Robust Blind Spectral Estimation in the Presence of Impulsive Noise

Joel Thomas Kees

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Electrical Engineering

A. A. (Louis) Beex, Chair

Joseph M. Ernst, Co-Chair

William C. Headley

Harpreet S. Dhillon

February 5, 2019

Blacksburg, Virginia

Keywords: Spectrum Sensing, Power Line Communication, Welch Method, Multitaper Method, Middleton Class A Noise, Impulsive Noise, Robust Estimation

Robust Blind Spectral Estimation in the Presence of Impulsive Noise

Joel Thomas Kees

ABSTRACT

Robust nonparametric spectral estimation includes generating an accurate estimate of the Power Spectral Density (PSD) for a given set of data while trying to minimize the bias due to data outliers. Robust nonparametric spectral estimation is applied in the domain of electrical communications and digital signal processing when a PSD estimate of the electromagnetic spectrum is desired (often for the goal of signal detection), and when the spectrum is also contaminated by Impulsive Noise (IN). Power Line Communication (PLC) is an example of a communication environment where IN is a concern because power lines were not designed with the intent to transmit communication signals. There are many different noise models used to statistically model different types of IN, but one popular model that has been used for PLC and various other applications is called the Middleton Class A model, and this model is extensively used in this thesis. The performances of two different nonparametric spectral estimation methods are analyzed in IN: the Welch method and the multitaper method. These estimators work well under the common assumption that the receiver noise is characterized by Additive White Gaussian Noise (AWGN). However, the performance degrades for both of these estimators when they are used for signal detection in IN environments. In this thesis basic robust estimation theory is used to modify the Welch and multitaper methods in order to increase their robustness, and it is shown that the signal detection capabilities in IN is improved when using the modified robust estimators.

Robust Blind Spectral Estimation in the Presence of Impulsive Noise

Joel Thomas Kees

GENERAL AUDIENCE ABSTRACT

One application of blind spectral estimation is blind signal detection. Unlike a car radio, where the radio is specifically designed to receive AM and PM radio waves, sometimes it is useful for a radio to be able to detect the presence of transmitted signals whose characteristics are not known ahead of time. Cognitive radio is one application where this capability is useful. Often signal detection is inhibited by Additive White Gaussian Noise (AWGN). This is analogous to trying to hear a friend speak (signal detection) in a room full of people talking (background AWGN). However, some noise environments are more impulsive in nature. Using the previous analogy, the background noise could be loud banging caused by machinery; the noise will not be as constant as the chatter of the crowd, but it will be much louder. When power lines are used as a medium for electromagnetic communication (instead of just sending power), it is called Power Line Communication (PLC), and PLC is a good example of a system where the noise environment is impulsive. In this thesis, methods used for blind spectral estimation are modified to work reliably (or robustly) for impulsive noise environments.

Acknowledgments

I would like to acknowledge my advisor and committee chair: Dr. A. A. (Louis) Beex. Through his patient and wise guidance he has taught me to be a better researcher, writer, and professional. He has demonstrated an amazing commitment to my professional success through the time and effort he has invested in me. He always seemed to know how to encourage me when I needed it and exactly how hard to push me.

I would like to acknowledge my co-chair Dr. Joseph Ernst for aiding and encouraging me all the way up to and through the midnight submission deadline of my first published paper, my committee member Dr. William (Chris) Headley for all the wise graduate school advice he gave me and his excellent leadership and example as the principle investigator to the Hume Center project I worked on, and my committee member Dr. Harpreet Dhillon for teaching me almost everything I know about statistics and always being patient with me when I sought out answers to mathematical questions.

I would like to acknowledge the faculty, staff, and students that compose the Hume Center for National Security and Technology for all their support, advice, and encouragement. I would like specifically to acknowledge Dr. Alan Michaels for his professional guidance as I sought a career path.

I would like to acknowledge my extremely supportive family for all their academic

encouragement and especially for their spiritual encouragement during the worst of my wrist health issues during the first year of graduate school. I would like specifically to acknowledge my parents for the stability they provided while raising me and for their hand in forming me into someone capable of achieving this goal. I would like to thank my brother and sister-in-law for providing wonderful housing and the much needed social engagement during my first year of graduate school.

I would like to thank my home church, Blacksburg Christian Fellowship, and specifically Chris Faith, Bill Gibbs, and Steve Hodges for their spiritual mentorship.

I would like to thank most of all the Lover of my soul, my Savior, and the Lord of my life, Jesus—Who showers His gifts upon me (as evidenced by the aforementioned acknowledgements), Who sustains me spiritually, academically, and physically, and Who has prepared me for (and blessed me with) a job with LGS Innovations.

Contents

Contents vi

List of Acronyms ix

List of Figures xi

List of Tables xix

1 Introduction 1

1.1 Outline and Contributions 5

2 The Middleton Class A Noise Model 7

2.1 Definition of Class A Noise 8

2.2 Overall Power of Class A Noise 9

2.3 Gaussian Factor 11

2.4 Overlap Index 12

2.5 In-Phase and Quadrature Representation 13

3 Spectral Estimation Methods **16**

3.1 Overview of Spectral Estimation 16

3.2 The Welch Method 21

3.2.1 Overview of Welch Method 21

3.2.2 Making Welch more Robust 28

3.3 The Multitaper Method 32

3.3.1 Overview of Multitaper Method 33

3.3.2 Making Multitaper more Robust 39

4 Results **43**

4.1 Measurement Definitions and Default Parameters 43

4.2 Welch Results 48

4.3 Multitaper Results 76

4.4 Future Work 84

5 Conclusion **85**

Appendix A Supporting Math for Class A Model **88**

A.1 Demonstrate that the Class A PDF is a PDF 88

A.2 Second Moment of Rayleigh Distributed Random Variable 90

A.3 Two Useful Views of a Random Variable Satisfying a Given PDF 92

A.4 Method for Computing Second Moments of Middleton Class A Noise 94

A.5 Second Moment Class A Noise Envelope 95

A.6 Mapping Class A Complex Envelope and Phase Representation to In-Phase and Quadrature Representation 96

A.7 Showing Class A IQ Samples are Dependent and Uncorrelated 98

Appendix B Estimation Proofs **100**

B.1 Wiener-Khinchin Theorem 100

Bibliography **103**

List of Acronyms

ACF	AutoCorrelation Function
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
CEP	Complex Envelope and Phase
CDF	Cumulative Distribution Function
CR	Cognitive Radio
DFT	Discrete Fourier Transform
DPSS	Discrete Prolate Spheroidal Sequence
DSA	Dynamic Spectrum Access
DTFT	Discrete-Time Fourier Transform
E_b/N_0	energy per bit to noise power spectral density ratio
ED	Energy Detection
FCC	Federal Communications Commission
FFT	Fast Fourier Transform
GN	Gaussian Noise
IID	Independent and Identically Distributed
IN	Impulsive Noise
IQ	In-phase and Quadrature

IRLS	Iteratively Re-weighted Least-Squares
M3TWM	Median-Modified MultiTaper Welch Method
MTM	MultiTaper Method
MTWM	MultiTaper Welch Method
NB	NarrowBand
NGN	Non-Gaussian Noise
PLC	Power Line Communication
PDF	Probability Distribution Function
PU	Primary User
PSD	Power Spectral Density
16-QAM	16-state Quadrature Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying
ROC	Receiver Operating Characteristic
RP	Random Process
RV	Random Variable
SNR	Signal-to-Noise Ratio
SS	Spectrum Sensing
SU	Secondary User
TMMWM	Taper-Median-Modified Welch Method
WOSA	Welch's/Weighted Overlapped Segment Averaging
WSS	Wide-Sense Stationary
w.p.	With Probability

List of Figures

3.1 A recreation of Welch's original graphic depicting the setup of his method. Rather than taking one Discrete Fourier Transform (DFT) of a time series, he breaks the series into multiple smaller segments, computes the DFTs, and then averages. 22

3.2 Example of the simple periodogram method for a 256 sample complex baseband Quadrature Phase-Shift Keying (QPSK) signal in AWGN at E_b/N_0 of 10 decibels. Even at E_b/N_0 of 10 decibels the variance of the noise highly distorts the signal when using the simple periodogram method. 23

3.3 Example of the simple periodogram method for a complex baseband QPSK signal in AWGN at E_b/N_0 of 10 decibels. Even with a factor of 64 as many samples as in Figures 3.2, the variance caused by the noise does not decrease. 24

3.4 Example of the Welch method for a complex baseband QPSK signal in AWGN at E_b/N_0 of 10 decibels. By comparing Figure 3.3 with this figure, one can observe the low variance estimate attainable via the Welch estimator since the same number of QPSK samples is used to produce both Power Spectral Density (PSD) estimates. 25

3.5 The q function for the Huber estimator expressed by (3.19d). The smaller residuals are evenly weighted, but starting at $\pm\gamma$ the q function begins to down-weight the residuals. This makes the estimator more robust against outliers while still extracting location information efficiently from the samples near the estimate of location. The estimate of location is iteratively attained when the weighted residuals are minimized. 31

3.6 Discrete Prolate Spheroidal Sequence (DPSS) taper (left) and tapered QPSK In-phase and Quadrature (IQ) (right). Tapers are generated with $NW = 2.5$ and $N = 512$. 35

3.7 Spectral window of the DPSS tapers in Figure 3.6 (left) and cumulative spectral windows (right). 37

4.1 Example of traditional, median-modified, and Huber-modified Welch estimates in AWGN. All three estimators perform similarly. 47

4.2 Example of traditional, median-modified, and Huber-modified Welch estimates in Middleton Class A noise. The traditional estimator is not robust against outliers caused by impulsive noise, but the modified versions (which are nearly indistinguishable) are. 47

4.3 The range of Middleton Class A noise parameter pairs (Γ, A) found in the literature. The rectangular box represents the range examined in this thesis. 49

4.4 P_m for varying Γ and A for the first five values of m. Since P_m does not depend on Γ, each column is the same. As A increases (progressing down a column), P_m decays more slowly, which means the Class A noise will be less Gaussian (because the weight of P_0 is decreasing) though also less impulsive (as will be made clear in Figure 4.5). 51

4.5 σ_m^2 for varying Γ and A for the first 10 values of m. From the plots and (2.2), one can see that σ_m^2 is approximately proportional to m and inversely proportional to A. Recall that the power from a Class A Random Variable (RV) is not a function of Γ or A (2.7). Since the power of the mth Rayleigh RV increases as A decreases, the probability of selecting a higher power Rayleigh RV is simultaneously decreased by P_m (refer back to Figure 4.4) to maintain constant power. Additionally, as Γ changes, σ_m^2 shifts in such a way as to maintain constant power. 52

4.6 Complex envelope of Class A noise for varying A and Γ. As A increases (progression down the columns) the noise becomes less impulsive but retains the same underlying Gaussian power *and* retains the same non-Gaussian power (trading impulsiveness for density). As Γ increases the noise also becomes less impulsive; while retaining constant total power, the underlying Gaussian Noise (GN) power increases relative to the Non-Gaussian Noise (NGN) power (best illustrated by the third row). 54

4.7 Rayleigh Probability Distribution Function (PDF) (blue) and the PDFs of Class A envelopes (orange). Samples characterized by any of the PDFs have a power of unity. 56

4.8 Average performance of the three Welch PSD estimators (traditional-blue; median-red; Huber-orange) for QPSK signals with Class A noise from Figure 4.6. Recall that performance is measured by effective Signal-to-Noise Ratio (SNR) (the higher the better). With the large number of averaged segments and constant $E_b/N_0^{(\mathrm{T})} = 3$, the traditional Welch can adequately detect the signal. However, the robust estimators can improve upon the traditional estimator in highly impulsive environments (upper-left plots). 58

4.9 Distribution of effective SNR for 100 iterations for each estimator (traditional-blue; median-red; Huber-orange) in each plot. Performance centers around the averages shown in Figure 4.8 except every effective SNR above 9 decibels has been placed in the bin at 10 decibels. The top two rows contain scenarios where the traditional estimator does not detect the signal, but the robust estimators do. 60

4.10 Complex envelope of Class A noise varies according to Γ and A. Whereas the total power is held constant in Figure 4.6, in this figure the underlying Gaussian power is constant at $E_b/N_0^{(\mathrm{G})} = 5$ (assuming the noise is applied to a signal with $E_b = 1$). Notice as Γ increases, the total noise decreases because $E_b/N_0^{(\mathrm{G})}$ is directly proportional to $E_b/N_0^{(\mathrm{T})}$ and (for small Γ) roughly inversely proportional to Γ (2.8) and (4.3). The total noise power is constant for the plots within a given column because $E_b/N_0^{(\mathrm{G})}$ does not depend on A. 62

4.11 Average performance of the three Welch PSD estimators (traditional-blue; median-red; Huber-orange) for QPSK signals with Class A noise from Figure 4.10. Since the $E_b/N_0^{(\mathrm{T})}$ increase as Γ increases, the ability to detect the signal is better for the columns on the right. The bottom-left plots are examples of highly non-Gaussian environments with low impulsiveness, such that no estimator can detect the signal. The top-left plots have the same $E_b/N_0^{(\mathrm{T})}$ as the corresponding plots below, but are impulsive enough that the robust estimators detect the signal even though the traditional estimator cannot. . . 63

4.12 Distribution of effective SNR for 100 iterations for each estimator (traditional-blue; median-red; Huber-orange) in each plot. Similar to Figure 4.9, the estimators perform comparably with the major exception of under the condition of the second row ($A = 0.001$) and to a lesser extent the row above that, where the robust estimators outperform the traditional one. . . . 65

4.13 Receiver Operating Characteristic (ROC) plots generated using 1000 iterations for each estimator (traditional-blue; median-red; Huber-orange) in each plot. Note that the $E_b/N_0^{(\mathrm{T})}$ varies depending on the values of Γ and A to avoid generating ROC curves that are too linear or form a right angle. . . 67

4.14 Distribution of effective SNR for 100 iterations for each estimator (traditional-blue; median-red; Huber-orange) in each plot. Notice that this figure looks almost identical to Figure 4.12. 69

4.15 Distribution of effective SNR for 100 iterations for each estimator (traditional-blue; median-red; Huber-orange) in each plot. A segment length of 64 is used instead of the default length, 256. Notice that the plots in the middle row (primarily to leftmost two) show performance improvement for the robust methods. 71

4.16 ROC plots generated using 1000 iterations for each estimator (traditional-blue; median-red; Huber-orange) in each plot. The conditions are also equal to those used in Figure 4.13 except the segment length is 64 for the simulation conditions in this figure rather than the default length of 256. Note that the performance for the robust methods has increased for the left three plots in the middle row (compared to Figure 4.13). 73

4.17 The range of Middleton Class A noise parameter pairs (Γ, A) found in the literature. The rectangular box represents the range examined in this thesis. The blue and grey areas cover the regions in which the robust Welch estimators outperform the traditional Welch estimator using a segment length of 256 or 64, respectively. The dark blue area is the region of overlap. 75

4.18 Demonstration of how multitapering affects variance versus bias trade-off. Using additional tapers (orange) instead of only the first taper (blue) slightly increases the bias but drastically decreases the variance for each estimator—single segment (top), overlapped averaging (middle), and overlapped with median (bottom). 78

4.19 PSD estimation variance for the four estimators used to generate the middle and bottom plots of Figure 4.18. The dashed lines are the average variances for each estimator across the frequency bins. For $P = 7$, applying the MultiTaper Method (MTM) to the Welch estimator reduces the variance by 75% (blue to red). Since the median is a less efficient estimator than the mean for the Chi-Squared distribution, the variance for the Taper-Median-Modified Welch Method (TMMWM) is almost double the variance of the Welch estimator (blue to yellow). However, applying the MTM to the TMMWM reduces the variance by almost 80% (yellow to purple). 79

4.20 Distribution of effective SNR for 100 iterations for each estimator (MultiTaper Welch Method (MTWM)-blue; TMMWM-red; Median-Modified MultiTaper Welch Method (M3TWM)-orange) in each plot. The $E_b/N_0^{(\mathrm{T})} = 3$ for the simulation conditions for each plot. For certain pairs of (Γ, A) (primarily the values that correspond to the highlighted plots in the second and third row), the M3TWM (orange) is shown to be more robust than its non-robust counterpart (MTWM, blue). However, the robust non-multitapered benchmark (TMMWM, red) is more robust than the M3TWM (as seen in the middle row). This relationship roughly equates to a robustness versus variance trade-off. 81

4.21 Distribution of effective SNR for 100 iterations for each estimator (MTWM-blue; TMMWM-red; M3TWM-orange) in each plot. The $E_b/N_0^{(\mathrm{G})} = 5$ for the simulation conditions for each plot. Similar to Figure 4.20, the estimators perform comparably with the exception of under the conditions of the highlighted plots and to a lesser extent the top row, where the TMMWM (red) is more robust than the M3TWM (orange) which is much more robust than the MTWM (blue). 82

List of Tables

3.1 Nonparametric Spectral Estimation Methods 42

4.1 Default Simulation Parameters 46

Chapter 1

Introduction

The application of spectral estimation is seemingly limitless. This is increasingly true in a digitized world where information and data are perennially stored, transferred, manipulated, and processed. It is often desirable to know certain characteristics of data such as size, location, scale, frequency, etc. An inquiry of how data varies over time (frequency) is by nature a spectral estimation question. A few examples of when knowing the frequency characteristics of data is important include: encoding human speech [1], audio equalization, geoscience [2], [3], economics [4], cognitive radio [5], and many additional facets of communication systems.

Spectral estimation applied to communication systems is the topic of interest of this thesis. More specifically, the topic called Spectrum Sensing (SS). In layman's terms, SS can be thought of as holding up an antenna and trying to discern characteristics of the electromagnetic spectral environment and any observable signals propagating in the spectrum. SS can be used for applications such as signal detection, noise estimation, and channel estimation. There are various approaches to SS such as wavelet decomposition,

cyclostationarity, and nonparametric spectral estimation [6]. Wavelets can be used to extract various features via correlation with an orthogonal set of finite-length return-to-zero signals (wavelets). Cyclostationary detection exploits periodicity of mean and autocorrelation in the spectrum and can be used to extract certain features (such as baud rate of an unknown signal). Nonparametric spectral estimation involves generating an estimate of the Power Spectral Density (PSD), is comparatively computationally light, and can be used for applications such as blind signal detection via Energy Detection (ED). Since applications requiring quick blind spectral estimation are of interest in this thesis, two popular methods for performing blind nonparametric spectral estimation will be analyzed: (1) the Welch method developed by Peter Welch in 1967 [7] and (2) the MultiTaper Method (MTM) developed by David Thomson in 1982 [8], [9]. Relative to the wavelet based and cyclostationarity based methods, these two nonparametric spectral estimation methods are computationally efficient and completely blind, but suffer from limitations such as an inability to detect signals below the noise floor. The Welch method and MTM are natural choices for spectral estimation applications, such as Power Line Communication (PLC) and Cognitive Radio (CR), which require noise floor estimation and ED.

Like spectral estimation, robust estimation also has seemingly endless application. Robust estimation is the attempt to estimate parameters that characterize any given set of data without allowing data outliers to bias the estimate; outlier detection is another term that is closely related to robust estimation. In robust estimation theory, data outliers are always considered undesirable, but for other applications outliers give useful information, rendering outlier detection very useful. Robust estimation theory and outlier detection are widely applied in domains such as financial fraud detection, robotics, and PLC. For the previous examples, outliers could be caused by fraud, low quality sensors, and caustic noise environments, respectively. This demonstrates that outliers manage to contaminate all kinds

of data in various ways, and that correctly handling outliers is important.

In signal processing applied to communications, outliers can arise in the noise when the environment is characterized by impulsive noise (a type of Non-Gaussian Noise (NGN)); Additive White Gaussian Noise (AWGN) (which is often assumed—sometimes unwisely) does not produce outliers because the Gaussian distribution decays sufficiently fast. NGN environments can manifest for multiple reasons, such as caustic transmission mediums as in PLC or interference caused by other transmitters. Nonparametric power spectral estimation can be a good method for modeling some NGN environments or signal detection [10], but Impulsive Noise (IN) has the potential to prohibit these tasks. Therefore robust nonparametric spectral estimation of the electromagnetic spectrum is one application generated by the confluence of spectral estimation theory and robust estimation theory; and it is the focus of this thesis.

One potential application for robust nonparametric spectral estimation is using robust ED for CR. CR is one proposed solution for the spectrum scarcity problem. Traditionally, the Federal Communications Commission (FCC) has allocated static frequency bands for different wireless technologies. With the large growth rate of electrical communication systems, spectrum scarcity is a problem. However, studies have shown that the spectrum is often idle in space, time, or frequency. One method to resolve these inefficiencies is enabling Dynamic Spectrum Access (DSA) via CR [11] .

DSA allows each band to remain allocated for its Primary Users (PUs), while devices from Secondary Users (SUs) are allowed to transmit in the PUs's band so long as the SUs do not interfere with the PUs. These SUs fill the spectrum holes created by idle PUs, thereby increasing the spectrum efficiency for each band [11]. One difficulty with this approach is that the SUs need to be able to detect the PUs. That is, the devices need to be able to sense when a PU is communicating in the same space, time, and frequency as the SU, so that the

SU can switch to another spectrum hole or cease transmission. One promising venue for implementing DSA is the TV band [12].

Because the environment in which a given CR is deployed is unknown, it is desirable to consider robust SS. Robust signal detection for CR has been attempted using Bayesian statistics [13] and kernel theory [14]. Partial results for attempting quick robust ED in CR using nonparametric methods based on the Welch method are found in [15] and the same idea was applied to geoscience in [3], but the author has found no prior attempt at increasing the robustness of the MTM.

Another application for robust nonparametric spectral estimation is PLC. There are many advantages to being able to transmit information through power lines in addition to power. Applications such as remote metering, redundant protection communication between substations, and networking household devices are three realized applications [16].

Unfortunately PLC suffers from caustic noise due to the absence of communications as a design parameter for power lines. PLC noise is generally categorized into the following five categories: (1) low power colored background noise; (2) NarrowBand (NB) interference from broadband radio stations; (3) periodic IN asynchronous to mains caused by switched-mode power supplies; (4) periodic IN synchronous to mains caused by switching rectifier diodes in electrical appliances; and (5) asynchronous IN caused by switching from power network equipment [17], [18].

Various modulation methods have been proposed to effectively transmit within the PLC environment [19], [20]. However, to design an effective modulation scheme for PLC, the noise statistics must be accurately estimated. Since some PLC noise is synchronous to the mains, the noise statistics vary with the main cycle and cyclostationary noise models have been suggested [21]. A cyclostationary model based on experimental data has been proposed

in [22] where the noise follows a Gaussian Probability Distribution Function (PDF) whose variance is periodic with the mains cycle. Interestingly, if the model in [22] is sampled randomly, the resulting set of samples follows the Middleton PDF. The Nakagami PDF has also been used to model PLC noise [18]. It is noted in [21] that since the characteristics of PLC noise vary across location and time of day, nonparametric spectral estimation methods (which generalize better than parametric methods) are preferred over parametric methods.

From a signal processing perspective, IN and its effects can be processed in the time-domain or the frequency-domain. Often the time-domain is used, unless the goal is to correct signal estimation bias [23], which is one of the central goals of this thesis. IN is characterized by the pulse amplitude, pulse width, and inter-arrival time [18]. It is important to understand how an estimation technique is affected by these three parameters. The Middleton Class A model used in this thesis has been used to model PLC IN [24], [25], and the model can generate IN with varying amplitude and inter-arrival time—though it requires generalization to a time correlated model to vary the pulse width [26].

1.1 Outline and Contributions

Chapter 2 contains an overview of the Middleton Class A noise model. This is a flexible NGN model that will be used to test performance of the robust nonparametric spectral estimation methods proposed. The model encompasses a range of impulsiveness and deviations from the Gaussian distribution.

Chapter 3 contains an overview of PSD estimation methods, overviews of the Welch method and the MTM, and the description of the robust versions of these methods presented in this thesis. In this chapter, the ways in which the Welch method and MTM differ from other spectral estimation methods, and why they are chosen as the methods to adapt for

increased robustness is explained.

Chapter 4 contains the simulation results that demonstrate the performance of the robust PSD methods presented in this thesis, and how they compare to their respective non-robust versions. In addition to robustness, there are some results and discussion about estimation characteristics such as variance, bias, and computational complexity.

Chapter 5 contains concluding remarks and a summary of the contributions.

Appendix A supplements Chapter 2 with detailed mathematical derivations and explanations for understanding the Middleton Class A noise model.

Appendix B supplements Chapter 3 by showing the Wiener-Khinchin theorem. The Wiener-Khinchin theorem is a fundamental theorem that shows the equivalence between two popular mathematical definitions of the PSD.

Chapter 2

The Middleton Class A Noise Model

Accurate noise estimation is imperative for designing optimal communication systems. Specifically, a receiver can leverage statistical characteristics of both a transmitted signal and noise in order to drive the Bit Error Rate (BER) lower. Often noise is assumed to be Additive White Gaussian Noise (AWGN). In most environments, and especially tame environments, this is a good assumption because the central limit theorem applied to thermal noise results in noise that is Gaussian distributed. However, the noise in some environments cannot be accurately modeled as AWGN. Noise is often caused by noise sources external to the receiver, and simple examples include co-channel interference [27], snapping shrimp [28], gas emissions [29], microwave ovens [30], and various Power Line Communication (PLC) noise [17]. Since there is no reasonable chance that the PLC environment will attain a calmer noise profile, the PLC environment is a good motivation for properly modeling Non-Gaussian Noise (NGN). The discussion here primarily focuses on noise related to the PLC application.

One common method for modeling the manifestation of NGN is to select an NGN model—such as one of David Middleton's (Class A, Class B, or Class C) models [29],

the Symmetric α-Stable model [31], or even a simple two-term Gaussian mixture model (the Middleton Class A model is an infinite term Gaussian mixture model)—and tune the available parameters to fit the experimental statistics of the given environment. In the context of PLC, the literature tended to favor Class A noise because Class A noise is more tractable than Class B and Class C but is more tunable and adaptable than the Gaussian mixture model. The Symmetric α-Stable model does not seem to be as popular as Class A—specifically in the PLC context.

2.1 Definition of Class A Noise

Class A noise is designed to model NarrowBand (NB) NGN. Class A noise can be fully defined by using statistical distributions to represent the complex envelope E and phase Φ which are assumed to be independent. The probabilistic distribution of these Random Variables (RVs), as defined by Middleton [29], is shown below. Note that $f_E(\epsilon)$ is shown to be a legitimate Probability Distribution Function (PDF) in Appendix A.1, while $f_\Phi(\phi)$ is legitimate by inspection (that is, $f_\Phi(\phi)$ is non-negative and integrates to one).

$$\begin{aligned} E \sim f_E(\epsilon) &\triangleq \sum_{m=0}^{\infty} P_m \frac{\epsilon}{\sigma_m^2} e^{-\frac{\epsilon^2}{2\sigma_m^2}}, \quad \epsilon \in [0,\infty) \\ \Phi \sim f_\Phi(\phi) &\triangleq \frac{1}{2\pi}, \quad \phi \in [0,2\pi) \end{aligned} \tag{2.1}$$

The parameters for (2.1) are defined as follows:

$$\begin{aligned} P_m &\triangleq \frac{e^{-A}A^m}{m!} \\ 2\sigma_m^2 &\triangleq \sigma^2 \frac{m/A + \Gamma}{1+\Gamma} \\ \sigma^2 \in [0,\infty), \quad A \in [0,\infty), \quad \Gamma &\in [0,\infty) \end{aligned} \tag{2.2}$$

The three fundamental parameters that fully constrain the model are described below.

- σ^2 (total power): The overall power of the narrowband noise.
- Γ (Gaussian factor): The ratio of the underlying Gaussian Noise (GN) power and the excess noise power (i.e., the NGN power).
- A (overlap index): The impulsiveness of the noise. A larger A results in noise statistics that are more impulsive, while a smaller A results in noise statistics that are closer to purely Gaussian; as A approaches 0 the model becomes purely jointly complex Gaussian.

To gain a better understanding of the model, these three parameters will be discussed in detail in the following three sections.

2.2 Overall Power of Class A Noise

In the author's opinion, it is extraordinarily useful to develop a relationship between the envelope E of Class A noise and the Rayleigh RV. Understanding this relationship is helpful when writing a program to generate Class A noise or (as shown in this section) for deriving the overall power of the Class A model. Recall that the Rayleigh RV characterizes the complex envelope of purely jointly complex GN. The definition of a Rayleigh RV is shown below.

$$R \sim f_R(r;\sigma) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}, \quad r \in [0,\infty) \tag{2.3}$$

Note that the power of a Random Process (RP) generated from a Rayleigh RV (the second moment) is $2\sigma^2$ as shown in Appendix A.2.

It is helpful to view the PDF of the envelope of Class A noise ($f_E(\epsilon)$) as a infinite convex combination of the PDFs of Rayleigh RVs with differing powers (σ_m^2). By defining an infinite number of Rayleigh RVs $R_m \sim f_R(r;\sigma_m)$, (2.1) can be rewritten as (2.4) and consequently as (2.5) by applying the proof shown in Appendix A.3.

$$E \sim f_E(\epsilon) = \sum_{m=0}^{\infty} P_m f_R(\epsilon;\sigma_m), \quad \epsilon \in [0,\infty) \tag{2.4}$$

$$E = R_m \quad \text{w.p.} \quad P_m \tag{2.5}$$

Since the second moment of a Rayleigh RV is known, the second moment of E (the total power of Class A noise) can easily be computed by taking advantage of the relationship between the two RVs (this is shown in Appendix A.4). Therefore, the power of Class A noise can be written as:

$$\mathbb{E}\left\{E^2\right\} = \sum_{m=0}^{\infty} P_m 2\sigma_m^2 \tag{2.6}$$

Equation (2.6) can be simplified by substituting the definition of P_m and σ_m^2 as defined by Middleton (shown in (2.2)). See Appendix A.5 for the detailed calculation. The result is:

$$\mathbb{E}\left\{E^2\right\} = \sigma^2 \tag{2.7}$$

The latter result demonstrates the significance of how Middleton defined his parameters. He defines P_m and σ_m^2 in such a way that regardless of the overlap index (A) and the Gaussian factor (Γ), the overall power of a RP generated by the Class A RV is σ^2. Note that this is *not* the variance of E. Variance is the second *central* moment and is expressed for E as $\mathbb{E}\left\{(E - \mathbb{E}\{E\})^2\right\}$. The power of a RV is often computed using the variance. More precisely, however, the power equals the second moment; if the mean of a RV is zero, then the variance (the second central moment) reduces to the second moment (otherwise the variance is not

equal to the power). Variance is also often a useful parameter for understanding the scale of a RV. But for distributions like the Rayleigh or the Class A, the second moment is a more useful parameter of scale. Therefore, the variance of the Class A RV is not discussed in this thesis.

2.3 Gaussian Factor

The Class A model is designed to model a mixture of GN and NGN. Every receiver adds a layer of GN produced by physical phenomena such as thermal noise. NGN usually results from an external source and is additive to the GN. Looking at (2.5) one can observe that the envelope of Class A samples is obtained by drawing from an infinite set of Rayleigh RVs with differing power. By noting that the power of the Rayleigh RVs increases with m (2.2), intuition can be gained regarding the physical significance of R_0—the lowest power and most likely Rayleigh RV to be drawn. This is intended to model the *underlying* Gaussian process (recall that a complex Gaussian RV has a Rayleigh distributed envelope). Each R_m for $m \neq 0$ models the underlying GN *and* the additional power from a NGN process.

With this interpretation of the Class A model, the power contributed by the underlying GN process is merely the power of R_0 which is:

$$\sigma_G^2 \triangleq \mathbb{E}\left\{R_0^2\right\} = 2\sigma_0^2 = \sigma^2 \frac{\Gamma}{1+\Gamma} \tag{2.8}$$

where σ_G^2 is the overall power of the GN process. Any additional power generated from Class A samples is attributed to the power generated by the NGN, as defined in (2.9)

$$\Omega_{NG} \triangleq \sigma^2 - \sigma_G^2 = \sigma^2 \frac{1}{1+\Gamma} \tag{2.9}$$

Algebraic manipulation then shows that the ratio of the power contributed by the GN (σ_G^2) to the power contributed by the NGN (Ω_{NG}) is Γ. That is:

$$\frac{\sigma_G^2}{\Omega_{NG}} = \Gamma \tag{2.10}$$

With this understanding of Γ, new insight about σ_m^2 can be gained. Multiplying the definition of $2\sigma_m^2$ in (2.2) by Ω_{NG}/Ω_{NG} yields:

$$2\sigma_m^2 = \frac{m}{A}\Omega_{NG} + \sigma_G^2 \tag{2.11}$$

This form plainly shows that the power of the R_0 Rayleigh RV is σ_G^2, and thereby represents the power of the underlying GN. Every R_m for $m > 0$ increases linearly in power from the Gaussian base according to the scaling factor Ω_{NG}/A.

2.4 Overlap Index

Unlike the overall power, σ^2, and the Gaussian Factor, Γ, an intuitive understanding of the Overlap Index cannot be gained as easily from analytic and quantitative analysis of A. However, the Overlap Index can be analyzed in light of how it qualitatively affects two properties of the Class A model: (1) the probabilistic distribution controlling which Rayleigh RVs are selected as seen in (2.5) and (2) the power of these Rayleigh RVs.

1. Looking at how the Rayleigh RVs are selected according to P_m in (2.5) and how P_m is partially defined in terms of A in (2.2), one can begin to analyze how A affects the selection of the Rayleigh RVs and consequently the impulsiveness of Class A noise. As A increases, the relative weight of P_m for large m increases compared to the relative

weight of P_m for small m. This results in the selection of higher power Rayleigh RVs more often which aids in increasing the impulsiveness.

2. As mentioned previously, the Overlap Index also affects the power of the mth Rayleigh RV. This can most clearly be seen in (2.2). As A increases, the power of the mth Rayleigh RV ($2\sigma_m^2$) decreases.

From the analysis of the overall power, it was seen that A does not affect the overall power. Therefore, even as the Overlap Index adjusts the weight of the selection parameter P_m, it simultaneously adjusts the power of the mth Rayleigh RVs to maintain a constant overall power.

2.5 In-Phase and Quadrature Representation

Often Class A noise is defined and analyzed in its In-phase and Quadrature (IQ) form rather than Complex Envelope and Phase (CEP) [26], [32]–[36]. In Middleton's first paper defining Class A noise [29], he begins with the CEP form as done here. The primary reason for this is that unlike purely complex GN, the IQ components of Class A noise are *dependent* [37]. This property is very relevant for simulations involving synthetically generated Class A noise because direct generation of Class A noise in its IQ representation becomes very difficult. It is instead best to generate the CEP samples independently and then map these samples to IQ samples (which are dependent).

Because the IQ representation for Class A noise is so prevalent in the literature and occasionally useful, the process of mapping from CEP to IQ will be shown followed by analysis of the IQ distributions. Since E and Φ are independent [29], the joint PDF can be

expressed as follows:

$$f_{E,\Phi}(\epsilon,\phi) = f_E(\epsilon)f_\Phi(\phi) = \sum_{m=0}^{\infty} P_m \frac{\epsilon}{2\pi\sigma_m^2} e^{-\frac{\epsilon^2}{2\sigma_m^2}} \tag{2.12}$$

Using the following variable mapping, the IQ representation can be computed.

$$\begin{aligned} X &= E\cos(\Phi) \\ Y &= E\sin(\Phi) \end{aligned} \tag{2.13}$$

This mapping yields the following joint PDF.

$$\begin{aligned} f_{X,Y}(x,y) &= \sum_{m=0}^{\infty} P_m \frac{1}{2\pi\sigma_m^2} e^{-\frac{x^2+y^2}{2\sigma_m^2}} \\ x \in (-\infty,\infty), &\quad y \in (-\infty,\infty) \end{aligned} \tag{2.14}$$

Integrating out X or Y yields the marginal PDFs shown below.

$$\begin{aligned} X \sim f_X(x) &= \sum_{m=0}^{\infty} P_m \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{-\frac{x^2}{2\sigma_m^2}} \\ Y \sim f_Y(y) &= \sum_{m=0}^{\infty} P_m \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{-\frac{y^2}{2\sigma_m^2}} \end{aligned} \tag{2.15}$$

Note that the detailed calculations of this variable transform are given in Appendix A.6.

Though previously stated, it is important to note that X and Y are uncorrelated yet dependent if one desires to generate IQ Class A noise samples, so a more detailed explanation of this property is merited. By definition, two RVs are uncorrelated if their covariance (definition shown in (2.16)) is zero.

$$\mathrm{Cov}(X,Y) \triangleq \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] \tag{2.16}$$

Two RVs are defined as being independent if their joint PDF is equal to the product of their marginal PDFs (note that an equivalent definition uses CDFs instead of PDFs). It can be seen that X and Y are dependent by observing that the joint PDF in (2.14) is not equal to the multiplication of the marginal PDFs in (2.15). Appendix A.7 explicitly shows that X and Y are uncorrelated as well as showing their dependence by giving a more rigorous demonstration of the aforementioned inequality.

Chapter 3

Spectral Estimation Methods

Power spectral estimation is succinctly defined in the following statement: "From a finite record of a stationary data sequence, estimate how the total power is distributed over frequency [38]." In the context of this thesis, the "stationary data sequence" will be a digitized communications signal modulated with Quadrature Phase-Shift Keying (QPSK) in the presence of various types of noise. The "finite record" will be a snapshot of this noisy signal.

3.1 Overview of Spectral Estimation

All spectral analysis methods are either *parametric* or *nonparametric*. Parametric methods assume a particular model for a data sequence, and then estimate the parameters of the assumed model. This method is accurate when (1) the data sequence can be closely approximated by the parameterized model (across the span of possible parameters) and (2) the best parameters can be accurately determined. Nonparametric (or classical) methods do

not assume a model but attempt to estimate the Power Spectral Density (PSD) of a sequence, which yields a desirable visual representation of how the data's power is distributed according to frequency. Nonparametric methods only assume the signal is Wide-Sense Stationary (WSS), and are therefore more blind—which is ideal when there is no *a priori* knowledge about the signal. Since blind estimation is the concern of this thesis, nonparametric methods are investigated. It should be noted that QPSK is used in this thesis as a stand-in for a generic unknown WSS signal; none of the unique QPSK signal features are utilized during estimation. Analysis of how different WSS signals affect the nonparametric spectral estimation methods discussed in this thesis is briefly performed in Chapter 4.

A PSD is a function of frequency whose magnitude represents signal power; the PSD depicts the prominent (and absent) oscillations in a data sequence. Mathematically, the PSD of a Random Process (RP) can be defined in more than one way [39]. The mathematical definition that best captures the physical description of the PSD is given by (3.1) and will be referred to as the *intuitive* definition of the PSD.

$$\phi^{(\mathrm{i})}(f) \triangleq \lim_{M\to\infty} \mathbb{E}\left\{ \frac{1}{2M+1} \left| \sum_{n=-M}^{M} x[n] e^{-j2\pi f n} \right|^2 \right\}, \quad f \in \left[-\frac{1}{2}, \frac{1}{2}\right] \tag{3.1}$$

This definition can be interpreted as applying the following operations working from $x[n]$ outward:

1. Discrete-Time Fourier Transform (DTFT): convert time-domain samples into a frequency-domain function
2. Magnitude squared: convert a frequency-domain function into energy across frequency
3. Normalize: convert energy into average power
4. Expectation: convert the PSD of *one* stochastic realization into a PSD that

characterizes an "expected" realization based on the *statistics* of the RP

5. Limit: convert the PSD of the truncated RP into a PSD of the true RP (because RPs are generally considered to have infinite support in the time domain).

However, most often[1] the PSD of an RP is defined by (3.2), which is mathematically equivalent to (3.1) if the AutoCorrelation Function (ACF) (3.2a) decreases sufficiently fast [40].

$$r[\ell] \triangleq \mathbb{E}\{x[n]x^*[n-\ell]\} \tag{3.2a}$$

$$\phi^{(\mathrm{wk})}(f) = \sum_{\ell=-\infty}^{\infty} r[\ell]e^{-j2\pi f\ell}, \quad f \in \left[-\frac{1}{2}, \frac{1}{2}\right] \tag{3.2b}$$

The theorem that proves (3.1) and (3.2b) are equal is known as the Wiener-Khinchin theorem (sometimes referred to as the Wiener-Khinchin-Einstein theorem). It was first stated in a two-page sketch by Albert Einstein in 1914 [41], then proved for a deterministic signal by Norbert Wiener in 1930 [42], [43], and finally proved for a stochastic signal by Aleksandr Khinchin in 1934 [44][2]. This theorem is proved in Appendix B.1. Somewhat confusingly, the particular equation (3.2b) is also referred to as the Wiener-Khinchin theorem.

Nonparametric methods are further categorized into *periodogram* methods and *correlogram*[3] methods [38]. These methods estimate the PSD using *one finite* realization of an RP. Reducing the intuitive PSD definition (3.1) by applying the restriction of one length N finite realization (i.e., removing the limit and expectation) yields the periodogram

[1] In the literature, (3.2) is a more popular PSD definition than (3.1). For example, the first equation in Kay's book on spectral estimation [40] is (3.2) (using different notation).

[2] Though deterministic signals are of no interest here, it should be noted that the theorem also works for the deterministic case. However, the deterministic case requires slightly different definitions for both the ACF and the "intuitive" form of the PSD.

[3] The correlogram method is more commonly referred to as the Blackman-Tukey method—first developed in [45].

estimate (3.3b) [38], [40].

$$X(f) = \sum_{n=0}^{N-1} x[n]e^{-j2\pi fn} \tag{3.3a}$$

$$\hat{\phi}^{(\mathrm{p})}(f) = \frac{1}{N}|X(f)|^2 \tag{3.3b}$$

Similarly, reducing the Wiener-Khinchin PSD definition (3.2) in the same manner as above yields the correlogram estimate (3.4b) which is equal to (3.3) under weak conditions [38].

$$\hat{r}[\ell] = \begin{cases} \frac{1}{N}\sum_{n=\ell}^{N-1} x[n]x^*[n-\ell] & \text{for} \quad \ell = 0,\, 1,\, \ldots,\, N-1 \\ \hat{r}^*[-\ell] & \text{for} \quad \ell = -(N-1),\, -(N-2),\, \ldots,\, -1 \end{cases} \tag{3.4a}$$

$$\hat{\phi}^{(\mathrm{c})}(f) = \sum_{\ell=-(N-1)}^{N-1} \hat{r}[\ell]e^{-j2\pi f\ell} \tag{3.4b}$$

For both (3.3) and (3.4), $n = 0, \ldots, N-1$ and $f \in \left[-\frac{1}{2}, \frac{1}{2}\right]$.

Fully discrete computation is common for spectral estimation due to the digital nature of computers and the power of the Fast Fourier Transform (FFT). When applying this computation constraint to the computation of the periodogram and correlogram, the DTFT operations are replaced with FFT operations by replacing the continuous variable f with the discrete variable k using $f = \frac{k}{N}$. The variable k represents the frequency bins, whose widths are determined by the sampling frequency and the FFT length. The discrete periodogram and correlogram are shown in (3.5b) and (3.6)

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi kn/N} \tag{3.5a}$$

$$\hat{\phi}^{(\mathrm{p})}[k] = \frac{1}{N}|X[k]|^2 \tag{3.5b}$$

$$\begin{aligned}\hat{\phi}^{(\mathrm{c})}[k] &= \sum_{\ell=-(N-1)}^{N-1} \hat{r}[\ell] e^{-j2\pi k\ell/N} \\ &= \sum_{\ell=0}^{N-1} \hat{r}[\ell] e^{-j2\pi k\ell/N} + \sum_{\ell=0}^{N-1} \hat{r}^*[\ell] e^{j2\pi k\ell/N} - \hat{r}[0] \\ &= 2\,\mathrm{Re}\left\{ \sum_{\ell=0}^{N-1} \hat{r}[\ell] e^{-j2\pi k\ell/N} \right\} - \hat{r}[0]\end{aligned} \tag{3.6}$$

where $k = 0, \ldots, N-1$.

The periodogram based methods are generally preferred because (1) the periodogram is easier to compute[4] and (2) the correlogram's spectral window is the Discrete Fourier Transform (DFT) of the data window *convolved with itself*, whereas the periodogram's spectral window is the DFT of only the data window—thereby constituting an easier design process for the periodogram's spectral window. Any PSD estimator that utilizes the periodogram belongs to the periodogram family of estimators. The simplest in this family is the periodogram (3.5). Unfortunately the periodogram suffers from severe noise constraints which cannot be reduced by changing parameters [40], so other PSD estimators with better noise mitigation are often use.

More complex methods based on the periodogram have been designed to help mitigate noise interference. These methods include the Daniell method [38], Bartlett method [46], [47], Welch method[5] [7], and Multitaper method [8], [9]. The Daniell method is the oldest of the four, not well-known, and will not be analyzed[6]. The Welch method is a generalized version of the Bartlett method, and probably the most popular nonparametric method after

[4]To compute the correlogram, one must compute a correlation (3.4a) (which would be performed directly or with two FFTs depending on the length) and an FFT (3.6). To compute the periodogram, one must only compute a single FFT (3.5a) and then square the magnitude (3.5b).

[5]Sometimes referred to as Welch's/Weighted Overlapped Segment Averaging (WOSA) [9], [39], [48].

[6]The Daniell method first computes a length $\tilde{N}$ periodogram by zero padding a length N data sequence (where $\tilde{N} \gg N$). Utilizing the fact that the frequency bin estimates are asymptotically uncorrelated, an averaging window is then used to reduce the noise variance. The window size is sufficiently small such that the power estimate is nearly constant across a set of windowed bins.

the standard periodogram. The Multitaper method is the most recent of the four and is by far the most theoretically and computationally complex. In this thesis, both the Welch and the Multitaper method will be analyzed and modified to increase robustness.

3.2 The Welch Method

A noise mitigating PSD estimation approach was designed by Peter Welch in 1967 and can be found in [7]. Understanding Welch's method is easiest when viewed as a generalized version of the Bartlett method. The Bartlett method decreases noise variance in its PSD estimate by computing the periodogram for each consecutive subsequence in a data sequence. These periodograms are then averaged. If the signal of interest and interfering noise are second-order stationary processes (and the interfering noise samples are uncorrelated), then the averaging operation decreases the noise variance.

Peter Welch generalized Bartlett's method in two ways: (1) he incorporated different spectral windows (Bartlett used only the rectangular window) and (2) he allowed for overlapping subsequences rather than consecutive subsequences.

3.2.1 Overview of Welch Method

In Peter Welch's original paper [7], the time series data, $x[n]$, of length N is parsed into M overlapped segments, $x_1[n]$, $x_2[n]$, ..., $x_M[n]$, each with length L. Each segment overlaps its adjacent segment(s) by $L - D$ samples. A graphic depicting this setup is shown in Figure 3.1. These time series segments are then windowed before their corresponding windowed periodograms are computed. Since for this thesis the analysis of multiple signals—near in frequency and with differing power—is not of interest, the windowing term can be dropped

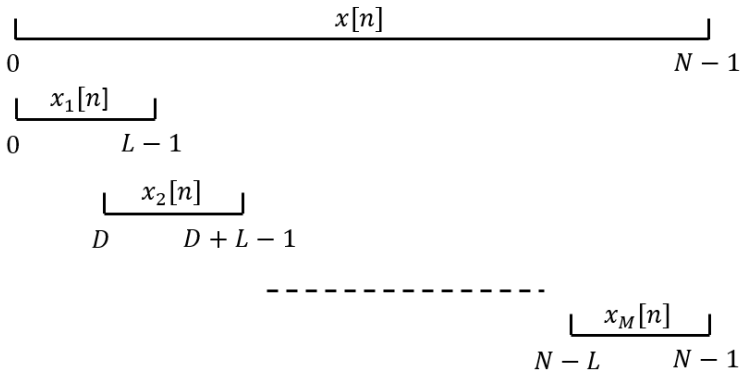

Figure 3.1: A recreation of Welch's original graphic depicting the setup of his method. Rather than taking one DFT of a time series, he breaks the series into multiple smaller segments, computes the DFTs, and then averages.

(which results in the use of a rectangular window) with negligible effect on the robust methods presented later. Furthermore, the relative effective signal to noise power (as opposed to absolute power levels) is the main concern in Energy Detection (ED) because thresholding is often used to determine if a signal is present. Therefore, the normalizing terms have been omitted and (unless explicitly stated) every final PSD estimate shown in this thesis is normalized such that the maximum bin value is 0 decibels without loss of generality. The final simplified Welch PSD estimate is computed by averaging the standard periodograms (3.7)

$$\hat{\phi}^{(\mathrm{w})}[k] = \frac{1}{M} \sum_{m=1}^{M} \hat{\phi}_m^{(\mathrm{p})}[k] \tag{3.7}$$

where $\hat{\phi}_m^{(\mathrm{p})}[k]$ is the periodogram (3.5b) of the mth time series segment that comprises the samples $x_m[n]$ for $n = 0, \ldots, L-1$.

The Welch method accurately estimates signals in the presence of Additive White Gaussian Noise (AWGN). To demonstrate the accurate performance of the Welch method, the performance of the simplest PSD estimation method, the periodogram (3.3), will be shown first. Figures 3.2 and 3.3 show the periodogram estimator for 256 and 8192 samples

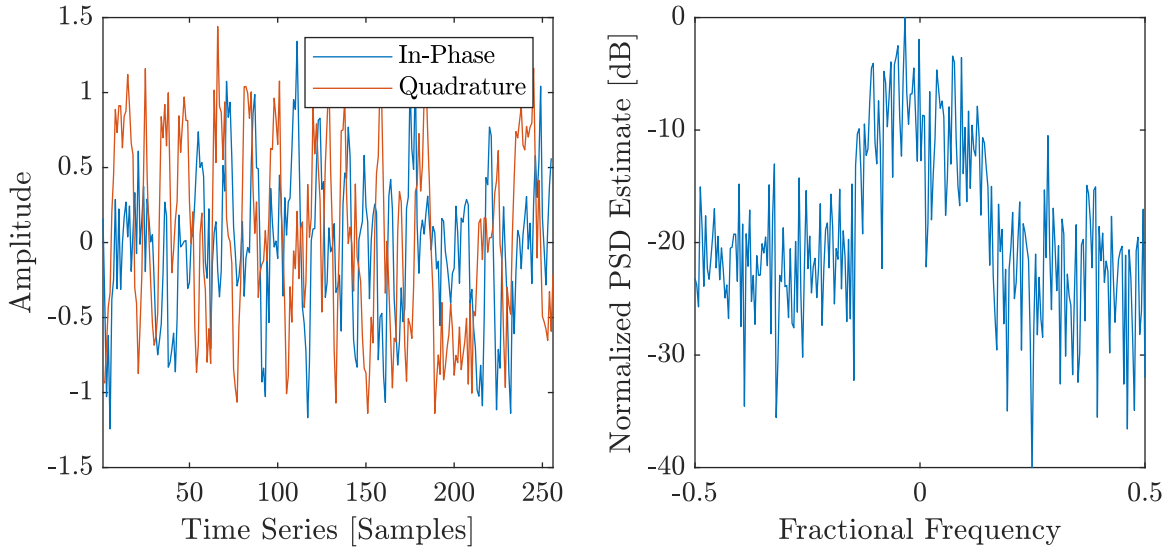

Figure 3.2: Example of the simple periodogram method for a 256 sample complex baseband QPSK signal in AWGN at E_b/N_0 of 10 decibels. Even at E_b/N_0 of 10 decibels the variance of the noise highly distorts the signal when using the simple periodogram method.

of a QPSK signal in AWGN. Though the PSD estimate with more samples produces a slightly clearer visual representation of the signal, mathematically the noise variance does not decrease as the number of samples increases [38].

To contrast, the low variance and clearly defined signal bandwidth for the Welch method can be seen in Figure 3.4 where the FFT size (window length), L, is the same as in Figure 3.2 but the total number of samples, N, is the same as in Figure 3.3. The overlap (here and for the remainder of this thesis) is $D - L = L/2$ because the literature most commonly uses 50% overlap [49]–[55]. These parameters result in $M = 62$ averaged periodograms (which is an M roughly one order of magnitude larger than the literature tends to use [50], [51], but it is nevertheless illustrative).

Now the statistical performance of the Welch method is analyzed. Assume that a received signal $r[n]$ comprises the stochastic transmitted signal $s[n]$ that is contaminated by AWGN

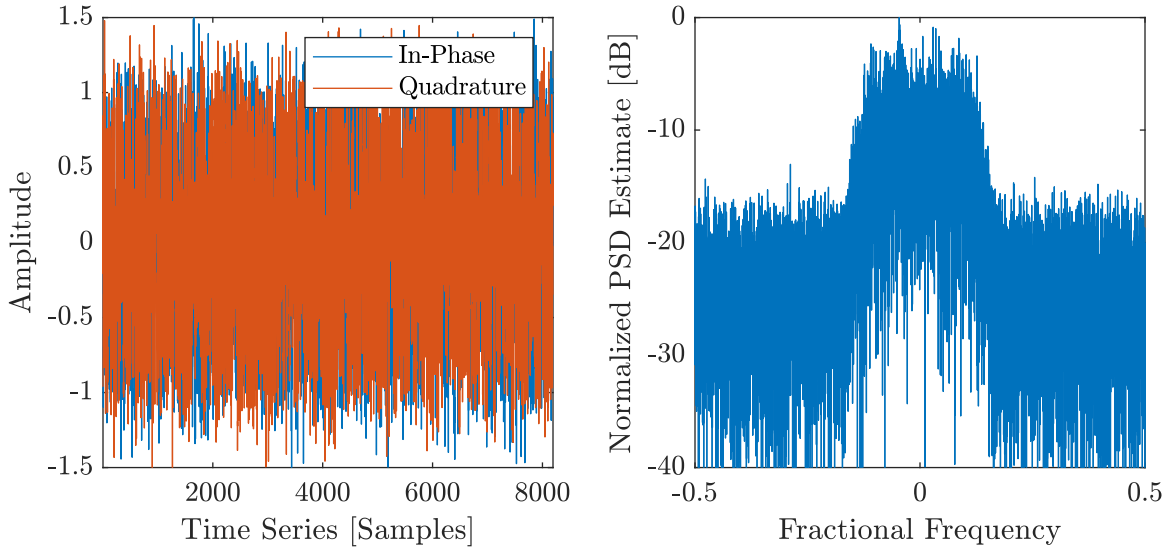

Figure 3.3: Example of the simple periodogram method for a complex baseband QPSK signal in AWGN at E_b/N_0 of 10 decibels. Even with a factor of 64 as many samples as in Figures 3.2, the variance caused by the noise does not decrease.

$a[n]$, so that the received signal is expressed as:

$$r[n] = s[n] + a[n] \tag{3.8}$$

The received signal is partitioned into M overlapping segments denoted:

$$r_m[n] = s_m[n] + a_m[n] \tag{3.9}$$

The stochastic process $a[n]$ is a wide sense stationary complex Gaussian RP with zero-mean and variance σ_a^2, ($a[n] \sim \mathcal{N}(0, \sigma_a^2)$ for each n). Keeping in mind that the DFT is a linear operator, the Fourier transform of each $r_m[n]$ is $R_m[k] = S_m[k] + A_m[k]$, where $S_m[k]$ is the DFT of the transmitted signal's mth segment and $A_m[k]$ is still a complex Gaussian RP with

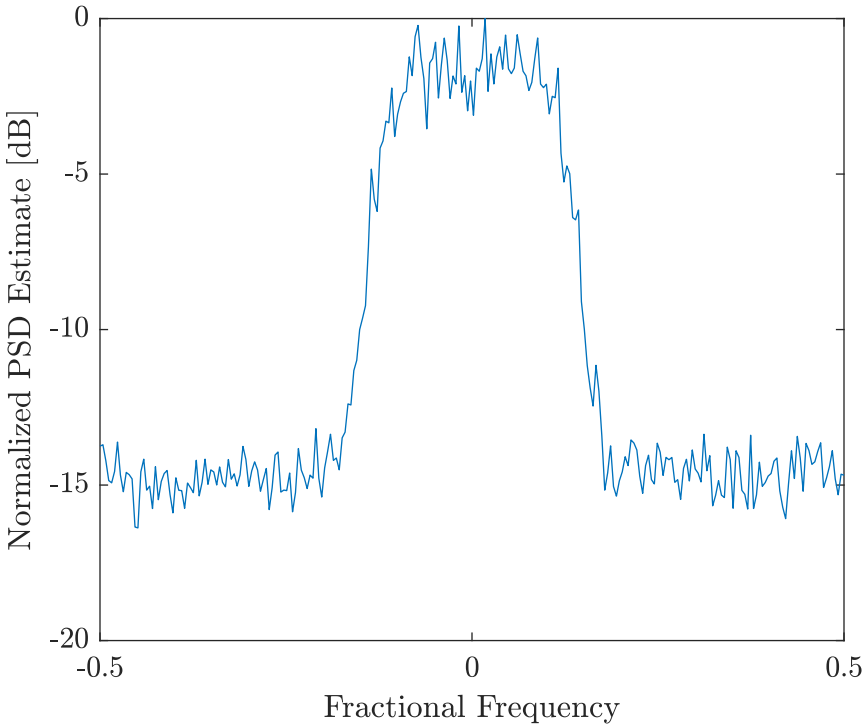

Figure 3.4: Example of the Welch method for a complex baseband QPSK signal in AWGN at E_b/N_0 of 10 decibels. By comparing Figure 3.3 with this figure, one can observe the low variance estimate attainable via the Welch estimator since the same number of QPSK samples is used to produce both PSD estimates.

zero-mean and variance σ_A^2. Using (3.3), the periodogram of $r_m[n]$ is expressed by (3.10)

$$\begin{aligned}\hat{\phi}_m^{(\mathrm{p})}[k] &= |S_m[k] + A_m[k]|^2 \\ &= (\mathrm{Re}[S_m[k]] + \mathrm{Re}[A_m[k]])^2 + (\mathrm{Im}[S_m[k]] + \mathrm{Im}[A_m[k]])^2 \\ &= |S_m[k]|^2 + |A_m[n]|^2 \\ &\quad + \underbrace{2(\mathrm{Re}[S_m[k]] \cdot \mathrm{Re}[A_m[k]] + \mathrm{Im}[S_m[k]] \cdot \mathrm{Im}[A_m[k]])}_{C_m[k]}\end{aligned} \tag{3.10}$$

where $|S_m[k]|^2$ is the energy of the transmitted signal for the mth segment as a function of frequency, $|A_m[k]|^2$ is the energy of the AWGN for the mth segment as a function of frequency, and $C_m[k]$ is the correlation term. Assuming that the signal of interest persists

throughout the entire set of N samples in a given Welch estimate, then the averaged $\hat{\phi}_m^{(\mathrm{p})}[k]$ terms in (3.7) will be statistically equivalent for any mth index across all periodograms. The noise variation comes from the $|A_m[k]|^2$ and $C_m[k]$ terms.

It can be shown that $|A_m[k]|^2$ is a scaled Chi-Squared RP with two degrees of freedom. The Chi-Squared Random Variable (RV) is defined in (3.11) as $Q \sim \chi_p^2$ and is equal to the sum of the square of p Independent and Identically Distributed (IID) standard normal Gaussian RVs ($Z_i \sim \mathcal{N}(0,1)$). The value of p represents the degrees of freedom of the Chi-Squared distribution, which is the number of IID Z_i^2 summed. The expected value of the Chi-Squared distribution is also shown in (3.11) for general p degrees of freedom.

$$Q \sim \chi_p^2, \quad Q = \sum_{i=1}^{p} Z_i^2, \quad \mathbb{E}[Q] = p \tag{3.11}$$

Using the definition of a Chi-Squared RV, the RP $|A_m[k]|^2$ is shown to be a scaled Chi-Squared RP with two degrees of freedom in (3.12).

$$\begin{aligned} |A_m[k]|^2 &= \mathrm{Re}[A_m[k]]^2 + \mathrm{Im}[A_m[k]]^2 \\ &= G_{\mathrm{Re}}^2[k] + G_{\mathrm{Im}}^2[k] \\ &= \sigma_A^2\, Z_{\mathrm{Re}}^2[k] + \sigma_A^2\, Z_{\mathrm{Im}}^2[k] \\ &= \sigma_A^2\, \tilde{Q}[k] \end{aligned} \tag{3.12}$$

where

$$\begin{aligned} G_{\mathrm{Re}}[k], G_{\mathrm{Im}}[k] &\sim \mathcal{N}(0, \sigma_A^2) \\ Z_{\mathrm{Re}}[k], Z_{\mathrm{Im}}[k] &\sim \mathcal{N}(0,1) \\ \tilde{Q}[k] &\sim \chi_2^2 \end{aligned} \tag{3.13}$$

Assuming $s_m[n]$ and $a_m[n]$ are independent, then $C_m[k]$ follows a zero-mean Gaussian distribution because $\mathrm{Re}[A_m[k]]$ and $\mathrm{Im}[A_m[k]]$ are independent zero-mean Gaussian RVs and both $\mathrm{Re}[S_m[k]]$ and $\mathrm{Im}[S_m[k]]$ are stationary. Therefore, $C_m[k]$ is a linear combination of independent zero-mean Gaussian RVs which yields a zero-mean Gaussian RV. The expectation of the traditional Welch PSD estimate of a signal in the presence of AWGN can now be computed as shown in (3.14).

$$\begin{aligned}\mathbb{E}\left\{\hat{\phi}_m^{(\mathrm{w})}[k]\right\} &= \mathbb{E}\left\{|S_m[k]|^2 + |A_m[k]|^2 + C_m[k]\right\} \\ &= \mathbb{E}\left\{|S_m[k]|^2\right\} + \mathbb{E}\left\{|A_m[k]|^2\right\} + \mathbb{E}\left\{C_m[k]\right\} \\ &= \mathbb{E}\left\{|S_m[k]|^2\right\} + 2\sigma_A^2\end{aligned} \tag{3.14}$$

The first line is obtained by taking the expectation of (3.10). The second line utilizes the fact that the expectation operator is linear. The third line utilizes the insight gained from (3.11) and (3.12) as well as the previous discussion of $C_m[k]$.

The Welch method takes the sample mean across multiple periodograms by their corresponding indices. As the number of averaged periodograms, M, increases, the Welch estimate approaches its expected value (3.14). The rate at which the estimate converges to the expected value can be computed using the asymptotic variance, which is a function of the estimation method (in this case, the sample mean) and the probability density function (in this case, the Gaussian and Chi-Squared distributions). To this end, (3.7) combined with (3.10) produce an expanded form of the Welch method shown in (3.15).

$$\begin{aligned}\hat{\phi}^{(\mathrm{w})}[k] &= \frac{1}{M}\sum_{i=1}^{M}\left[|S_m[k]|^2 + |A_m[k]|^2 + C_m[k]\right] \\ \hat{\phi}_{\mathrm{dB}}^{(\mathrm{w})}[k] &= 10\log_{10}(\hat{\phi}^{(\mathrm{w})}[k])\end{aligned} \tag{3.15}$$

A detailed analysis of the asymptotic variance of $\hat{\phi}^{(\mathrm{w})}[k]$ is beyond the scope of this thesis; however, it is worth mentioning that the sample mean is the optimal estimator of location for Gaussian distributed samples. This indicates that the asymptotic convergence of the averaged $C_m[k]$ tends to 0 is optimal under the traditional Welch method. In robust estimation theory parlance, the sample mean is the most *efficient* estimator for Gaussian distributed samples but it is not *robust*. In the next section, methods are explored for making the Welch estimator more robust to guard against more caustic noise environments.

3.2.2 Making Welch more Robust

As shown in the previous section, the Welch PSD estimation method works well under the assumption that a signal is in the presence of AWGN. However, using the traditional Welch method when a signal is in the presence of impulsive noise (such as Middleton Class A noise) causes performance degradation. One can think about this from a qualitative perspective: any periodogram section containing even a single large spike can be modeled as a linear combination of the signal of interest, AWGN, and an impulse. The DFT of an impulse is a constant. Therefore, any periodogram segment containing one or more noisy impulses in time (which will be termed "contaminated") results in a periodogram with a large constant offset. When all the segment periodograms are averaged, the contaminated ones can drastically bias the overall Welch estimate.

These contaminated periodograms produce outliers in the index-wise frequency bin distributions. Identifying and handling these outliers appropriately can recover signals in the PSD estimate that would be invisible with the traditional averaging of the Welch method. Two estimators of location are used to replace the index-wise averaging in the traditional Welch method: (1) the median, a highly robust and theoretically simple estimator and (2) the

Huber, a popular maximum-likelihood robust estimator that requires iterative computation. These new modified Welch methods are referred to as the "median-modified Welch" and the "Huber-modified Welch," respectively.

The median-modified Welch method is defined as follows: for a given Welch size of N samples, each mth individual periodogram is computed and placed row-wise in an $M \times L$ matrix $W_\phi[m,k]$ as shown in (3.16). One can also represent the matrix by the column vectors $\hat{B}_k[m]$ where each $\hat{B}_k[m]$ comprises the kth frequency bin values from each periodogram, $m = 1, \ldots, M$ (3.16).

$$W_\phi^{(\mathrm{p})}[m,k] = \begin{bmatrix} \hat{\phi}_1^{(\mathrm{p})}[k] \\ \hat{\phi}_2^{(\mathrm{p})}[k] \\ \vdots \\ \hat{\phi}_m^{(\mathrm{p})}[k] \\ \vdots \\ \hat{\phi}_M^{(\mathrm{p})}[k] \end{bmatrix} = \begin{bmatrix} \hat{B}_1^{(\mathrm{p})}[m] & \hat{B}_2^{(\mathrm{p})}[m] & \cdots & \hat{B}_k^{(\mathrm{p})}[m] & \cdots & \hat{B}_L^{(\mathrm{p})}[m] \end{bmatrix} \tag{3.16}$$

The number of rows is equal to the number of periodograms, M (i.e., $m = 1, \ldots, M$); the number of columns is equal to the length of each periodogram, L (i.e., $k = 1, \ldots, L$). Each column is then sorted (in either ascending or descending order—both equivalent for the purpose of locating the median) to produce $W_S[m,k]$ shown in (3.17).

$$W_S^{(\mathrm{p})}[m,k] = \mathrm{SORT}(W_\phi^{(\mathrm{p})}[m,k],\, m) \tag{3.17}$$

The median-modified Welch estimate is then produced by averaging the values at the two

middle indices (because L is even) of each sorted column vector (3.18).

$$\hat{\phi}^{(\mathrm{m})}[k] = \frac{1}{2}\left(W_S^{(\mathrm{p})}\left[\frac{M}{2}, k\right] + W_S^{(\mathrm{p})}\left[\frac{M}{2}+1, k\right]\right) \tag{3.18}$$

This is the median value for each $\hat{B}_k[m]$.

Starting from (3.16), the Huber-modified Welch applies the Iteratively Re-weighted Least-Squares (IRLS) algorithm to each frequency bin vector, $\hat{B}_k[m]$. For a given frequency bin column vector $\hat{B}_k[m]$, the values $b_1, \ldots, b_M$ represent the mth periodogram's frequency bin estimates (3.19a). The current residuals are denoted by $r_m^{(j)}$ and represent the difference between the measured values b_m and the current estimate of location $\hat{\theta}^{(j)}$ (3.19b). The sample median is used for the first estimate of location, $\hat{\theta}^{(1)}$. The estimate of location at iteration $j+1$ is recomputed iteratively by taking a weighted average of the current residuals $r_m^{(j)}$ (3.19c). The Huber weighting function used in (3.19c) is denoted by $q(r)$ and shown in (3.19d).

$$\hat{B}_k^{(\mathrm{p})}[m] = \begin{bmatrix} b_1 & b_2 & \cdots & b_m & \cdots & b_M \end{bmatrix}^T \tag{3.19a}$$

$$r_m^{(j)} = b_m - \hat{\theta}^{(j)} \tag{3.19b}$$

$$\hat{\theta}^{(j+1)} = \frac{\sum_{m=1}^{M} b_m \, q(r_m^{(j)})}{\sum_{m=1}^{M} q(r_m^{(j)})} \tag{3.19c}$$

$$q(r) = \begin{cases} 1 & \text{for} \quad |r| \le \gamma \\ \gamma \cdot \frac{\mathrm{sign}(r)}{r} & \text{for} \quad |r| > \gamma \end{cases} \tag{3.19d}$$

Every estimator of location uses a weighting function $q(r)$ (e.g., the sample mean weights every residual equally). The goal of a weighting function is to modify the influence of samples to gain robustness against outliers while still extracting as much information from the samples as possible (retaining estimation efficiency). The weighting function, $q(\cdot)$, is

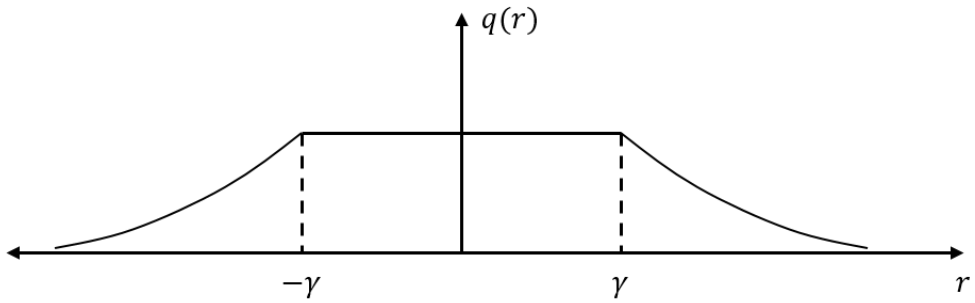

Figure 3.5: The q function for the Huber estimator expressed by (3.19d). The smaller residuals are evenly weighted, but starting at $\pm\gamma$ the q function begins to down-weight the residuals. This makes the estimator more robust against outliers while still extracting location information efficiently from the samples near the estimate of location. The estimate of location is iteratively attained when the weighted residuals are minimized.

unique for each estimator. For the Huber estimator, the mathematical representation of the q function (note there is one tunable parameter, γ) is shown in (3.19d) and a plot of the q function is provided in Figure 3.5.

One important characteristic that differs between these spectral estimation methods is the computational complexity. Performing an average is computationally easier than a median because the median requires sorting. Assuming the sorting algorithm used is $\mathcal{O}(N \cdot \log_2(N))$, such as quick sort or merge sort, then the order of computational effort to perform the traditional Welch and the median-modified Welch are:

$$\begin{aligned}
\mathcal{O}(\hat{\phi}^{(\mathrm{w})}[k]) &= \underbrace{ML\log_2(L)}_{\text{FFTs}} + \underbrace{LM}_{\text{mean}} \\
&= ML(\log_2(L) + 1) \\
\mathcal{O}(\hat{\phi}^{(\mathrm{m})}[k]) &= \underbrace{ML\log_2(L)}_{\text{FFTs}} + \underbrace{LM\log_2(M)}_{\text{median}} \\
&= ML(\log_2(L) + \log_2(M))
\end{aligned} \tag{3.20}$$

This results in an increase in computation over the traditional method by the factor:

$$\frac{\mathcal{O}(\hat{\phi}^{(\mathrm{m})}[k])}{\mathcal{O}(\hat{\phi}^{(\mathrm{w})}[k])} = \frac{\log_2(L) + \log_2(M)}{\log_2(L) + 1} \tag{3.21}$$

Generally the DFT length, L, is larger—even larger by two orders of magnitude—than the number of periodograms, M, for a given Welch estimate. Therefore the increase in computational complexity will likely be tolerable in a realizable system.

The complexity of the Huber-modified Welch is not as simple to quantify. The Huber estimation process does not require sorting, but the Huber estimator solves for its estimate of location iteratively. The IRLS algorithm used here has a linear rate of convergence, and the number of iterations is dependent on the stop condition. After each iteration, the condition in (3.22) is checked.

$$\frac{\left|\hat{\theta}^{(j+1)} - \hat{\theta}^{(j)}\right|}{\hat{\theta}^{(j+1)}} < C_{\mathrm{stop}} \tag{3.22}$$

If the above condition is met, then $\hat{\theta}^{(j+1)}$ is assumed to be close enough to the theoretical convergence point and is selected as the estimate of location. For the conditions used to generate the results in Chapter 4, if the stop constant is $C_{\mathrm{stop}} = 0.01$ then most often 2 or 3 iterations are required; if $C_{\mathrm{stop}} = 0.001$ then most often 3 or 4 iterations are required. In [3] 3–5 iterations are needed.

3.3 The Multitaper Method

As stated in the introduction, the MultiTaper Method (MTM) was develop by David Thomson and first published in 1982 [8]. Similar to Welch's method, the MTM is a nonparametric spectral estimation technique that uses modified periodogram averaging to reduce variance; but unlike Welch's method, it does not use overlapped segmentation to

obtain loosely uncorrelated estimates. The MTM instead uses orthogonal tapers (which are similar to windows) to generate loosely uncorrelated modified periodograms.

3.3.1 Overview of Multitaper Method

The MTM uses tapers instead of conventional windows for spectral shaping and bias control. Window functions are even, positive, and their middle value is the maximum value (at unity) [56]. The traditional MTM uses Discrete Prolate Spheroidal Sequences (DPSSs)[7] as the tapers which can be odd or even and can contain both negative and positive weights. A set of tapers is fully defined by the time-bandwidth product, NW, and the data length, N. The integer $P = 2NW - 1$ is the number of DPSS tapers for a given (N, W) pair and $P \in \{1, \ldots, N\}$ (but since NW is usually small such as 2.5 [9], 4 [48], or 6, then P is also small). The passband of the spectral window for each taper is on the fractional frequency interval $[-W, W]$, where $W = NW/N$.

If the eigenvalue-eigenvector equation (3.23) comprises an $N \times N$ A matrix whose (n_1, n_2) elements are $\sin(2\pi W(n_2 - n_1))/(\pi(n_2 - n_1))$, then the eigenvectors, $\boldsymbol{g}_\zeta$, are called DPSSs where $n_1, n_2, \zeta = 0, \ldots, N - 1$.

$$A\boldsymbol{g}_\zeta = \lambda_\zeta \boldsymbol{g}_\zeta \tag{3.23}$$

The eigenvalues are real and have the properties $1 > \lambda_0 > \lambda_1 > \cdots > \lambda_{N-1} > 0$ and $\lambda_0, \ldots, \lambda_{P-1}$ are near unity while λ_ζ quickly decreases toward zero for $\zeta \geq P$. The DPSSs form a basis for $\mathbb{R}^N$. Most of the energy in the bandwidth $2W$ is concentrated in the first P eigenvectors (rendering the remaining $N - P$ eigenvectors negligible) which is convenient because computing P modified periodograms is far more desirable than computing N modified periodograms. The selection of W determines what is effectively the main-lobe

[7] Other families of tapers have been suggested [57].

width of the spectral window (which is the same for each taper), but also scales proportionally to P. The first P eigenvectors form the DPSS tapers ($h_p[n] = \boldsymbol{g}_p$) used to compute the P taper-modified periodograms (3.24).

$$\hat{\phi}_p^{(\mathrm{t})}(f) \triangleq \frac{1}{N}\left|\sum_{n=0}^{N-1} h_p[n]x[n]e^{-j2\pi fn}\right|^2 \tag{3.24}$$

The relationship between continuous and discrete estimation is analyzed when the periodogram and correlogram are defined in the beginning of this chapter, so the discrete version of the MTM is given in (3.25) and all further spectral estimation methods shown will be discrete.

$$\hat{\phi}_p^{(\mathrm{t})}[k] = \frac{1}{N}\left|\sum_{n=0}^{N-1} h_p[n]x[n]e^{-j2\pi kn/N}\right|^2 \tag{3.25}$$

A few properties of these tapers enable the MTM. As stated at the beginning of this section, the tapers are orthogonal to one another—which can be observed in the left-side plots in Figure 3.6 where $NW = 2.5$ and $N = 512$. These parameters result in $P = 2NW - 1 = 4$ tapers, each of which is applied to the In-phase and Quadrature (IQ) of the same QPSK signal (shown in the right-side plots) to demonstrate the term $h_p[n]x[n]$ in (3.25).

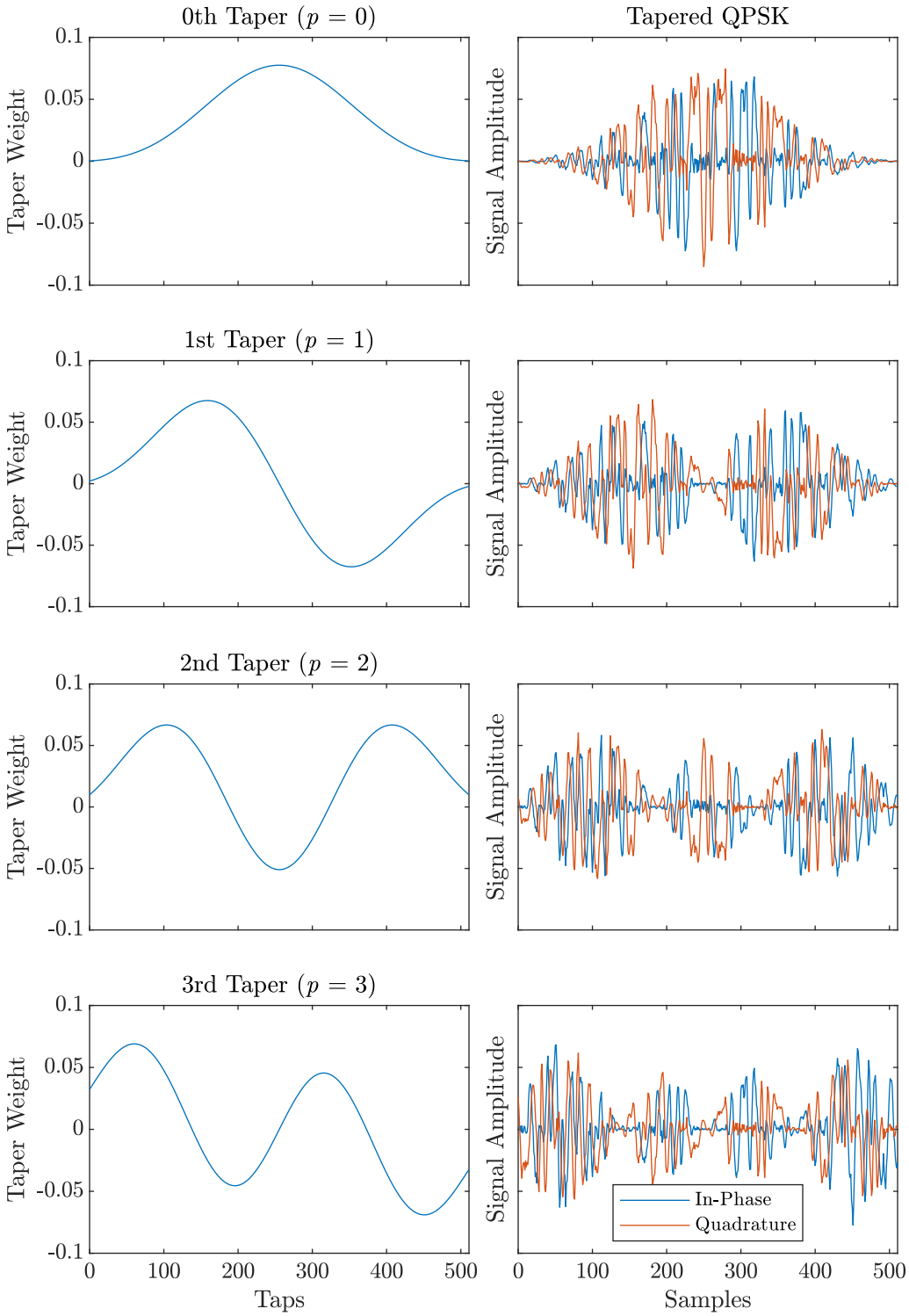

Figure 3.6: DPSS taper (left) and tapered QPSK IQ (right). Tapers are generated with $NW = 2.5$ and $N = 512$.

Furthermore, the frequency concentration for all the windows, or—in windowing terminology—the main-lobe width, is equal. This can be observed in the spectral windows shown in Figure 3.7 (whose corresponding tapers were shown in Figure 3.6). An $NW = 2.5$ and $N = 512$ result in $W = 0.00448$. A vertical dashed line is drawn in the spectral window plots at W to show that the energy is concentrated on $[0, W]$. The cumulative effect of these windows is also shown in Figure 3.7 where one can observe that the cumulative passband approaches a constant as more spectral windows are averaged. The spectral window of the pth taper is shown in the left-side plots, and the corresponding right-side plots show the average of the 0th–pth spectral windows. Lastly, notice that the spectral leakage increases with p. Since this is undesirable, techniques to mitigate total leakage for the MTM are discussed later.

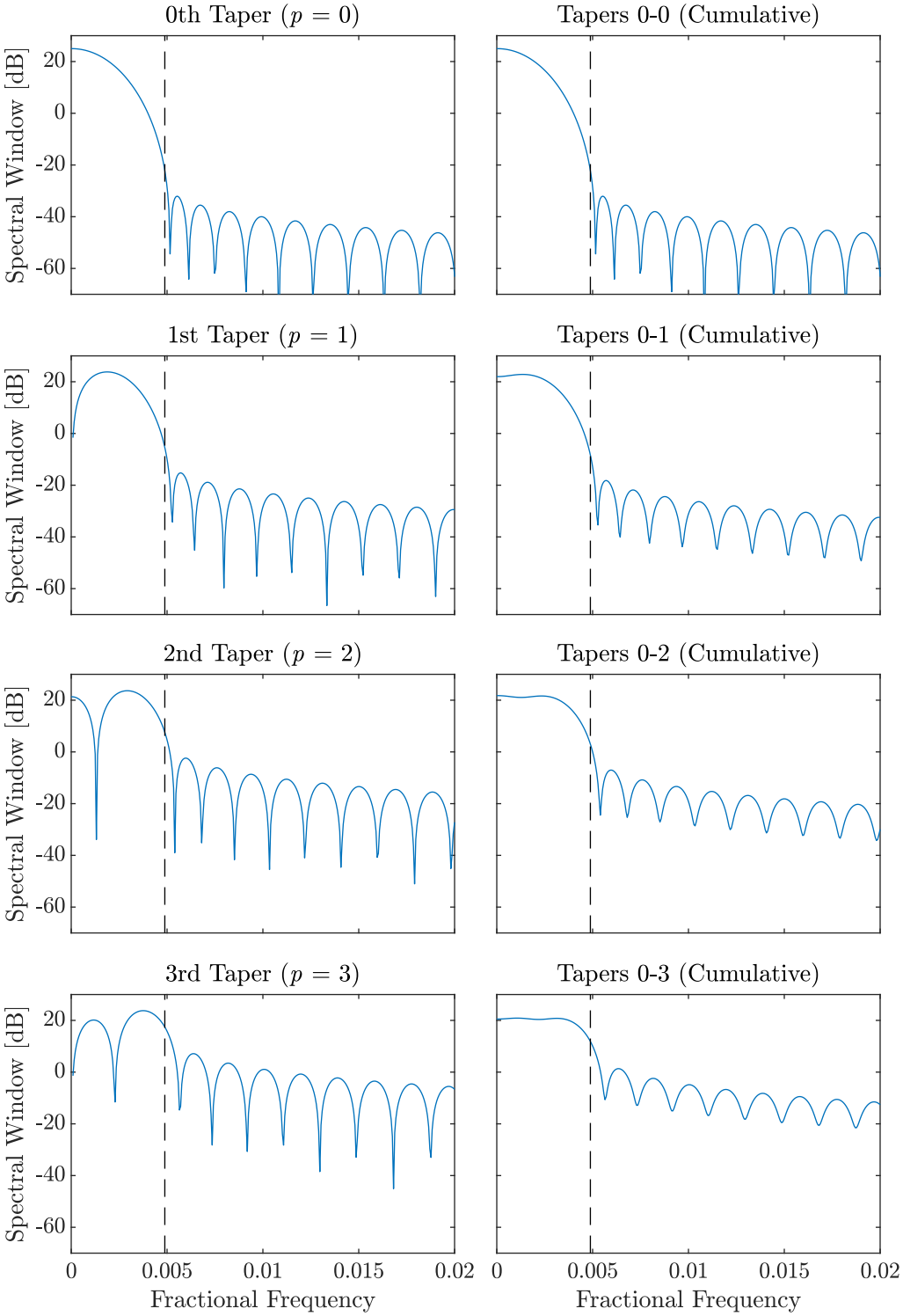

Figure 3.7: Spectral window of the DPSS tapers in Figure 3.6 (left) and cumulative spectral windows (right).

The three standard methods for combining the P taper-modified periodograms into a standard multitaper spectral estimate are outlined below [9].

1. *Unity* (3.26a): The unity method weights each $\hat{\phi}_p^{(\mathrm{t})}[k]$ equally by averaging. This is the simplest method, but is generally not used because the following two methods are less biased.

2. *Eigenvalue* (3.26b): The eigenvalue method weights each $\hat{\phi}_p^{(\mathrm{t})}[k]$ according to the corresponding eigenvalue, λ_p. In other words, this method computes a weighted average. The eigenvalues are natural weights because they correspond to the strength of the signal's projection onto the basis function (taper).

3. *Adaptive* (3.26c): The adaptive method weights each $\hat{\phi}_p^{(\mathrm{t})}[k]$ according to the bias properties of the corresponding tapers [48]. This method is called the adaptive method because the weights are adapted based on the variance of the signal—which means the corresponding weights for two different estimates can differ.

$$\hat{\phi}^{(\mathrm{mt,u})}[k] \triangleq \frac{1}{P}\sum_{p=0}^{P-1}\hat{\phi}_p^{(\mathrm{t})}[k] \tag{3.26a}$$

$$\hat{\phi}^{(\mathrm{mt,e})}[k] \triangleq \frac{\sum_{p=0}^{P-1}\lambda_p\hat{\phi}_p^{(\mathrm{t})}[k]}{\sum_{p=0}^{P-1}\lambda_p} \tag{3.26b}$$

$$\hat{\phi}^{(\mathrm{mt,a})}[k] \triangleq \frac{\sum_{p=0}^{P-1}b_p^2[k]\lambda_p\hat{\phi}_p^{(\mathrm{t})}[k]}{\sum_{p=0}^{P-1}b_p^2[k]\lambda_p} \tag{3.26c}$$

The unity method is mathematically defined but almost never used because intelligent weighting is preferred. The adaptive method reduces bias best but is computationally complex, requires an estimate of the noise variance, and the small advantage gained is unnecessary for this thesis. For these reasons, the eigenvalue method is used in this thesis.

3.3.2 Making Multitaper more Robust

Perhaps the most intuitive method for making the MTM more robust would be to replace the combination techniques—average (for unity method) or weighted average (for eigenvalue and adaptive methods)—with a more robust estimate (such as the median or the Huber). The issues with this are twofold:

1. By design, the spectral windows of the tapers combine to produce a spectral window with a very flat main lobe and a sharp cut off (which yields the accurate frequency resolution of $2W$). One needs to use caution when tampering with this process.

2. The robust approach to the Welch method works because it is based on the qualitative principle that certain segments are more contaminated than others, and robust estimation techniques are used to down-weight or ignore these highly biased estimates caused by high contamination. Since the MTM uses the same samples for each $\hat{\phi}_p^{(\mathrm{t})}[k]$, it is likely that each taper-modified periodogram will be biased similarly.

Instead of attempting to increase the robustness of the MTM, the advantage of orthogonal tapering (from the MTM) is applied to overlapped segmentation (from the Welch method) into a combined approach of Thomson's method and Welch's method. The most straightforward combination of these approaches is called the MultiTaper Welch Method (MTWM) and is defined in (3.27)

$$\hat{\phi}^{(\mathrm{mtw})}[k] \triangleq \frac{1}{M} \sum_{m=1}^{M} \hat{\phi}_m^{(\mathrm{mt,e})}[k] \tag{3.27}$$

where the periodogram (3.5b) in the Welch method (3.7) is replaced with the MTM (3.26b). A reminder for understanding the above and below equations: recall that M denotes the

number of segments and L denotes segment length (for any PSD estimation method that uses overlapped segmentation).

It will be observed later that the trade-off between the Welch method and the MTWM is the classic variance versus bias trade-off.

By analogy, the process used to turn the Welch method into the median-modified Welch can be used to convert the MTWM into a more robust Median-Modified MultiTaper Welch Method (M3TWM). In order to write the mathematical expression for the M3TWM, an intermediate estimator needs to be defined first: by replacing the periodogram with the taper-modified periodogram in the median-modified Welch described in (3.16)–(3.18), the Taper-Median-Modified Welch Method (TMMWM) is defined as shown in (3.28).

$$W_{\phi,p}^{(\mathrm{t})}[m,k] = \begin{bmatrix} \hat{\phi}_{p,1}^{(\mathrm{t})}[k] \\ \hat{\phi}_{p,2}^{(\mathrm{t})}[k] \\ \vdots \\ \hat{\phi}_{p,m}^{(\mathrm{t})}[k] \\ \vdots \\ \hat{\phi}_{p,M}^{(\mathrm{t})}[k] \end{bmatrix} = \begin{bmatrix} \hat{B}_{p,1}^{(\mathrm{t})}[m] & \hat{B}_{p,2}^{(\mathrm{t})}[m] & \cdots & \hat{B}_{p,k}^{(\mathrm{t})}[m] & \cdots & \hat{B}_{p,L}^{(\mathrm{t})}[m] \end{bmatrix} \tag{3.28a}$$

$$W_{S,p}^{(\mathrm{t})}[m,k] = \mathrm{SORT}(W_{\phi,p}^{(\mathrm{t})}[m,k],\, m) \tag{3.28b}$$

$$\hat{\phi}_{p}^{(\mathrm{tmmw})}[k] = \frac{1}{2}\left(W_{S,p}^{(\mathrm{t})}\left[\frac{M}{2},k\right] + W_{S,p}^{(\mathrm{t})}\left[\frac{M}{2}+1,k\right]\right) \tag{3.28c}$$

The M3TWM can now be defined in terms of (3.28c) as:

$$\hat{\phi}^{(\mathrm{m3tw})}[k] \triangleq \frac{\sum_{p=0}^{P-1} \lambda_p \hat{\phi}_{p}^{(\mathrm{tmmw})}[k]}{\sum_{p=0}^{P-1} \lambda_p} \tag{3.29}$$

The M3TWM utilizes the robustness gained from the median-modified Welch and then

attains variance reduction through orthogonal tapering from the MTM. The cost is bias and computational effort. The computational cost is increased by a factor of P–the selection of which also results in a bias versus variance trade-off ($P = 4$ is used in this thesis).

In this chapter many different methods for nonparametric spectral estimation are outlined. However, all the methods center around the concept of overlapped segmentation, orthogonal tapering, and robustness. It is quite easy to forget which method is which, so to this end, Table 3.1 shows each method in the order it is presented in this thesis. The table catalogs all nonparametric methods (aside from the correlogram) shown. For each method the official name, notation, equation, description, and (if applicable) how the method builds upon previous methods, are listed. The estimators in bold are the robust estimators defined in this thesis, and the performance of the robust estimators is compared to the performance of the non-robust estimators in Class A noise in Chapter 4.

Table 3.1: Nonparametric Spectral Estimation Methods

Name	**Definition**		**Description**
Periodogram	$\hat{\phi}^{(\mathrm{p})}[k]$	3.5	The square magnitude of the DFT (the simplest method).
Welch Method	$\hat{\phi}^{(\mathrm{w})}[k]$	3.7	Segments the data into M overlapped segments, computes a *periodogram* for each segment, and then computes the mean for each frequency bin.
Median-modified Welch method	$\hat{\phi}^{(\mathrm{m})}[k]$	3.18	Same as the *Welch method* except replaces the frequency bin estimate of location (mean) with the median.
Huber-modified Welch method	$\hat{\phi}^{(\mathrm{h})}[k]$	3.19[a]	Same as the *Welch method* except replaces the frequency bin estimate of location (mean) with the Huber estimator.
Taper-modified periodogram	$\hat{\phi}_p^{(\mathrm{t})}[k]$	3.25	Same as the *periodogram* except weights the data using the pth DPSS taper prior to computing the DFT.
MultiTaper Method (MTM)	$\hat{\phi}^{(\mathrm{mt,e})}[k]$	3.26b	The eigenvalue weighted average of P *taper-modified periodograms*.
MultiTaper Welch Method (MTWM)	$\hat{\phi}^{(\mathrm{mtw})}[k]$	3.27	Same as the *Welch method* except the *periodogram* of the mth segment is replaced by the *MTM*.
Taper-Median-Modified Welch Method (TMMWM)	$\hat{\phi}_p^{(\mathrm{tmmw})}[k]$	3.28	Same as the *median-modified Welch method* except the *periodogram* for each segment is replaced with the *taper-modified periodogram*.
Median-Modified MultiTaper Welch Method (M3TWM)	$\hat{\phi}^{(\mathrm{m3tw})}[k]$	3.29	Same as the *MTM* except the pth *taper-modified periodogram* is replaced with the *TMMWM*.

[a] The Huber estimator requires use of an iterative algorithm, so a concise mathematical definition is not given, but the algorithmic process is described by (3.19a)–(3.19d).

Chapter 4

Results

New robust nonparametric spectral estimation techniques based on the Welch method and the MultiTaper Method (MTM) were discussed in Chapter 3. In this chapter, the performance of these methods is experimentally tested by attempting signal detection in noise environments characterized by Middleton Class A noise over a broad range of parameters. The Class A noise model is very flexible and it is expected that certain parameters yield noise characteristics for which each estimation method will perform similarly. The goal is to discern if there are any realistic environments in which performance can be improved with the robust methods.

4.1 Measurement Definitions and Default Parameters

Recall that energy-based signal detection is the motivation for increasing the robustness of the Welch method and the MTM. In order to test the robustness of the modified version of these methods, their signal detection capabilities in various noise environments must be

measured. The performance of these methods can be quantitatively measured by using them to compute Power Spectral Density (PSD) estimates of signals modulated using Quadrature Phase-Shift Keying (QPSK) in various noise environments and then measuring the effective Signal-to-Noise Ratios (SNRs) of the estimates. The effective SNR of a PSD estimate is defined here as the median amplitude over the bins within the signal bandwidth to the median amplitude over the bins outside the signal bandwidth and transition bands (the noise floor) (4.1).

$$\mathrm{SNR}_{\mathrm{Eff,dB}} \triangleq \mathrm{median}(\{\hat{\phi}^{(\mathrm{w})}_{\mathrm{dB}}[k]\}_{k\in\mathrm{BW}}) - \mathrm{median}(\{\hat{\phi}^{(\mathrm{w})}_{\mathrm{dB}}[k]\}_{k\notin\mathrm{BW}}) \tag{4.1}$$

Note that effective SNR is a metric only defined in this thesis due to the utility it provides in quantitatively measuring the robustness of the PSD estimation methods. Furthermore, it is safe to assume that the effective SNR is not equal to the true SNR. Indeed, if the modified methods are robust, then the effective SNR should be higher than the true SNR because mitigating bias caused by Impulsive Noise (IN) increases effective SNR.

In this thesis two definitions are used for energy per bit to noise power spectral density ratio (E_b/N_0) in the context of Class A noise. In the first, N_0 represents the total noise power of the Class A noise; that is, if one set of samples drawn from a Gaussian distribution and another set of samples drawn from a Class A distribution produce noise PSDs with equal magnitude, then applying either of these sets to the same signal would produce the same E_b/N_0. Since this definition is based on the total power of Class A noise, it is denoted more explicitly using an uppercase T in (4.2).

$$E_b/N_0^{(\mathrm{T})} \triangleq E_b/N_0 \tag{4.2}$$

In the second definition the N_0 represents the Gaussian power. Recall that the Gaussian

factor (2.10) relates the power of the underlying Gaussian Noise (GN) to the power of the Non-Gaussian Noise (NGN). The relationship between the total noise power of a Class A Random Process (RP) and the power contributed by the underlying Gaussian component is shown in (2.8) and used to define the second definition of E_b/N_0 (which uses an uppercase G to denote Gaussian) in (4.3).

$$E_b/N_0^{(\mathrm{G})} \triangleq E_b/N_0 \frac{1+\Gamma}{\Gamma} \tag{4.3}$$

Table 4.1 shows default simulation parameters used to generate the results. If a simulation parameter differs from the default, it will be stated explicitly in the text or shown on the relevant figure.

Examples of the modified Welch estimators compared to the traditional Welch estimator are shown in Figures 4.1 and 4.2 using the default values in Table 4.1.

Table 4.1: Default Simulation Parameters

Parameter Name	**Default Value**	**Units**
QPSK		
Energy-per-bit (E_b)	1	
Samples-per-symbol	4	samples/symbol
Pulse Shaping Filter		
Shape	root-raised cosine	
Roll-off factor	0.35	
Span	4	symbols
AWGN		
E_b/N_0	10	dB
Middle Class A		
$E_b/N_0^{(\mathrm{T})}$	3	dB
Gaussian factor (Γ)	0.001	
Overlap index (A)	0.001	
Welch		
FFT length (L)	256	samples
Number of segments (M)	10	
Overlap (D/L)	50	%
Window shape	rectangular	
IRLS		
γ	1.5	
Stop constant (C_{stop})	0.01	
Multitaper		
Taper length (N)	256	samples
Time-bandwidth product (NW)	4	samples $\cdot$ Hz
Number of Tapers (P)	7	

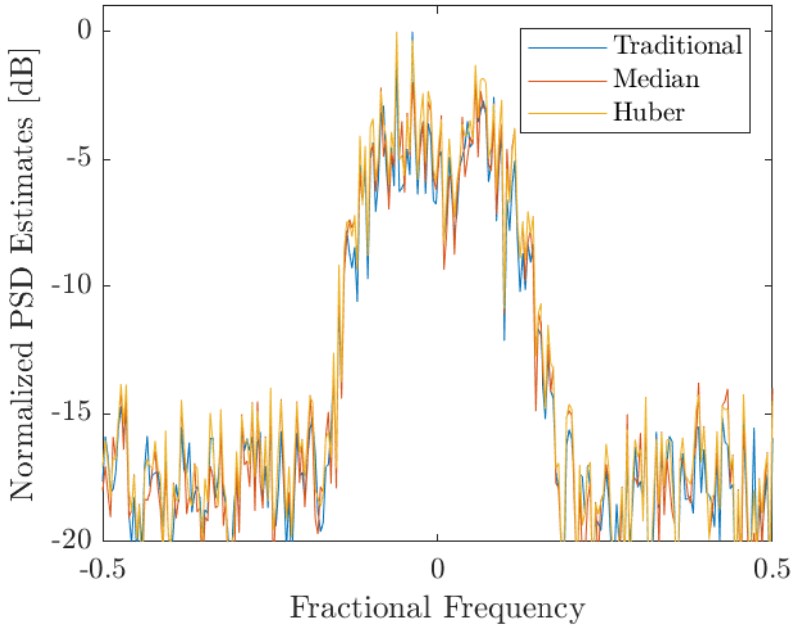

Figure 4.1: Example of traditional, median-modified, and Huber-modified Welch estimates in Additive White Gaussian Noise (AWGN). All three estimators perform similarly.

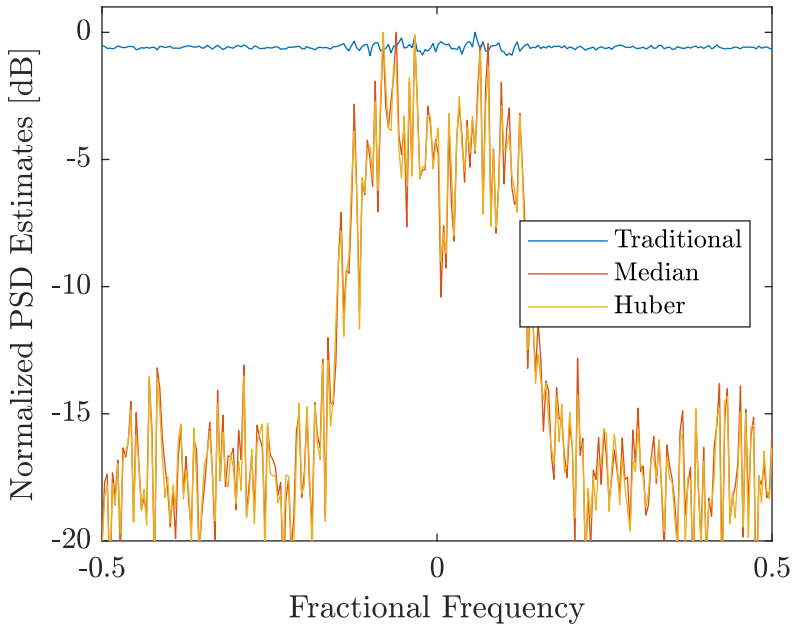

Figure 4.2: Example of traditional, median-modified, and Huber-modified Welch estimates in Middleton Class A noise. The traditional estimator is not robust against outliers caused by impulsive noise, but the modified versions (which are nearly indistinguishable) are.

For Figure 4.1 the complex baseband QPSK signal is in AWGN. For Figure 4.2 the same QPSK signal is used except with Middleton Class A noise. The values $A = \Gamma = 0.001$ taken from [32] at $E_b/N_0^{(\mathrm{G})} = 10$ decibels are used to generate Figure 4.2. Note in Figure 4.1 that the methods perform similarly but with highly impulsive noise, as shown in Figure 4.2, the traditional method breaks down.

4.2 Welch Results

With the aforementioned mathematical definitions and default parameters established, the performance analysis can be described. In the interest of comprehensive analysis, the Welch methods are tested across all reasonable Class A parameters. The literature was surveyed to determine the reasonable parameter space for the overlap index, A, and the Gaussian factor, Γ, and the results are shown in Figure 4.3 from [29], [32]–[34], [36], [58], [59].

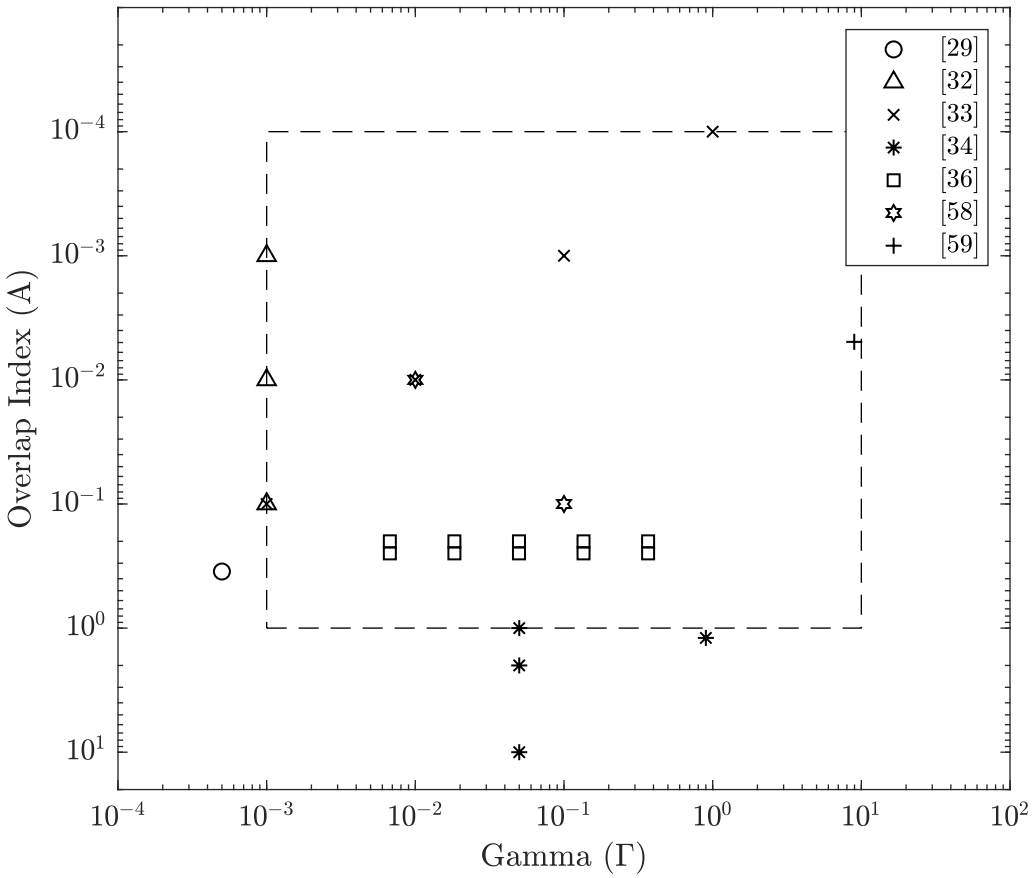

Figure 4.3: The range of Middleton Class A noise parameter pairs (Γ, A) found in the literature. The rectangular box represents the range examined in this thesis.

From (2.2) we can see that P_m is a function of $\{A, m\}$ while σ_m^2 is a function of $\{\sigma^2, A, \Gamma, m\}$. Logarithmically spaced values of A and Γ (five for each) that span the parameter space in Figure 4.3 are used to generate Figures 4.4 and 4.5, demonstrating how P_m and σ_m^2 vary across the relevant parameter space. Recall that P_m can be viewed as the weighting term that selects between Rayleigh Random Variables (RVs) of differing powers which compose the complex envelope of the Class A RV (2.5). Since P_m decreases quickly, the vertical axes for the plots in Figure 4.4 are $\log_{10}(P_m)$ instead of P_m. The first five values of m are sufficient to show the trend of P_m, but it should be noted that all simulations using

Class A noise use the first 10 values of m for an accurate approximation of the infinite series in the Probability Distribution Function (PDF) of the Class A RV (2.4). The parameter σ_m^2 controls the power (more scientifically, the second moment) of the mth Rayleigh RV (see (2.3), (2.5), and Appendix A.2). Since σ_m^2 is a linear function of σ^2, any value could be used to generate the plots in Figure 4.5 without altering the trends; here, the value $\sigma^2 = 1$ is used.

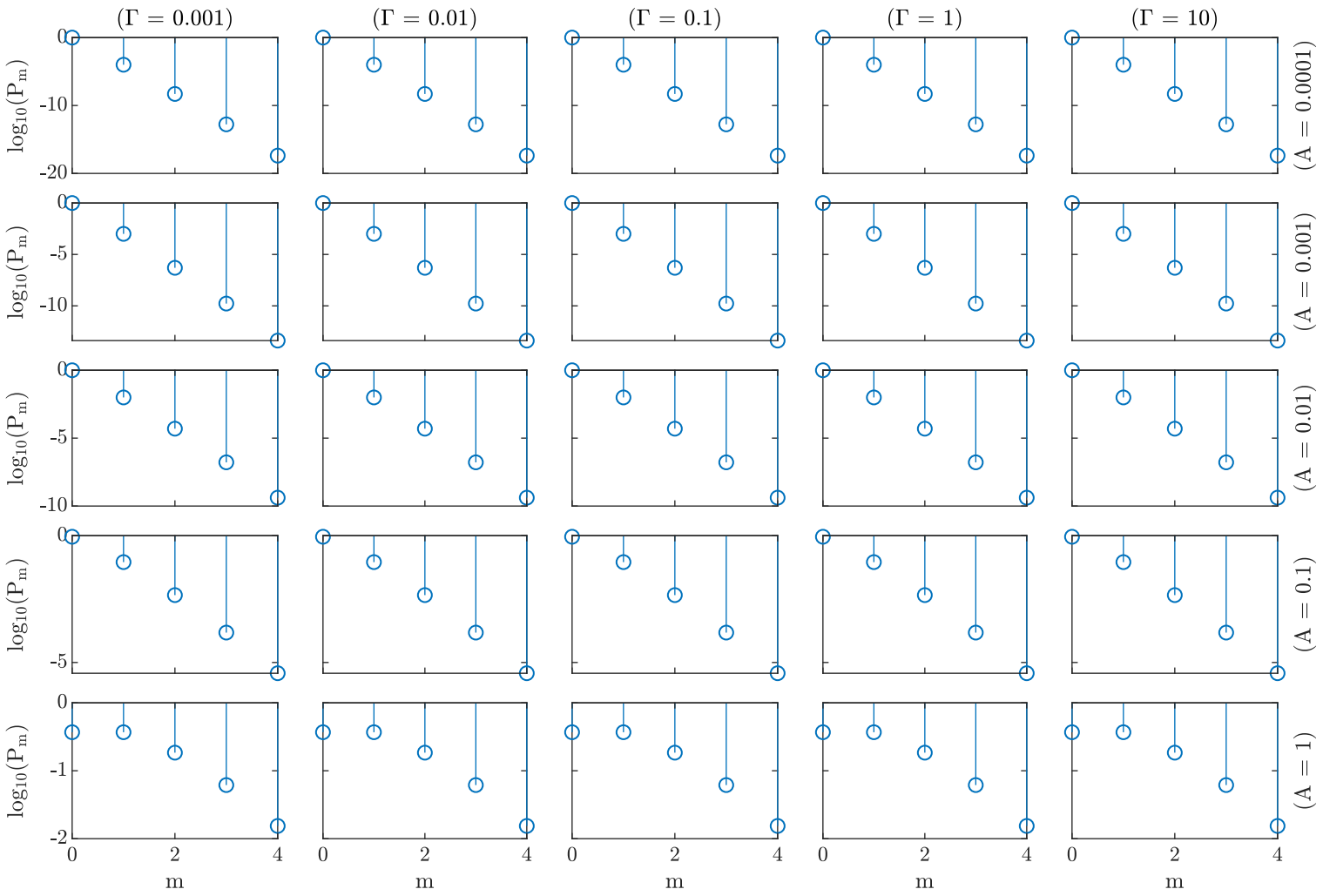

Figure 4.4: P_m for varying Γ and A for the first five values of m. Since P_m does not depend on Γ, each column is the same. As A increases (progressing down a column), P_m decays more slowly, which means the Class A noise will be less Gaussian (because the weight of P_0 is decreasing) though also less impulsive (as will be made clear in Figure 4.5).

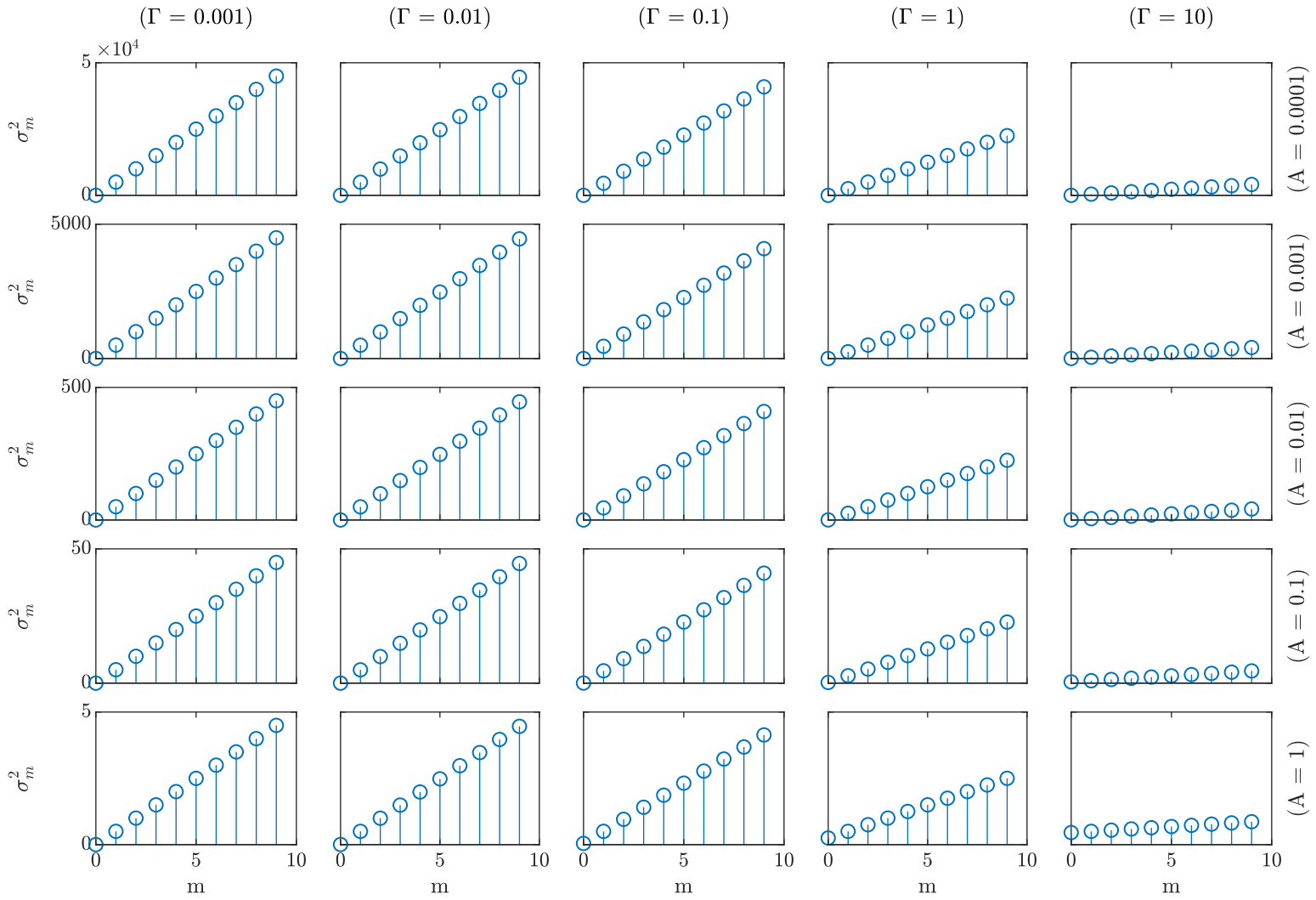

Figure 4.5: σ_m^2 for varying Γ and A for the first 10 values of m. From the plots and (2.2), one can see that σ_m^2 is approximately proportional to m and inversely proportional to A. Recall that the power from a Class A RV is not a function of Γ or A (2.7). Since the power of the mth Rayleigh RV increases as A decreases, the probability of selecting a higher power Rayleigh RV is simultaneously decreased by P_m (refer back to Figure 4.4) to maintain constant power. Additionally, as Γ changes, σ_m^2 shifts in such a way as to maintain constant power.

The understanding of how P_m and σ_m^2 vary across the span of Γ and A is sufficient to gain an intuitive sense of the statistics of the Class A complex envelope. Figure 4.6 shows the complex envelope for 10 000 samples using $\sigma^2 = 0.5$ which, if applied to a QPSK signal with E_b normalized to unity, would yield $E_b/N_0^{(\mathrm{T})} = 3$.

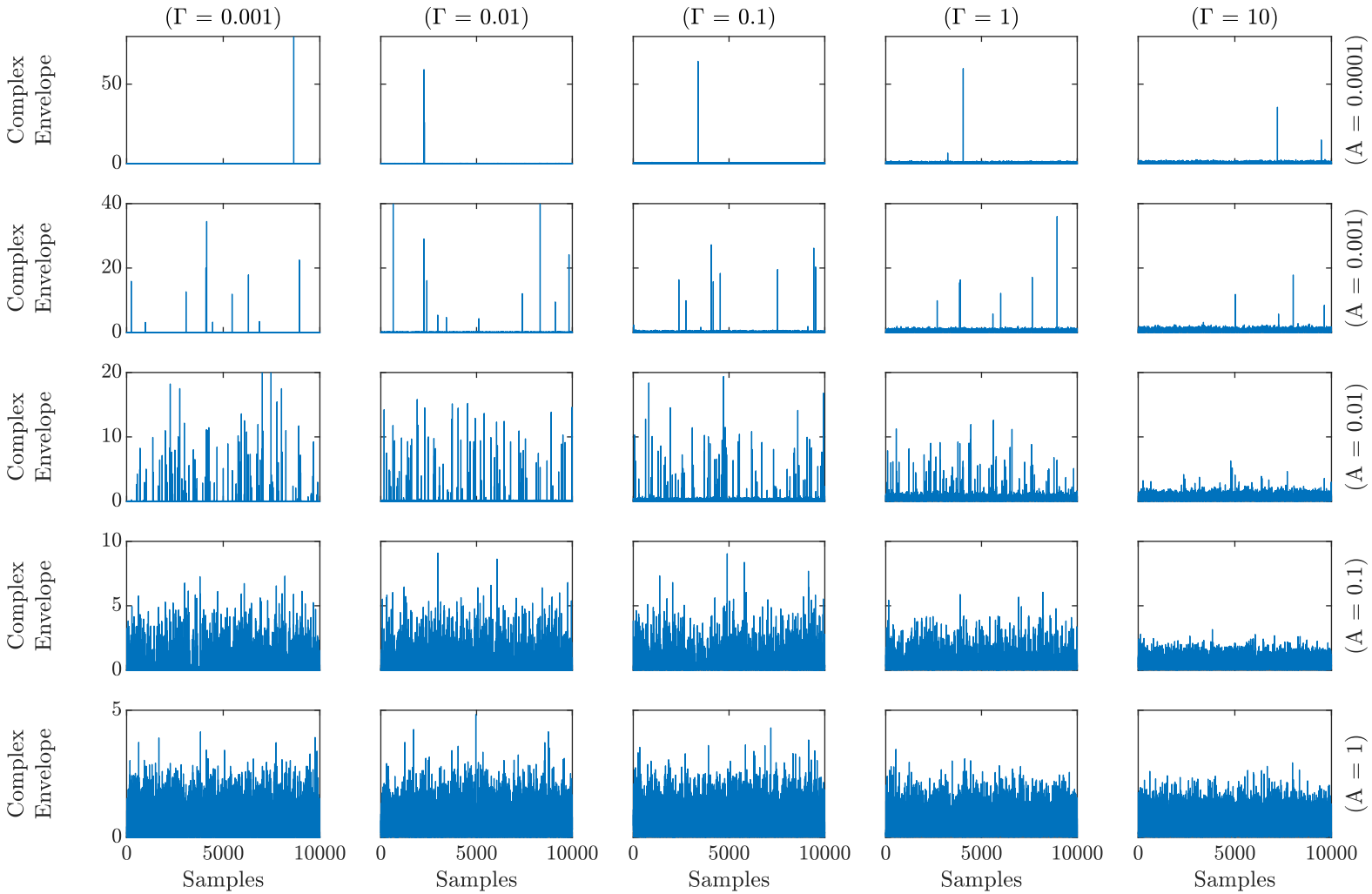

Figure 4.6: Complex envelope of Class A noise for varying A and Γ. As A increases (progression down the columns) the noise becomes less impulsive but retains the same underlying Gaussian power *and* retains the same non-Gaussian power (trading impulsiveness for density). As Γ increases the noise also becomes less impulsive; while retaining constant total power, the underlying GN power increases relative to the NGN power (best illustrated by the third row).

A more mathematical method to demonstrate how A and Γ affect the noise statistics is to look at the Class A PDF. As discussed in Section 2.5, it is best to analyze the PDF that characterizes the complex envelope of Class A noise (rather than the PDF that characterizes the projections onto the in-phase and quadrature components). Therefore, a fair comparison to AWGN would be to plot the Class A envelope against the Rayleigh PDF (because the Rayleigh distribution characterizes the complex envelope of AWGN). Figure 4.7 shows plots comparing the Rayleigh PDF (blue) and the Class A PDFs (orange) for different values of A and Γ. The PDFs are characterized by noise with unity power—that is, $\sigma^2 = 1$ for the Class A PDFs and $\sigma^2 = 0.5$ for the Rayleigh PDFs.

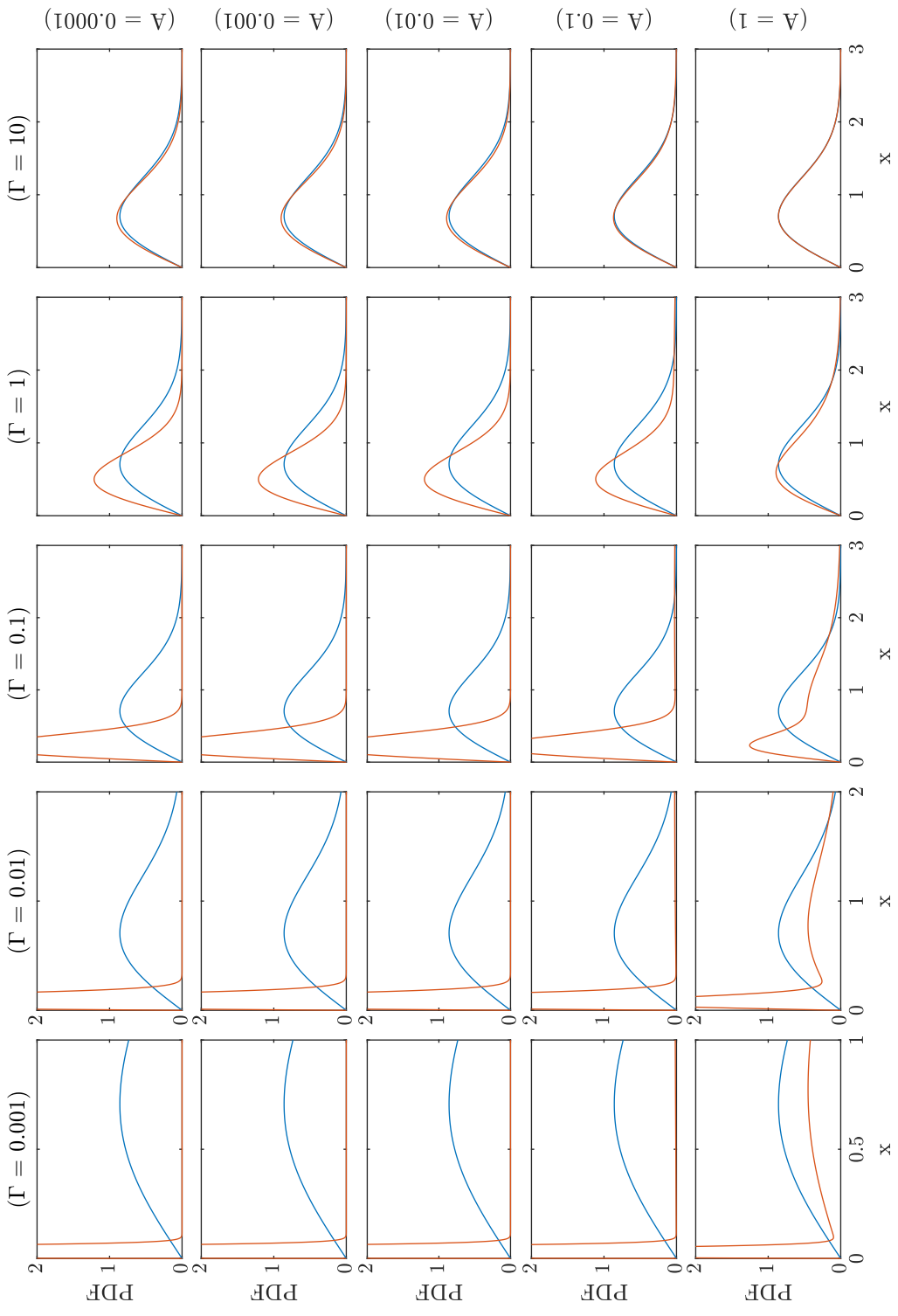

Figure 4.7: Rayleigh PDF (blue) and the PDFs of Class A envelopes (orange). Samples characterized by any of the PDFs have a power of unity.

Notice in Figure 4.7 that the shapes of the Class A PDFs tend toward the Rayleigh distribution as Γ increases (left to right). This is expected because the Gaussian factor, Γ, is the ratio of the underlying GN to the NGN (2.10). However, from the PDFs it is hard to see the change in the Class A PDF tails that reflects the increase in impulsiveness as A decreases (progression up the columns)—the change that is more clearly illustrated in Figure 4.6.

Each plot in Figure 4.8 shows the PSD estimate of a QPSK signal in noise represented by the corresponding plot in Figure 4.6 at $E_b/N_0^{(\mathrm{T})} = 3$. The number of segments used for the PSD estimates is $M = 500$. This is almost two orders of magnitude larger than what is commonly found in the literature. For example, in [50], [51] M is 7, in [55] M is 31, and in [52], [53] values one fewer than powers of two on the interval $[1, 255]$ are used. Therefore, the plots represent the average PSD estimates for each scenario. The remainder of the tricolor plots in this section uses blue for the traditional Welch estimator, red for the median-modified Welch estimator, and orange for the Huber-modified Welch estimator.

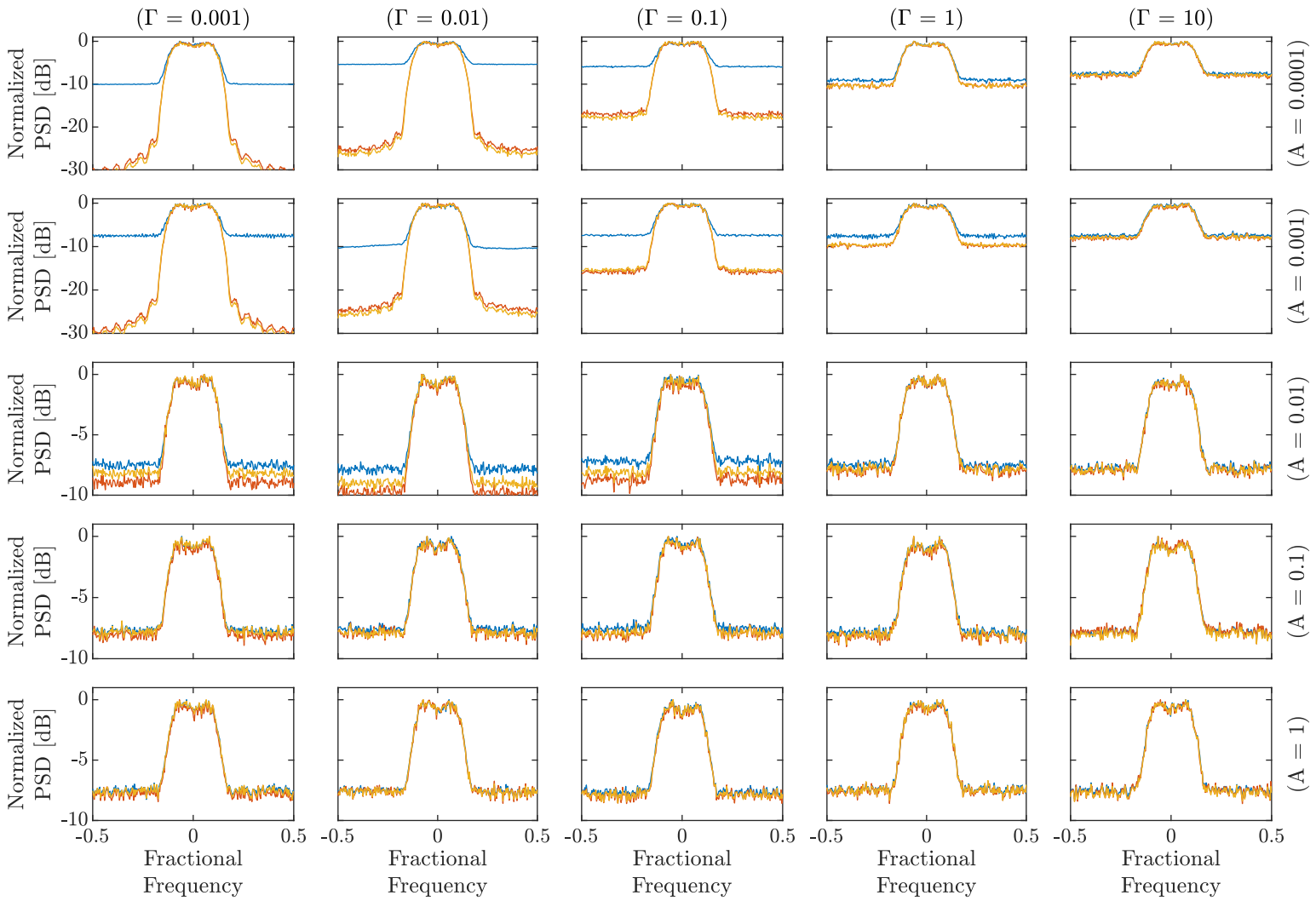

Figure 4.8: Average performance of the three Welch PSD estimators (traditional-blue; median-red; Huber-orange) for QPSK signals with Class A noise from Figure 4.6. Recall that performance is measured by effective SNR (the higher the better). With the large number of averaged segments and constant $E_b/N_0^{(\mathrm{T})} = 3$, the traditional Welch can adequately detect the signal. However, the robust estimators can improve upon the traditional estimator in highly impulsive environments (upper-left plots).

Notice in Figure 4.8 that the traditional estimator (blue) performs essentially the same for all 25 scenarios. With an $E_b/N_0^{(\mathrm{T})} = 3$ and $M = 500$ the SNR is high enough and the averaging is high enough that the traditional estimator can find the signal. From the middle row up, the robust methods begin to outperform the traditional estimator (the robust estimators attain a higher effective SNR). Even though the $E_b/N_0^{(\mathrm{T})}$ remains constant, the effective SNR increases for the robust estimators. This is because the robust estimators can mitigate the influence of the IN and the underlying GN is decreasing in power as A decreases.

Instead of showing average performance (as in Figure 4.8), the estimates in Figure 4.9 show performance using a practical number of segments ($M = 10$). With a small M and a segment length of only 256, the PSD estimates for a given (Γ, A) will statistically vary. To capture this variation, 100 Welch estimates are performed for each estimator and each (Γ, A) pair, and the effective SNRs are computed and shown in a histogram in Figure 4.9.

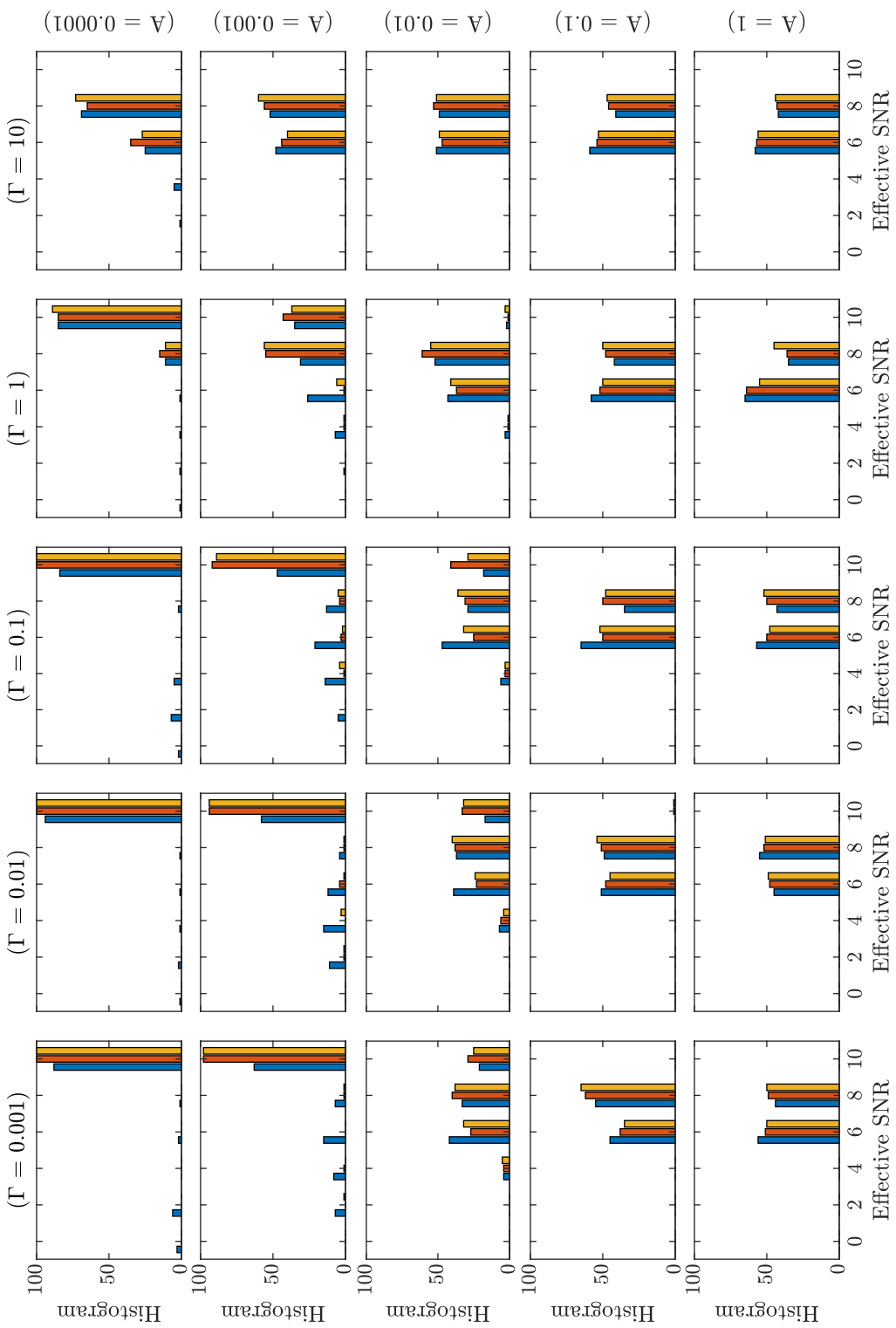

Figure 4.9: Distribution of effective SNR for 100 iterations for each estimator (traditional-blue; median-red; Huber-orange) in each plot. Performance centers around the averages shown in Figure 4.8 except every effective SNR above 9 decibels has been placed in the bin at 10 decibels. The top two rows contain scenarios where the traditional estimator does not detect the signal, but the robust estimators do.

Figure 4.9 does not show the high effective SNR that Figure 4.8 does because the maximum bin of each histogram is at 10 decibels. However, the robust estimators are still shown to be robust in Figure 4.9 in a way that cannot be seen with the high averaging in Figure 4.8: the bias of the traditional Welch estimator caused by the IN (and the robustness to this bias from the robust estimators) can be seen most clearly in the top two rows. There is even some bias distinction that can be seen in the middle row. Also, the traditional Welch estimator is more biased for smaller Γ as evidenced by the increase in effective SNR from left to right for the top three rows.

Rather than maintaining a constant $E_b/N_0^{(\mathrm{T})}$ as used for the results in Figures 4.6–4.9, for the next series of figures $E_b/N_0^{(\mathrm{G})}$ will be held constant at 5 decibels. This setup is designed to replicate the scenario where thermal noise causes constant underlying GN, but the statistics for the NGN contamination are varied. All other simulation parameters are equal between Figures 4.6 and 4.10, Figures 4.8 and 4.11, and Figures 4.9 and 4.12. Please refer to the figure titles to interpret the figures.

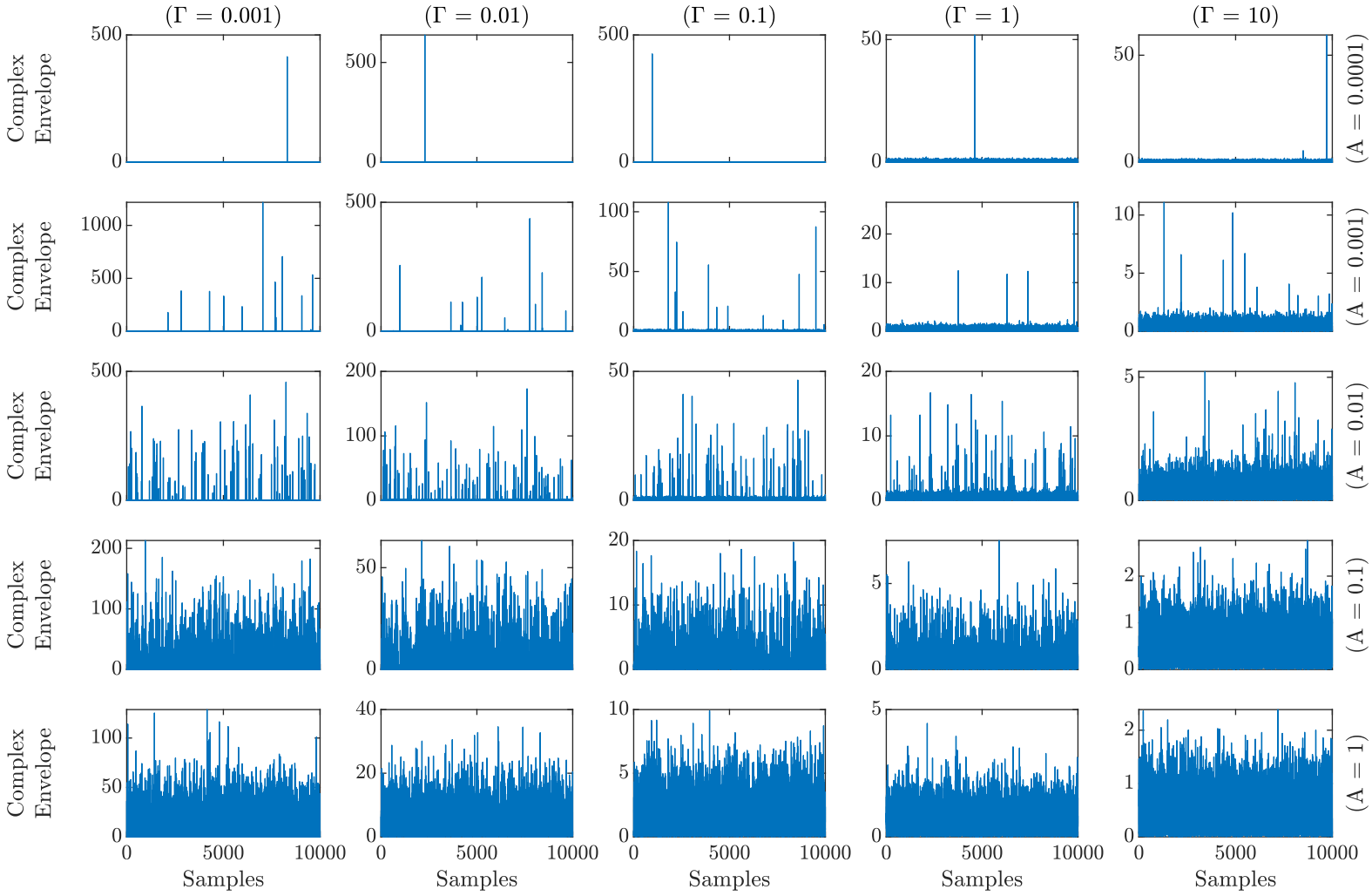

Figure 4.10: Complex envelope of Class A noise varies according to Γ and A. Whereas the total power is held constant in Figure 4.6, in this figure the underlying Gaussian power is constant at $E_b/N_0^{(\mathrm{G})} = 5$ (assuming the noise is applied to a signal with $E_b = 1$). Notice as Γ increases, the total noise decreases because $E_b/N_0^{(\mathrm{G})}$ is directly proportional to $E_b/N_0^{(\mathrm{T})}$ and (for small Γ) roughly inversely proportional to Γ (2.8) and (4.3). The total noise power is constant for the plots within a given column because $E_b/N_0^{(\mathrm{G})}$ does not depend on A.

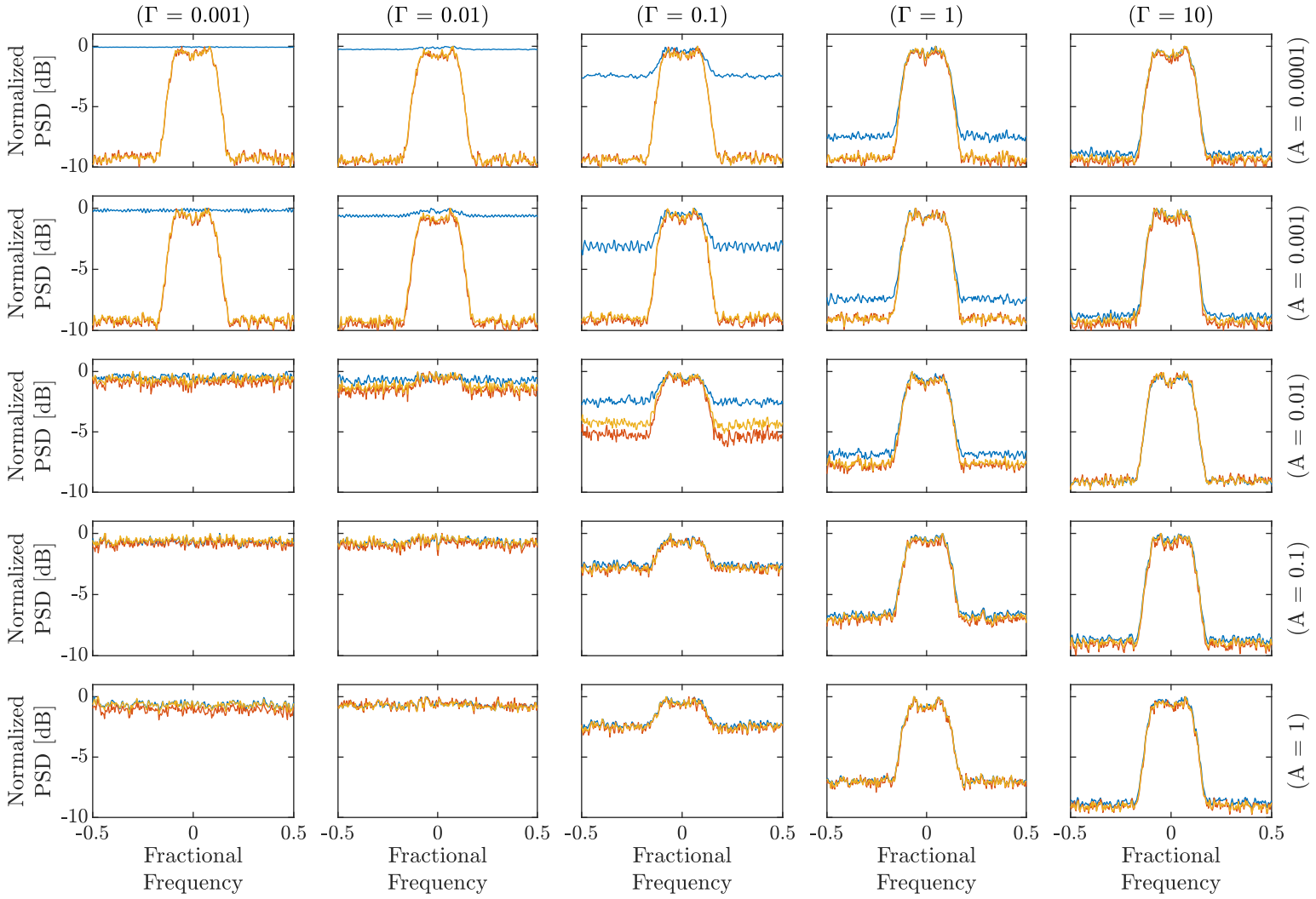

Figure 4.11: Average performance of the three Welch PSD estimators (traditional-blue; median-red; Huber-orange) for QPSK signals with Class A noise from Figure 4.10. Since the $E_b/N_0^{(\mathrm{T})}$ increase as Γ increases, the ability to detect the signal is better for the columns on the right. The bottom-left plots are examples of highly non-Gaussian environments with low impulsiveness, such that no estimator can detect the signal. The top-left plots have the same $E_b/N_0^{(\mathrm{T})}$ as the corresponding plots below, but are impulsive enough that the robust estimators detect the signal even though the traditional estimator cannot.

Since the noise characteristics shown in Figure 4.10 exhibit that the noise power decreases as Γ increases (left to right), then it is expected that the performance for the conditions in the right columns is better than for the conditions in the left columns. This is verified in Figure 4.11. Notice, as well, that—unlike in Figure 4.8—the effective SNR is capped at around 9 decibels in Figure 4.11 because the underlying GN is held constant for each plot ($E_b/N_0^{(\mathrm{G})} = 5$). As the total noise increases (from right to left), the QPSK signal cannot be detected by the traditional Welch estimator. The robust estimators, however, retain the ability to detect the QPSK signal in the high noise environment if the noise is impulsive (see top-left plots). This can also be seen by observing the consistent performance of the robust estimators across the top two rows.

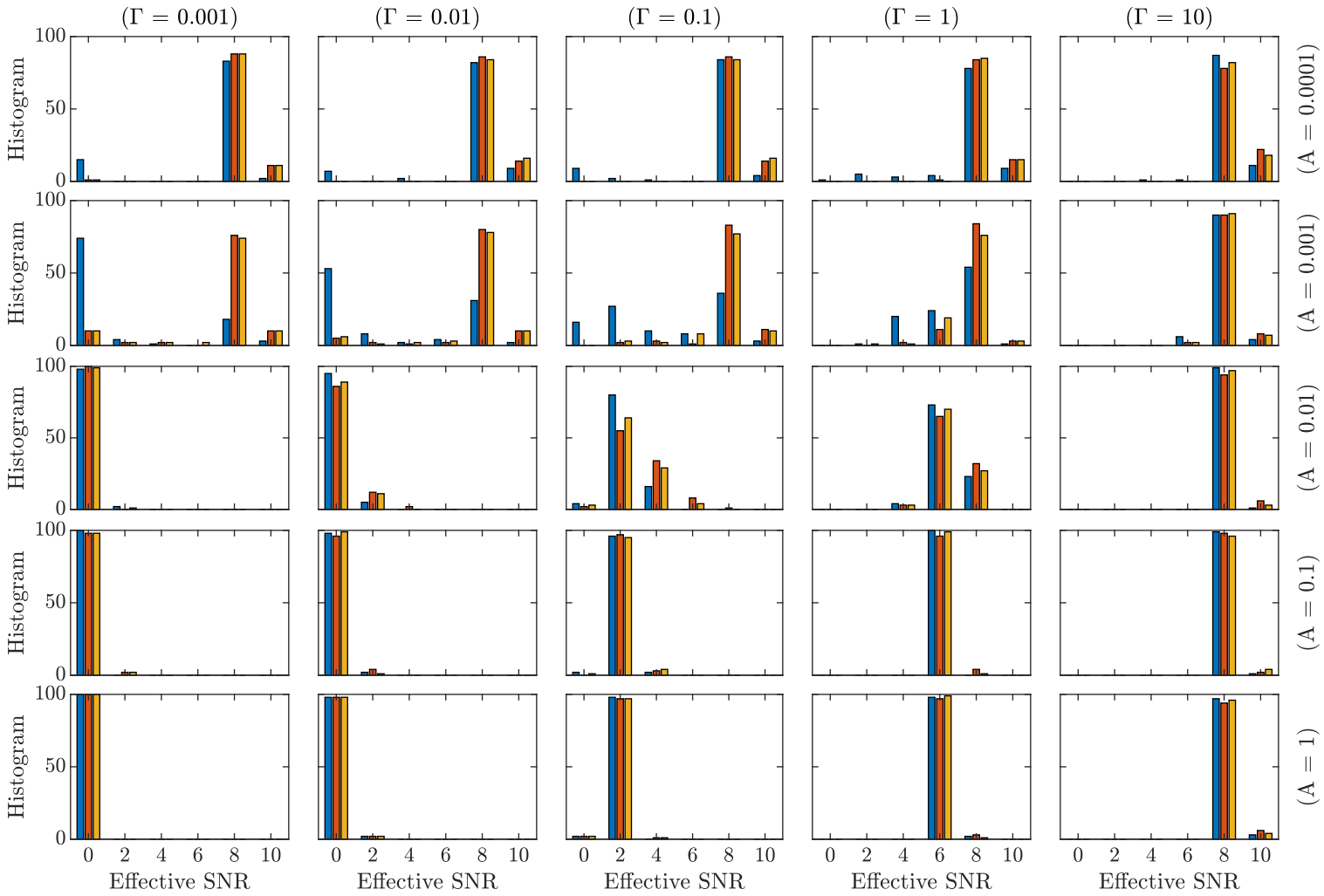

Figure 4.12: Distribution of effective SNR for 100 iterations for each estimator (traditional-blue; median-red; Huber-orange) in each plot. Similar to Figure 4.9, the estimators perform comparably with the major exception of under the condition of the second row ($A = 0.001$) and to a lesser extent the row above that, where the robust estimators outperform the traditional one.

As in Figure 4.9, the results in Figure 4.12 show the robust estimators outperforming the traditional Welch estimator under similar pairs of (Γ, A). Notice the observation from Figure 4.11—that the peak effective SNRs are around 9 decibels—can also be observed in Figure 4.12 in the rightmost column and top row.

By using effective SNR as the detection threshold, a qualitative sense of the probability of true detection can be gained from the histograms in Figures 4.9 and 4.12. The information shown in these histograms is similar to what is shown in Receiver Operating Characteristic (ROC) plots, except without the probability of false alarm. However, the probability of false alarm can be found by computing the effective SNR (i.e., using effective SNR as the detection threshold) on the noise samples alone. After obtaining the probabilities of false alarm, ROC plots can be generated. The ROC plots that result from this process are shown in Figure 4.13. If the SNR used to simulate a ROC plot is too low (or too high), then the ROC curves will each be linear (or form a right angle)—in which case any distinction in performance between the estimators cannot be ascertained. To prevent this, the $E_b/N_0^{(\mathrm{T})}$ differs among the plots depending on the values of Γ and A. The value of $E_b/N_0^{(\mathrm{T})}$ used for the simulation that generates a given plot is shown on that plot.

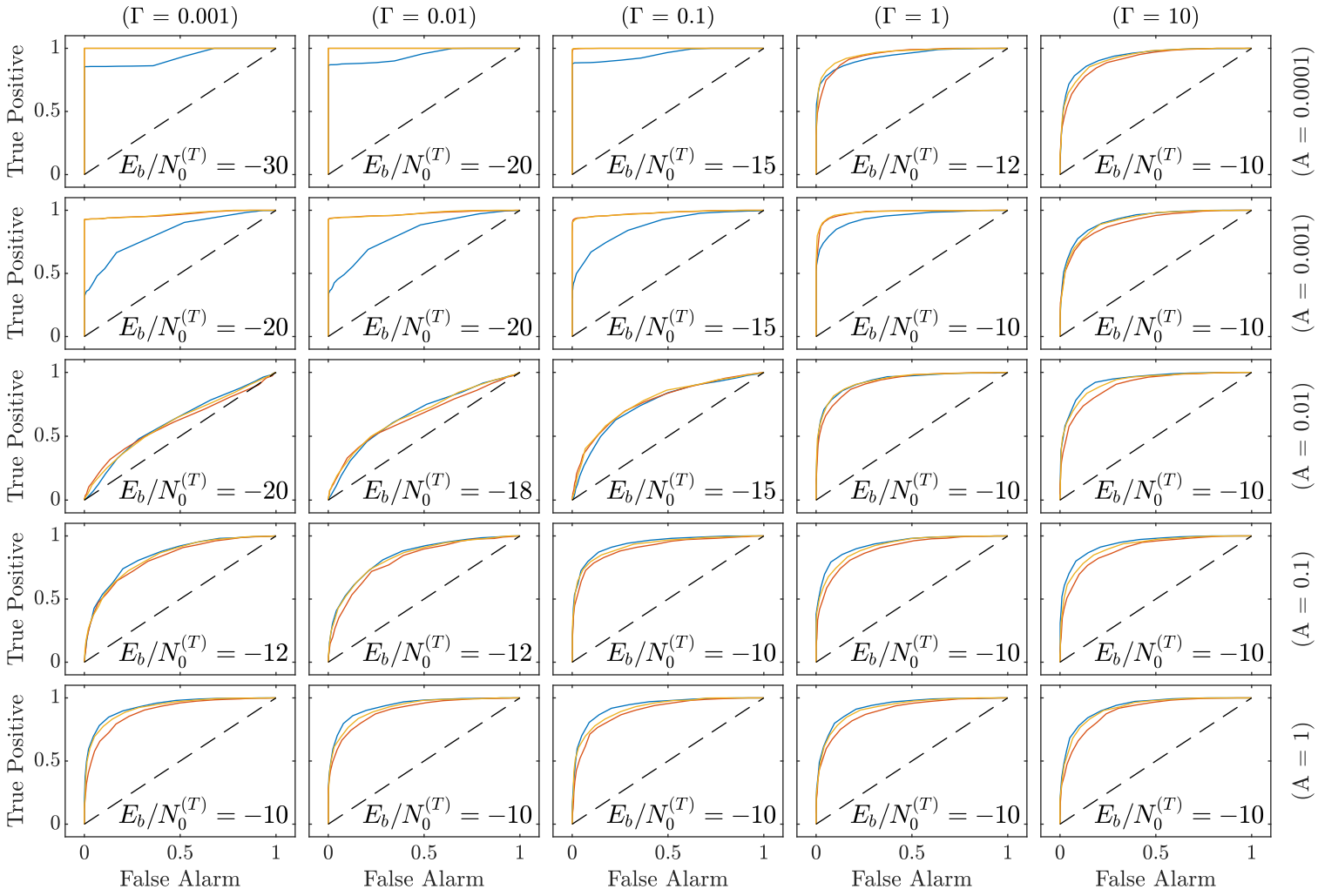

Figure 4.13: ROC plots generated using 1000 iterations for each estimator (traditional-blue; median-red; Huber-orange) in each plot. Note that the $E_b/N_0^{(\mathrm{T})}$ varies depending on the values of Γ and A to avoid generating ROC curves that are too linear or form a right angle.

It is demonstrated in Figure 4.13 that the robust estimators outperform the traditional Welch for the simulation conditions in the left-side plots of the top two rows. This is consistent with the histograms in Figures 4.9 and 4.12. Notice that in the other scenarios the traditional Welch estimator (blue) slightly outperforms the Huber-modified estimator (orange), which slightly outperforms the median-modified Welch estimator (red). This is an expected result in the less impulsive scenarios and demonstrates that the mean is a more optimal estimator of location than the Huber (which is more optimal than the median) for the noise conditions used to generate those particular plots.

In Chapter 3 it is remarked that these results are unaffected by the modulation scheme; no unique features of QPSK are utilized, and these results are expected to be equivalent for most modulation schemes. To support this hypothesis, Figure 4.14 was generated using the exact same parameters as for the simulation producing Figure 4.12 except the bits are modulated using 16-state Quadrature Amplitude Modulation (16-QAM) at $E_b/N_0^{(\mathrm{G})} = 2$ instead of QPSK at $E_b/N_0^{(\mathrm{G})} = 5$. A fair comparison requires the $E_b/N_0^{(\mathrm{G})}$ to be decreased by 3 decibels in order for the 16-QAM and QPSK signals to have equal power (i.e., the SNRs are equal).

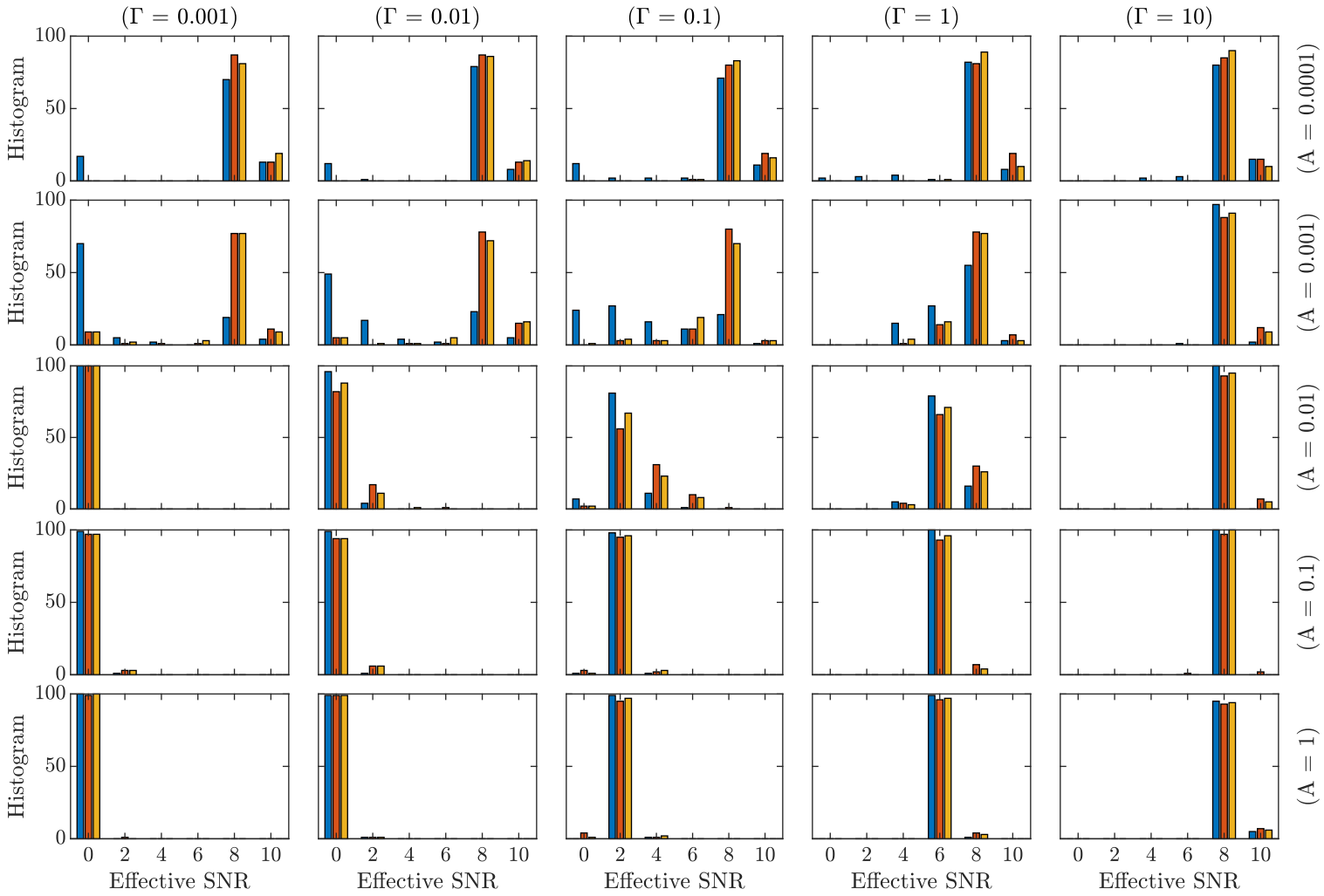

Figure 4.14: Distribution of effective SNR for 100 iterations for each estimator (traditional-blue; median-red; Huber-orange) in each plot. Notice that this figure looks almost identical to Figure 4.12.

Since the conditions for Figure 4.14 only differ from those used for Figure 4.12 in the modulation scheme used (16-QAM for Figure 4.14 and QPSK for Figure 4.12) and $E_b/N_0^{(\mathrm{G})}$ (to maintain equal SNRs), the differences between the results can be used to illuminate how modulation affects the robustness of the estimators. Since Figures 4.12 and 4.14 are produced using a stochastic simulation, the minor difference between the figures can reasonably be attributed to their stochastic nature. If the modulation scheme does affect the robustness of the PSD estimators, then the effect is small enough that it cannot be seen by comparing Figures 4.12 and 4.14; therefore, it is concluded that the robustness of the PSD estimators is hardly affected—or not affected—by modulation scheme.

The segment length is a very important consideration when computing a non-parametric PSD estimate. It is one parameter that controls the trade-off between low latency and fast computation time (for a shorter segment length) and frequency resolution (for a larger segment length). The particular segment length, L, can reasonably vary from 8 to 4096 (but is generally a power of two). For example, in [52], [53] L is analyzed on $[8, 1024]$ by powers of two; in [54] L is analyzed at lengths $\{128, 256, 512, 1024\}$; in [49] L is analyzed at lengths $\{1024, 2048, 4096\}$; and L is 64 in [50], [51], [55]. It is important to know that the performance is slightly altered depending on L. For shorter lengths, the robust methods can tolerate less impulsive noise (a larger A). Figure 4.15 is generated using the exact same scenario as used to generate the results in Figure 4.12 except with $L = 64$ instead of $L = 256$.

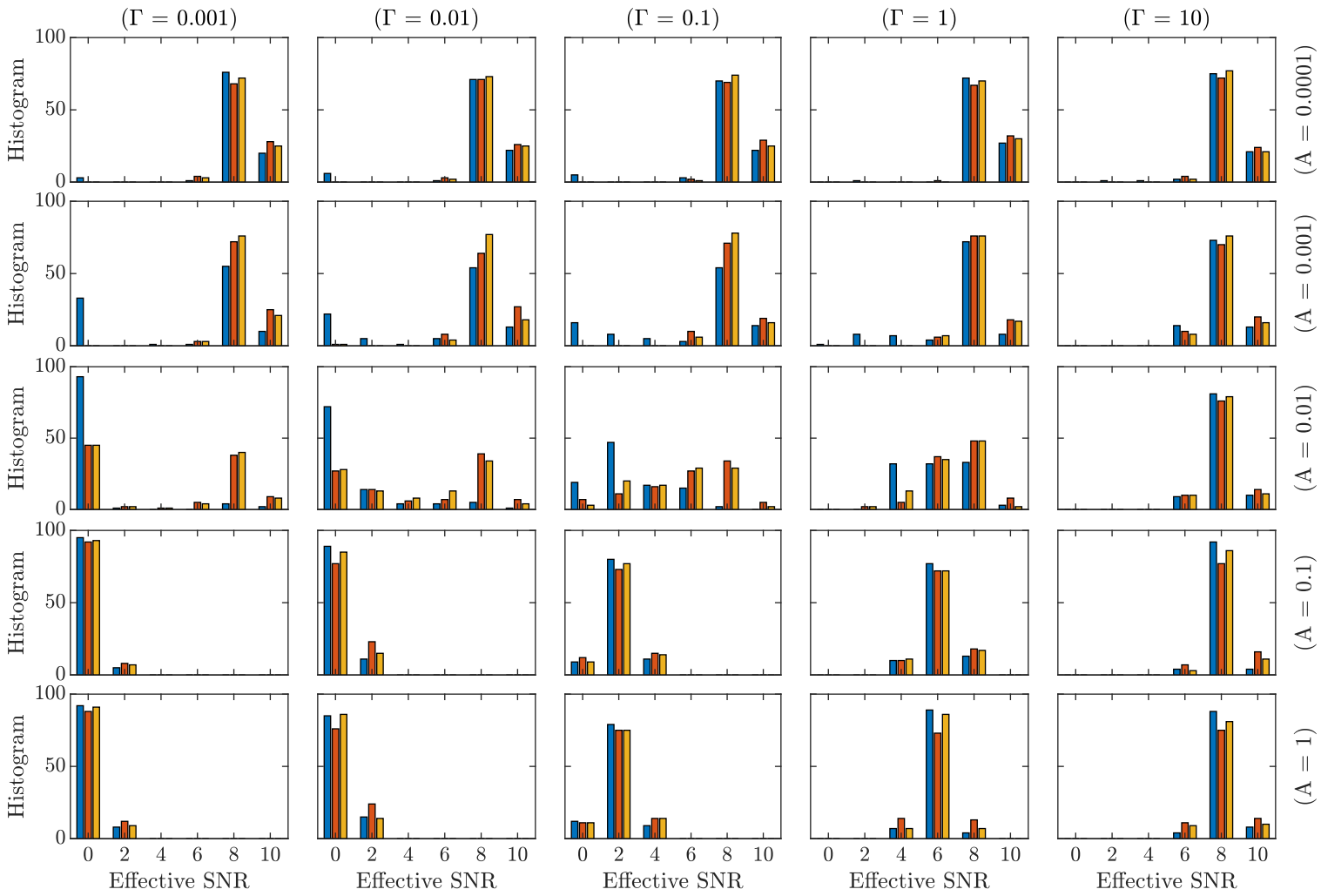

Figure 4.15: Distribution of effective SNR for 100 iterations for each estimator (traditional-blue; median-red; Huber-orange) in each plot. A segment length of 64 is used instead of the default length, 256. Notice that the plots in the middle row (primarily to leftmost two) show performance improvement for the robust methods.

The results in Figure 4.15 can be compared to the results in Figure 4.12 to see how segment length affects performance. Notice that a shorter segment length enables signal detection for less impulsive noise demonstrated by the third row ($A = 0.01$) in Figure 4.15 in contrast to the third row in Figure 4.12. This observation is important because it demonstrates that the parameters for the robust estimators can be selected to fit a given IN scenario by modifying the segment length.

The improvement for less impulsive noise is also seen in the ROC plots in Figure 4.16. The ROC plots in Figure 4.16 were generated using the exact same simulation conditions as used for Figure 4.13 except a segment length of 64 was used for the simulations used to generate Figure 4.16, instead of the default length of 256. The simulation conditions are equal to those used in Figure 4.15 except with different $E_b/N_0^{(\mathrm{T})}$.

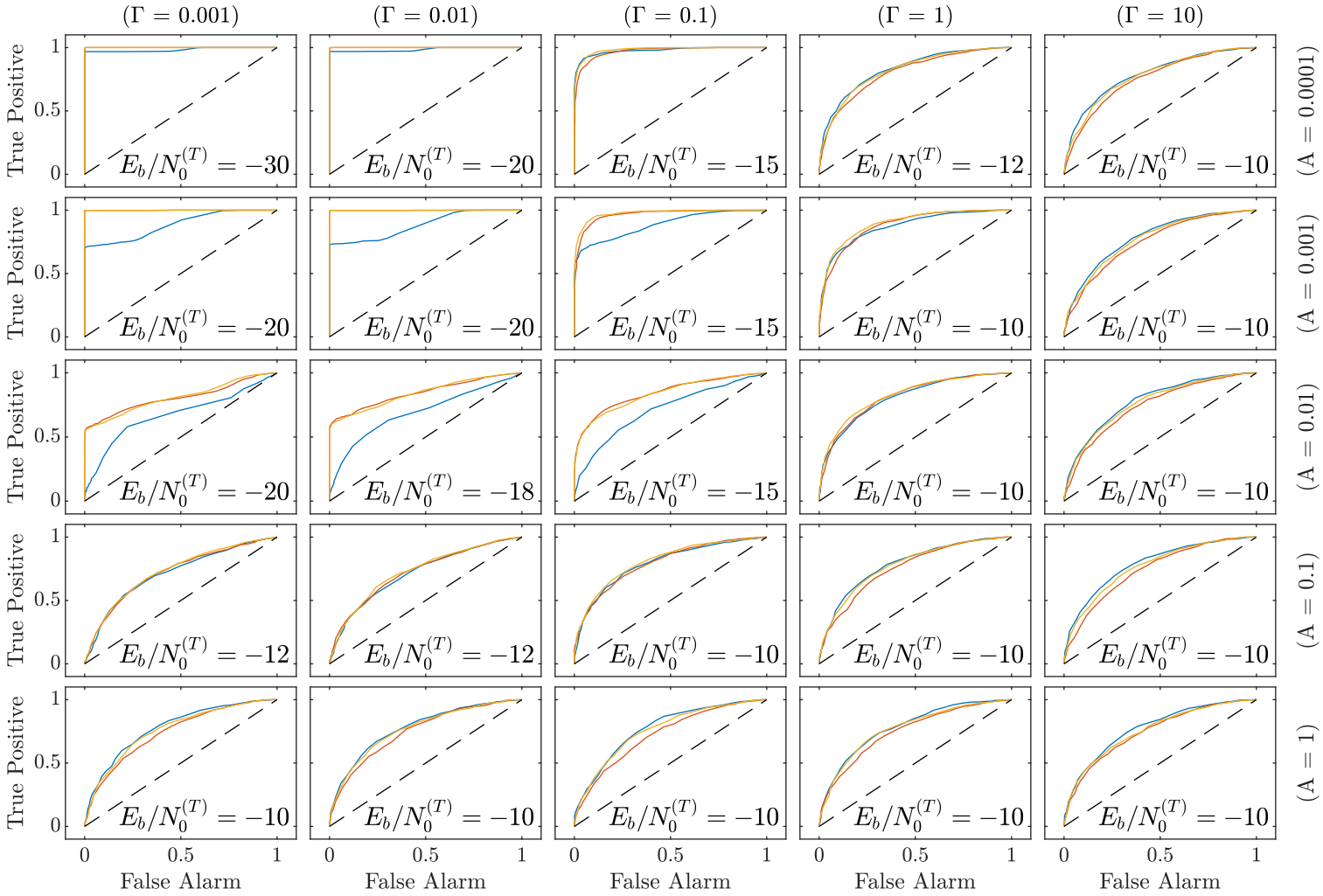

Figure 4.16: ROC plots generated using 1000 iterations for each estimator (traditional-blue; median-red; Huber-orange) in each plot. The conditions are also equal to those used in Figure 4.13 except the segment length is 64 for the simulation conditions in this figure rather than the default length of 256. Note that the performance for the robust methods has increased for the left three plots in the middle row (compared to Figure 4.13).

The results demonstrated in Figure 4.15 are confirmed in Figure 4.16: the conditions in which the robust estimators are advantageous are dependent on the segment length. For less IN (larger A), the segment length needs to be shorter in order to improve the detection robustness. This is most clearly seen by comparing the three leftmost ROC plots in the middle rows of Figures 4.13 and 4.16.

Using simple interpolation based on the ROC curves in Figures 4.13 and 4.16, the regions of increased performance can be estimated and are shown on the plot in Figure 4.3. Since Figures 4.13 and 4.16 are generated with logarithmically spaced values of Γ and A, the geometric mean is used to estimate the boundary between two consecutive values of Γ or A in which the robust estimators improve the performance for one set of (Γ, A) but not for an adjacent set. These regions of increased performance for the robust estimators are shown in Figure 4.17.

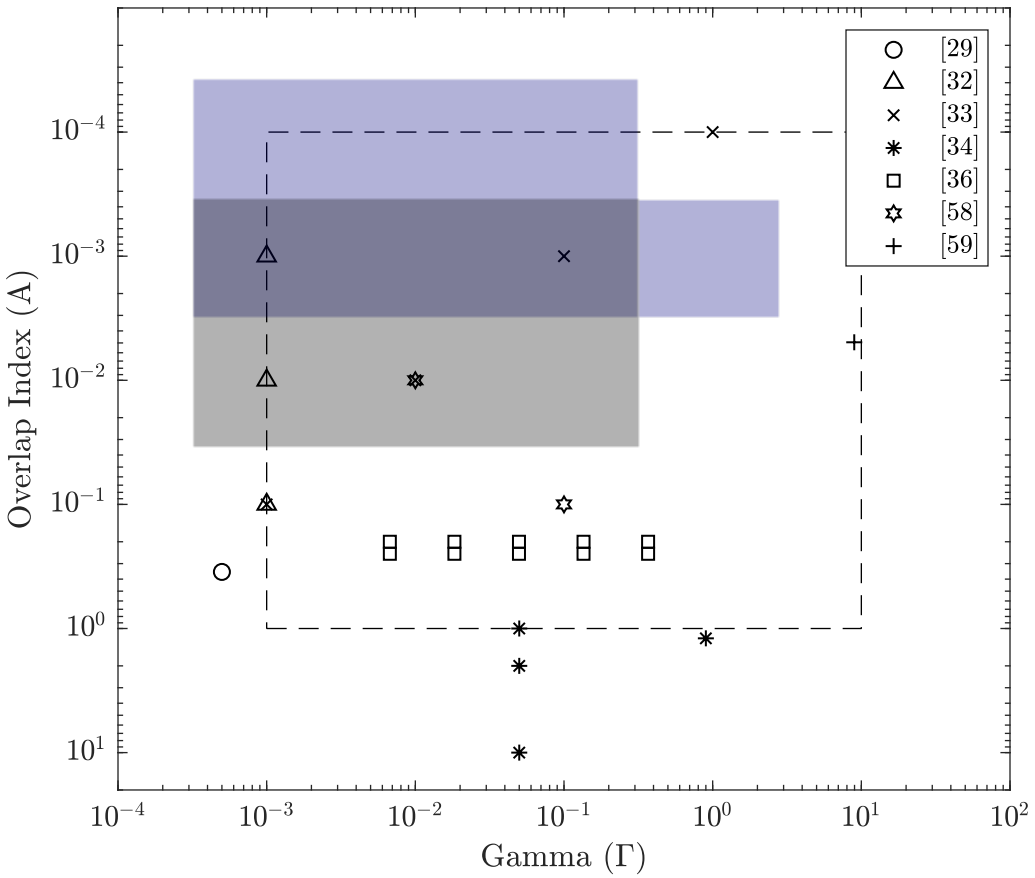

Figure 4.17: The range of Middleton Class A noise parameter pairs (Γ, A) found in the literature. The rectangular box represents the range examined in this thesis. The blue and grey areas cover the regions in which the robust Welch estimators outperform the traditional Welch estimator using a segment length of 256 or 64, respectively. The dark blue area is the region of overlap.

It is shown by Figure 4.17 that the segment length is an important parameter to consider when attempting robust estimation. Ideally the segment length can be chosen to increase the robustness for the region of interest—that is, the expected noise for a given application can be accurately characterized by the Class A noise produced using the values of Γ and A in one of the shaded regions in Figure 4.17. From a more general signal processing perspective, other factors influence the choice of segment length (namely frequency resolution), but it is important to realize that if robustness is a concern, then the segment length also affects the performance of the robust Welch estimators.

4.3 Multitaper Results

Before the robustness of any multitaper based method is analyzed, it is first useful to understand the traditional variance versus bias trade-off. To this end, Figure 4.18 has been created. The figure is designed to show a comparison between the three prior methods: periodogram, Welch method, and median-modified Welch method and the corresponding multitaper version of these methods: MTM, MultiTaper Welch Method (MTWM), and Median-Modified MultiTaper Welch Method (M3TWM). To make the comparisons fair, the prior methods should be tapered using the $p = 0$ taper to attain relatively similar spectral leakage properties between each comparison. Applying the taper converts the periodogram into the $p = 0$ taper-modified periodogram, the Welch method technically remains the Welch method (except windowed with the $p = 0$ taper)[1], and the median-modified Welch method is converted into the $p = 0$ Taper-Median-Modified Welch Method (TMMWM). Recall that Table 3.1 comprises these methods, their descriptions, and how they relate to one another.

[1]The official definition of the Welch method has a windowing term. The definition used throughout this thesis (3.7) does not show the windowing term (thereby using a rectangular window by default), but for the purposes of the comparison in Figure 4.18 a non-rectangular window has been used.

Note that because a bias comparison is being made, the PSD estimates cannot be normalized independently. Therefore, the multitaper estimates are normalized such that their maximum value is zero decibels, and the other methods are normalized relative to the corresponding multitaper method.

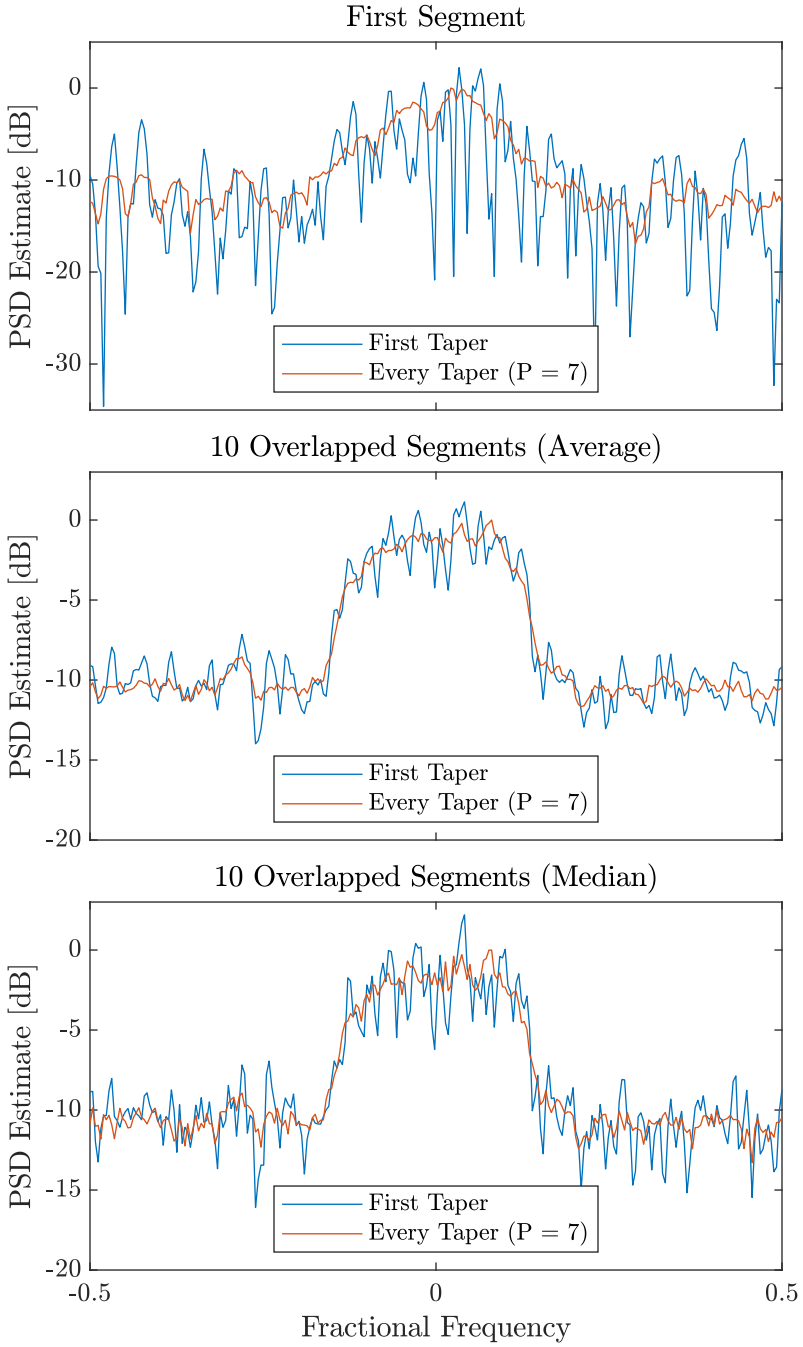

Figure 4.18: Demonstration of how multitapering affects variance versus bias trade-off. Using additional tapers (orange) instead of only the first taper (blue) slightly increases the bias but drastically decreases the variance for each estimator—single segment (top), overlapped averaging (middle), and overlapped with median (bottom).

The variance can be quantitatively analyzed by first computing 10 000 of each PSD estimate using the parameters in Figure 4.18, and then finding the variance for each estimator and frequency bin. The results of this procedure are shown in Figure 4.19.

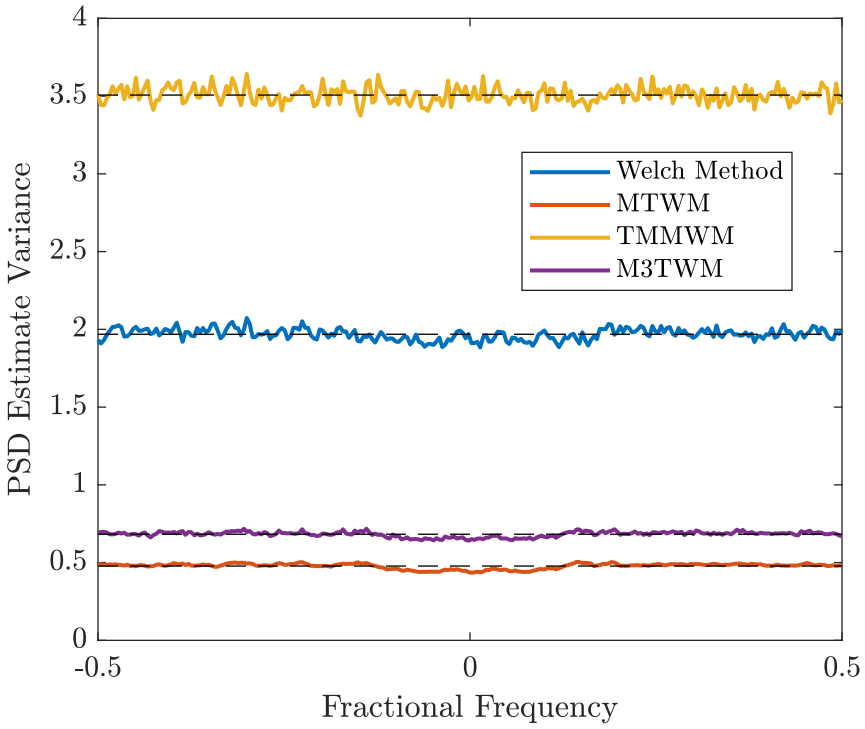

Figure 4.19: PSD estimation variance for the four estimators used to generate the middle and bottom plots of Figure 4.18. The dashed lines are the average variances for each estimator across the frequency bins. For $P = 7$, applying the MTM to the Welch estimator reduces the variance by 75% (blue to red). Since the median is a less efficient estimator than the mean for the Chi-Squared distribution, the variance for the TMMWM is almost double the variance of the Welch estimator (blue to yellow). However, applying the MTM to the TMMWM reduces the variance by almost 80% (yellow to purple).

Without getting lost in the details, Figure 4.19 shows that using multiple tapers—as opposed to a single taper or window—drastically decreases variance (but also increases computational effort). Additionally, the median is shown to be less efficient than the mean (for this scenario), and therefore supplanting the mean with the median increases the variance.

The performance of the multitaper based methods can now be analyzed. As in Figures 4.9 and 4.12, histograms and effective SNR are used to quantify performance metrics. Three methods will be compared: the MTWM, the M3TWM (the robust version of MTWM), and the $p = 0$ TMMWM (the robust non-multitapered benchmark method for the M3TWM). The simulation conditions for the plots in Figures 4.20 and 4.21 are the exact same as those use to generate Figures 4.9 and 4.12, respectively. The difference is that the MTWM (blue), TMMWM (red), and M3TWM (orange) estimators are used to generate Figures 4.20 and 4.21 instead of the traditional Welch, median-modified Welch, and Huber-modified Welch estimators as in Figures 4.9 and 4.12.

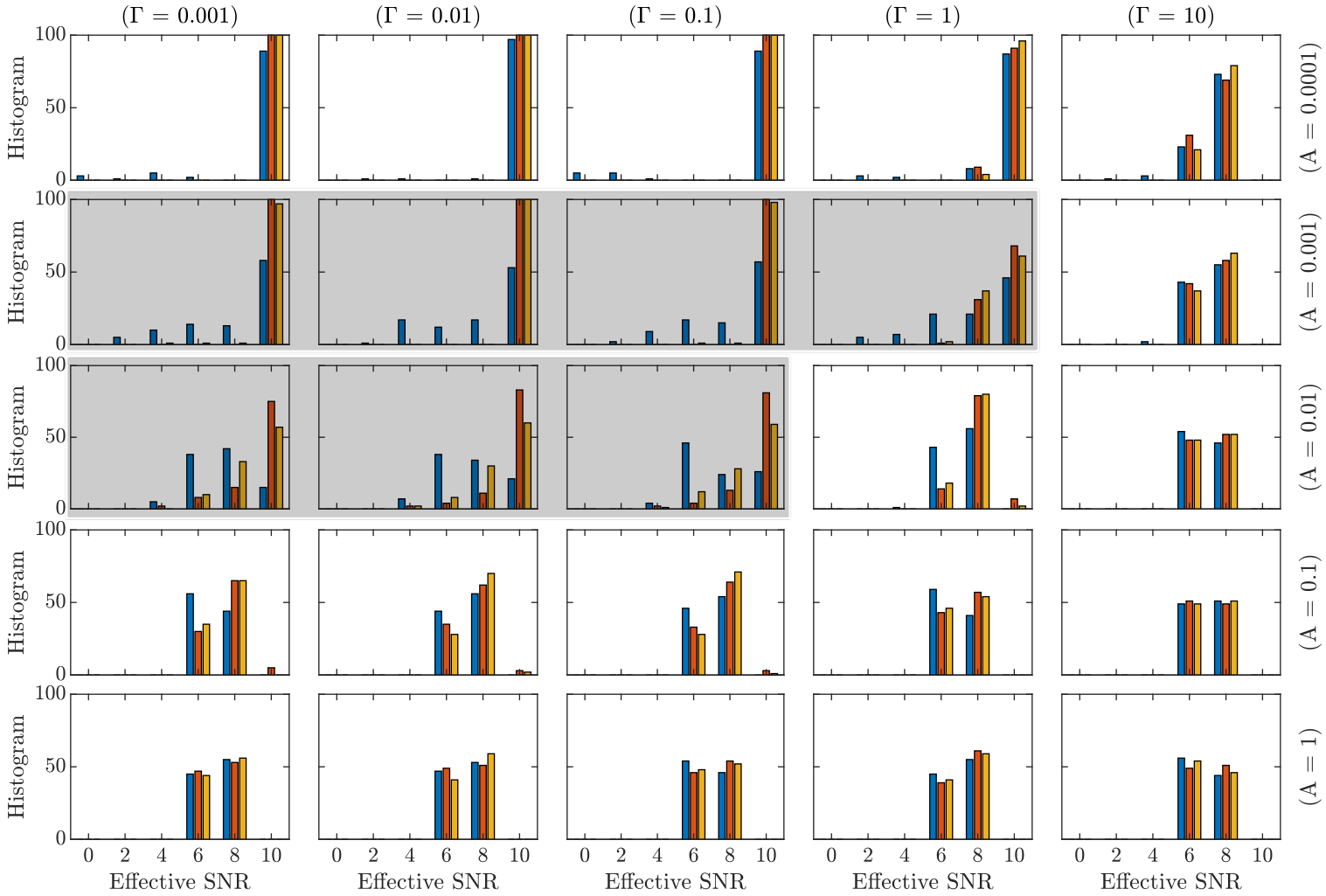

Figure 4.20: Distribution of effective SNR for 100 iterations for each estimator (MTWM-blue; TMMWM-red; M3TWM-orange) in each plot. The $E_b/N_0^{(\mathrm{T})} = 3$ for the simulation conditions for each plot. For certain pairs of (Γ, A) (primarily the values that correspond to the highlighted plots in the second and third row), the M3TWM (orange) is shown to be more robust than its non-robust counterpart (MTWM, blue). However, the robust non-multitapered benchmark (TMMWM, red) is more robust than the M3TWM (as seen in the middle row). This relationship roughly equates to a robustness versus variance trade-off.

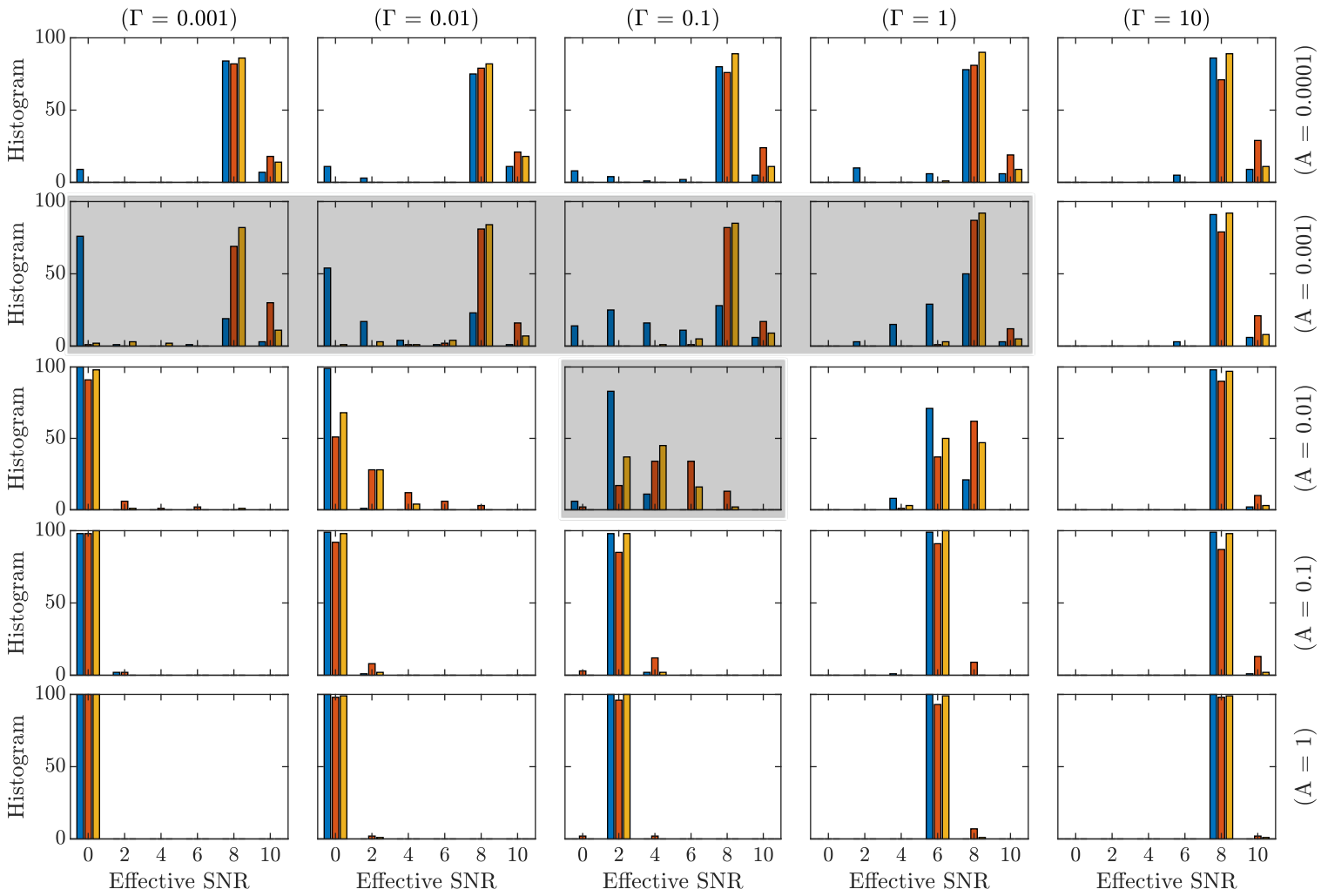

Figure 4.21: Distribution of effective SNR for 100 iterations for each estimator (MTWM-blue; TMMWM-red; M3TWM-orange) in each plot. The $E_b/N_0^{(\mathrm{G})} = 5$ for the simulation conditions for each plot. Similar to Figure 4.20, the estimators perform comparably with the exception of under the conditions of the highlighted plots and to a lesser extent the top row, where the TMMWM (red) is more robust than the M3TWM (orange) which is much more robust than the MTWM (blue).

Figures 4.20 and 4.21 show that using multiple tapers with M3TWM is nearly as robust as using only the first taper (TMMWM)—although it has been shown that the former estimator produces substantially lower variance estimates. As expected, using the median for obtaining the frequency bin estimates across the overlapped segments for each taper (as in M3TWM) is much more robust than using the mean (as in MTWM).

Keep in mind that the only difference between the conditions used to simulate Figures 4.20 and 4.21 is that $E_b/N_0^{(\mathrm{T})} = 3$ in Figure 4.20 and $E_b/N_0^{(\mathrm{G})} = 5$ in Figure 4.21. This is done for completeness, because sometimes it is useful to view performance across different pairs of (Γ, A) keeping the total power constant (Figure 4.20) and other times one might want to keep the underlying GN power constant (Figure 4.21). This difference causes an interesting effect that can be seen by comparing the top highlighted plots in Figures 4.20 to those in 4.21: in Figure 4.20 the non-robust *and* robust estimators perform consistently (albeit, differently) whereas in Figure 4.21 only the robust estimators perform consistently—the non-robust estimator (MTWM, blue) improves as Γ increases. This is because the $E_b/N_0^{(\mathrm{T})}$ is increasing with Γ in Figure 4.21. However, the robust estimators continue to maintain a relatively steady effective SNR even as the $E_b/N_0^{(\mathrm{T})}$ changes.

The results in Section 4.2 showed that the Welch can be made more robust by supplanting the mean with a more robust estimate of location. This was demonstrated by both the median and the Huber estimator. Across the reasonable set of Middleton Class A parameters each method either performed similarly in its detection capabilities, or the robust methods outperformed the traditional method. The primary costs for the robust methods are computational effort, which is seen to be a small increase relative to the effort of the traditional Welch. Additionally, the robust methods (especially the median) are less efficient (as defined by robust estimation parlance), and therefore result in higher variance estimates.

The results in Section 4.3 are produced by applying the concepts from the robust Welch

methods with the MTM. The addition of the MTM allows for a larger set of uncorrelated estimates (because each Welch segment produces P tapered-modified periodograms) which are combined to produce a final low variance PSD estimate. A brief analysis of the trade-offs between variance, bias, and computational effort was performed. Significant variance reduction can be achieved by applying MTM concepts to the robust Welch methods. The cost is bias and computational effort.

4.4 Future Work

As stated in the introduction, impulsive noise is characterized by the pulse amplitude, pulse width, and inter-arrival time [18]. Since a Middleton Class A noise process produces Independent and Identically Distributed (IID) noise samples, the pulse width cannot be modified unless the model is generalized with an additional time correlation parameter as in [26]. It is remarked in [26], [60], [61] that some Power Line Communication (PLC) noise is not accurately modeled by Class A noise since PLC noise is often correlated in time. Therefore, a natural extension to the work presented in this thesis would be to analyze the robust estimation methods using the model presented in [26]. The author intuits that increasing time correlation would improve the performance relative to the uncorrelated conditions (assuming the other parameters are kept equivalent). However, this has not been tested.

Chapter 5

Conclusion

The confluence of nonparametric spectral estimation and robust estimation theory finds application in the domain of electrical communications in Cognitive Radio (CR) and Power Line Communication (PLC). In this thesis it was shown that popular nonparametric estimation techniques—namely, the Welch method and the MultiTaper Method (MTM)—were not designed with robustness in mind. As designed, they produce Power Spectral Density (PSD) estimates with low variance and low bias in Additive White Gaussian Noise (AWGN) environments. By applying simple robust estimation techniques, robustness against Impulsive Noise (IN) can be achieved for these methods at the cost of computational effort and variance. This allows signal detection or more accurate non-impulsive noise spectral estimation by mitigating the influence of IN.

The Middleton Class A Non-Gaussian Noise (NGN) model is used to test the performance of the robust estimators presented. The Class A model can generate Independent and Identically Distributed (IID) NGN samples with varying pulse amplitude and inter-arrival time. The robust estimators are compared against the traditional versions in a variety of

noise environments that span most of the parameter space found in the literature. There are roughly three performance outcomes (in terms of robust signal detection) depending on the environment: all the estimators perform adequately, none of the estimators perform adequately, or the robust estimators outperform the traditional estimators.

It was found that as the noise becomes more impulsive, the separation between the detection capabilities increases between the robust estimators and the traditional estimators (in favor of the robust estimators); each estimator performs similarly for non-impulsive noise, regardless of whether or not the noise is Gaussian or non-Gaussian. Additionally, choosing smaller segment lengths increases robustness for IN characterized by shorter inter-arrival times, but this comes at the cost of frequency resolution; after accounting for different segment lengths, it is possible to increase the robustness of the Welch estimator for IN scenarios with a wider range of inter-arrival times and impulsiveness.

The Welch method is modified by replacing the averaging operation by more robust estimators of location (namely, the median and the Huber estimator). It is well-known that the median is very robust but not an efficient estimator of location for the chi-square distribution (which characterizes the distribution of frequency bin samples obtained by computing the periodogram of a Gaussian process). Therefore, the variance reduction intended by the Welch method is diminished when exchanging the averaging operation for more robust estimators of location. For this reason, estimators such as the Huber are often used because (for the distributions in this thesis) the Huber estimator is almost as robust as the median but substantially more efficient.

The concept of orthogonal tapering from the MTM is applied to the overlapped segmentation from the Welch method. When the same robustness concepts are also applied, a new estimator called the Median-Modified MultiTaper Welch Method (M3TWM) is developed. This estimator contrasts with the robust Welch estimators in that further variance

reduction can be achieved for a given set of samples. However, the M3TWM is slightly more biased (though not enough to matter in most applications); the real concern is that the computational effort increases. Depending on the parameters chosen, the increase could be anywhere from twofold to tenfold. More effort would result in lower variance but a wider spectral window (which results in spectral leakage—bias).

Appendix A

Supporting Math for Class A Model

A.1 Demonstrate that the Class A PDF is a PDF

If for all $\{\epsilon, A, \Gamma, \sigma^2\} \in \mathbb{R}_0^+$ it can be shown that both $f_E(\epsilon) \geq 0$ and $\int_{-\infty}^{\infty} f_E(\epsilon)\mathrm{d}x = 1$, then $f_E(\epsilon)$ is a legitimate Probability Distribution Function (PDF).

By inspection one can see that $f_E(\epsilon) \geq 0$. Showing the second condition:

$$\begin{aligned}\int_{-\infty}^{\infty} f_E(\epsilon)\,\mathrm{d}\epsilon &= \int_0^{\infty} f_E(\epsilon)\,\mathrm{d}\epsilon \\ &= \int_0^{\infty} \sum_{m=0}^{\infty} P_m \frac{\epsilon}{\sigma_m^2} e^{-\frac{\epsilon^2}{2\sigma_m^2}}\,\mathrm{d}\epsilon \\ &= \sum_{m=0}^{\infty} P_m \int_0^{\infty} \frac{\epsilon}{\sigma_m^2} e^{-\frac{\epsilon^2}{2\sigma_m^2}}\,\mathrm{d}\epsilon\end{aligned} \tag{A.1}$$

The integral can be solved with the following u substitution: $u \triangleq \frac{\epsilon^2}{2\sigma_m^2}$, $\mathrm{d}u = \frac{\epsilon}{\sigma_m^2}\mathrm{d}\epsilon$.

$$\begin{aligned}
\int_{-\infty}^{\infty} f_E(\epsilon)\,\mathrm{d}\epsilon &= \sum_{m=0}^{\infty} P_m \int_0^{\infty} e^{-u}\,\mathrm{d}u \\
&= \sum_{m=0}^{\infty} P_m \left(-e^{-u}\Big|_0^{\infty}\right) \\
&= \sum_{m=0}^{\infty} \frac{e^{-A}A^m}{m!} \\
&= e^{-A} \underbrace{\sum_{m=0}^{\infty} \frac{A^m}{m!}}_{=e^A} \\
&= 1
\end{aligned} \tag{A.2}$$

Therefore, $f_E(\epsilon)$ is an actual PDF.

A.2 Second Moment of Rayleigh Distributed Random Variable

A Rayleigh random variable R is defined by the following PDF.

$$R \sim f(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}, \quad r \in [0, \infty) \tag{A.3}$$

The second moment can be computed by definition.

$$\begin{aligned} \mathbb{E}\left\{R^2\right\} &= \int_{-\infty}^{\infty} r^2 f(r)\,\mathrm{d}r \\ &= \int_{0}^{\infty} r^2 f(r)\,\mathrm{d}r \\ &= \int_{0}^{\infty} r^2 \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}\,\mathrm{d}r \end{aligned} \tag{A.4}$$

The integral can be simplified using integration by parts with the following assignments for u and $\mathrm{d}v$.

$$\begin{aligned} \int_a^b u\,\mathrm{d}v &= uv\Big|_a^b - \int_a^b v\,\mathrm{d}u \\ u \triangleq r^2 \quad &\Rightarrow \quad \mathrm{d}u = 2r\,\mathrm{d}r \\ \mathrm{d}v \triangleq \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}\,\mathrm{d}r \quad &\Rightarrow \quad v = -e^{-\frac{r^2}{2\sigma^2}} \end{aligned} \tag{A.5}$$

Using the above integration by parts yields:

$$\begin{aligned}
\mathbb{E}\left\{R^2\right\} &= -r^2 e^{-\frac{r^2}{2\sigma^2}}\Big|_0^\infty - \int_0^\infty -2re^{-\frac{r^2}{2\sigma^2}}\,\mathrm{d}r \\
&= 2\sigma^2 \int_0^\infty \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}\,\mathrm{d}r \\
&= 2\sigma^2 \underbrace{\int_0^\infty f_R(r)\,\mathrm{d}r}_{=1} \\
&= 2\sigma^2
\end{aligned} \tag{A.6}$$

where the last integral evaluates to unity because (1) $f_R(r)$ is the PDF of a Rayleigh Random Variable (RV) which is defined on the interval $[0, \infty)$ and (2) integration over the domain of any PDF is always unity.

A.3 Two Useful Views of a Random Variable Satisfying a Given PDF

Define N arbitrary independent RVs $Z_1, \ldots, Z_N$ that follow PDFs $f_{Z_1}(x), \ldots, f_{Z_N}(x)$, respectively, each with arbitrary mean and variance. Define two new RVs X and Y as follows:

$$\begin{aligned} X \sim f_X(x) &\triangleq \sum_{n=1}^{N} a_n f_{Z_n}(x) \\ Y &\triangleq Z_n \quad \text{w.p.} \quad a_n \\ a_n &\in [0,1] \\ \sum_{n=1}^{N} a_n &= 1 \end{aligned} \tag{A.7}$$

Show that the PDFs of X and Y are equal. That is, X and Y have the same statistics.

Proof. By definition, the Cumulative Distribution Function (CDF) of X is:

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f_X(\hat{x})\,\mathrm{d}\hat{x} \\ &= \int_{-\infty}^{x} \sum_{n=1}^{N} a_n f_{Z_n}(\hat{x})\,\mathrm{d}\hat{x} \\ &= \sum_{n=1}^{N} a_n \int_{-\infty}^{x} f_{Z_n}(\hat{x})\,\mathrm{d}\hat{x} \\ &= \sum_{n=1}^{N} a_n F_{Z_n}(x) \end{aligned} \tag{A.8}$$

The CDF of Y can be computed via the total probability theorem as follows:

$$\begin{aligned} F_Y(y) &= P[Y \leq y] \\ &= \sum_{n=1}^{N} P[Y \leq y \mid Y = Z_n] P[Y = Z_n] \\ &= \sum_{n=1}^{N} P[Z_n \leq y] a_n \\ &= \sum_{n=1}^{N} a_n F_{Z_n}(y) \end{aligned} \tag{A.9}$$

Since the CDF of X is equal to the CDF of Y, their PDFs are also equal. This result is useful since the RV that characterizes the complex envelope of Class A noise can be expressed in the form of X, so therefore that RV can also be expressed in the form of Y. This application is also suitable for the RVs that characterize the IQ components of Class A noise. The author has found it ideal to perform mathematical manipulation upon RVs using the form of X while using the form of Y is ideal for simulation. □

A.4 Method for Computing Second Moments of Middleton Class A Noise

Define N independent RVs $X_1, \ldots, X_N$ that follow PDFs $f_1(x), \ldots, f_N(x)$ each with arbitrary mean and with second moments $2\sigma_1^2, \ldots, 2\sigma_N^2$, respectively. (This odd notation for the second moments is motivated by the notation for the second moment of the Rayleigh RV shown in Appendix A.2). Define a new RV Y s.t.

$$\begin{aligned} Y \sim f(y) &\triangleq \sum_{n=1}^{N} a_n f_n(y) \\ a_n &\in [0,1] \\ \sum_{n=1}^{N} a_n &= 1 \end{aligned} \tag{A.10}$$

By definition, the second moment of Y can be computed as shown below.

$$\begin{aligned} \mathbb{E}[Y^2] &= \int_{-\infty}^{\infty} y^2 f(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} y^2 \sum_{n=1}^{N} a_n f_n(y)\,\mathrm{d}y \\ &= \sum_{n=1}^{N} a_n \int_{-\infty}^{\infty} y^2 f_n(y)\,\mathrm{d}y \\ &= \sum_{n=1}^{N} a_n 2\sigma_n^2 \end{aligned} \tag{A.11}$$

A.5 Second Moment Class A Noise Envelope

Compute the power of Class A noise using the complex envelope, $E \sim f_E(\epsilon)$. This is simply the second moment of E—the random variable that represents the magnitude of the complex envelope of Class A noise.

$$\mathbb{E}\left\{E^2\right\} = \sum_{m=0}^{\infty} P_m 2\sigma_m^2 \tag{A.12}$$

Substituting $P_m = \frac{e^{-A}A^m}{m!}$ and $2\sigma_m^2 = \sigma^2 \frac{\frac{m}{A}+\Gamma}{1+\Gamma}$ as shown in (2.2) as part of the Class A model yields:

$$\begin{aligned}
\mathbb{E}\left\{E^2\right\} &= \sum_{m=0}^{\infty} \frac{e^{-A}A^m}{m!}\sigma^2 \frac{\Gamma + \frac{m}{A}}{1+\Gamma} \\
&= \frac{\sigma^2 e^{-A}}{1+\Gamma}\left[\Gamma \sum_{m=0}^{\infty} \frac{A^m}{m!} + \sum_{m=0}^{\infty} \frac{A^m}{m!}\frac{m}{A}\right] \\
&= \frac{\sigma^2 e^{-A}}{1+\Gamma}\left[\Gamma e^A + \sum_{m=1}^{\infty} \frac{A^{(m-1)}}{(m-1)!}\right] \\
&= \frac{\sigma^2 e^{-A}}{1+\Gamma}\left[\Gamma e^A + e^A\right] \\
&= \sigma^2
\end{aligned} \tag{A.13}$$

A.6 Mapping Class A Complex Envelope and Phase Representation to In-Phase and Quadrature Representation

The joint PDF of the complex envelope and phase for Class A noise is shown in (2.12) and shown below.

$$f_{E,\Phi}(\epsilon,\phi) = f_E(\epsilon) f_\Phi(\phi) = \sum_{m=0}^{\infty} P_m \frac{\epsilon}{2\pi\sigma_m^2} e^{-\frac{\epsilon^2}{2\sigma_m^2}} \tag{A.14}$$

Using the following variable mapping will map the complex envelope (E) and phase (Φ) RVs to in-phase and quadrature RVs. This is analogous to mapping polar coordinates to rectangular coordinates.

$$\begin{aligned} X &\triangleq E\cos(\Phi) \\ Y &\triangleq E\sin(\Phi) \end{aligned} \tag{A.15}$$

The random variable assignment in (A.15) implies the following mapping between the (x, y) and (ϵ, ϕ) domains.

$$\begin{aligned} x &= g_1(\epsilon,\phi) = \epsilon\cos(\phi) \\ y &= g_2(\epsilon,\phi) = \epsilon\sin(\phi) \\ \epsilon &= h_1(x,y) = \sqrt{x^2+y^2} \\ \phi &= h_2(x,y) = \tan^{-1}\left(\frac{y}{x}\right) \end{aligned} \tag{A.16}$$

When the mapping from one domain to another is well defined, variable transforms are often performed using the Jacobian. The Jacobian for this variable mapping is shown below.

$$\mathcal{J}(\epsilon,\phi) = \det\begin{bmatrix} \frac{\partial x}{\partial \epsilon} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial \epsilon} & \frac{\partial y}{\partial \phi} \end{bmatrix} = \det\begin{bmatrix} \cos(\phi) & -\epsilon\sin(\phi) \\ \sin(\phi) & \epsilon\cos(\phi) \end{bmatrix} = \epsilon \tag{A.17}$$

Now the in-phase and quadrature joint PDF can be written as:

$$\begin{gathered} f_{X,Y}(x,y) = f_{E,\Phi}(h_1(x,y), h_2(x,y)) \cdot \frac{1}{|\mathcal{J}(\epsilon,\phi)|} = \sum_{m=0}^{\infty} P_m \frac{1}{2\pi\sigma_m^2} e^{-\frac{x^2+y^2}{2\sigma_m^2}} \\ x \in (-\infty,\infty), \quad y \in (-\infty,\infty) \end{gathered} \tag{A.18}$$

A.7 Showing Class A IQ Samples are Dependent and Uncorrelated

Though some readers might be comfortable concluding that the IQ components of Class A noise are dependent by mere inspection of (2.14) and (2.15), this attribute is explicitly shown here. To show that the IQ components of Class A noise are dependent, it only needs to be demonstrated that the joint PDF shown in (2.14) is not equal to the product of its marginal PDFs, which are shown in (2.15), for any valid set $\{x, y, \sigma^2, \Gamma, A\}$. Selecting $x = y = 0$ and $\sigma^2 = \Gamma = A = 1$ facilitates easy simplification. The joint PDF simplifies to:

$$\begin{aligned} f_{X,Y}(0,0)\Big|_{\sigma^2=\Gamma=A=1} &= \sum_{m=0}^{\infty} \frac{e^{-1}}{m!} \frac{2}{\pi(m+1)} \\ &= \frac{2}{\pi e} \sum_{m=0}^{\infty} \frac{1}{(m+1)!} \\ &= \frac{2(e-1)}{\pi e} \\ &\approx 0.402\ldots \end{aligned} \tag{A.19}$$

Performing the same procedure for the multiplication of the marginal PDFs yields:

$$\begin{aligned} [f_X(0) \cdot f_Y(0)]\Big|_{\sigma^2=\Gamma=A=1} &= \left[\sum_{m=0}^{\infty} \frac{e^{-1}}{m!} \sqrt{\frac{2}{\pi(m+1)}}\right]^2 \\ &= \frac{2}{\pi e^2} \left[\sum_{m=0}^{\infty} \frac{1}{m!\sqrt{m+1}}\right]^2 \\ &\approx 0.381\ldots \end{aligned} \tag{A.20}$$

Showing that the joint PDF is not equal to the product of the marginals for this one set of values is sufficient to demonstrate dependence.

Using the definition of covariance in (2.16), the uncorrelated nature of Class A IQ samples can quickly be verified. This is shown below.

$$\begin{aligned}
\mathrm{Cov}(X,Y) &\triangleq \mathbb{E}[(X - \underbrace{\mathbb{E}[X]}_{=0})(Y - \underbrace{\mathbb{E}[Y]}_{=0})] \\
&= \mathbb{E}[XY] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{X,Y}(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \sum_{m=0}^{\infty} P_m \frac{1}{2\pi\sigma_m^2} e^{-\frac{x^2+y^2}{2\sigma_m^2}}\,\mathrm{d}x\,\mathrm{d}y \\
&= \sum_{m=0}^{\infty} P_m \underbrace{\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{-\frac{x^2}{2\sigma_m^2}}\,\mathrm{d}x}_{=0} \underbrace{\int_{-\infty}^{\infty} y \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{-\frac{y^2}{2\sigma_m^2}}\,\mathrm{d}y}_{=0} \\
&= 0
\end{aligned} \tag{A.21}$$

It can be seen that $\mathbb{E}[X]$ and $\mathbb{E}[Y]$ are zero because their respective PDFs are even functions. The fourth line is obtained by substitution of (2.14). The infinite sum prevents (2.14) from being factored into a function of x and a function of y, so the order of the sum and double integral are switched to facilitate factoring the integrand (and consequently easy simplification). The last integrals all equal zero because the integrals are expectations of zero mean Gaussian random variables with variances σ_m^2. Therefore, since the covariance of X and Y is zero, the IQ samples are uncorrelated.

Appendix B

Estimation Proofs

B.1 Wiener-Khinchin Theorem

For convenience, the intuitive definition of the Power Spectral Density (PSD) (3.1) is shown in (B.1),

$$\phi^{(\mathrm{i})}(f) \triangleq \lim_{M\to\infty} \mathbb{E}\left\{ \frac{1}{2M+1} \left| \sum_{n=-M}^{M} x[n] e^{-j2\pi f n} \right|^2 \right\}, \quad f \in \left[-\frac{1}{2}, \frac{1}{2}\right] \tag{B.1}$$

the AutoCorrelation Function (ACF) (3.2a) is shown in (B.2),

$$r[\ell] \triangleq \mathbb{E}\{x[n]x^*[n-\ell]\} \tag{B.2}$$

and the Wiener-Khinchin theorem (3.2b) is shown in (B.3)

$$\phi^{(\mathrm{wk})}(f) = \sum_{\ell=-\infty}^{\infty} r[\ell] e^{-j2\pi f \ell}, \quad f \in \left[-\frac{1}{2}, \frac{1}{2}\right] \tag{B.3}$$

where $x[n]$ is a Wide-Sense Stationary (WSS) Random Process (RP).

The Wiener-Khinchin theorem states that the PSD, as intuitively defined by (B.1), is equal to (B.3)[1]. Often the particular equation (B.3) is referred to as the Wiener-Khinchin theorem, but the theorem properly links (B.1) to (B.3).

Rewriting the squared magnitude operation in (B.1) and utilizing the complex-conjugate operator yields:

$$\phi^{(\mathrm{i})}(f) = \lim_{M\to\infty} \mathbb{E}\left\{ \frac{1}{2M+1} \sum_{n=-M}^{M} \sum_{m=-M}^{M} x[n]x^*[m]e^{-j2\pi f(n-m)} \right\} \tag{B.4}$$

Since the expectation operator is linear, the order of the expectation and summation operators can be swapped. Substituting in (B.2) (with a change of variable) gives:

$$\phi^{(\mathrm{i})}(f) = \lim_{M\to\infty} \frac{1}{2M+1} \sum_{n=-M}^{M} \sum_{m=-M}^{M} r[n-m]e^{-j2\pi f(n-m)} \tag{B.5}$$

When the addends of the summation are viewed as the $[n, m]$ elements of an $(2M+1)\times(2M+1)$ dimensional matrix, then the following equality becomes more easily understandable.

$$\sum_{n=-M}^{M} \sum_{m=-M}^{M} g[n-m] = \sum_{\ell=-2M}^{2M} (2M+1-|\ell|)g[\ell] \tag{B.6}$$

Substituting (B.6) into (B.5) and rearranging yields:

$$\phi^{(\mathrm{i})}(f) = \lim_{M\to\infty} \sum_{\ell=-2M}^{2M} \left(1 - \frac{|\ell|}{2M+1}\right) r[\ell]e^{-j2\pi f\ell} \tag{B.7}$$

[1]Technically the theorem can also refer to the opposite: the PSD is defined by (B.3) and then (B.1) is shown to be valid [43]. This is closer to how Wiener originally defined it [43], but the definition used in this thesis is used in [40] and seems to be more popular.

If the ACF decays sufficiently rapidly [40], then B.7 simplifies to:

$$\phi^{(\mathrm{i})}(f) = \sum_{\ell=-\infty}^{\infty} r[\ell] e^{-j2\pi f \ell} \tag{B.8}$$

One can now see how the intuitive definition of the PSD (B.1) leads to the Wiener-Khinchin theorem (B.3).

Bibliography

[1] M. Parchami, W. Zhu, B. Champagne, and E. Plourde, "Recent developments in speech enhancement in the short-time fourier transform domain," *IEEE Circuits and Systems Magazine*, vol. 16, no. 3, pp. 45–77, Aug. 2016.

[2] X. Xie, P. Du, J. Xia, and J. Luo, "Spectral indices for estimating exposed carbonate rock fraction in karst areas of southwest china," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 9, pp. 1988–1992, Sep. 2015.

[3] A. D. Chave, D. J. Thomson, and M. E. Ander, "On the robust estimation of power spectra, coherences, and transfer functions," *Journal of Geophysical Research: Solid Earth*, vol. 92, no. B1, pp. 633–648, Jan. 1987.

[4] R. Gençay, F. Selçuk, and B. Whitche, *An introduction to wavelets and other filtering methods in finance and economics*. San Diego: Academic Press, 2002.

[5] H. Arslan and T. Yücek, "Spectrum sensing for cognitive radio applications," in *Cognitive Radio, Software Defined Radio, and Adaptive Wireless Systems*, H. Arslan, Ed. Dordrecht: Springer Netherlands, 2007, pp. 263–289.

[6] 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, Jan. 2009.

[7] P. Welch, "The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, Jun. 1967.

[8] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, Sep. 1982.

[9] D. B. Percival and A. T. Walden, *Spectral analysis for physical applications: Multitaper and conventional univariate techniques*. Cambridge, England: Cambridge University Press, 1993.

[10] S. A. Kassam, *Signal detection in non-gaussian noise*. New York City, NY: Springer Publishing, 1988.

[11] G. Staple and K. Werbach, "The end of spectrum scarcity [spectrum allocation and utilization]," *IEEE Spectrum*, vol. 41, no. 3, pp. 48–52, Mar. 2004.

[12] M. J. Marcus, "Unlicensed cognitive sharing of tv spectrum: The controversy at the federal communications commission," *IEEE Communications Magazine*, vol. 43, no. 5, pp. 24–25, May 2005.

[13] J. Shen, Y. Liu, S. Liu, J. Gao, G. Xie, and C. Chi, "Robust energy detection based on bayesian estimation for cognitive radio," in *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, Nov. 2008, pp. 1–5.

[14] A. Margoosian, J. Abouei, and K. N. Plataniotis, "An accurate kernelized energy detection in gaussian and non-gaussian/impulsive noises," *IEEE Transactions on Signal Processing*, vol. 63, no. 21, pp. 5621–5636, Nov. 2015.

[15] J. T. Kees, J. M. Ernst, W. C. Headley, and A. A. L. Beex, "Robust blind spectral estimation in the presence of non-gaussian noise," in *MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM)*, Oct. 2017, pp. 629–634.

[16] S. Galli, A. Scaglione, and Z. Wang, "For the grid and through the grid: The role of power line communications in the smart grid," *Proceedings of the IEEE*, vol. 99, no. 6, pp. 998–1027, Jun. 2011.

[17] M. Zimmermann and K. Dostert, "Analysis and modeling of impulsive noise in broad-band powerline communications," *IEEE Transactions on Electromagnetic Compatibility*, vol. 44, no. 1, pp. 249–258, Feb. 2002.

[18] H. Meng, Y. L. Guan, and S. Chen, "Modeling and analysis of noise effects on broadband power-line communications," *IEEE Transactions on Power Delivery*, vol. 20, no. 2, pp. 630–637, Apr. 2005.

[19] J. Lin, T. Pande, I. H. Kim, A. Batra, and B. L. Evans, "Time-frequency modulation diversity to combat periodic impulsive noise in narrowband powerline communications," *IEEE Transactions on Communications*, vol. 63, no. 5, pp. 1837–1849, May 2015.

[20] K. M. Rabie and E. Alsusae, "On improving communication robustness in plc systems for more reliable smart grid applications," *IEEE Transactions on Smart Grid*, vol. 6, no. 6, pp. 2746–2756, Nov. 2015.

[21] M. Nassar, A. Dabak, I. H. Kim, T. Pande, and B. L. Evans, "Cyclostationary noise modeling in narrowband powerline communication for smart grid applications," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2012, pp. 3089–3092.

[22] M. Katayama, T. Yamazato, and H. Okada, "A mathematical model of noise in narrowband power line communication systems," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1267–1276, Jul. 2006.

[23] S. V. Vaseghi, *Advanced digital signal processing and noise reduction*, Second. Hoboken, NJ: Wiley, 2000.

[24] J. Matanza, S. Alexandres, and C. Rodriguez-Morcillo, "Difference sets-based compressive sensing as denoising method for narrow-band power line communications," *IET Communications*, vol. 7, no. 15, pp. 1580–1586, Oct. 2013.

[25] R. L. Itagi, K. P. Vittal, and U. Sripati, "A low snr approach to substation communication using powerline for emi reduction," in *2012 Asia-Pacific Symposium on Electromagnetic Compatibility*, May 2012, pp. 453–456.

[26] G. Ndo, F. Labeau, and M. Kassouf, "A markov-middleton model for bursty impulsive noise: Modeling and receiver design," *IEEE Transactions on Power Delivery*, vol. 28, no. 4, pp. 2317–2325, Oct. 2013.

[27] K. Gulati, B. L. Evans, J. G. Andrews, and K. R. Tinsley, "Statistics of co-channel interference in a field of poisson and poisson-poisson clustered interferers," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, pp. 6207–6222, Dec. 2010.

[28] A. Mahmood and M. Chitre, "Modeling colored impulsive noise by markov chains and alpha-stable processes," in *OCEANS 2015 - Genova*, May 2015, pp. 1–7.

[29] D. Middleton, "Statistical-physical models of electromagnetic interference," *IEEE Transactions on Electromagnetic Compatibility*, vol. EMC-19, no. 3, pp. 106–127, Aug. 1977.

[30] H. Kanemoto, S. Miyamoto, and N. Morinaga, "A study on modeling of microwave oven interference and optimum reception," in *1998 IEEE EMC Symposium. International Symposium on Electromagnetic Compatibility. Symposium Record (Cat. No.98CH36253)*, vol. 1, Aug. 1998, 57–62 vol.1.

[31] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proceedings of the IEEE*, vol. 81, no. 7, pp. 986–1010, Jul. 1993.

[32] R. Prasad, A. Kegel, and A. de Vos, "Performance of microcellular mobile radio in a cochannel interference, natural, and man-made noise environment," *IEEE Transactions on Vehicular Technology*, vol. 42, no. 1, pp. 33–40, Feb. 1993.

[33] T. Shongwey, A. J. H. Vinck, and H. C. Ferreira, "On impulse noise and its models," in *18th IEEE International Symposium on Power Line Communications and Its Applications*, Mar. 2014, pp. 12–17.

[34] M. Singh, A. Amruthakala, M. L. Sudheer, and A. S. Murthy, "Analysis of noise in broadband powerline communications (b-plc) in frequency range of 150khz–30mhz," in *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)*, Oct. 2017, pp. 101–106.

[35] E. Xu, S. Shi, D. Chen, and X. Gu, "An flom based signal detection approach under middleton class a noise," in *2017 IEEE/CIC International Conference on Communications in China (ICCC)*, Oct. 2017, pp. 1–4.

[36] K. Yan, H. Zhou, X. Zhang, and H. Xiao, "Novel robust spectrum sensing in impulsive noise environment," in *2015 10th International Conference on Communications and Networking in China (ChinaCom)*, Aug. 2015, pp. 766–770.

[37] S. Miyamoto, M. Katayama, and N. Morinaga, "Performance analysis of qam systems under class a impulsive noise environment," *IEEE Transactions on Electromagnetic Compatibility*, vol. 37, no. 2, pp. 260–267, May 1995.

[38] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Upper Saddle River, NJ: Pearson Prentice Hall, 2005.

[39] T. P. Bronez, "On the performance advantage of multitaper spectral analysis," *IEEE Transactions on Signal Processing*, vol. 40, no. 12, pp. 2941–2946, Dec. 1992.

[40] S. M. Kay, *Modern spectral estimation: Theory and application*. Upper Saddle River, NJ: Pearson Prentice Hall, 1988.

[41] T. Kailath, "Norbert wiener and the development of mathematical engineering," in *Communications, Computation, Control, and Signal Processing: A tribute to Thomas Kailath*, A. Paulraj, V. Roychowdhury, and C. D. Schaper, Eds. New York City, NY: Springer Publishing, 1997, pp. 35–64.

[42] N. Wiener, "Generalized harmonic analysis," *Acta Math.*, vol. 55, pp. 117–258, 1930.

[43] D. C. Champeney, *A handbook of fourier theorems*. Cambridge, England: Cambridge University Press, 1987.

[44] A. Khinchin, "Korrelationstheorie der stationären stochastischen prozesse," *Mathematische Annalen*, vol. 109, no. 1, pp. 604–615, Dec. 1934.

[45] R. B. Blackman and J. W. Tukey, *The measurement of power spectra from the point of view of communications engineering*. Mineola, NY: Dover Publications, 1959.

[46] M. S. Bartlett, "Smoothing periodograms from time-series with continuous spectra," *Nature*, vol. 161, pp. 686–687, May 1948.

[47] ——, "Periodogram analysis and continuous spectra," *Biometrika*, vol. 37, no. 1/2, pp. 1–16, Jun. 1950.

[48] S. Haykin, D. J. Thomson, and J. H. Reed, "Spectrum sensing for cognitive radio," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 849–877, May 2009.

[49] K. K. Parhi and M. Ayinala, "Low-complexity welch power spectral density computation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 61, no. 1, pp. 172–182, Jan. 2014.

[50] H. C. So, Y. T. Chan, Q. Ma, and P. C. Ching, "Comparison of various periodograms for sinusoid detection and frequency estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 3, pp. 945–952, Jul. 1999.

[51] Y. T. Chan, Q. Ma, H. C. So, and R. Inkol, "Evaluation of various fft methods for single tone detection and frequency estimation," in *CCECE '97. Canadian Conference on Electrical and Computer Engineering. Engineering Innovation: Voyage of Discovery. Conference Proceedings*, vol. 1, May 1997, pp. 211–214.

[52] H. Iwata, K. Umebayashi, S. Tiiro, Y. Suzuki, and J. J. Lehtomäki, "Optimum welch fft segment size for duty cycle estimation in spectrum awareness system," in *2015 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, Mar. 2015, pp. 229–234.

[53] ——, "A study on welch fft segment size selection method for spectrum awareness," in *2016 IEEE Wireless Communications and Networking Conference*, Apr. 2016, pp. 1–6.

[54] N. Manshouri, M. Maleki, and T. Kayikçioğlu, "Power spectrum analysis of eeg for watching 2d amp; 3d videos and resting state," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, May 2018, pp. 1–4.

[55] K. Umebayashi, R. Takagi, N. Ioroi, Y. Suzuki, and J. J. Lehtomäki, "Duty cycle and noise floor estimation with welch fft for spectrum usage measurements," in *2014 9th International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)*, Jun. 2014, pp. 73–78.

[56] f. j. harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proceedings of the IEEE*, vol. 66, no. 1, pp. 51–83, Jan. 1978.

[57] T. P. Bronez and D. S. Brown, "Alternate windows for multi-window spectral analysis," in *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, Mar. 1992, pp. 429–432.

[58] M. Andrei, L. Trifina, and D. G. Tarniceriu, "Capacity of middleton class-a impulsive noise channel with binary input," May 2015, pp. 1291–1298.

[59] H. Kanemoto, S. Miyamoto, and N. Morinaga, "A study on modeling of microwave oven interference and optimum reception," in *1998 IEEE EMC Symposium. International Symposium on Electromagnetic Compatibility. Symposium Record (Cat. No.98CH36253)*, vol. 1, Aug. 1998, 57–62 vol.1.

[60] Q. Shan, I. A. Glover, R. C. Atkinson, S. A. Bhatti, I. E. Portugues, P. J. Moore, R. Rutherford, M. d. Fátima Queiroz Vieira, A. M. N. Lima, and B. A. de Souza, "Estimation of impulsive noise in an electricity substation," *IEEE Transactions on Electromagnetic Compatibility*, vol. 53, no. 3, pp. 653–663, Aug. 2011.

[61] Q. Shan, S. Bhatti, I. A. Glover, R. Atkinson, P. J. Moore, I. E. Portugues, and R. Rutherford, "Noise amplitude distribution of impulsive noise from measurements in a power substation," in *2009 44th International Universities Power Engineering Conference (UPEC)*, Sep. 2009, pp. 1–5.