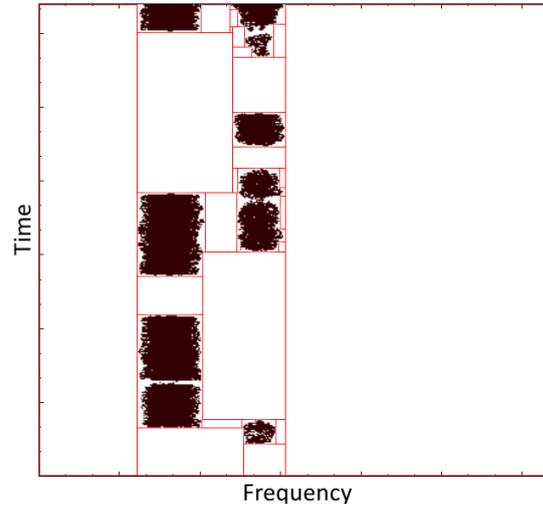


Figure 5.16: MWIS Partitioning of Spectrum Occupancy Grid \mathbf{X}

The MWIS problem itself, is NP-Complete, however due to the holes within the rectangular container (i.e. the incumbents) the resulting problem becomes NP-Hard and thus the maximum set can only be approximated and a maximum set is unknown. Figure 5.16 illustrates a possible solution to the MWIS problem.

We extend the previous definition of \mathbf{w} to a rectangle R on a 2D grid. We create each rectangle $R \in \mathbf{R}$ defined by a quintuple of frequency, bandwidth, duration, frame index and weight: $R = (R_f, R_b, R_d, R_n, R_m)$.

The set of rectangles are chosen to cover the largest possible contiguous area of the container per rectangle, minimize the total number of rectangles and maintaining that each rectangle forms part of an independent set. A set $\mathbf{Q} \subseteq \mathbf{R}$ is an independent set iff no pair of rectangles in \mathbf{Q} intersect: $\{Q, Q' \in \mathbf{Q} \mid Q \neq Q', Q \cap Q' = \emptyset\}$

The DBP of an opportunity is used to calculate its weight R_m . Weights are subject to minimum thresholds (B, D) to reduce partitioning fragmentation $\{R_m = R_b \cdot R_d \mid R_b \geq B, R_d \geq D\}$.

5.6.2 Brute Force Approach

The first approach toward determining the MWIS, employs a brute force approach for generating rectangles. A list of all possible rectangles spanning the spectrum was generated with no concern for overlapping rectangles. Once all possible rectangles were found, a conflict graph was generated that represents the rectangles as nodes and their overlapping regions as conflicts (edges).

To construct the conflict graph \mathbf{G} , for each $R \in \mathbf{R}$ we create a graph $\mathbf{G}(\mathbf{V}, \mathbf{E})$ where each node $v \subseteq \mathbf{V}$ inherits the quintuple from its representation in \mathbf{R} , $v \mapsto R$.

For $v \cap v' \neq \emptyset$, an overlap is present, thus creating a conflict between the two regions spanned by v and v' . A conflict edge $\{e \in \mathbf{E} \mid e = (v, v')\}$ indicates this conflict.

To find a maximal independent set $\mathbf{V}' \subseteq \mathbf{V}$, the condition where $\mathbf{E} = \emptyset$ must be met. More than one maximal set can be found in \mathbf{V} . The nodes within \mathbf{V}' must be selected appropriately so no nodes overlap. To find the largest maximal set, $\{\mathbf{V}'' \subseteq \mathbf{V} \mid \mathbf{V}'' = \max \sum_{i \in \mathbf{V}'} i_m \forall \mathbf{V}'\}$ must be met. \mathbf{V}'' is the largest weighted partitioning of \mathbf{X} .

We initially chose a largest node replacement strategy to create a maximal set from \mathbf{V} . The largest node in \mathbf{V} is selected, conflicting nodes are removed according to \mathbf{E} , then next largest node is selected and repeated until $\mathbf{E} = \emptyset$. This approach however was found to be extremely time and computationally intensive, thus we opted for a more time efficient approach.

To reduce the computational complexity, a second method to eliminate the need for a conflict graph was explored via a procedural ‘greedy’ algorithm.

5.6.3 Procedural Generation Approach

In the procedural generation approach, whitespace detection is performed procedurally, iterating through each time slice of \mathbf{X} , generating rectangles based on whitespaces present

in previous time slices. The algorithm prioritises contiguous regions of WSRBs in bandwidth, dependent on thresholds B and D , and then opts to maximise the duration of each rectangle before a decision is made to confirm the partitioning. The resulting set of whitespace opportunities are then used for evaluation.

Figure 5.17 highlights the main functions of the partitioning algorithm. Contiguous regions of WSRBs are partitioned due to incumbent events or noise that is greater than the defined threshold at the energy detection stage. In each spectrum frame (time step), the algorithm identifies ‘chunks’ of spectrum with contiguous WSRBs in frequency.

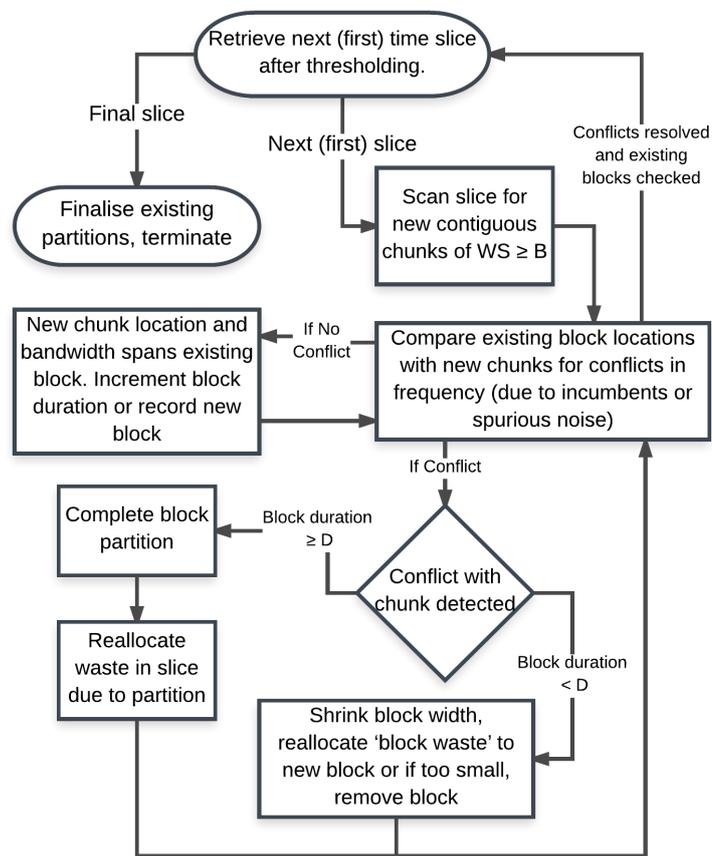


Figure 5.17: Flowchart of Partitioning Algorithm Main Functions, where B is Minimum Bandwidth and D is Minimum Duration

Once a minimum number of sequential time steps (greater than the time threshold) are spanned by the chunk, the bandwidth component of the chunk is converged to (partitioned). This convergence initialises the ‘block’ and defines the maximum bandwidth that this currently identified opportunity will possess, regardless of time. The block will only be terminated to form a cluster if an incumbent event (or spurious noise) is detected within current time step of the bandwidth spanned by the block. This termination point now determines the time component of the newly partitioned block, confirming the assignment of bandwidth and duration of that cluster.

In the case where the total time steps are below a time threshold, then the WSRBs tracked by that current block will be reallocated into two or more blocks in an effort to minimise orphaned WSRBs due to the incumbent event.

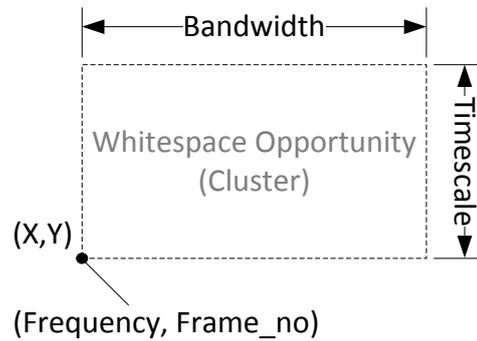


Figure 5.18: Recording Structure of WSRB Cluster

The algorithm tracks the timescale (y dimension, in discrete time steps of τ), starting frequency (beginning x coordinate), bandwidth (x dimension, in discrete frequency bins with width β), whitespace (unique counts of WSRBs for error checking) and frame number (ending y coordinate). This successfully encodes the coordinates, dimensions and area of each partitioned whitespace opportunity as a contiguous cluster of WSRBs. In combination with the configuration file 4.1 that is generated when the measurements are

taken, the actual frequencies and whitespace events in time can be determined. Figure 5.18 illustrates how partitioned WSRB clusters are structured and stored.

This approach proved to be much more useful when evaluating candidate bands as the partitioning is effectively performed as new frames are made available, providing near realtime performance (limited by data read speeds), reducing the overall time for candidate evaluation. As this is done as a post processing step, realtime performance in this context is somewhat less important than the tremendous speedup offered when compared to finding the maximum independent set of a conflict graph. An example snippet of the output from the partitioning algorithm is as follows:

Listing 5.1: Partitioning Algorithm Output File

```
// cluster_dump_greedy_465MHz.csv
timescale , frequency , bandwidth , whitespace , frame_no
...
24 , 49256 , 534 , 12816 , 91
19 , 100438 , 2970 , 56430 , 91
39 , 103429 , 1090 , 42510 , 91
92 , 26697 , 9624 , 885408 , 92
34 , 13119 , 2150 , 73100 , 93
52 , 52825 , 432 , 22464 , 93
74 , 77227 , 5992 , 443408 , 94
38 , 85075 , 111 , 4218 , 96
...
```

There are some drawbacks however with this algorithm. The foremost issue is that weights of each cluster are not optimal. One dimension (bandwidth) is selected before the

second dimension (time), which effectively limits the degrees of freedom of the packing algorithm, resulting in ‘long’ clusters in time, rather than ‘wide’ clusters in frequency. Figure 5.19 represents a re-generated psuedo-waterfall plot, which is generated by taking a partitioned window file and displaying which whitespace opportunities were captured, and the remaining incumbent events that were not captured.

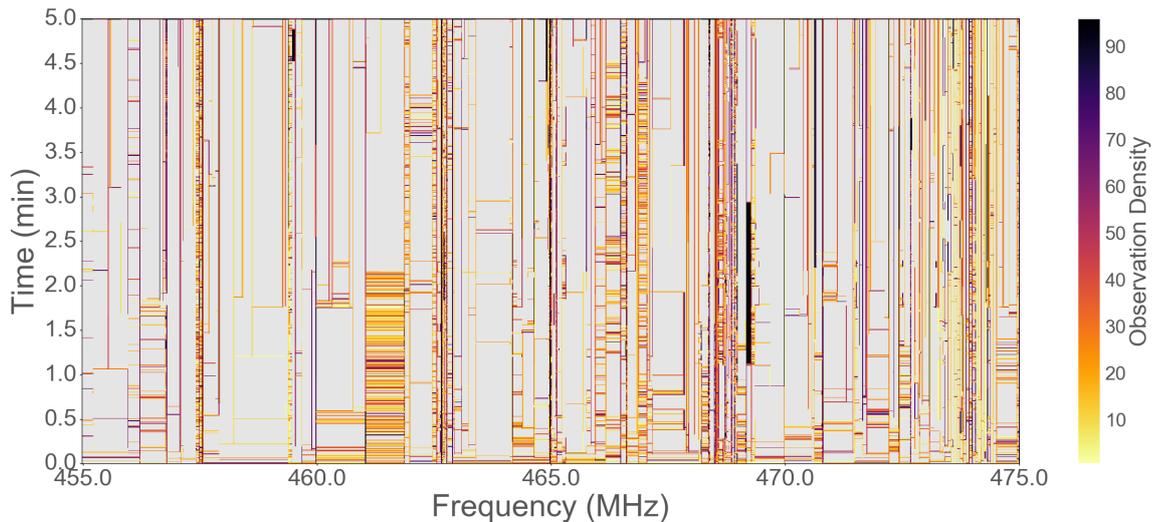


Figure 5.19: A detailed look at the packing algorithm performance, Figure 5.22 represents the underlying spectrum for which this packing was performed on. Grey regions indicate 100% whitespace, coloured regions indicate WSRBs that are either not packed due to fragmentation or incumbent events.

Figure 5.19 shows how the packing algorithm typically generates long clusters due to how the algorithm is implemented. It is, however, preferable that the clusters do remain long, as: once a bandwidth is selected then that becomes the known frequency range for an opportunity, the longer that that opportunity now persists for, the higher value and more practical the given cluster will be. Thus it is functionally more desirable to have thin (in frequency) and long (in time) clusters than wide (in frequency) and short (in time) clusters. It was also found that a minimum duration and bandwidth

needed to be selected such that the smallest whitespace cluster had a viable functional value. This minimum value was selected at 12.5kHz bandwidth and 50ms timescale, the implications and performance of which can be found in Section 5.7. If a single WSRB was the minimum size viable cluster, significant granulation of the spectrum would have been packed, resulting in a substantially lower overall functional value being achieved.

The work in [135] established that the MWIS with holes problem is indeed NP-Hard and can be solved in polynomial time, however no bounds for the problem are presented in the text. In the proposed procedural solution, as one dimension takes precedence over the other, it is clear that there will be an inefficiency with respect to achieving a MWIS. However, by converging to one dimension first across all elements in the set in a somewhat sensible manner, this improves the computability of the problem considerably, resulting in a reasonable and useful solution to the MWIS with holes problem in linear time.

It is important to note that certain vertical bins within Figure 5.19 do not perfectly line up with the corresponding vertical bin in Figure 5.22. This is primarily due to how the binning is displayed in Figure 5.22 as there are 131072 horizontal bins compressed into an approximate 1000 pixel X axis, causing a substantial loss in visual fidelity. In the re-generated psuedo-waterfall plot Figure 5.19, particular nuances within the packing algorithm and how contiguous regions of whitespace are packed combined with the computationally intensive mechanism required to generate from the windowed dataset and requires further substantial effort to resolve completely.

5.7 Whitespace Opportunity Distributions

The results from the partitioning algorithm offers considerable additional insight toward the whitespace structure of a candidate band when compared to conventional spectrum occupancy and duty cycle representations. This method is also completely scalable for observation duration (hours, weeks, months), data permitting.

The partitioned whitespace clusters are presented on a 2D log hexbin plot, with log density of clusters at given bandwidth and duration intervals. Hexagonal bins (hexbins) were chosen as the preferred binning method for displaying whitespace clusters. In whitespace structure investigation, the opportunities need to be plotted logarithmically due to the large range between several millisecond opportunities and several hour or longer opportunities. Additionally, the number of observations is very dense so each discrete opportunity dimension cannot be represented without an infinite level of resolution.

When binning highly dense datasets for presentation, it is essential that the distance between the extremes of the ranges covered by each bin is minimised. The most appropriate shape to minimise the distance between all points within a region is a circle, however circles of the same size cannot tessellate, wasting space and adding ambiguity when the data is presented. A hex more closely approximates a circle with the added benefit of tessellation, making hex shaped bins a superior decision when compared to square (or discrete pixel) bins for the representation of whitespace clusters graphically.

Each hex presents the frequency of observation and spanned time and frequency ranges of all discretely partitioned clusters. Histograms are present on the X and Y axis to add visibility to the distribution of opportunities in duration and bandwidth. In this work we refer to the displaying of whitespace structure under this novel scheme as a ‘Whitespace Opportunity Distribution’ (WOD) presented in Figure 5.21 and 5.23.

5.7.1 WOD Evaluation on 465MHz and 880MHz Whitespace Structure

Figure 5.20, 5.21 and 5.22, 5.23 are generated from two separate 5 minute observation periods. Figure 5.20 and 5.22 present the spectrum occupancy plots and Figure 5.21 and 5.23 are the results of the procedural partitioning strategy, for the 3G/4G and 12.5/25kHz datasets respectively. Figure 5.21 and 5.23 compare the whitespace opportunities present in the aforementioned 5 minute observations. The partitioning threshold is set to $B = 12.5\text{kHz}$ and $D = 50\text{ms}$ to reduce fragmentation of the spectrum, and to provide a minimum bound for ‘useful’ opportunities. Each plot comprises of approximately 30GB of spectrum samples.

The 3G/4G dataset first presented in Section 5.3, Figure 5.2, shows well defined guard bands, thus it should be observed that partitions directly relating to the guard band opportunities are present. In 5.21 there is indeed a clear clustering of opportunities around 500kHz in bandwidth, as expected. This clustering maps directly to the guard band present between the upper and lower band, with durations that span the entire observation period, confirming that the algorithm performs as expected in the extreme.

The 465MHz band in Section 5.3, Figure 5.1, shows a significant amount of incumbent activity, highlighted by the ‘max’ value series. It is important to note that within that same bin range, a value in the range of the noise floor is also captured. Thus it can be concluded that all the activity happening within that band is primarily at or below the bandwidth spanned by each bin (approximately 12.5kHz), as expected.

In Figure 5.22 the observed density of the incumbent activity is again due to the bandwidth spanned by each bin and the nominal bandwidth allocated to each spectrum licensee in that band. This provides an interesting challenge to assess the packing performance of the algorithm within a spurious, but consistently occupied band, which can be

seen in Figure 5.19.

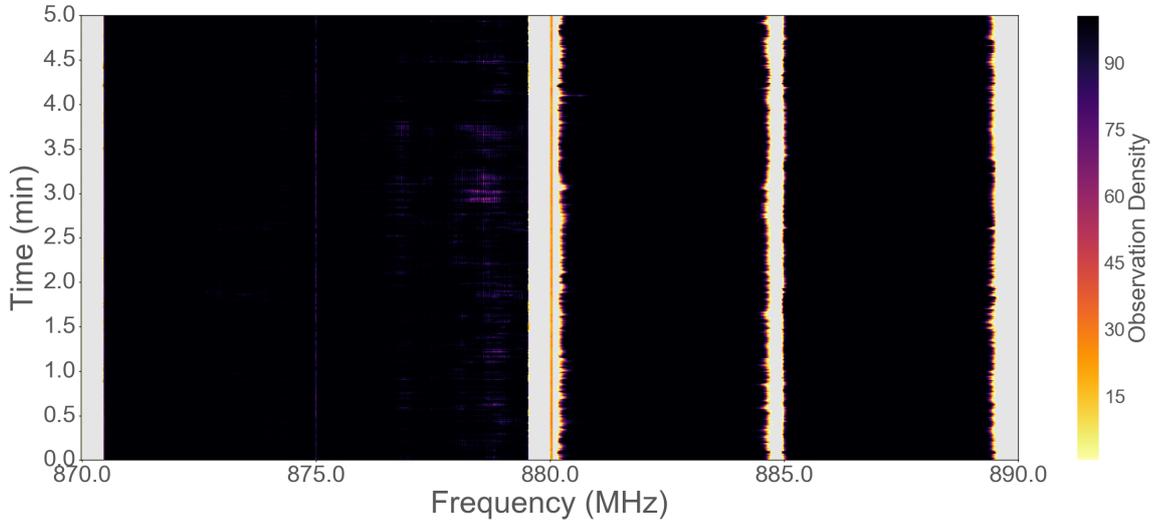


Figure 5.20: Spectrum Occupancy for 3G/4G Mobile Band, 880MHz centre frequency, 20MHz bandwidth (effective), 5 minute observation period

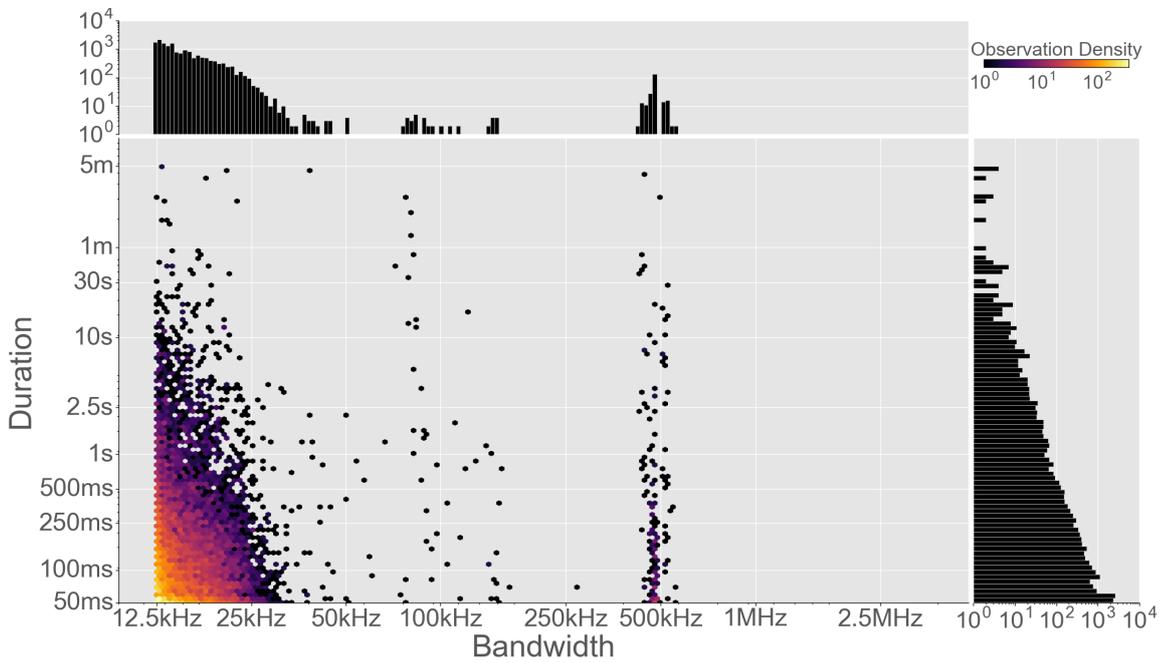


Figure 5.21: Whitespace Opportunity Distribution for 3G/4G Mobile Band, 880MHz centre frequency, 20MHz bandwidth (effective), 5 minute observation period

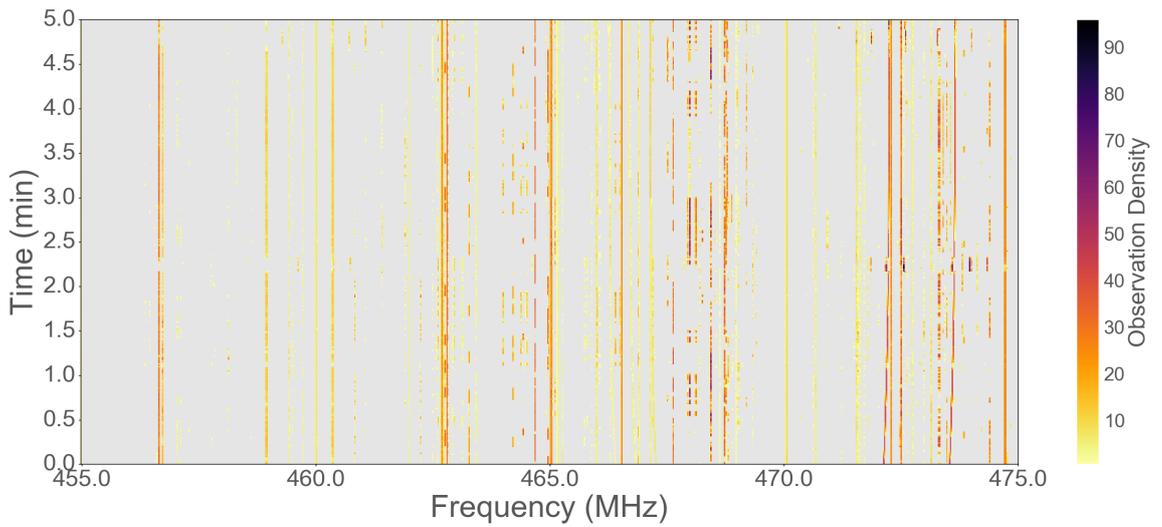


Figure 5.22: Spectrum Occupancy for Land Mobile/Fixed 12.5/25kHz Band, 465MHz centre frequency, 20MHz bandwidth (effective), 5 minute observation period

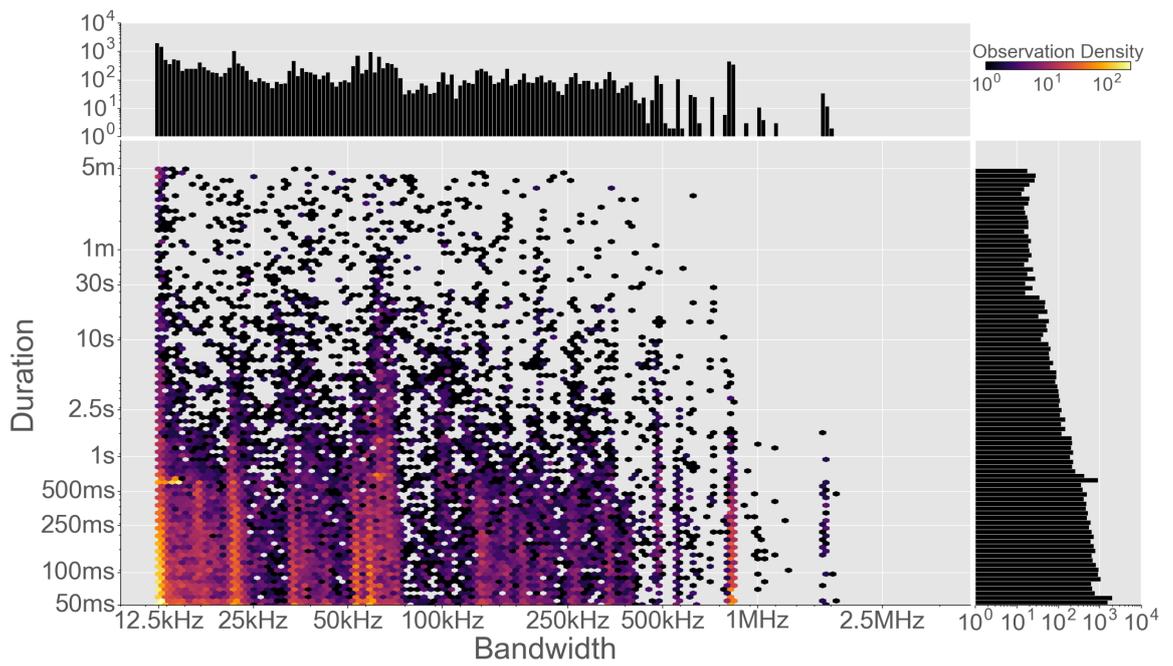


Figure 5.23: Whitespace Opportunity Distribution for Land Mobile/Fixed 12.5/25kHz Band, 465MHz centre frequency, 20MHz bandwidth (effective), 5 minute observation period

In Figure 5.21 and 5.23, it can be seen that there is a high concentration of smaller duration and bandwidth opportunities in general for reasons in Section 5.6.3. Notable banding at particular bandwidth intervals is also observed as expected. This banding of opportunities correlates closely to the gaps between concurrently communicating devices and/or high traffic channels present within the observed spectrum in Figure 5.20 and 5.22 respectively.

The concentration of low bandwidth and low duration opportunities in Figure 5.21 are found due to the ‘feathered’ (or fragmented) edges of the upper band as highlighted by the low observation density of secondary spectrum in Figure 5.20.

The following figures are obtained by evaluating the discrete whitespace resources spanned by the packing algorithm and secondly, a count of all the observed discrete whitespace resources in the dataset is also counted. These values are then used to compute the percentage coverage of the partitioned set, and the total whitespace resources spanning the observed spectrum.

The percentage of available resources in Figure 5.23 with respect to the total spectrum resources is 99.41%. Only .59% of the observed band comprised incumbent activity, leaving 99.41% of the spectrum resources unutilised. The partitioned set covered 97.97% of those available resources. Therefore 1.44% of available resources were not spanned by the partitioned set.

Conversely, the available resources in Figure 5.21 is only 11.95% of the total spectrum resources, with the partitioned set covering 78.75% of those resources. This suggests that significant resources (21.25%) exist below the selected threshold.

It can be seen in Figure 5.23 that there are considerable whitespace opportunities. Bandwidths of: 40, 70, 100, 250 and 500kHz have prominent densities, however the durations are concentrated below 1 second, reducing the value of these opportunities. There are numerous opportunities at bandwidths greater than 250kHz that exist for almost the

entire observation period which would provide excellent candidates for low bandwidth secondary spectrum devices. However for a secondary spectrum device, designed to function efficiently with sub second opportunities, this scenario presents a significant concentration of spectrum resources to exploit.

By comparison, the limited opportunities in Figure 5.21 would not present a viable candidate for secondary spectrum usage, other than within the guard bands (which are designated to be devoid of transmitters). Observing only the partitions above 250kHz and 1 second in Figure 5.23, the total resources available above this threshold are 72.75%. Thus there is significant whitespace available at bandwidths and durations that can be deemed useful for realistic secondary spectrum devices, with only a quarter of the remaining resources below this threshold.

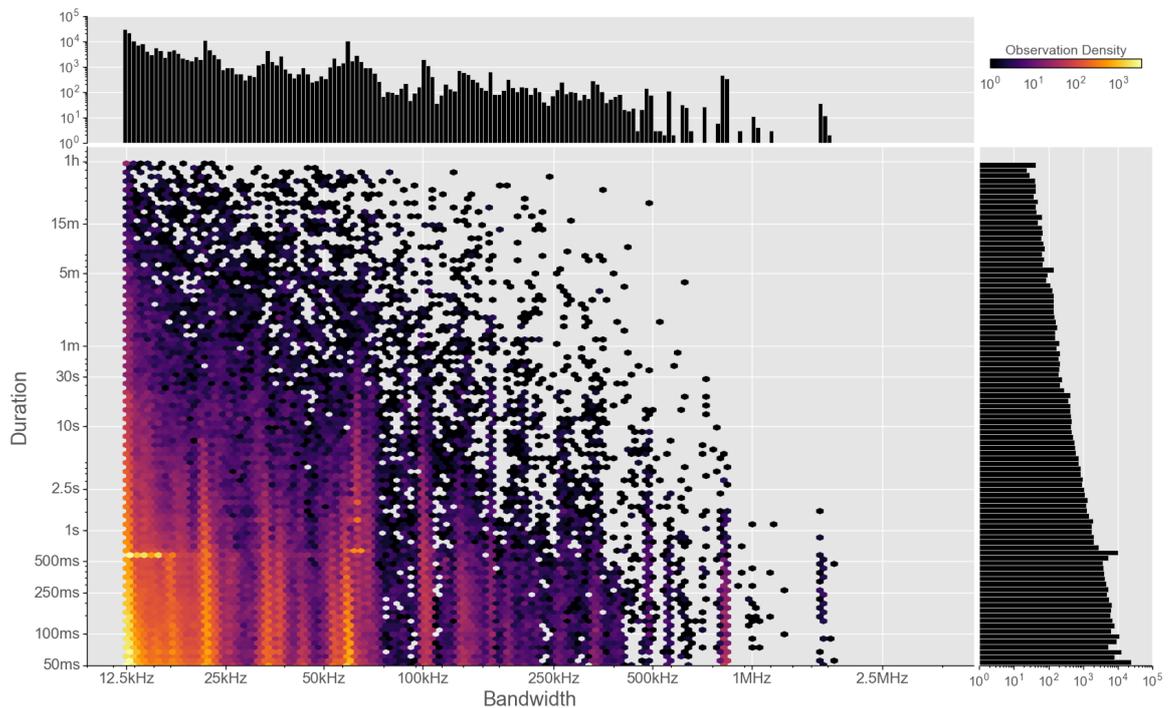


Figure 5.24: Whitespace Opportunity Distribution for Land Mobile/Fixed 12.5/25kHz Band, 465MHz centre frequency, 20MHz bandwidth (effective), 1 hour observation period

Figure 5.24 is generated from one hour of continuous observations from the same dataset as Figure 5.23. There is a substantial self-similarity with respect to the concentration of particular bandwidth and timescale whitespaces across both figures. This suggests that the spectrum environment over the one hour period possesses a relatively static or self-similar distribution of whitespace structure with respect to the five minute observation. Thus it can be concluded that there is some predictive power that could be exploited from the smaller duration observations with respect to the longer duration observation.

5.8 Summary

This Chapter has introduced and developed a method to quantify secondary spectrum resources, to enable unprecedented evaluation of secondary spectrum opportunities within a candidate band.

Several candidate bands were observed in Section 5.3, for which to perform secondary spectrum analysis upon. These observations were made possible from the development of the high resolution spectrum measurement system detailed in Chapter 4.

The notion of functional whitespace was introduced in Section 5.5, presenting a method to evaluate discrete secondary spectrum opportunities for further analysis. Section 5.6 expanded upon the functional whitespace concept, devising a method to cluster discrete secondary spectrum opportunities to maximise the functional value of observed clusters within a candidate band.

The partitioning algorithm in Section 5.6 enables a new way to evaluate secondary spectrum whitespaces. By taking a quantitative approach to evaluating secondary spectrum, the timescales and expected bandwidths for secondary spectrum communications can now be known, with the potential secondary spectrum resources that exist at those

timescales and bandwidths also made known.

The concept of a Whitespace Opportunity Distribution is finally presented in Section 5.7, as a novel method for performing secondary spectrum whitespace evaluation enabling unprecedented granular evaluation of the secondary spectrum resources contained within a candidate band. WODs are constructed via the application of a novel procedural spectrum partitioning algorithm, which clusters detected secondary spectrum whitespaces in a somewhat optimal way to quantify the intrinsic value of the whitespace contained within the aforementioned band.

Whitespace Opportunity Distributions are a key tool going forward for performing a deep analysis of the secondary spectrum resources contained within a band of interest, and at the same time, this wealth of information can be communicated and compared intuitively between other bands of interest. This presents itself as an exceptionally useful tool for the identification of future secondary spectrum candidate bands. The resulting analysis can also be used as an important tool for informing the expected timescale and bandwidth parameters for future secondary spectrum devices.

Chapter 6

Intel Project Patent Submissions

6.1 Introduction

This Chapter covers several successful patent submissions during the collaborative works with the Intel Corporation. The four patents listed here are in chronological order of filing. This ordering is different to the date at which the patents were awarded due to numerous factors within the patenting process.

Section 6.2 details a formal method for creating secondary spectrum license agreements within an LSA and SAS context. These agreements are designed to function with a novel agreement matching and negotiation engine termed the ‘Service Level Management Entity’, which is a broker that evaluates incumbent secondary spectrum access proposals and secondary user access requests (proposals). The SLME then creates a dynamic sharing agreement between two or more compatible proposals, formalising and enforcing a managed secondary spectrum sharing procedure.

Section 6.3 provides an enhancement to the existing LSA and SAS architectures by rigorously defining the regions of hardware ownership within each architecture. The regions of ownership and the appropriate interfaces required to support the correct function

of the LSA and SAS system, while providing incentives for third parties and MNOs alike to maintain their own LSA and SAS capable hardware, without the risk of exposing their internal network configuration is presented. This solution also provides a scalable architecture for the management and enforcement of sharing agreements, within the appropriate domains.

Section 6.4 develops a novel concept of capturing secondary spectrum opportunities before a formal control and connection set up sequence has to be performed, within a short range (such as Device to Device, Vehicle to Vehicle, Machine to Machine, or Internet of Things) context. A detailed mechanism for capturing secondary spectrum opportunities, while provide protections for device energy consumption and ensuring security of the payload transmitted via the secondary spectrum opportunity is provided.

Section 6.5 examines channel duplex communication modes (Time Division and Frequency Division) and assesses them for their performance within particular spectrum use-cases. Scenarios where a base station has the capabilities of using both TDD or FDD duplex modes are presented (via Carrier Aggregation or otherwise). The access to both duplex schemes is a feature of the LSA and SAS systems, that needs to be optimally managed. Solutions are provided to optimise which duplex mode should be used to optimally suit the challenges of the spectrum use-case with the objective of minimising the energy consumption of wireless devices attached to the base station and maximising overall channel capacity.

The patent filings as a collaborative work, are difficult to attribute specific concepts and ideas to individual authors, as often the final submission is a fusion of collaborative ideas. The ordering of the authors however, is indicative of the proportion of the contribution each individual was responsible for in each respective filing.

6.2 Methods and Devices for Shared Spectrum Allocation

This patent [12] is developed in the context of LSA and SAS (see Section 2.5.3). At the time of filing, the negotiation mechanism and specifics of a Sharing Agreement were not considered closely, and are assumed to be left to some off-line process not explicitly defined by the relevant sharing systems. The solution proposed by this invention, proposes a highly efficient framework for automating and formalising the negotiation process, with added enhancements to facilitate near real-time sharing agreement modification between the Primary User(s) (Incumbents), Secondary User(s) and spectrum regulator/authority.

The primary contributions of this solution are: an automated negotiation mechanism that can be initiated by Incumbents and/or licensee(s); specific definition for the terms of a sharing agreement; dynamic time and availability dependency of those sharing agreements; and a method to facilitate matching and evaluation of compatibility between negotiation terms.

6.2.1 Agreement Matching and Negotiation

This work proposes the addition of a Sharing Level Management Entity (SLME). The SLME is a separate entity within the overall LSA and SAS Controller(s) responsible for facilitating the negotiation and modification of sharing agreements. In Figs 6.1 and 6.2, the highlighted red components within the general LSA and SAS architectures identify the proposed modifications to facilitate spectrum license negotiations.

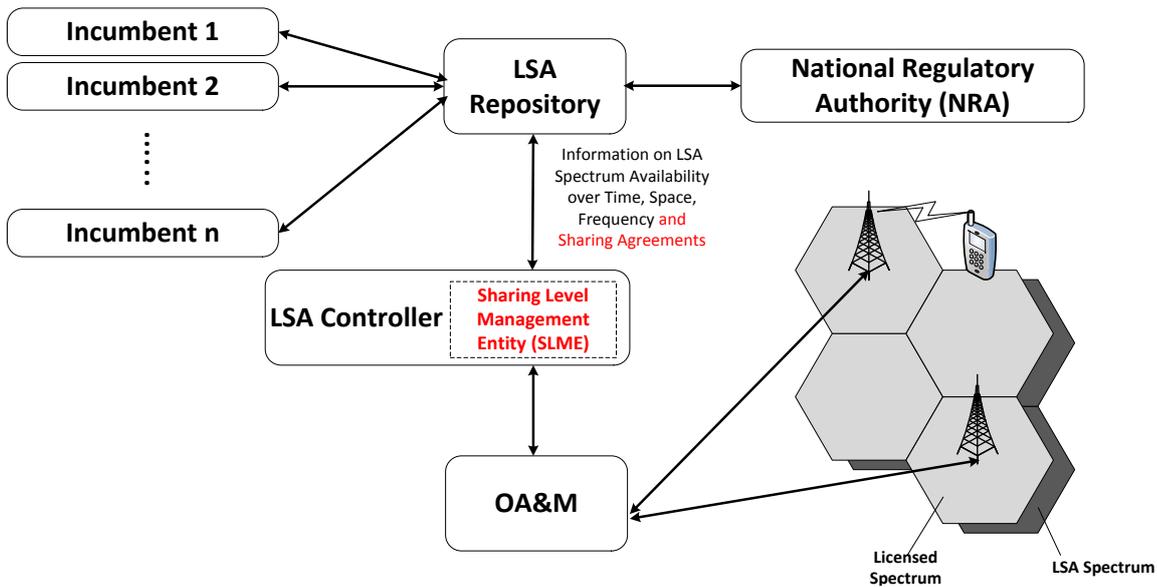


Figure 6.1: Modified Block Diagram of the LSA Architecture from Figure 2.13

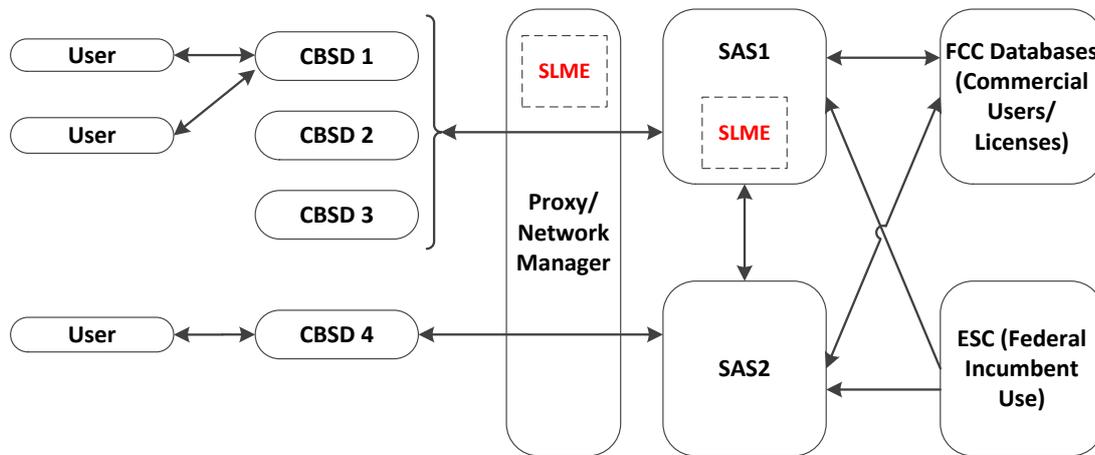


Figure 6.2: Modified Block Diagram of the SAS Architecture from Figure 2.14

Secondary Users and Incumbents are able to submit terms that they are willing to accept for a sharing agreement, the terms of such an agreement are outlined in Section 6.2.2. The SLME processes the proposed terms from one or more Secondary Users and assesses the compatibility of those terms with the terms proposed from one or more

Incumbents. These terms are compared together to assess the compatibility of Incumbent spectrum licenses, Secondary User access requests, and other spectrum access parameters (such as maximum output power, or geographical limits on device use) defined according to the LSA database or the particular responsible SAS entity. Once the terms are evaluated and matched, the specific terms of the license are sent to the licensee, and if the licensee accepts, the SLME will grant spectrum access to the Secondary User according to those terms.

6.2.2 Time and Availability Dependent Sharing Agreements

A key advantage that this solution provides is the capability to provide short time frame (from days to seconds) secondary spectrum sharing, under the existing LSA and SAS sharing frameworks. Sharing agreements in the proposed solution are to be submitted, evaluated and matched autonomously, eliminating long negotiation phases typically required under conventional spectrum sharing. This increase in agreement term evaluation and matching, provides the mechanism for an efficient ‘free market’ for spectrum sharing agreements, a desirable function for encouraging highly efficient secondary spectrum utilisation.

To enable a dynamic spectrum sharing scenario to be realisable, considerable effort must be spent to detail concretely the terms in which shared spectrum needs to be handled, as is evident in Section 2.5. A formal sharing agreement provides the necessary information to ensure that these details are captured and that undesired impacts due to interference and/or degraded quality of service for other spectrum users. To address this, a sharing agreement must comprise of the following:

- Geographic areas or regions that are valid under the agreement, including exclusion zones, protection zones (low transmission power allowed) and allowable interference levels to surrounding areas/regions.

- Overall maximum output power levels for communications, both in fixed and mobile cases.
- Time period(s) that the agreement is valid for, including statistical allowable interruptions where the Incumbent can retake their spectrum for brief (or otherwise) periods while the license is in effect. This also has the granularity of agreeing on different time period arrangements for various geographical locations previously outlined.
- Frequency range of operation, both contiguous and/or non-contiguous bands that can be accessed.
- Pricing information.
- Penalties for violation of agreement.
- Incumbent protection requirements, whether sensing is required before an access request can be granted, or if interrogating the geographical database is sufficient.
- Types of neighbouring services and their specific interference requirements or conditions.
- Particular services that are allowed to be deployed within the band (such as: fixed, mobile, aerospace, telemetry, etc).

6.2.3 Summary

The explicit nature of time dependent sharing agreements in combination with the overall LSA and SAS architectures ensures that the Incumbents interests will be protected. The SLME being integrated within the existing sharing systems, ensures that spectrum can

only used within the allowable parameters set out by the relevant spectrum regulation, spectrum databases and Incumbent requirements.

This work presents an improved solution and enhanced capability to render off-line binary sharing agreements that would have to have been agreed upon for secondary spectrum sharing to occur unnecessary and obsolete. By integrating a SLME, agreements no longer have to be explicitly binary, can involve multiple interested parties, enable real-time renegotiation, and enable the negotiation of sharing arrangements with parties that otherwise would not have met or been able to engage with each other.

6.3 Evolved Node-B, Local Controller and Method for Allocation of Spectrum for Secondary Usage

This patent [10] proposes a further development within the LSA and SAS architectures covered in Section 2.5.3, to provide a method for enabling the sharing of physical secondary spectrum radio infrastructure (Base Stations (BS) ¹) owned by either an incumbent or third party, between several interested incumbents and/or spectrum license holders. The patent also enables the control domains of the LSA and SAS system to be explicitly separated, such that only necessary information to facilitate effective secondary spectrum management is exchanged, without exposing the internal configuration (infrastructure, clients, device configurations, access strategies, access patterns etc) of an MNO or third party's network to the greater LSA/SAS system and vice versa.

The primary function of the LSA/SAS system is to facilitate the transfer of spectrum ownership for secondary spectrum access. There should be no restrictions placed by the LSA/SAS network in how either the Incumbent or SU should utilize the spectrum when ownership of the spectrum is transferred (unless otherwise defined under a sharing agreement). Thus the LSA/SAS system itself should be limited to informing the SU when the spectrum is available, specific details of the sharing agreement, and when the spectrum is retaken by the Incumbent.

The SU should be responsible for providing information of the state of the secondary spectrum to the LSA/SAS system in accordance with the particular licensing agreement to prove SU compliance and ensure Primary User protection, rather than the LSA/SAS system itself being required to constantly monitor and police the spectrum.

¹In this context a Base Station is used to refer to any kind of physical radio infrastructure, mobile or static, that broadcasts a wireless communications network, such as a Macro Cell, Micro Cell, Access Point etc within a secondary spectrum apportionment

It is foreseeable that competing service providers would not want to make each other aware of their particular secondary spectrum licensing and access strategies. Thus only necessary localised and obfuscated secondary spectrum access constraints need to be informed to the relevant infrastructure, while not exposing the greater spectrum licensing agreement (or agreements).

The solution proposed addresses this concern with a novel modification to the existing LSA and SAS architectures and details the mechanisms required to ensure anonymisation of secondary spectrum access by interested parties (with an appropriate license) to eliminate conflicts of interest and enable sharing of physical infrastructure.

6.3.1 Control Domains

The core concept introduces a distributed domain control approach, with respect to how spectrum sharing agreements are processed and enforced. This distributed approach introduces the ‘Regional Controller’ (RC) and ‘Local Controller’ (LC) system blocks to facilitate distributed control.

The novel system level blocks defined in this patent are the RC, LC the interfaces S4 and S5 and the explicit domain separation. The RC and LC concept is to be implemented as either an extension or complete replacement for the somewhat arbitrary ‘LSA Controller’ block within the reference architecture. While the SAS case is not explicitly defined here, the same concept can be logically extended by integrating the concept within the proxy/network manager and SAS entity blocks (See Figure 2.14, covered in the original patent document).

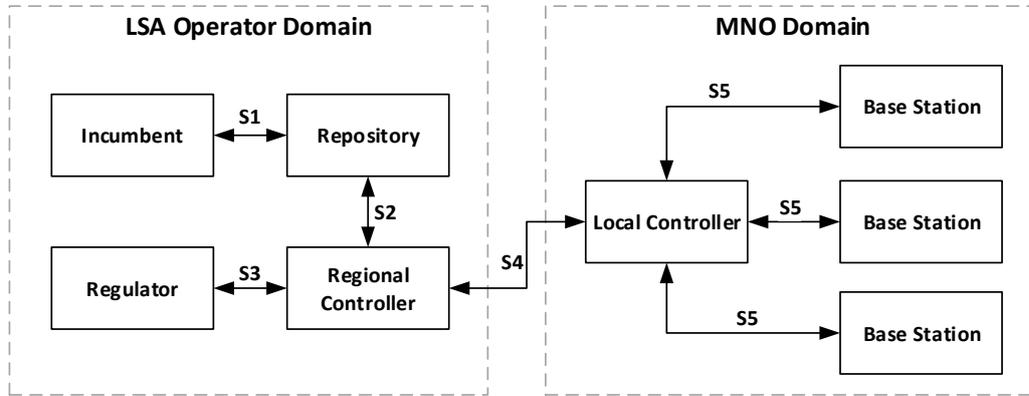


Figure 6.3: System Domain Diagram for the ‘Regional Controller’, ‘Local Controller’ and Interfaces

The core system components within the system domain diagram in Figure 6.3 and the interfaces S1, S2 and S3 are analogous to the interfaces between the appropriate system blocks in the reference architecture, Figure 2.13. The interfaces and system blocks in the LSA and SAS reference architectures are discussed in Section 2.5.3. Interfaces S4 and S5 are detailed in Section 6.3.2.

The Regional Controller (RC) is under the ownership of the LSA operator. The RC aggregates and maintains the licensing agreements that pertain to the geographical region for which the RC serves (termed the Regional Controller Service Area (RCSA)). The RC is responsible for converting licensing agreement(s) into ‘local access policies’ for individual Local Controllers (LC) located within the service area that the RC is responsible for.

Each LC is owned by either an MNO or third party, and is responsible for enforcing the particular local access policy as provided to it by the RC. A local access policy defines how secondary spectrum can be accessed by the the devices attached to the LC. These policies are dependent on the licensing agreements in effect for the geographical area spanned by the LC . A single LC may be responsible for one or more co-located BS’s/cells in a given

geographic area, termed a Local Controller Control Group (LCCG) (see Section 6.3.3).

The local access policy is effectively an obfuscated subset of spectrum sharing agreements, restricted to the discrete region spanned by the LC. These policies are filtered down from the total set of secondary spectrum access policies (regional access policies), that are maintained by the RC. The local access policies only provide instruction on how secondary spectrum is to be accessed and by whom it can be accessed. This simplification of spectrum management is important, as by only having local access policies visible in the MNO domain, LC's are effectively unaware of the particular sharing agreements that in effect but are still able to correctly manage the spectrum in accordance with the agreements.

To ensure compliance with the regional policy, spectrum access decisions (i.e. granting secondary spectrum access to a secondary user that requests it) made by a LC at various points in time are sent to the RC to validate the compatibility of that decision with the regional access policy.

6.3.2 Interfaces

RC to LC Interface (S4)

Policy updates from the RC to connected LCs is performed over this interface which also supports notification of incumbent retake events in accordance with the relevant sharing agreements. Incumbent retake events are high priority messages that are highly latency sensitive, whereas policy updates have a much lower latency requirement.

Spectrum access decisions will be reported from the LC to the RC with the time and frequency band(s) currently occupied by the LCCG. Depending on the particular licensing agreements, the RC will request LC REM data to enable the RC to construct a regional REM to assist with network planning and more efficient secondary spectrum allocation

via refining local access policies.

The only information made visible to the RC is REM information of the LCCG and when a secondary resource is engaged by that LCCG. The LC also has no visibility of anything behind the RC, such as other RC's, the Repository or other Incumbents connected to the LSA/SAS system.

This interface provides anonymisation of the LSA/SAS network from the MNO and anonymisation of the MNO from the LSA/SAS network. As such, the RC must trust the information being sent from the LC and the LC must trust the information sent from the RC without question. Every command from the RC to the LC should be immutable, thus forcing any BS under the LCCG to comply. This is necessary to ensure synchronization of policy updates and is critical for incumbent retake events. This established chain of trust, enables the anonymisation of MNO (or third party) network configurations without negatively impacting on the services able to be provided via the LSA/SAS system.

LC to BS Interface (S5)

Connections between the LC and appropriate BS infrastructure is handled via the S5 interface. As the LC has knowledge of who is connected to it and where they are connected within the LCCG, spectrum allocations can be awarded to the appropriate holder of a license agreement under the geographical region of the relative infrastructure.

Spectrum sensing information will be transmitted over this link from the BS by interrogating appropriate connected UEs when required. Only the received signal strength, and geographical location data points tied to the sensing information will be transmitted to the LC to enable mapping of the radio environment.

This offers an additional level of granularity for spectrum sharing, where spectrum resources can be allocated on a per cell basis, rather than a greater area where everything under that area is affected. Knowledge of cell overlap is also provided to the LC over

this interface via spectrum sensing measurements and measurements performed when commissioning the LC.

BS's will not be made aware of the local access policy or the spectrum environment, they are only signaled when spectrum of a particular band is available or unavailable by the LC (and specific non identifiable details of the access policy, such as allowable transmit power, mobility, duration of availability, etc). It is up to the BSs discretion for how the spectrum awarded should be utilised and scheduled among the devices connected to it (within the confines of the local access policy). BSs may be asked for sensing information when required per the local access policy of the LC before spectrum resources are made available.

The BS must also inform the LC when it has engaged an available spectrum resource. An engaged resource does not mean that the BS is actually accessing or utilizing that spectrum, it means that the current lease of that resource has been transferred to the secondary user. Engagements do not have to be made immediately upon notification of available spectrum resources, the opportunity to engage the spectrum is available until the secondary resource is retaken by the incumbent. Only the BS knows what it is doing with the spectrum when the lease of spectrum has been transferred to it. Thus, random spectrum access strategies may be performed by secondary users to further obfuscate their access patterns if required.

6.3.3 Control Groups

A LC that is connected to multiple secondary spectrum BS entities, forms a Local Controller Control Group (LCCG) as seen in Figure 6.4. This group is a cluster of BS's that have overlapping cells and/or exist within a particular geographical region where coordination between those cells is desirable and difficult to ensure a high QoS by using the slower RC link for interference and mobility coordination between the secondary

spectrum BS entities. All BSs under the LCCG can either be entities that enable access secondary spectrum resources or are entities without secondary spectrum access capabilities. Both of these entity types can be requested by the LC to provide additional real-time radio environment information for the LCCG. This is achieved by each BS in the LCCG requesting connected secondary spectrum capable devices to perform sensing of the secondary spectrum environment and then report the measurements.

As the RC has visibility over a number of LCs and the interactions between neighbouring LCs, the RC can be utilised for coordination at the LC edge for interference management and cross LC coordination/mobility. This is necessary as the LCs are expected to be operating in isolation and only act upon the specific policy updates provided by the RC and relevant information provided by their respective control group.

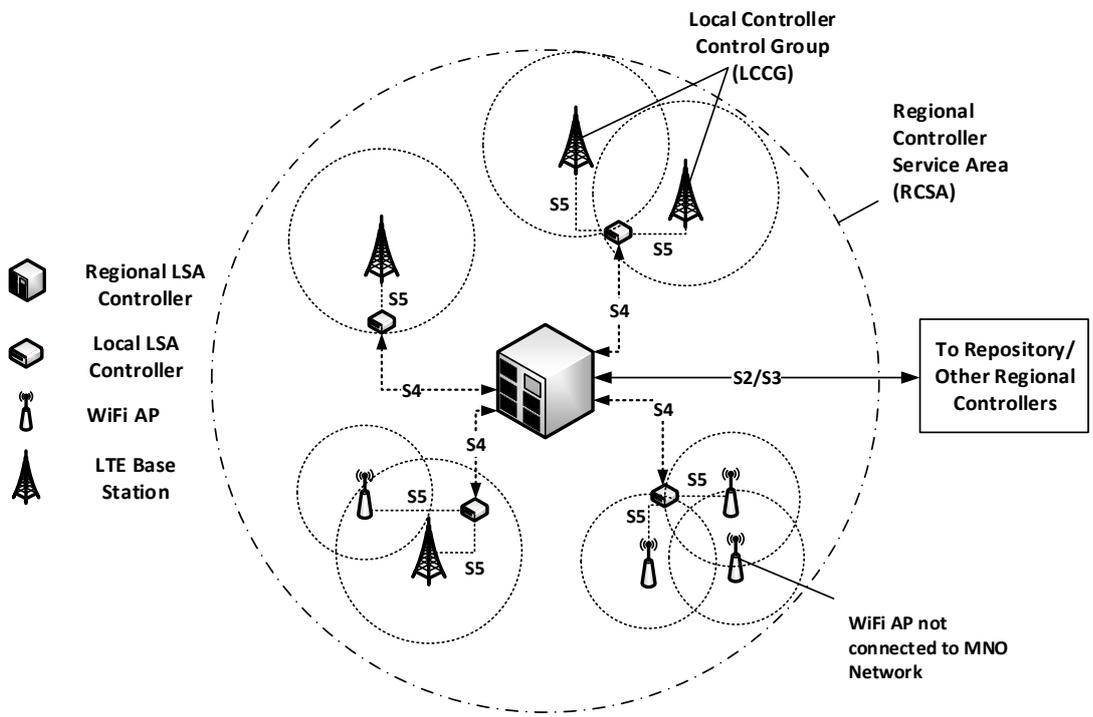


Figure 6.4: System Diagram Detailing Control Groups and Regional Service Areas

6.3.4 Summary

By separating the LSA/SAS Controller into Regional and Local Controllers operating within defined domains, secondary spectrum can be better managed while preserving a level of anonymity and coordination between MNO networks and the LSA/SAS operator. This approach additionally opens up the possibility for third party infrastructure operators to provide anonymised secondary spectrum access in accordance with any secondary spectrum license, without being required to know the explicit details of the license itself.

6.4 Signal Buffering for Licensed Shared Access (LSA) technology

Stemming from the investigation of control plane latency and its impacts upon the availability of secondary spectrum in Section 3.3.1. A method to reduce the time cost of accessing secondary spectrum, in both managed and unmanaged scenarios is developed. This method is additionally developed with the transceiver circuitry for a portable device required to achieve the desired function forming part of this patent.

Looking back at Figure 3.2, performing low level processing at the front end module and analog front end, provides orders of magnitude improvement in latency and power consumption. If a decision to ignore a given signal or signals, can be made at the analog level and not higher up in the digital front end, baseband and MAC layers, then there are substantial power consumption savings that can be captured by avoiding the processing of signals that were not intended for that device in the first place.

The simple whitespace model presented in Section 3.4.1, shows that to access an arbitrary secondary spectrum whitespace opportunity (in lieu of a dedicated control channel), a control plane phase must take place before the data payload can be transmitted. A scenario where a dedicated control channel typically does not exist, is within a Device to Device (D2D), Vehicle to Vehicle (V2V), or Internet of Things (IoT) context, where two (or more) devices are attempting to send messages between each other without the aid of a third party coordinator.

The control plane phase, is an unavoidable function required for the correct function of a wireless data communication. As it is unavoidable it can be thought of as an unavoidable latency or ‘access latency penalty’ that is incurred every time a device wants to transmit some payload from an unconnected state.

As there is no guarantee that a whitespace opportunity will exist in the future, it makes

sense to make the most of any opportunities that are currently present. This system is designed with that intent, to maximise current secondary spectrum opportunities, without having to incur latency penalties from ‘unavoidable’ set up times due to control signaling.

6.4.1 B2TF Protocol and System Architecture

The B2TF protocol and system architecture detailed in the patent [9], aims to capture geographically local secondary spectrum opportunities and use that as a medium to circumvent the access latency penalty incurred by the requirement of control plane signaling before user plane signaling under conventional signaling schemes. A method to send a payload of interest simultaneously (and/or before) the required control plane signaling to negotiate a connection is devised for use within a secondary spectrum context.

As an example: in a Device to Device (D2D) scenario, a transmitter may want to send a payload to a nearby receiver. This payload may be quite large, and by transmitting it over a primary spectrum network (LTE for example) there may be data charges for both parties, and it would cause a substantial load to the network. As the devices are in close proximity it would be preferable to transfer the payload directly between the devices over a local network (Wifi, Bluetooth, ZigBee) or available secondary spectrum.

In the secondary spectrum case, the transmitter may have detected a significant amount of secondary spectrum that is available for it to broadcast the payload over directly, rather than using a primary spectrum network (such as LTE). The transmitter first transmits a (previously agreed upon, in this case protocol defined) activation sequence, to inform the receiver that it is about to transmit a payload. The payload follows shortly after that. To avoid nearby devices with similar capabilities from snooping on the payload, the transmitter encrypts the payload, and then transmits the key over a primary spectrum network, informing the network of the intended receiver. The receiver that has now already captured the payload (or is currently in the process of receiving it) receives

the key and proceeds to decode the payload. This way the traffic over the primary network can be minimised, and short timescale secondary spectrum opportunities become viable to perform (effectively) latency free communications in a D2D context. Figure 6.5 highlights this general premise of the system, including an example activation sequence followed by the payload, with the control signaling and recipient identification handled by a primary spectrum network:

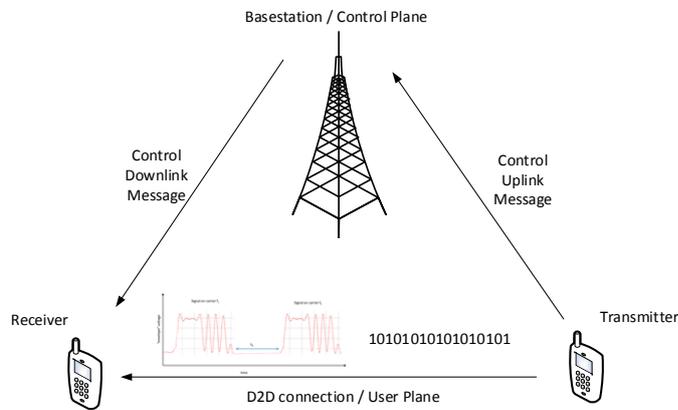


Figure 6.5: Diagram of the B2TF User and Control Plane Signaling

The B2TF (Back To The Future) protocol enables the previously outlined scenario to the extreme with the interest of exploiting short timescale secondary spectrum opportunities. If control signaling would first have had to have been performed before the opportunities could be accessed, would result in those opportunities being missed. Figure 6.6 outlines the system architecture that facilitates the aforementioned D2D secondary spectrum access scenario, with the objective of minimising the additional energy consumption on the device (due to this capability and with no impact on conventional operation and device latency).

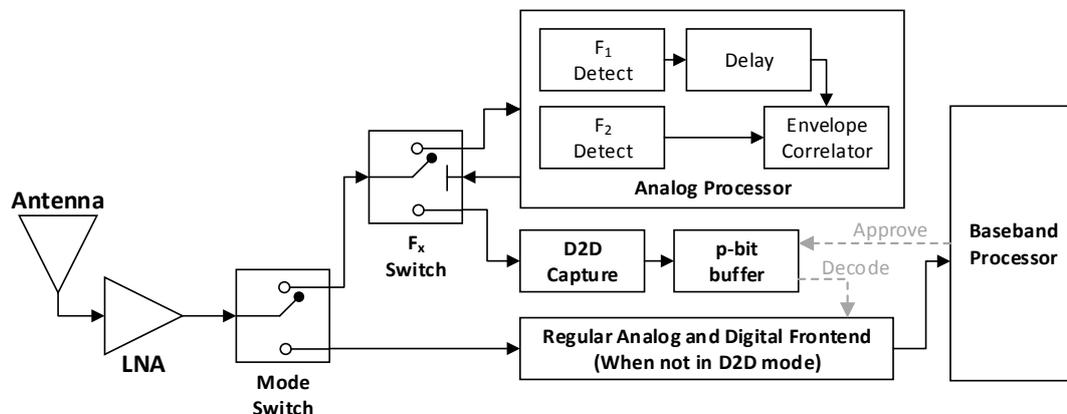


Figure 6.6: Diagram of the B2TF Protocol Analog Processing and Decode Stages

F_1 and F_2 are separate frequencies where an activation sequence is transmitted by the transmitting D2D device (device T). This sequence is broadcast on F_1 and then after some delay, is broadcast again on F_2 . Once the analog processor on the receiver D2D device (device R) detects these two sequences, via the delay offset and envelope correlator, the F_x switch is toggled as the transmitter has signaled that it had an incoming payload that needs to be captured. Device T then transmits a short encrypted payload right after the activation sequence, with the intent that a second D2D device (or any device within the vicinity of the transmission) has already detected the activation sequence and is waiting for the payload. Device R captures this payload via the D2D capture stage, which is achieved by directly sampling the spectrum for a short period. The directly sampled spectrum is then buffered for possible later retrieval. Once the payload is buffered, the device R requires a control signal with a decode key to be broadcast over a primary network (LTE, WiFi or otherwise), to decrypt the payload. Device T will simultaneously, perform control plane operation with a primary network, and notify the network that the B2TF message sent is intended for a particular recipient and the decryption key for the payload is transmitted over the primary network to the recipient. At device R, if after a

fixed time has elapsed and a control signal is not received, the p-bit buffer will be cleared and the F_x switch is reset. If the control signal and decode key is received, then the payload will be released from the p-bit buffer to the front end to complete the processing of the payload and retrieve the data contained within it.

Effectively, this solution reduces the previously aforementioned access latency penalty to 0 (or negative), which is highly desirable in a dynamic spectrum environment where secondary spectrum opportunities may only exist briefly. The solution also has limited impact on additional energy consumption, and requires minimal additional electronics to perform its function.

6.4.2 Summary

A given D2D device employing this system will have access to short time scale secondary spectrum opportunities, that would previously have been unable to be captured by conventional means, and have the added advantage of receiving user plane communications before a control channel needs to be negotiated. By only releasing captured D2D communications that are intended for a given device to the front end signal processing and decoding electronics, the proposed solution performs its operation in an energy efficient manner by eliminating the possibility of decoding effectively useless or unintended data. This system could have significant application in future D2D, M2M or IoT technologies operating in heavily constrained wireless environments, where secondary spectrum opportunities are highly dynamic and every potential bit that can be packed into a given spectrum is required.

6.5 Methods for Performing Mobile Communications Between Mobile Terminal Devices, Base Stations and Network Control Devices

This patent [11] pertains to the idea of balancing user data among FDD and TDD carriers depending on the user context. Conventional communication networks typically utilise wither a TDD or FDD duplex scheme to manage the uplink and downlink communication channels between a BS and UE. Section 2.4.3 covers the background material on different channel duplexing and channel aggregation strategies commonly employed in existing high throughput mobile systems.

The strengths of TDD are due to its simpler beam forming, potentially lower hardware power consumption and the ability to asymmetrically allocate capacity to the uplink and downlink. The strengths of FDD are due to reduced latency (when compared to a TDD system), superior performance in high mobility environments and longer transmission ranges. Thus as FDD and TDD both have different properties, it is desirable to optimise between these capabilities in the granular use case of a user to obtain the best possible system latency and power consumption performance.

In State-of-the-Art solutions, there is no optimised mapping of user data onto FDD and TDD carriers, particularly in the context of a FDD/TDD carrier aggregation. Additionally, dynamic changing of the duplex mode between FDD and TDD within a cell or spatial stream also does not exist. Typically, spectrum and time resources are just filled up as the data comes in and as the resources are available and carriers are locked to either a time or frequency duplex mode. This may typically lead to a sub-optimum usage of the available resources. In this approach, the new degree of freedom of being able to map user data either on the FDD or TDD carrier is optimally exploited and thus the overall

system efficiency is improved.

This patent is also of particular interest with respect to the LSA architecture. The LSA system operates at 2.3 to 2.4GHz in Europe, which is an LTE TDD band. The typical deployment of LTE services in Europe are FDD. Similarly, in the US, the SAS system operated at 3.6GHz which is also a TDD band, where the typical LTE deployments use FDD duplexing. Thus it is essential for the success of LSA and SAS that carrier aggregation be enabled across different duplex schemes and at the same time, to maximise the utility provided by the different duplex schemes.

6.5.1 Duplex Balancing Schemes

The following Figures illustrate several balancing schemes that can be exploited to optimise between TDD and FDD operation within a given cell.

In Figure 6.7, a user has a carrier aggregation capable handset and is initially allocated into a TDD channel. After some time, the requirements of the user has changed, either due to payload asymmetry, channel performance, energy consumption requirements etc, the user is allocated to the FDD band, as it presents a more suitable duplex scheme for the current scenario. This is repeated over time, as the requirements of the user varies.

In this particular case, the user is not simultaneously allocated to both the FDD and TDD channels, and is switched between them. This could either be due to the capabilities (software, hardware or a combination) of the users handset, or base station load balancing. In essence, from the perspective of the user, the duplex mode has simply toggled.

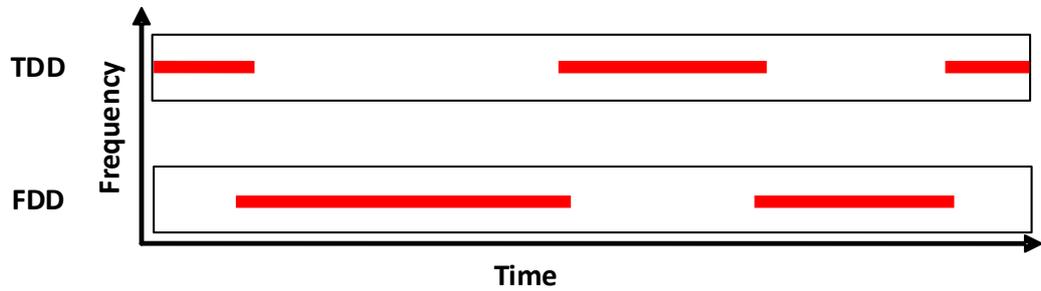


Figure 6.7: Carrier Aggregation with Switched Balancing

Figure 6.8, portrays another carrier aggregation scenario, where a handset is able to access both the TDD and FDD CA channels simultaneously, as determined by the base station scheduling and load balancing.

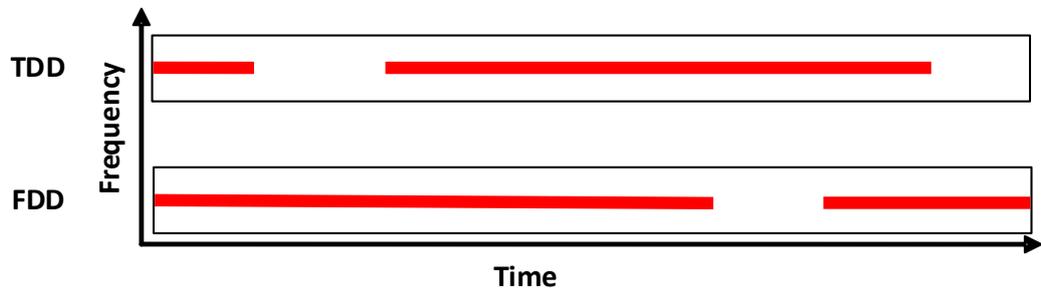


Figure 6.8: Carrier Aggregation with Simultaneous Balancing

Figure 6.9 shows a scenario where there is no carrier aggregation present, however the base station is able to dynamically optimise the duplex mode, as a function of all connected users to that particular base station (or cell or sector). The base station will aggregate the performance metrics of all the connected devices, and will make decisions to change the duplex mode based on the best overall outcome for the connected users.



Figure 6.9: Single Channel Duplex Optimisation

Figure 6.10 presents a scenario where the single channel duplex optimisation from Figure 6.9 is extrapolated to the spatial domain via a MIMO channel. In this case a user is not switched in frequency, the cell/spatial stream changes the local duplex mode for that cell/spatial stream to better suit the users of that resource. The balancing can be performed for all spatial streams on all appropriate frequency channels used by the BS and connected UEs.

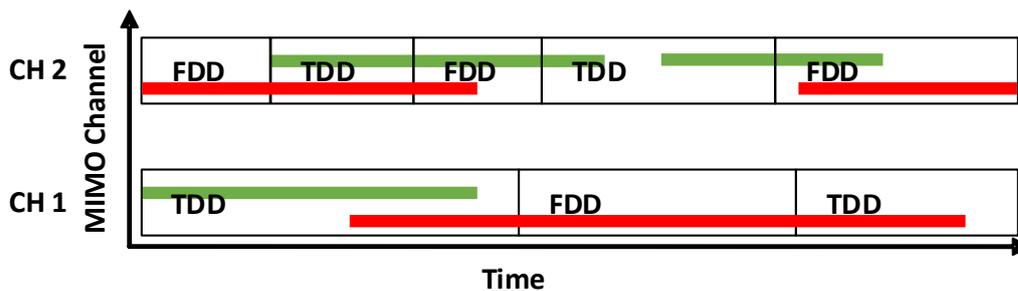


Figure 6.10: Spatial Optimisation

It is foreseeable that this can be extended to cells with sectorized antennas or even phased array antennas. For the sectorized case, each sector can be set to either TDD or FDD as per Figures 6.8 and 6.9. Spatial balancing can be further exploited if each each

sector has independent MIMO antennas. In the phased array case, it is foreseeable that basestations will implement massive MIMO antenna schemes where it will be possible for a different FDD or TDD stream to be the duplex mode for each Sub-Cell (SC) within each basestations cell/sector highlighted in Figure 6.11.

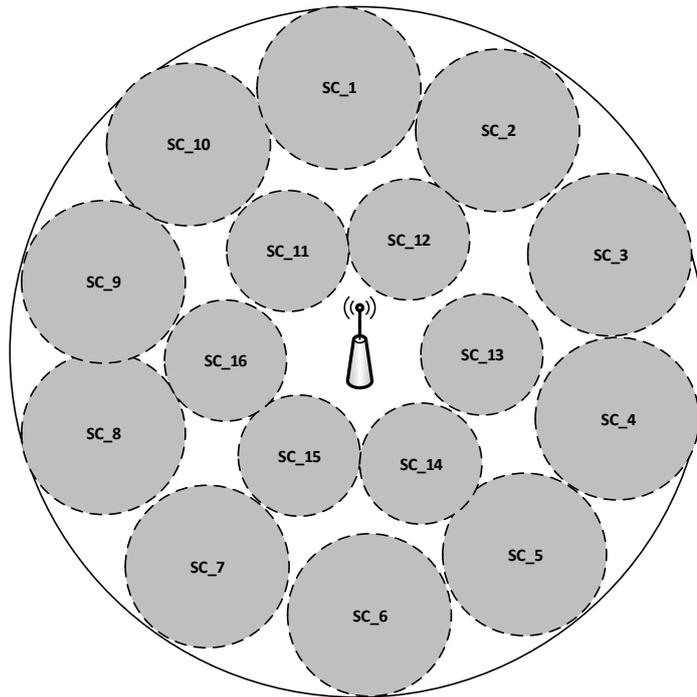


Figure 6.11: Phased Array SCs within Base Station Cell

The concept of a SC is to further increase the resolution of how a basestation can manage spatial resources at a much higher granularity than simply enforcing a single duplex mode (or duplex scheme) for the entire cell. SCs can scale to the point that an individual balanced/aggregated channel(s) can be provided for every device connected to the basestation.

Note that Figure 6.11 is for illustrative purposes and the size and position of SCs can be arbitrary.

6.5.2 Applications for Dynamic Carriers

To provide a brief set of examples as to why TDD is preferred over FDD (and vice versa), various communications scenarios are presented with a recommendation for the duplex scheme that best addresses the scenario:

- A favourable communication context. i.e. target devices are in close geographic proximity to the base station and channel propagation conditions are good with minimal fading:
 - Only a TDD channel for data transmission is used. There is no energy allocated to the FDD channel.
 - TDD can also be used for its capacity to dynamically reallocate uplink and downlink resources for asymmetric traffic contexts.
- A slightly challenging communication context. i.e. some of the target devices are not in geographic proximity, propagation conditions are reasonably good, with moderate fading in mobility contexts:
 - Only a TDD channel is used for conveying data or alternatively, a major portion of the data traffic allocated to the TDD channel (for example, 90%) while a smaller portion of the data traffic allocated to the FDD channel (for example, 10%).
 - The splits could also be modified according to a throughput constraint on a given channel, with the second channel used for spectrum ‘offloading’. In this case, some users will have to be migrated, thus it would be ideal to select the users to migrate that would benefit from the migration
- A challenging communication context. i.e. some of the target devices are not in geographic proximity, the propagation conditions are challenging and some devices are in aggressive fading channels:

- A split of data traffic between the TDD and FDD channels is preferred. Approximately the same percentage of payload traffic is allocated to each channel (50 - 50 split). Users can be randomly or optimally placed across one or both of the channels.
- Particularly in a scenario where the TDD and FDD channels are significantly different in carrier frequency, the lower frequency carrier would be preferred in a more ‘hostile’ environment.
- A highly challenging communication context. i.e. target devices are not in geographic proximity, the propagation conditions are highly challenging, numerous devices are in shielding or shadowing situations:
 - Only a FDD channel is used for the data traffic. There is no energy or data allocated to the TDD channel, due to its poor performance in this context.

6.5.3 Summary

At a high level, Carrier Aggregation between TDD and FDD bands is a necessity for capturing the best performance from future LSA and SAS systems and mixed duplex Carrier Aggregation systems as a whole.

In future it is predicted that both duplex modes will be made more available to users, thus it is important to consider under which scenarios should one duplex mode be preferred over the other, as there are indeed scenarios that benefit one duplex method over the other. With both duplex modes at the disposal of a MNO, it is in the best interests of the operator from a power consumption, latency and spectrum efficiency perspective to optimally allocate connected devices across their network with the most suitable duplex mode.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

Prior to this work, secondary spectrum is often presented simply by the presence (or abundance) of all secondary spectrum observed with respect to the incumbent activity. Classical abundance measures such as duty cycle or spectrum occupancy can be used in some capacity to indicate the mere presence of secondary spectrum, but provide no insight to the structure of secondary spectrum resources observed. This lack of insight is even more pronounced as timescales approach zero.

For a physical secondary spectrum capable device attempting to harness secondary spectrum, it is essential to have an understanding and hence an expectation of the underlying secondary spectrum resources structure. By knowing the structure of the resources within a candidate band, the viability of that band for supporting secondary spectrum communications can be determined with greater insight. Additionally, it can be determined whether a secondary spectrum device itself is capable of harnessing the particular spectrum resources in question.

As an analogy: geological surveys are always conducted prior to the construction of

a mine, in an effort to determine the structure and abundance of a mineral resource to optimise its mode of extraction. Similarly, a survey of secondary spectrum structure too needs to be performed prior to optimally capturing the secondary spectrum resources. However, unlike a geological study that likely has not and will not change for millenia, a geological survey is largely immutable, and completely determinable. By comparison, a survey of secondary spectrum resources presents an NP-Hard problem that can also vary considerably over time, requiring constant re-evaluation and can never be completely known. While the discovery of resources is the common objective of both geological and spectrum surveys, the results of a spectrum survey is completely unlike anything presented in the geological analogy.

Chapter 3 investigated the role of timescale and the impact of timescale on secondary spectrum observation and exploitation. The timescale at which a device operates significantly impacts the devices' energy consumption and latency metrics, and by extrapolation the devices' ability to efficiently operate within a secondary spectrum context. The delay of a systems response to a change in the environment or the lag of the system is effectively the minimum timescale at which the system operates.

Chapter 4 detailed a low cost, and highly capable high resolution measurement system for which to perform long duration and large instantaneous bandwidth spectrum measurements. A novel filter was designed and implemented to significantly improve the performance of the energy detection method used to classify whitespace from incumbent activity.

This capability enabled a thorough investigation into the structure of secondary spectrum whitespace to be conducted. Chapter 5 successfully developed a method to evaluate the smallest unit of a secondary spectrum resource, the 'WhiteSpace Resource Block' (WSRB). Multiple WSRBs are clustered and their resulting structure (due to the clustering) is used to calculate a 'functional value' of the cluster for which whitespace can

be compared and evaluated. This method is then expanded to determine the greater structure of secondary spectrum resources within a candidate band. A novel partitioning algorithm is developed analyse the high resolution, continuous spectrum observations to provide insight into how whitespace is structured within the band. By intelligently clustering the secondary spectrum resources detected, information about the structure of secondary spectrum is able to be made known.

Chapter 6 detailed several novel improvements to current and future secondary spectrum management systems. These significant contributions are largely attributed to the investigation into the impacts of optimal timescales on secondary spectrum sharing, and the impact of timescales on device power consumption and latency of operation.

The successfully awarded patents and publications detailed in this work are a testament to the significant interest behind secondary spectrum and the opportunities and engineering challenges that it presents. The investigation into optimal timescales and the quantification of secondary spectrum, provides significant new insight and methods for the development of future secondary spectrum management systems, and secondary spectrum evaluation.

7.2 Future Work

Several key areas of the investigation present opportunities for further investigation and improvement. These items of future work are as follows:

1. Stemming from the RoC investigation in Section 4.4.2, the filter defined in Section 4.6, lacks a concrete evaluation for the misdetection and false detection rates. In the absence of a ‘known’ set of incumbent signals for which to perform the evaluation, an alternative to performing this evaluation can be performed via Monte Carlo methods to simulate a reasonable spectrum environment for which to assess the classification

of the filter.

2. From Section 5.6, a closed form for the upper bound of the partitioning algorithms performance, was not formally obtained, and would be an interesting point of further investigation to concretely evaluate the performance of the partitioning algorithm. Particular improvements to the partitioning algorithm, could also be investigated such as re-packing to find a greater maximal set, or optimising across all bandwidths and timescales to find a combination that results in a higher overall functional value extracted from the spectrum.
3. Section 3.4.3 lacks a closed mathematical form for computing risk and evaluating the performance of decisions. Additional development for the time access risk thought experiment can also be performed, by integrating partitioning results from spectrum observations to calculate a PSD of actual whitespace and incumbent events to further optimise projected energy consumption.
4. As a further extension to Section 3.4.3: Partitioned spectrum datasets can be used to model the underlying spectrum this model can then be stepped forward over time with current state spectrum observations to determine whether the model possesses any predictability. If so, a method to exploit the model to reduce power consumption and latency metrics by reducing sensing periods and/or making more intelligent decisions of where and when to access the secondary spectrum can be formulated.
5. As a further investigation this partitioning algorithm detailed in Section 5.6 can be run in realtime with relatively low computational overhead, making it an interesting solution for deployment on distributed sensor nodes. These observation can then be used to construct a secondary spectrum resource radio environment map, combining the merits of the Whitespace Opportunity Distributions, over a geographical

region, rather than a single observation point. This will have numerous interesting implications for spectrum cartography.

Bibliography

- [1] J. McMillan, “Selling spectrum rights,” *The Journal of Economic Perspectives*, vol. 8, no. 3, pp. 145–162, 1994.
- [2] P. Cramton, “Spectrum auctions,” in *Handbook of Telecommunications Economics: Structure, regulation, and competition*, M. Cave, S. Majumdar, and I. Vogelsang, Eds. Amsterdam: Elsevier, 2002, p. 37.
- [3] M. Wellens, J. Wu, and P. Mahonen, “Evaluation of spectrum occupancy in indoor and outdoor scenario in the context of cognitive radio,” in *Cognitive Radio Oriented Wireless Networks and Communications, 2007. CrownCom 2007. 2nd International Conference on*. IEEE, 2007, pp. 420–427.
- [4] M. Lopez-Benitez and F. Casadevall, “Empirical time-dimension model of spectrum use based on a discrete-time markov chain with deterministic and stochastic duty cycle models,” *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 6, pp. 2519–2533, July 2011.
- [5] D. Cabric, S. Mishra, and R. Brodersen, “Implementation issues in spectrum sensing for cognitive radios,” in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Eighth Asilomar Conference on*, vol. 1, Nov 2004, pp. 772–776 Vol.1.

- [6] **P. M. Rixon** and M. Heimlich, “Power and latency limitations in secondary spectrum reuse for mobile and home wireless systems,” *2014 International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pp. 480–485, 2014.
- [7] A. Wyglinski, M. Nekovee, and T. Hou, *Cognitive Radio Communications and Networks: Principles and Practice*. Elsevier Science, 2009. [Online]. Available: <https://books.google.com.au/books?id=d3HBCmFEemicC>
- [8] **P. M. Rixon** and M. C. Heimlich, “Quantifying secondary spectrum whitespace opportunity distributions,” in *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, June 2017, pp. 640–646.
- [9] M. D. Mueck, M. Heimlich, E. Dutkiewicz, G. Fang, **P. Rixon**, and B. A. Jayawickrama, “Signal Buffering for Licensed Shared Access (LSA) Technology,” Patent TW201 631 995 (A). [Online]. Available: https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20160901&DB=&locale=en_EP&CC=TW&NR=201631995A
- [10] M. D. Mueck, **P. Rixon**, M. Heimlich, E. Dutkiewicz, and C. Drewes, “Evolved Node-B, Local Controller and Method for Allocation of Spectrum for Secondary Usage,” Patent WO2017052517 (A1). [Online]. Available: https://worldwide.espacenet.com/publicationDetails/biblio?II=2&ND=3&adjacent=true&locale=en_EP&FT=D&date=20170330&CC=WO&NR=2017052517A1
- [11] M. D. Mueck, M. Heimlich, **P. Rixon**, B. A. Jayawickrama, E. Dutkiewicz, C. Drewes, and G. Fang, “Methods for Performing Mobile Communications Between Mobile Terminal Devices, Base Stations and Network Control Devices,” Patent WO2017052869 (A1). [Online]. Available: https://worldwide.espacenet.com/publicationDetails/biblio?II=2&ND=3&adjacent=true&locale=en_EP&FT=D&date=20170330&CC=WO&NR=2017052869A1

- [//worldwide.espacenet.com/publicationDetails/biblio?II=3&ND=3&adjacent=true&locale=en_EP&FT=D&date=20170330&CC=WO&NR=2017052869A1](https://worldwide.espacenet.com/publicationDetails/biblio?II=3&ND=3&adjacent=true&locale=en_EP&FT=D&date=20170330&CC=WO&NR=2017052869A1)
- [12] M. D. Mueck, C. Drewes, E. Dutkiewicz, B. A. Jayawickrama, **P. Rixon**, G. Fang, M. Heimlich, and S. Srikanteswara, “Methods and Devices for Shared Spectrum Allocation,” Patent WO2018004641 (A1). [Online]. Available: https://worldwide.espacenet.com/publicationDetails/biblio?II=0&ND=3&adjacent=true&locale=en_EP&FT=D&date=20180104&CC=WO&NR=2018004641A1
- [13] *Methods for Determining National Long Term Strategies For Spectrum Utilization*, ITU, ITU-R SM.2015.
- [14] J. Proakis and M. Salehi, *Digital Communications*, ser. McGraw-Hill International Edition. McGraw-Hill, 2008. [Online]. Available: <https://books.google.com.au/books?id=ksh0GgAACAAJ>
- [15] C. Phillips, D. Sicker, and D. Grunwald, “A survey of wireless path loss prediction and coverage mapping methods,” *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 255–270, First 2013.
- [16] T. Yucek and H. Arslan, “A survey of spectrum sensing algorithms for cognitive radio applications,” *IEEE Communications Surveys Tutorials*, vol. 11, no. 1, pp. 116–130, First 2009.
- [17] G. J. Foschini, “Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas,” *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–59, Autumn 1996.
- [18] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, “Massive mimo for next generation wireless systems,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.

- [19] A. A. M. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 2, pp. 128–137, February 1987.
- [20] (2017) International Telecommunications Union. ITU. [Online]. Available: <http://www.itu.int/en/Pages/default.aspx>
- [21] (2017) Master International Frequency Register. ITU. [Online]. Available: <http://www.itu.int/en/ITU-R/terrestrial/broadcast/Pages/MIFR.aspx>
- [22] (2017) ITU regions. ITU. [Online]. Available: <http://www.itu.int/net/ITU-R/index.asp?category=information&rlink=emergency-bands&lang=en>
- [23] (2017) Radio Spectrum Allocation. Federal Communications Commission. [Online]. Available: <https://www.fcc.gov/engineering-technology/policy-and-rules-division/general/radio-spectrum-allocation>
- [24] (2017) Spectrum management. Office for Communications. [Online]. Available: <https://www.ofcom.org.uk/spectrum/spectrum-management>
- [25] (2017) EU's spectrum policy framework. European Commission. [Online]. Available: <https://ec.europa.eu/digital-single-market/en/eus-spectrum-policy-framework>
- [26] (2017) Radiofrequency spectrum regulation. ACMA. [Online]. Available: <http://www.acma.gov.au/theACMA/About/The-ACMA-story/Regulating/radiofrequency-spectrum-regulation>
- [27] *Radio Regulations*, 2016 ed., ITU, 2016, vol. 1.
- [28] Spectrum auctions list. ACMA. [Online]. Available: <http://acma.gov.au/Industry/Spectrum/Radiocomms-licensing/Spectrum-licences/spectrum-auctions-list-spectrum-planning-acma>

- [29] (2015) Auctions summary. FCC. [Online]. Available: http://wireless.fcc.gov/auctions/default.htm?job=auctions_all
- [30] P. Cramton, "Spectrum auction design," *Review of Industrial Organization*, vol. 42, no. 2, pp. 161–190, 2013.
- [31] *Radiocommunications Act 1992*, Australian Government, sec. 60.
- [32] (2017) Australian radiofrequency spectrum allocations chart. ACMA. [Online]. Available: <http://www.acma.gov.au/~media/Spectrum%20Engineering/Information/pdf/Spectrum%20Chart%202017%20pdf>
- [33] (2017, Feb.) Cisco visual networking index: Global mobile data traffic forecast update, 2016–2021. Cisco. [Online]. Available: <https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf>
- [34] (2017, Nov.) Ericsson mobility report - november 2017. Ericsson. [Online]. Available: <https://www.ericsson.com/assets/local/mobility-report/documents/2017/ericsson-mobility-report-november-2017-central-and-eastern-europe.pdf>
- [35] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5g of terrestrial mobile telecommunication," *Communications Magazine, IEEE*, vol. 52, no. 5, pp. 36–43, May 2014.
- [36] (2015, Mar.) 5G Candidate Band Study. OFCOM. [Online]. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0014/31910/qa-report.pdf
- [37] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannis, E. Björnson, K. Yang, C. I, and A. Ghosh, "Millimeter wave communications

- for future mobile networks,” *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1909–1935, Sept 2017.
- [38] (2018, Apr.) ITU Radiocommunication Sector. ITU-R. [Online]. Available: <https://www.itu.int/en/ITU-R/information/Pages/default.aspx>
- [39] (2018, Feb.) ECC and ETSI. CEPT Electronics Communications Committee. [Online]. Available: <https://www.cept.org/ecc/ecc-and-etsi>
- [40] (2018, Apr.) About ETSI. European Telecommunications and Standards Institute. [Online]. Available: <http://www.etsi.org/about>
- [41] (2018, Apr.) About 3GPP. 3GPP. 3rd Generation Partnership Project. [Online]. Available: <http://www.3gpp.org/about-3gpp>
- [42] (2017, Jan.) IEEE 802 LAN/MAN Standards Committee. IEEE. [Online]. Available: <http://www.ieee802.org/>
- [43] J. Zander and S.-L. Kim, *Radio Resource Management for Wireless Networks*. Artech House, 2001.
- [44] E. Dahlman, S. Parkvall, and S. Johan, *4G: LTE/LTE-Advanced for mobile broadband*. Academic Press is an imprint of Elsevier, 2014.
- [45] ———, *4G, LTE-Advanced Pro and the Road to 5G*. Elsevier, 2016.
- [46] *Data Networks and Open System Communications - Open Systems Interconnection - Model and Notation*, ITU-T, Jun. 1994, t-REC-X.200 Sec. 6.
- [47] L. Cimini, “Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing,” *IEEE Transactions on Communications*, vol. 33, no. 7, pp. 665–675, Jul 1985.

- [48] (2017) Communication - OFDM. Sharetechnote. [Online]. Available: http://www.sharetechnote.com/html/Communication_OFDM.html
- [49] Y. Zhao and S. G. Haggman, "Sensitivity to doppler shift and carrier frequency errors in ofdm systems-the consequences and solutions," in *Proceedings of Vehicular Technology Conference - VTC*, vol. 3, Apr 1996, pp. 1564–1568 vol.3.
- [50] (2008) LTE PHY Spec. Samsung. [Online]. Available: <https://www.slideshare.net/allabout4g/lte-rel-8-physical-layer>
- [51] (2018) LTE resource grid. [Online]. Available: http://niviuk.free.fr/lte_resource_grid.html
- [52] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile ofdm system," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 3, pp. 1122–1128, Aug 1998.
- [53] P. W. C. Chan, E. S. Lo, R. R. Wang, E. K. S. Au, V. K. N. Lau, R. S. Cheng, W. H. Mow, R. D. Murch, and K. B. Letaief, "The evolution path of 4g networks: Fdd or tdd?" *IEEE Communications Magazine*, vol. 44, no. 12, pp. 42–50, Dec 2006.
- [54] G. Yuan, X. Zhang, W. Wang, and Y. Yang, "Carrier aggregation for lte-advanced mobile communication systems," *IEEE Communications Magazine*, vol. 48, no. 2, pp. 88–93, February 2010.
- [55] H. Lee, S. Vahid, and K. Moessner, "A survey of radio resource management for spectrum aggregation in lte-advanced," *IEEE Communications Surveys Tutorials*, vol. 16, no. 2, pp. 745–760, Second 2014.

- [56] (2014, Feb.) Ecc report 205 licensed shared access (lsa). CEPT Electronics Communications Committee. FM53. [Online]. Available: <http://www.erodocdb.dk/Docs/doc98/official/pdf/ECCREP205.PDF>
- [57] J. Khun-Jush, P. Bender, B. Deschamps, and M. Gundlach, "Licensed shared access as complementary approach to meet spectrum demands: Benefits for next generation cellular systems," in *ETSI Workshop on reconfigurable radio systems*, 2012.
- [58] (2016, Dec.) Wireless telecommunications bureau and office of engineering and technology conditionally approve seven spectrum access system administrators for the 3.5ghz band. FCC. [Online]. Available: https://apps.fcc.gov/edocs_public/attachmatch/DA-16-1426A1_Rcd.pdf
- [59] M. D. Mueck, S. Srikanteswara, and B. Badic. (2015, oct) Spectrum sharing: Licensed shared access (LSA) and spectrum access system (SAS). Intel Corporation. [Online]. Available: <http://www.intel.com/content/dam/www/public/us/en/documents/white-papers/spectrum-sharing-lsa-sas-paper.pdf>
- [60] (2015, Oct.) Reconfigurable Radio Systems (RRS); System architecture and high level procedures for operation of Licensed Shared Access (LSA) in the 2,300 MHz - 2,400 MHz band . ETSI. TS 103 235 v1.1.1. [Online]. Available: http://www.etsi.org/deliver/etsi_ts/103200_103299/103235/01.01.01_60/ts_103235v010101p.pdf
- [61] (2018, May) Requirements for commercial operation in the u.s. 3550-3700 mhz citizens broadband radio service band. WINNF. [Online]. Available: https://workspace.winnforum.org/higherlogic/ws/public/document?document_id=6531&wg_id=CBRS-Specs
- [62] C. Cordeiro, K. Challapali, D. Birru, and S. Shankar, "Ieee 802.22: the first worldwide wireless standard based on cognitive radios," in *First IEEE International Sym-*

- posium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, Nov 2005, pp. 328–337.
- [63] A. Mishra and D. Johnson, *White Space Communication: Advances, Developments and Engineering Challenges*, ser. Signals and Communication Technology. Springer International Publishing, 2014. [Online]. Available: <https://books.google.com.au/books?id=G0vPBAAAQBAJ>
- [64] J. van de Beek, J. Riihijarvi, A. Achtzehn, and P. Mahonen, “Tv white space in europe,” *IEEE Transactions on Mobile Computing*, vol. 11, no. 2, pp. 178–188, Feb 2012.
- [65] (2012, Apr.) In the matter of Unlicensed Operation in the TV Broadcast Bands - Additional Spectrum for Unlicensed Devices Below 900 MHz and in the 3 GHz Band. FCC. Third Memorandum Opinion and Order, 12-36. [Online]. Available: https://apps.fcc.gov/edocs_public/attachmatch/FCC-12-36A1.pdf
- [66] (2015, Feb.) Implementing TV White Spaces. OFCOM. [Online]. Available: https://www.ofcom.org.uk/__data/assets/pdf_file/0034/68668/tvws-statement.pdf
- [67] A. B. Flores, R. E. Guerra, E. W. Knightly, P. Ecclesine, and S. Pandey, “Ieee 802.11af: a standard for tv white space spectrum sharing,” *IEEE Communications Magazine*, vol. 51, no. 10, pp. 92–100, October 2013.
- [68] (2017, May) Digital dividend Auction - Results. ACMA. [Online]. Available: <https://www.acma.gov.au/theACMA/digital-dividend-auction-results>
- [69] (2008, Mar.) Auction 73: 700 MHz Band. FCC. [Online]. Available: <https://www.fcc.gov/auction/73>

- [70] (2011) Report on Collective Use of Spectrum (CUS) and other spectrum sharing approaches. Radio Spectrum Policy Group. [Online]. Available: http://rspg-spectrum.eu/wp-content/uploads/2013/05/rspg11_392_report_CUS_other_approaches_final.pdf
- [71] (2017, Jan.) Reconfigurable Radio Systems (RRS); Information elements and protocols for the interface between LSA Controller (LC) and LSA Repository (LR) for operation of Licensed Shared Access (LSA) in the 2,300 MHz - 2,400 MHz band. ETSI. TS 103 379 v1.1.1. [Online]. Available: http://www.etsi.org/deliver/etsi_ts/103300_103399/103379/01.01.01_60/ts_103379v010101p.pdf
- [72] (2013, Jan.) Electromagnetic compatibility and Radio spectrum Matters (ERM); System Reference document (SRdoc); Mobile broadband services in the 2,300 MHz - 2,400 MHz frequency band under Licensed Shared Access regime. ETSI. TR 103 113 v1.1.1. [Online]. Available: http://www.etsi.org/deliver/etsi_tr/103100_103199/103113/01.01.01_60/tr_103113v010101p.pdf
- [73] (2014, Oct.) Reconfigurable Radio Systems (RRS); System requirements for operation of Mobile Broadband Systems in the 2,300 MHz - 2,400 MHz band under Licensed Shared Access (LSA). ETSI. TS 103 154 v1.1.1. [Online]. Available: http://www.etsi.org/deliver/etsi_ts/103100_103199/103154/01.01.01_60/ts_103154v010101p.pdf
- [74] (2018, Mar.) LSA Implementation. CEPT ECC. [Online]. Available: <https://www.cept.org/ecc/topics/lsa-implementation>
- [75] (2015, Apr.) Amendment of the Commission's Rules with Regard to Commercial Operations in the 3550-3650 MHz Band. FCC. Report and Order and

- Second Further Notice of Proposed Rulemaking. [Online]. Available: https://apps.fcc.gov/edocs_public/attachmatch/FCC-15-47A1.pdf
- [76] R. I. Lackey and D. W. Upmal, "Speakeasy: the military software radio," *IEEE Communications Magazine*, vol. 33, no. 5, pp. 56–61, May 1995.
- [77] E. Buracchini, "The software radio concept," *IEEE Communications Magazine*, vol. 38, no. 9, pp. 138–143, Sep 2000.
- [78] (2018) Software Defined Radio. Lime microsystems. LimeSDR and LimeSDR Mini. [Online]. Available: <http://www.limemicro.com/products/software-defined-radio/>
- [79] (2017) USRP n210. Ettus Research. [Online]. Available: <https://www.ettus.com/product/details/UN210-KIT>
- [80] R. Bagheri, A. Mirzaei, S. Chehrazi, M. E. Heidari, M. Lee, M. Mikhemar, W. Tang, and A. A. Abidi, "An 800-mhz - 6-ghz software-defined wireless receiver in 90-nm cmos," *IEEE Journal of Solid-State Circuits*, vol. 41, no. 12, pp. 2860–2876, Dec 2006.
- [81] A. A. Abidi, "The path to the software-defined radio receiver," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 5, pp. 954–966, May 2007.
- [82] E. Buracchini, "The software radio concept," *IEEE Communications Magazine*, vol. 38, no. 9, pp. 138–143, Sep 2000.
- [83] J. Mitola, "Cognitive radio for flexible mobile multimedia communications," in *Mobile Multimedia Communications, 1999.(MoMuC'99) 1999 IEEE International Workshop on*. IEEE, 1999, pp. 3–10.

- [84] E. Hossain and V. Bhargava, *Cognitive Wireless Communication Networks*. Springer US, 2007. [Online]. Available: <https://books.google.com.au/books?id=IBjbJqmlH4AC>
- [85] W. A. Gardner, "Exploitation of spectral redundancy in cyclostationary signals," *IEEE Signal Processing Magazine*, vol. 8, no. 2, pp. 14–36, April 1991.
- [86] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "A survey on spectrum management in cognitive radio networks," *IEEE Communications magazine*, vol. 46, no. 4, 2008.
- [87] G. I. Tsiropoulos, O. A. Dobre, M. H. Ahmed, and K. E. Baddour, "Radio resource allocation techniques for efficient spectrum access in cognitive radio networks," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 824–847, Firstquarter 2016.
- [88] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb 2005.
- [89] M. Bkassiny, Y. Li, and S. K. Jayaweera, "A survey on machine-learning techniques in cognitive radios," *IEEE Communications Surveys Tutorials*, vol. 15, no. 3, pp. 1136–1159, Third 2013.
- [90] (2017, Sep.) Continuing moore's law. Intel. [Online]. Available: <https://newsroom.intel.com/newsroom/wp-content/uploads/sites/11/2017/09/mark-bohr-on-continuing-moores-law.pdf>
- [91] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, "Band parameters for III-V compound semiconductors and their alloys," *Journal of Applied Physics*, vol. 89, pp. 5815–5875, Jun. 2001.

- [92] (2018, Sep.) Intel custom foundry process technology. Intel. [Online]. Available: <https://www.intel.com/content/www/us/en/foundry/process-technology.html>
- [93] (2018, Sep.) Tsmc logic technology. TSMC. [Online]. Available: <http://www.tsmc.com/english/dedicatedFoundry/technology/logic.htm>
- [94] (2018, Sep.) Samsung semiconductor business overview. Samsung. [Online]. Available: <https://www.samsungfoundry.com>
- [95] (2018, Sep.) Qorvo products. Qorvo. [Online]. Available: <https://www.qorvo.com/products>
- [96] (2018, Sep.) Skyworks products. Skyworks. [Online]. Available: <http://www.skyworksinc.com/ProductsLanding.aspx>
- [97] (2018, Sep.) Broadcom products. Broadcom. [Online]. Available: <https://www.broadcom.com/products/>
- [98] (2018, Mar.) Samsung galaxy s9 teardown. TechInsights. [Online]. Available: <http://www.techinsights.com/about-techinsights/overview/blog/samsung-galaxy-s9-teardown/>
- [99] (2014, Sep.) Samsung v-nand technology. Samsung. [Online]. Available: https://www.samsung.com/us/business/oem-solutions/pdfs/V-NAND_technology_WP.pdf
- [100] H. Esmailzadeh, E. Blem, R. S. Amant, K. Sankaralingam, and D. Burger, “Dark silicon and the end of multicore scaling,” in *2011 38th Annual International Symposium on Computer Architecture (ISCA)*, June 2011, pp. 365–376.
- [101] S. Albers and A. Antoniadis, “Race to idle: New algorithms for speed scaling with a sleep state,” pp. 1266–1285, 01 2012.

- [102] (2018, Jan.) Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Architecture description. ETSI. TS 136 401 v15.0.0. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/36_series/36.401/36401-f00.zip
- [103] (1972) Apollo experience report - lunar module communications system. NASA. Technical Note TN D-6974. [Online]. Available: <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19720023255.pdf>
- [104] N. Yanduru, D. Griffith, S. Bhagavatheeswaran, C.-C. Chen, F. Dulger, S.-J. Fang, Y.-C. Ho, and K. M. Low, “A wcdma, gsm/gprs/edge receiver front end without interstage saw filter,” in *Radio Frequency Integrated Circuits (RFIC) Symposium, 2006 IEEE*, June 2006, pp. 4 pp.–.
- [105] M. Nilsson, S. Mattisson, N. Klemmer, M. Anderson, T. Arnborg, P. Caputa, S. Ek, L. Fan, H. Fredriksson, F. Garrigues, H. Geis, H. Hagberg, J. Hedestig, H. Huang, Y. Kagan, N. Karlsson, H. Kinzel, T. Mattsson, T. Mills, F. Mu, A. Mårtensson, L. Nicklasson, F. Oredsson, U. Ozdemir, F. Park, T. Pettersson, T. Pålhlsson, M. Pålsson, S. Ramon, M. Sandgren, P. Sandrup, A. Stenman, R. Strandberg, L. Sundström, F. Tillman, T. Tired, S. Uppathil, J. Walukas, E. Westesson, X. Zhang, and P. Andreani, “A 9-band wcdma/edge transceiver supporting hspa evolution,” in *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International*, Feb 2011, pp. 366–368.
- [106] S. Rodriguez, A. Rusu, and M. Ismail, “Wimax/lte receiver front-end in 90nm cmos,” in *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*, May 2009, pp. 1036–1039.

- [107] L.-H. Li, F.-L. Lin, and H.-R. Chuang, "Complete rf-system analysis of direct conversion receiver (dcr) for 802.11a wlan ofdm system," *Vehicular Technology, IEEE Transactions on*, vol. 56, no. 4, pp. 1696–1703, July 2007.
- [108] F. Khan, *LTE for 4G mobile broadband: air interface technologies and performance*. Cambridge University Press, 2009.
- [109] A. Ghosh, J. Zhang, J. G. Andrews, and R. Muhamed, *Fundamentals of LTE*. Pearson Education, 2010.
- [110] M. Laner, P. Svoboda, P. Romirer-Maierhofer, N. Nikaein, F. Ricciato, and M. Rupp, "A comparison between one-way delays in operating hspa and lte networks," in *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2012 10th International Symposium on*, May 2012, pp. 286–292.
- [111] (2012, Nov.) 3gpp tr 25.911 v11.0.0 release 11. ETSI. [Online]. Available: <http://www.etsi.org/>
- [112] A. Jensen, M. Lauridsen, P. Mogensen, T. Sørensen, and P. Jensen, "Lte ue power consumption model: For system level energy and performance optimization," in *Vehicular Technology Conference (VTC Fall), 2012 IEEE*, Sept 2012, pp. 1–5.
- [113] V. Valenta, R. Maršálek, G. Baudoin, M. Villegas, M. Suarez, and F. Robert, "Survey on spectrum utilization in europe: Measurements, analyses and observations," in *2010 Proceedings of the Fifth International Conference on Cognitive Radio Oriented Wireless Networks and Communications*, June 2010, pp. 1–5.
- [114] (2017) Microsoft spectrum observatory. Microsoft. [Online]. Available: <http://observatory.microsoftspectrum.com/>

- [115] (2017) Google spectrum database. Microsoft. [Online]. Available: <https://www.google.com/get/spectrumdatabase/>
- [116] G. Heinzel, A. Rüdiger, and R. Schilling. (2002) Spectrum and spectral density estimation by the discrete fourier transform (dft), including a comprehensive list of window functions and some new at-top windows. [Online]. Available: https://holometer.fnal.gov/GH_FFT.pdf
- [117] M. TRETHERWEY, “Window and overlap processing effects on power estimates from spectra,” *Mechanical Systems and Signal Processing*, vol. 14, no. 2, pp. 267 – 278, 2000. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0888327099912748>
- [118] S. G. Johnson and M. Frigo, “A modified split-radix fft with fewer arithmetic operations,” *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 111–119, Jan 2007.
- [119] H. Urkowitz, “Energy detection of unknown deterministic signals,” *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, April 1967.
- [120] F. F. Digham, M. Alouini, and M. K. Simon, “On the energy detection of unknown signals over fading channels,” *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 21–24, Jan 2007.
- [121] S. Atapattu, C. Tellambura, and H. Jiang, *Energy Detection for Spectrum Sensing in Cognitive Radio*. Springer, 2014. [Online]. Available: https://link.springer.com/chapter/10.1007/978-1-4939-0494-5_2
- [122] K. Arshad and K. Moessner, “Robust spectrum sensing based on statistical tests,” *IET Communications*, vol. 7, no. 9, pp. 808–817, June 2013.

- [123] J. G. Proakis and D. K. Manolakis, *Digital Signal Processing (4th Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.
- [124] T. Fawcett, “An introduction to roc analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861 – 874, 2006, rOC Analysis in Pattern Recognition.
- [125] (2012, nov) 3gpp tr 36.913 v11.0.0 release 11. ETSI. [Online]. Available: <http://www.etsi.org/>
- [126] (2007) ADS62P4X. Texas Instruments. [Online]. Available: <http://www.ti.com/lit/ds/symlink/ads62p45.pdf>
- [127] (2017) VERT900 Antenna. Ettus Research. [Online]. Available: <https://www.ettus.com/product/details/VERT900>
- [128] (2018) Matlab Product Page. MathWorks. [Online]. Available: <https://www.mathworks.com/products/matlab.html>
- [129] (2018) High Definition Software Defined Radio. HSDSDR. [Online]. Available: <http://www.hdsdr.de/>
- [130] (2018) GNU Radio - The Free and Open Software Radio Ecosystem. GNU Radio. [Online]. Available: <https://www.gnuradio.org/>
- [131] (2018) FFTW3. [Online]. Available: <http://www.fftw.org/>
- [132] (2017, apr) Citizens Band Radio Service (CBRS). FCC. [Online]. Available: <https://www.fcc.gov/wireless/bureau-divisions/mobility-division/citizens-band-radio-service-cbrs#block-menu-block-4>
- [133] (2014, apr) 12.5 Citizen band radio service (CBRS) repeater (400 MHz CBRS band). ACMA. [Online]. Available: <https://www.acma.gov>

au/Industry/Spectrum/Spectrum-planning/Current-APs-info-and-resources/
citizen-band-radio-service-repeater-400-mhz-cbrs-band

- [134] R. Tandra and A. Sahai, “Snr walls for signal detection,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 1, pp. 4–17, Feb 2008.
- [135] S. Durocher and S. Mehrabi, “Computing partitions of rectilinear polygons with minimum stabbing number,” in *International Computing and Combinatorics Conference*. Springer, 2012, pp. 228–239.
- [136] M. M. Halldórsson, “Approximations of weighted independent set and hereditary subset problems,” *Journal of Graph Algorithms and Applications*, vol. 4, no. 1, pp. 1–16, 2000.
- [137] D. L. Donoho and P. B. Stark, “Uncertainty principles and signal recovery,” *SIAM J. Appl. Math.*, vol. 49, no. 3, pp. 906–931, Jun. 1989.
- [138] P. Chalermsook and J. Chuzhoy, “Maximum independent set of rectangles,” in *Proceedings of the twentieth Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, 2009, pp. 892–901.