UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Impulsive noise cancellation and channel estimation in power line communication systems

Deep Shrestha

UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

Impulsive Noise Cancellation and Channel Estimation in Power Line Communication Systems

Ph.D. Dissertation

Author

Deep Shrestha

Advisors

Dr. Miquel Payaró

Dr. Xavier Mestre

This work is supported in part by the European Commission 7th Framework Program under grant agreement no. 607774 Advanced Communications and Information Processing in Smart Grid Systems (ADVANTAGE) and by the Generalitat de Catalunya under grants 2014 SGR 1551, 2014 SGR 1567 and 2017 SGR 891.

Abstract

Power line communication (PLC) is considered as the most viable enabler of the smart grid. PLC exploits the power line infrastructure for data transmission and provides an economical communication backbone to support the requirements of smart grid applications. Though PLC brings a lot of benefits to the smart grid implementation, impairments such as frequency selective attenuation of the high-frequency communication signal, the presence of impulsive noise (IN) and the narrowband interference (NBI) from closely operating wireless communication systems, make the power line a hostile environment for reliable data transmission. Hence, the main objective of this dissertation is to design signal processing algorithms that are specifically tailored to overcome the inevitable impairments in the power line environment.

First, we propose a novel IN mitigation scheme for PLC systems. The proposed scheme actively estimates the locations of IN samples and eliminates the effect of IN only from the contaminated samples of the received signal. By doing so, the typical problem encountered while mitigating the IN is avoided by using passive IN power suppression algorithms, where samples besides the ones containing the IN are also affected creating additional distortion in the received signal.

Apart from the IN, the PLC transmission is also impaired by NBI. Exploiting the duality of the problem where the IN is impulsive in the time domain and the NBI is impulsive in the frequency domain, an extended IN mitigation algorithm is proposed in order to accurately estimate and effectively cancel both impairments from the received signal. The numerical validation of the proposed schemes shows improved BER performance of PLC systems in the presence of IN and NBI.

Secondly, we pay attention to the problem of channel estimation in the

power line environment. The presence of IN makes channel estimation challenging for PLC systems. To accurately estimate the channel, two maximum-likelihood (ML) channel estimators for PLC systems are proposed in this thesis. Both ML estimators exploit the estimated IN samples to determine the channel coefficients. Among the proposed channel estimators, one treats the estimated IN as a deterministic quantity, and the other assumes that the estimated IN is a random quantity. The performance of both estimators is analyzed and numerically evaluated to show the superiority of the proposed estimators in comparison to conventional channel estimation strategies in the presence of IN. Furthermore, between the two proposed estimators, the one that is based on the random approach outperforms the deterministic one in all typical PLC scenarios. However, the deterministic approach based estimator can perform consistent channel estimation regardless of the IN behavior with less computational effort and becomes an efficient channel estimation strategy in situations where high computational complexity cannot be afforded.

Finally, we propose two ML algorithms to perform a precise IN support detection. The proposed algorithms perform a greedy search of the samples in the received signal that are contaminated by IN. To design such algorithms, statistics defined for deterministic and random ML channel estimators are exploited and two multiple hypothesis tests are built according to Bonferroni and Benjamini and Hochberg design criteria. Among the proposed estimators, the random ML-based approach outperforms the deterministic ML-based approach while detecting the IN support in typical power line environment.

Hence, this thesis studies the power line environment for reliable data transmission to support smart grid. The proposed signal processing schemes are robust and allow PLC systems to effectively overcome the major impairments in an active electrical network. The efficient mitigation of IN and NBI and accurate estimation of channel enhances the applicability of PLC to support critical applications that are envisioned for the future electrical power grid.

Acknowledgements

As I am finalizing my thesis, I cannot refrain myself from expressing my deepest gratitude towards my advisors Dr. Miquel Payaró and Dr. Xavier Mestre. At this moment, I want to look back to this three year journey and would like to thank for the kind support, professional guidance and utter motivation that I received from both of you. Thank you for training me to become a good researcher. I thoroughly enjoyed this journey as your student.

During my PhD research, I had wonderful opportunities to visit the Institute of Networked and Embedded Systems (NES), Klagenfurt, and the Corporate Research Center - ABB, Vasteras. I would like to thank Prof. Andrea Tonello from NES for hosting me during my three month research visit. I would also like to thank Dr. Zhibo Pang from ABB for hosting me for five months and allowing me to collaborate with his research group.

I would also like to express my sincere gratitude towards my family. I dedicate a huge part of this thesis to my father Mr. Pawan Bhakta Shrestha, my mother Mrs. Raju Shrestha, my brother Mr. Sunny Shrestha and last but not the least, my better half Mrs. Sunakshi Shrestha. Thank you for your endless support and patience.

Finally, I would like to thank everyone who directly or indirectly helped me to make this thesis a success for me.

List of Acronyms

AC	Alternating Current
AMI	Advanced Metering Infrastructure
AMR	Automatic Meter Reading
ARIB	Association of Radio Industries and Businesses
AWGN	Additive white Gaussian Noise
BB-PLC	Broadband Power Line Communication
BER	Bit Error Rate
BF	Bonferroni
BH	Benjamini and Hochwald
CENELEC	Comité Européen de Normalisation Électrotechnique
CFR	Channel Frequency Response
CIR	Channel Impulse Response
CP	Cyclic Prefix
CRB	Cramér-Rao Bound
CS	Compressive Sensing
DEG	Distributed Energy Generator
DEMS	Distributed Energy Management System
DML	Deterministic Maximum-Likelihood
DPSK	Differential Phase Shift Keying
ED	Eigenvalue Decomposition
FA	False Alarm
FCC	Federal Communications Commission
FDBI	Frequency Domain Block Interleaving
FDR	False Discovery Rate

FEC	Forward Error Correction
FFT	Fast Fourier Transform
FMR	False Missed Target Rate
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
HV	High Voltage
i.i.d	Independent and Identically Distributed
IDFT	Inverse Discrete Fourier Transform
IN	Impulsive Noise
INR	Impulsive Noise to Background Noise Power Ratio
ISR	Interference to Signal Power Ratio
LOS	Line of Sight
LS	Least Squares
LV	Low Voltage
MDL	Minimum Descriptive Length
MIMO	Multiple Input Multiple Output
ML	Maximum Likelihood
MMSE	Minimum Mean Squared Error
MSE	Mean Squared Error
MT	Missed Target
MUSIC	Mutiple Signal Classification
MV	Medium Voltage
NBI	Narrowband Interference
NB-PLC	Narrowband PLC
OFDM	Orthogonal Frequency Division Multiplexing
PHY	Physical Layer
PLC	Power Line Communication
PRIME	Power Line Intelligent Metering Evolution
QAM	Quadrature amplitude Modulation
RES	Renewable Energy Source
RML	Random Maximum-Likelihood
SISO	Single Input Single Output
SNR	Signal to Noise Power Ratio
TDBI	Time Domain Block Interleaving

TDP	Time Domain Periodogram
TSE	True Support Estimation
TWACS	Two-Way Automatic Communication System
UMTS	Universal Mobile Telecommunications System
WiMAX	Worldwide Interoperability for Microwave Access
WLAN	Wireless Local Area Network

Notation

x	A scalar.
$\mathbf{x}$	A column vector.
$\mathbf{X}$	A matrix.
$\mathbf{X}^{-1}$	Inverse of matrix $\mathbf{X}$.
$\mathcal{X}$	A set.
$\lvert\mathcal{X}\rvert$	Cardinality of set $\mathcal{X}$.
$\lvert\cdot\rvert$	Absolute value.
$[\cdot]$	An entry of a matrix or vector.
$\mathrm{diag}(\mathbf{x})$	Matrix with the entries of vector $\mathbf{x}$ in its diagonal.
$\mathbf{X}^T$	Transpose of matrix $\mathbf{X}$.
$\mathbf{X}^H$	Hermitian of matrix $\mathbf{X}$.
log	Logarithm of base 2.
$\mathbb{E}[\cdot]$	Expected value.
$\det(\mathbf{X})$	Determinant of matrix $\mathbf{X}$.
$\mathbf{I}_N$	Identity matrix of dimension $N \times N$.
$\lVert\cdot\rVert$	Norm.
$\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Circularly symmetric complex Gaussian distributed vector with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.
$\mathrm{tr}[\mathbf{X}]$	Trace of matrix $\mathbf{X}$ unless explicitly defined otherwise.
$\mathcal{R}(\mathbf{x})$	$\left[\mathrm{Re}^T(\mathbf{x}), \mathrm{Im}^T(\mathbf{x})\right]^T$ unless defined otherwise.
$f(x)$	A function of x.
$\nabla f(\mathbf{x})$	A column vector denoting the Gradient of function $f(\mathbf{x})$ with respect to $\mathbf{x}$.

$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}^*}$	Derivative of function $f(\mathbf{x})$ with respect to $\mathbf{x}^*$, "$*$" denoting the conjugate of complex vector $\mathbf{x}$.
$\frac{\partial f(x,y,z)}{\partial x}$	Derivative of function $f(x, y, z)$ with respect to $\mathbf{x}$.
$\mathbb{H}f(\mathbf{x})$	Hessian matrix of function $f(\mathbf{x})$ with respect to $\mathbf{x}$.
$\mathbf{P}_{\mathcal{A}}$	is a diagonal selection matrix of dimension $N \times N$ with ones at its entries denotes by the IN support set $\mathcal{A}$.
$\mathbf{P}_{\mathcal{A}}^{\perp}$	$\mathbf{I}_N - \mathbf{P}_{\mathcal{A}}$.
$\rightarrow$	Converges to.
$\triangleq$	Defined as.
$\sim$	Equivalence relation.
$\bigcup$	Set union.
$\subset$	Subset.
$\in$	Set membership.
$N \backslash \mathcal{A}$	All those elements of N that are not in $\mathcal{A}$.
lim	Limit.
lim sup	Limit superior.
lim inf	Limit inferior.
$\sum$	Summation.
$\oint$	Contour integral.

Contents

List of Figures

Chapter 1

Introduction

The technological evolution that we are witnessing today would not have been possible without the innovations in electrical energy production and distribution technologies. In fact, we make extensive use of electrical energy powered devices to make our daily life easier. As a consequence, the growth in energy demand has increased significantly and has outpaced the rate at which the electrical energy can be produced only by exploiting conventional power generation strategies. Recent practices integrate distributed energy generators (DEGs) that exploit renewable energy sources (RESs) to produce electrical energy to the existing grid to fulfill increased energy demand [1]. The addition of RES-based DEGs alongside the conventional fossil fuel-based generators transforms the traditional unidirectional power flow electrical network into a bi-directional power flow network [2]. This transformation of the power grid instigates complexity on its safe, efficient, and reliable operation [3, 4]. In such a heterogeneous network, an optimum balance between the energy generation and demand, real-time control of the grid operation, and wide area monitoring of the power network needs to be done to maintain grid stability and achieve efficient delivery of the power to the end consumers. Hence to achieve performance and operation of such granularity, the concept of smart grid has emerged [5].

Smart grid envisions to modernize the traditional electrical power system by fostering sustainable models of energy production, distribution, and usage [6]. As its key features, it will incorporate pervasive communication for realtime monitoring of the complex grid operations and will inherit distributed autonomous control and management capabilities to the electrical power grid for its efficient operation. Moreover, a successful smart grid implementation

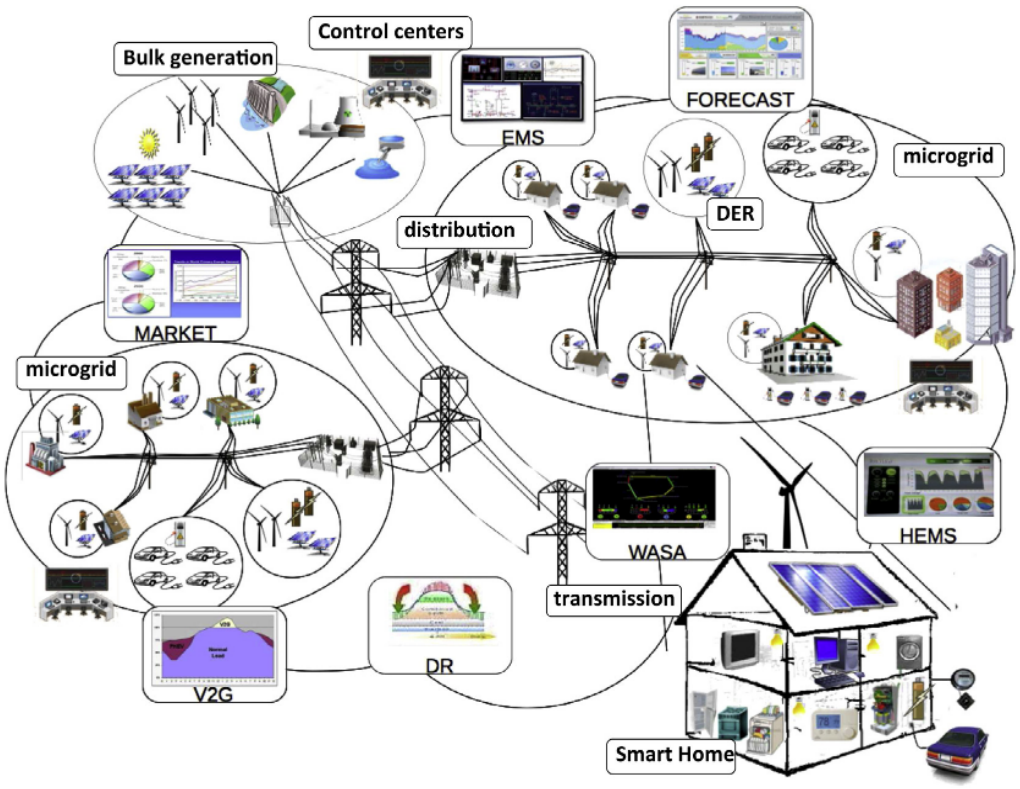


Figure 1.1: Schematic diagram of a fully operational smart grid [8].

will essentially [7]:

- enable efficient deployment of massive DEGs that exploit RESs to produce electrical energy;
- enhance efficiency, resilience, and sustainability of the traditional electrical power grid by incorporating real-time distributed intelligence enabling automated protection, optimization, and control functions;
- provide customers with information about energy usage pricing and allow them to minimize their consumption by optimizing the usage time and pattern of their appliances;
- enable real-time time situational awareness of the grid status and operation by deploying smart metering and monitoring systems.

To successfully implement the smart grid, as shown in Fig. 1.1, extensive deployment of sensors, actuators and metering devices to collect information from all the relevant components of the electrical grid in real-time is required. The data collected from such devices, typically containing energy usage patterns, the status of the DEGs and the components of the grids, needs to

be securely and reliably delivered to distributed energy management systems (DEMSs). Upon reception, the DEMS is then responsible for analyzing the data to study the network load profile and forecast energy demand, identify faults and disturbances in the network and make power system fault rectifications, control and operation optimization decisions and execute them [8].

Application	Communication Technology
Substation Automation	WiMAX, WLAN
Transmission Line Monitoring	Cellular technologies based on GSM/GPRS/UMTS
Home Energy Management	ZigBee, Bluetooth, Power line communication
Automatic Metering	Power line communication
Demand Response Management	Power line communication

Table 1.1: Potential communication technologies to support smart applications [2].

Smart grid envisions to entail innovative services and applications by exploiting the fine-grained information collected from all the relevant parts of the electrical network for efficient power delivery from the generators to the end consumers. To name some of the key smart grid applications, substation automation, transmission line monitoring, home energy management, advanced metering infrastructure, wide area situational awareness, demand response management, outage management, and distribution automation are of utmost importance to achieve the smart grid goals [2]. All these applications demand diverse communication requirements. To identify suitable communication technologies to provide the communications backbone for the smart grid, quantitative and qualitative characteristics of the candidate technologies play vital roles [7, 9–13]. The quantitative characteristics define the latency[1],

[1]An algorithm for precise clock synchronization in low latency communication system has been proposed in [14].

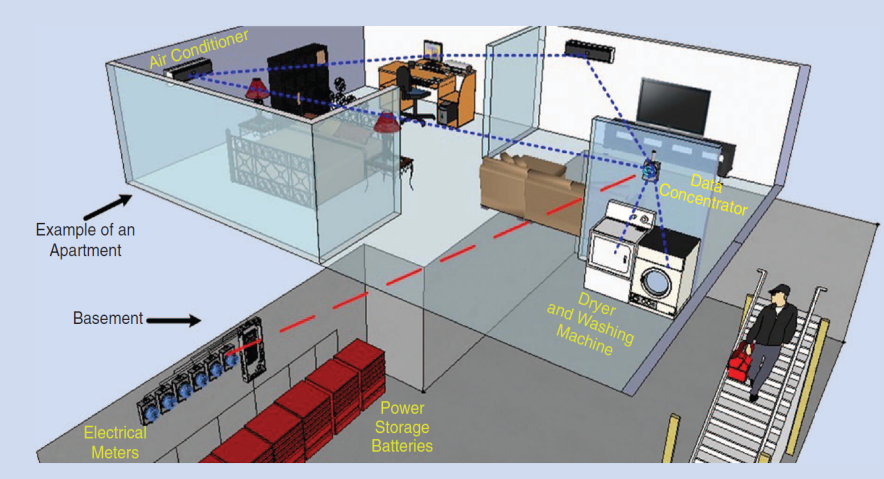


Figure 1.2: Location of energy consumption metering devices in high-rise apartments [15].

reliability, and bandwidth offered by the candidate communication technology [9–11]. In contrast, the qualitative requirements define the scalability, interoperability, flexibility, and security of the communication technology under consideration [7,9,12,13]. As shown in Table. 1.1, depending upon the communication requirements, wireless communication technologies such as WiMAX and WLAN are considered suitable for the substation automation. Likewise, cellular technologies based on GSM, GPRS and UMTS are favored to monitor the status of transmission and distribution lines in the electrical power grid system. For applications like home energy management, short range communication technologies such as ZigBee and Bluetooth are considered as candidates. Similarly, power line communication (PLC) is considered for demand response management and automatic metering applications.

At present, the smart grid implementation is in its infancy, and only basic application such as automatic meter reading (AMR) to collect consumer energy consumption data is deployed. To realize a fully operational smart grid, the crucial phase of technology development of establishing a robust advanced metering infrastructure (AMI), however, needs to be surpassed as shown in Fig. 1.3. To achieve these milestones, proper exploitation of communication technologies is very crucial [8]. Among the available communication technologies, PLC exploits existing power lines in the electrical grid for data transmission and offers the most economical solution to the smart grid implementation [16]. Besides, the extensive outreach of power lines in an electrical network adds more value to PLC and makes it the most viable enabler of smart grid ap-

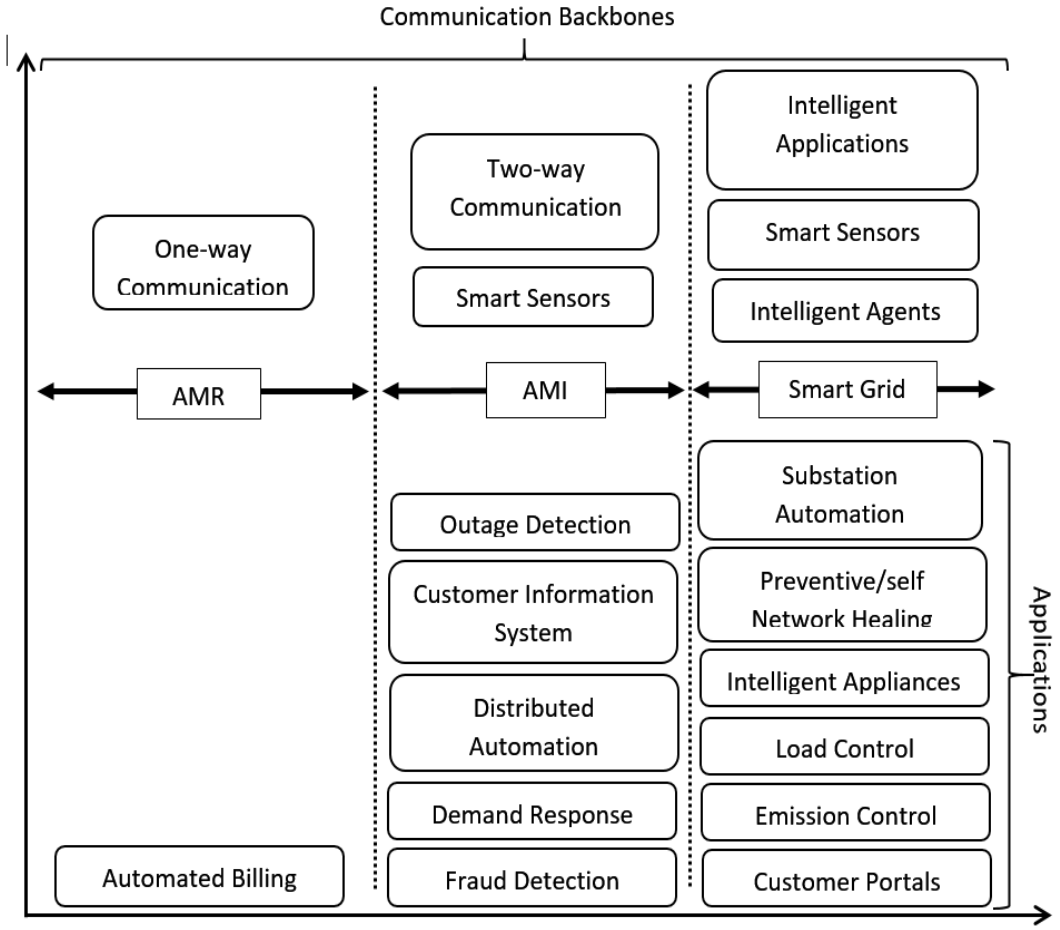


Figure 1.3: Smart grid deployment road map [2].

plication in comparison to other wired or wireless alternatives. Virtually all line powered devices in the network become targets of the value-added services that smart grid envisions to incorporate [17]. In addition to this, PLC provides an additional coverage boost to smart grid applications that other technologies such as wireless fail to provide [15]. In an urban area that is densely populated with people living in high-rises, metering instruments are installed in basements, as shown in Fig. 1.2. Wireless communication signals cannot penetrate and reach these underground areas and make wireless communication-based technologies less favorable for AMR applications. Besides, exploiting power lines also to transmit data, opens a door to realizing a single technology that is decoupled from other telecommunication technologies and can support smart grid applications in all domains of the electrical power grid.

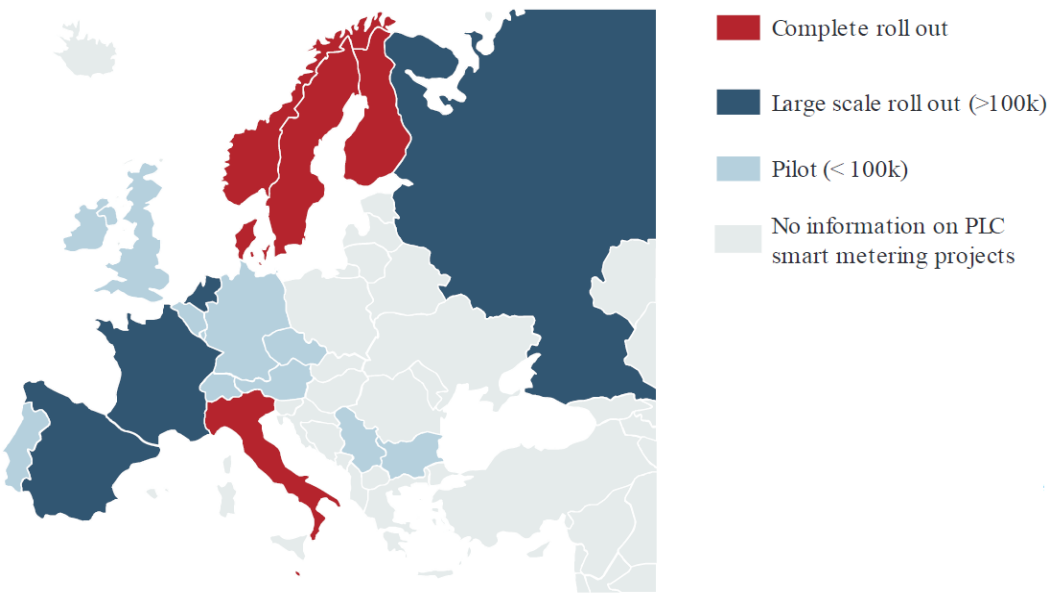


Figure 1.4: Power line communication based smart meter deployment graphical representation [18].

1.1 Motivation

Owing to the benefits that PLC can bring along, it is gaining a lot of attention from the utility service providers, who are in the front line of smart grid implementation [17]. The deployment of PLC devices to gather energy consumption data from the consumer premises is increasing rapidly, and a huge geographical area is already covered by PLC based metering devices to establish an extensive AMR infrastructure. A graphical representation of project rollouts to deploy PLC based metering devices is shown in Fig. 1.4 [18].

Extensive use of PLC based technologies, in recent deployments for AMR, not only signifies interest of utility service providers on this technology, but also shows an efficient approach through which we can reach the goal of establishing an AMI. Enhancing the use of already established AMR infrastructure and transforming it into a robust AMI by incorporating new and innovative features to perform power outage detection, establishing customer information systems, enabling distributed automation, performing fraud detection, and demand response management by exploiting PLC will lead towards realizing a fully operational smart grid. However, extending the application of PLC based technologies also to support critical smart grid applications is challenging. The

presence of random impulsive noise (IN) and narrowband interference (NBI) in the power line environment and the frequency selective attenuation of the communication signal transmitted through power lines hinder the application of PLC to support stringent requirements beyond those of AMR [19]. Therefore, to achieve efficient and reliable data transmission through power lines, robust and efficient signal processing algorithms that are tailored specifically to overcome impairments in harsh power line environment are required.

1.2 Thesis Outline

This thesis studies and analyzes signal processing techniques that deal with interference, noise, and channel frequency selectivity in PLC for smart grids. The main aim is to study impairments that make power line a harsh medium for data transmission and propose signal processing algorithms to overcome them.

In order to concisely elaborate the challenges that PLC systems have to face and the solutions that can enhance their applicability to support smart grid applications, this thesis is divided into five chapters. The current chapter has been dedicated to introducing the concept of smart grid and a brief motivation behind pursuing PLC for smart grid applications. The rest are dedicated to providing insights on modeling power lines as communication channels, on the inevitable impairments in power line environment and on the signal processing schemes that are proposed for reliable PLC. Furthermore, signal processing algorithms will be proposed to effectively overcome the impairments in power line environments. A brief outline of the remaining chapters follows:

Chapter 2

This chapter analyzes power lines in the electrical grid as communication channels. Since medium voltage (MV) distribution and low voltage (LV) access grids are considered as primary domains where smart grid applications will be deployed, in this chapter, we focus on:

- characterization of power lines in MV distribution and LV access grids as communication channels;
- impairments that make power lines in distribution and access grids a hostile medium for reliable data transmission.

Furthermore, given the uprising interest in using PLC for smart grids, signal processing schemes are proposed to overcome the power line environment impairments. In this chapter, we will provide a brief introduction to noise and interference mitigation schemes along with the channel estimation strategies that are richly discussed in PLC literature.

Chapter 3

In this chapter, we propose new IN mitigation and NBI cancellation schemes for PLC systems. The proposed schemes perform active estimation of the IN and NBI and cancel them from the received signal to provide noise mitigated samples for demodulation. They inherently exploit the fact that a part of the spectrum is not employed by PLC for data transmission (guard band). The proposed schemes improve the BER performance of the PLC systems by mitigating the effect of IN and NBI only from those samples in the received signal that are corrupted by these impairments. In brief, the goal of this chapter will be to:

- introduce the proposed IN estimation and cancellation scheme for PLC systems;
- numerically validate the proposed scheme's performance in a typical power line environment with IN;
- derive an extension of the proposed IN mitigation scheme for successive cancellation of simultaneously occurring IN and NBI during the PLC transmission;
- study the performance of the proposed algorithm numerically when the PLC transmission is impaired by both IN and NBI.

The signal processing algorithms proposed in this chapter along with their numerical validations are published in:

- **D. Shrestha**, X. Mestre and M. Payaró,"Asynchronous Impulsive Noise Mitigation Based on Subspace Support Estimation for PLC Systems", *IEEE International Symposium on Power Line Communications and its Applications*, Bottrop, Germany, 2016.
- **D. Shrestha**, A. Tonello, X. Mestre and M. Payaró, "Simultaneous Cancellation of Narrow Band Interference and Impulsive Noise in PLC Systems", *IEEE International Conference on Smart Grid Communications*, Sydney, Australia, 2016.

Chapter 4

The presence of IN in an active power line network hinders accurate estimation of the channel. The wideband contribution of such noise typically corrupts pilot symbols that are transmitted in an OFDM based system to assist the receiver to estimate the channel. Therefore, a robust and efficient channel estimator is required to perform precise channel estimation in harsh power line environment, such that the received symbols affected by frequency selectivity of power line channels can easily be equalized.

To precisely estimate the channel in the presence of IN, in this chapter, we propose two maximum-likelihood (ML) channel estimators for PLC systems. The proposed ML channel estimators exploit the estimated IN while estimating the channel. The first ML estimator considers IN as a random quantity and the second treats the estimated IN as a deterministic quantity. Both estimators show improved accuracy in channel estimation as compared to the conventional channel estimation. Moreover, the performance of proposed estimators is analyzed both analytically and numerically. In brief, the goal will be to:

- introduce the concept of ML channel estimation by exploiting the estimated IN in terms of support and amplitude;
- perform analysis and validation of the performance of the proposed channel estimators in a typical power line environment;
- provide simulation results of the proposed channel estimators in varied PLC scenarios.

The signal processing algorithms that are proposed in section 4.1 of this chapter along with the simulation results are published in:

- **D. Shrestha**, X. Mestre and M. Payaró,"Maximum-Likelihood Channel Estimation in Presence of Impulsive Noise for PLC Systems", *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Washington DC, USA, 2016.

and the joint ML CIR and IN estimator proposed in section 4.2 and has been published in:

- **D. Shrestha**, X. Mestre and M. Payaró,"On Channel Estimation for Power Line Communication Systems in the Presence of Impulsive Noise", Journal of Computers and Electrical Engineering, Vol. 72, Pages 406-419, 2018.

.

Chapter 5

The channel estimators proposed in chapter 4 rely on the estimated IN to estimate the channel. The numerical validation of the proposed estimators show that the accuracy of IN support detection has an effect on the precision of channel estimation. The optimum performance in channel estimation is achieved when the IN support is estimated correctly. In a situation when the IN support is underestimated or overestimated, the error in channel estimation increases proportionally to the number of underestimated or overestimated IN sample locations.

In order to perform precise detection of the IN support, in this chapter we propose novel IN support estimation algorithms. To do so, chapter 5 will focus on:

- exploiting the statistics introduced in chapter 4, used to design the ML channel estimators, to design ML IN support estimators;
- proposing an incremental greedy procedure to obtain the ML estimate of the IN support estimators;
- analyzing the performance of the proposed estimators;
- performing numerical validation of the proposed estimators in varied power line environment where IN with distinct power level is present.

The work in this chapter has been submitted to:

- X. Mestre, M. Payaró and **D. Shrestha**,"Maximum-Likelihood Detection of Impulsive Noise Support with Application to Channel Parameter Estimation", IEEE International Conference on Acoustics, Speech, and Signal Processing, 2019.
- **D. Shrestha**, X. Mestre and M. Payaró,"On Maximum-Likelihood Detection of the Impulsive Noise Support", IEEE Transactions on Signal Processing.

for review.

Chapter 6

Finally, we conclude this thesis by providing a brief conclusion and some final remarks on the signal processing algorithms proposed in this thesis. Furthermore, we also provide research directions that may be interesting to realize as the next generation of PLC systems for smart grid communication.

Chapter 2

PLC for Smart Grid Communications

The idea of exploiting power lines in the electrical grid also for communication is an old concept. The first power line communication (PLC) based systems were proposed in the early 1920s [17]. Since then, utility service providers have been using this technology for low data rate telecommunications. The primitive PLC systems, namely the Two-Way Automatic Communication System (TWACS) and Turtle, used the alteration of the electrical signal for communication. Both systems relied on the disturbance in the current waveform for the inbound communication, between the remote device and the control center, and the alteration of the voltage waveform for outbound communication, between the control center and the remote device. Offering data transmission capability of 60 bits in a second by TWACS and 1/1000 bit per second by Turtle, these systems were extensively used for remote meter reading and load control applications [20, 21]. With the advent of the smart grid, the scope of PLC has widened and modern PLC systems, both broadband PLC (BB-PLC) and narrowband PLC (NB-PLC), provide higher data transmission capabilities in comparison to the classical PLC. In particular, BB-PLC exploits a frequency band (2-30 MHz) and offers data transfer rates up to few hundred Mbps. The key application areas of BB-PLC are home automation and home energy management of the smart grid [17]. Conversely, NB-PLC, which operates in the CENELEC bands (3-148.5 kHz) or in FCC/ARIB bands extending up to 500 kHz, is gaining special attention for smart grid applications in medium voltage (MV) distribution and low voltage (LV) access grids [22]. Evolving from the interest of utility service providers, NB-PLC has emerged as a promising tech-

nology and over the years has gone through rapid development (a brief timeline showing the evolution of PLC technology is depicted in Fig. 2.1) [23]. There

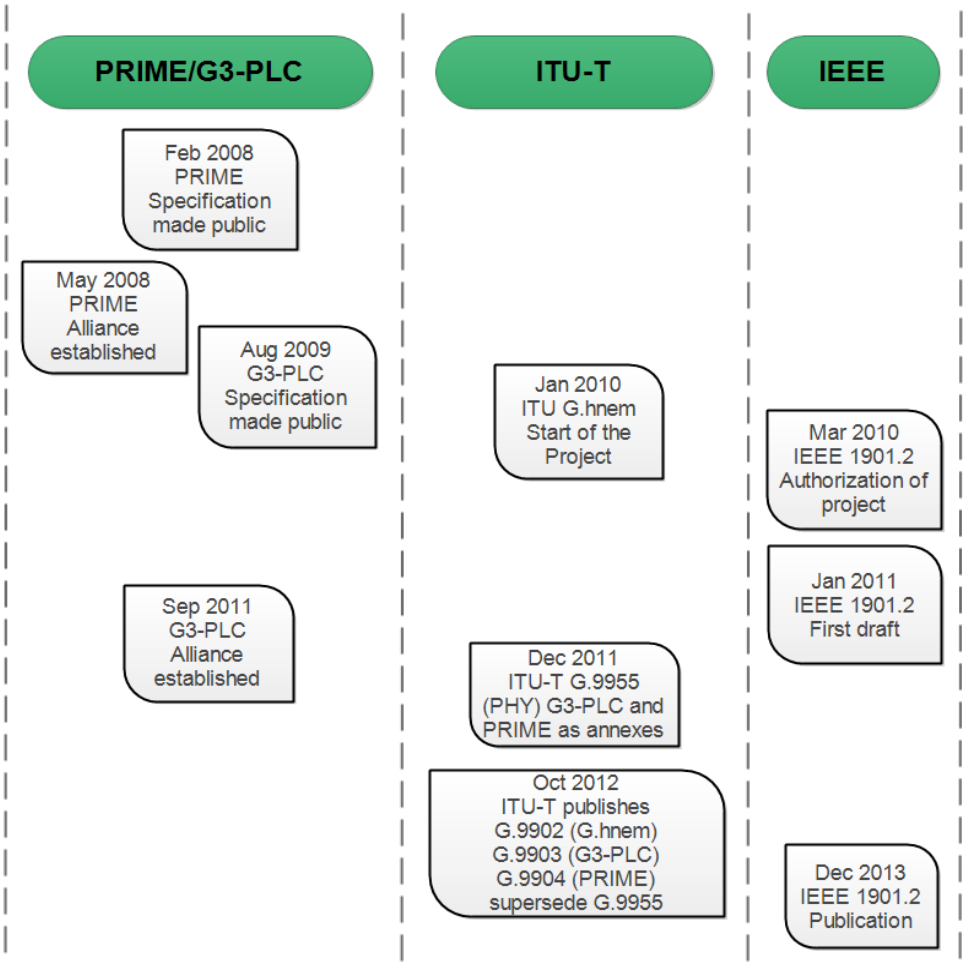


Figure 2.1: Development of high data rate PLC communication system standards for smart grid applications.

exist several NB-PLC standards, such as PRIME and G3-PLC (proposed by the alliance of the industries using PLC for automatic metering application), and ITU-G.9902 and 1901.2 from the standardizing bodies ITU-T and IEEE, respectively. At present, PLC systems based on these standards are extensively used to establish automatic meter reading (AMR) infrastructure. To enhance the applicability of PLC based technologies beyond AMR it is of utmost importance to understand the PLC environment and design signal processing algorithms tailored specifically to overcome its impairments.

To elaborate on the characteristics of PLC environment, in this chapter we will discuss the key behavior shown by power lines as communication channels.

Specifically, we will discuss the effect on high-frequency communication signals when transmitted through power lines. Furthermore, we will present channel modeling approaches to characterize power lines as communication channels. Apart from the effect of power line channels on high-frequency communication signals, in this chapter we will also elaborate on other impairments that turn an active electrical network into one of the most hostile environments for reliable data transmission. Later, we will introduce signal processing schemes that are proposed for PLC systems to overcome such impairments. We devote this chapter to defining the background on PLC and lay down the foundations to the research work that has been carried out during this Ph.D.

2.1 Power Lines as Communication Channels

A typical electrical grid has three domains. As shown in Fig. 2.2, a power grid is divided into high voltage (HV), medium voltage (MV), and low voltage (LV) networks. The HV part is called transmission grid and transports large quantities of high-voltage electricity from large power generation plants across large distances and delivers them to the MV distribution network. The MV part steps down the high-powered electrical energy from the HV transmission grid into lower voltages and distributes it among big to medium sized consumers. Further lowering the power, the LV part of the grid referred to as access grid, serves as a power delivery network and supplies electrical energy to the residential customers and small-sized enterprises [24]. Each domain of power grid represents a different communication environment for PLC. Recent field trials show that the HV transmission grid is the least noisy environment in comparison to the other domains of the grid. Furthermore, due to the less electrical branches and segregation from other parts of the power network, the HV transmission grid is a suitable environment for PLC [17, 25]. However, the multiple electrical branches and the impedance mismatches along an electrical path, and the presence of power converters along with other essential electro-mechanical components in the MV distribution and LV access grids make these domains harsh environments for PLC [26]. As MV distribution and LV access grids are the primary domains of smart grid applications [27], in the upcoming subsections we will elaborate on the behavior of power lines as a communication channel and their associated impairments.

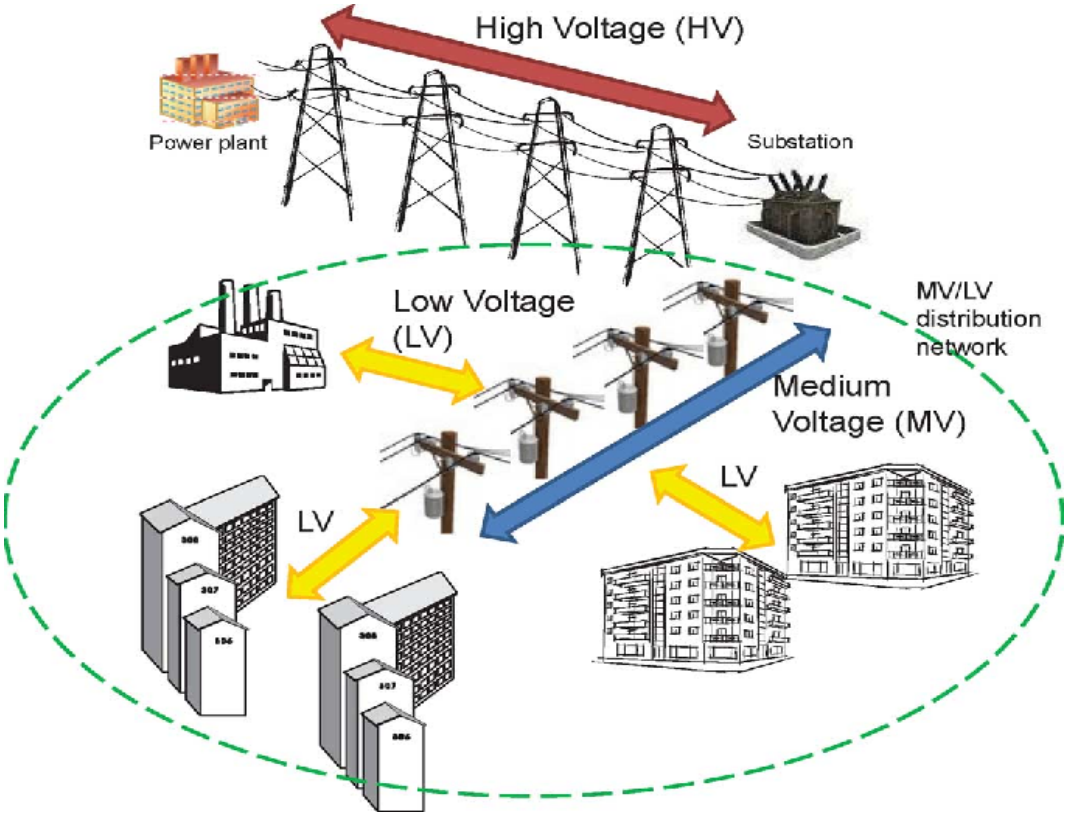


Figure 2.2: A typical electrical power grid network.

2.1.1 Channel Models

In the recent past years, a significant amount of effort has been made to characterize power lines in the distribution and access grids as communication channels. Understanding the behavior of power lines as data transmission media is of paramount importance to develop signal processing algorithms for PLC. Based on the proposed channel modeling approaches, power lines in an active electrical network can be modeled following three distinct approaches. Namely: top-down approach, bottom-up approach, and hybrid approach. In the upcoming subsections, we briefly elaborate on all three channel modeling approaches proposed for PLC.

2.1.1.1 Top-Down Approach

According to the top-down approach, the characterization of the power lines as a communication channel is done based on the net signal propagation phenomenon [28]. Typically, when a communication signal with a frequency higher than the fundamental electrical operating frequency (50 Hz or 60 Hz) is passed

through the power line in an active electrical network, it is reflected multiple times before reaching the receiving node. Due to the electrical branches and impedance mismatches along an electrical path, multiple copies of the transmitted signal are received at the receiving end via multiple paths. As a consequence, high-frequency communication signals are affected by a frequency selective channel [29]. To depict this property of power lines, the top-down approach treats communication channel as a black box and describes its transfer function characteristics by incorporating very few relevant parameters to the channel model. To base the power line communication channels on the fundamental physical effects, the values of the model parameters are determined by fitting them to the empirical data obtained via channel measurement campaigns.

To understand how this approach models power line communication channels, let us assume a typical single branch power line network as shown in Fig. 2.3. The nodes A and C in the network have matched impedances. The electrical branch originating at node B and terminating at node D divides this network into three distinct segments: (1), (2) and (3). In such a network, when a communication signal is transmitted from node A and is received at node C, the propagation of signal does not only take place along a line-of-sight (LOS) path between the transmitter and the receiver. Because of the reflection and transmission of the signal, due to intermediate discontinuities, there exist infinite propagation paths between the nodes A and C apart from the LOS path, i.e., A $\rightarrow$ B $\rightarrow$ C. Multiple copies of the transmitted communication signal via paths A $\rightarrow$ B $\rightarrow$ D $\rightarrow$ B $\rightarrow$ C, A $\rightarrow$ B $\rightarrow$ D $\rightarrow$ B $\rightarrow$ A $\rightarrow$ B $\rightarrow$ C and so on are received at the receiving nodes as the aggregation of the transmitted signal echoes. Denoting the reflection factors from: node D to B by r_{3D}, node B to D by r_{3B}, node B to A by r_{1B} and denoting the transmission factors from: node D to B by t_{3B}, node A to B by t_{1B}, each path i has a weighting factor g_i, representing the product of the reflection and transmission factors along the path. Moreover, the weighting factor of each path in power line is considered $|g_i| \leq 1$. The more transitions and reflections occur along a path, the smaller the weighting factor becomes. Furthermore, longer paths exhibit higher attenuation and contribute less to the overall signal received at the receiving end. Due to these reasons, the infinite number of paths between the transmitting end and the receiving end can be approximated using a finite number of propagation paths N to characterize a frequency selective multipath power line channel. In addition, the delay due to i^{th} path is related to

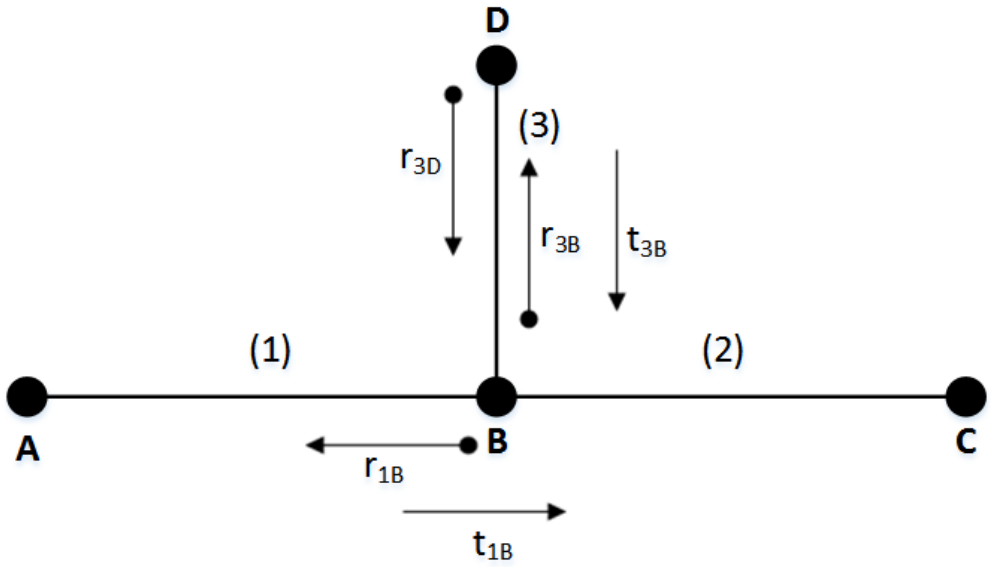


Figure 2.3: Multipath signal propagation through a cable with one tap.

the dielectric constant of the insulating material used in the power line and the signal propagation speed through the cable as $\tau_i = \frac{d_i\sqrt{\epsilon}}{c_o}$, where d_i is the length of the i^{th} propagation path, ϵ is the dielectric constant and c_o is the signal propagation speed. Also incorporating the attenuation of the communication signal $A(f,d)$ that increases with the propagation length and the transmission frequency, the signal components of the individual paths between A and C needs to be combined by superposition. Therefore, the frequency response between A and C following the top-down approach is expressed as [30]:

$$H(f) = \sum_{i=1}^{N} g_i A(f,d_i) e^{-j2\pi f\tau_i}, \tag{2.1}$$

where f denotes the communication signal frequency, N is the total number of multi-paths between transmitter and receiver, g_i is the gain of the i^{th} propagation path, τ_i is the signal propagation delay due to the i^{th} path and $A(f,d_i)$ is the attenuation of the communication signal due to the cable loss that depends on the transmitted communication signal frequency and the distance between the transmitter and the receiver following i^{th} propagation path.

2.1.1.2 Bottom-Up Approach

In contrast to the top-down approach, the bottom-up approach generates the power line channel transfer function (CTF) based on transmission-line theory.

This approach uses detailed knowledge of the power line network and characterizes the power line communication channels either in time domain or in frequency domain [31]. Typically detail information on large number of distributed components, such as cable types, number of joints and the connected loads, within the network is required to stetting up the network behavior. The CTF is then derived based on the scattering parameters or based on the four-pole impedance and admittance values of the network [32, 33]. Provided that all the relevant information on a typical power line network is available, bottom-up approach provides a deterministic characterization of the CTF.

A wide range of channel models based on the bottom-up approach are proposed to represent power lines as communication channels [34–37]. One such simplified channel model is proposed in [36]. Furthermore, an efficient transfer function computation for statistical PLC channel modeling is presented in [37].

2.1.1.3 Hybrid Approach

Following the hybrid method, the top-down and the bottom-up channel modeling approaches are exploited in conjunction to characterize power lines as a communication channel [24]. To generate the channel transfer function, first, the primary parameters referring to the power line impedance and admittance, and secondary parameters related to the propagation constant and the characteristic impedance of power lines are determined. Based on the evaluated values of these parameters, the end-to-end channel attenuation is characterized. To exploit the hybrid approach, three different methods (namely: multipath echo based model, scattering matrix method, and hybrid Smith chart) are proposed in [38].

2.1.1.4 Remarks on Channel Modeling Approaches

Both bottom-up and hybrid approaches provide a precise relationship between the channel effect on the communication signal and the physical characteristics of the power lines. Any relevant changes in the network configuration can be easily related to the channel transfer function. However, a price in terms of acquiring a priori information on the network topology and physical characteristics of the electrical cables needs to be paid to model the power line channel with such a degree of determinism. As a downside, the bottom-up and hybrid channel models cannot predict the behavior of power lines in a multitude of scenarios.

Given the fact that a broad variety of electrical network topologies and wiring practices exist in the real world, it becomes significantly important to have a channel model that can replicate multiple PLC scenarios. To ease the development of robust and reliable signal processing algorithms, it is more important to focus on the propagation phenomena of a communication signal in an active electrical network, rather than on the physical characteristics and parameters of the network. The top-down models best suit to fulfill this need. Hence, a multi-path channel model proposed in [30, 39] to express the channel transfer function of power line in a closed-form equation is used in this thesis. The used channel model is a particular version of the general channel model in (2.1) and is expressed as [39]:

$$H(f) = \sum_{i=1}^{N} g_i e^{-(\alpha_0+\alpha_1 f^k)d_i} e^{-j2\pi f(d_i/v_p)}, \tag{2.2}$$

where

- i is the index number of the path, with the shortest path indexed as $i = 1$;
- α_0, α_1 are the cable attenuation parameters;
- k is the exponent of the attenuation factor (typical values are between 0.5 and 1);
- g_i is the weighting factor of path i;
- d_i is the length of path i;
- v_p is the ratio of signal propagation speed and the dielectric constant of the electric cable.

The frequency response of a typical power line in the MV distribution grid following the channel model in (2.2) is shown in Fig. 2.4.

2.2 PHY of Modern PLC Systems

The standardization organizations propose CENELEC bands for communication over power line networks. The CENELEC band, which is divided into four sub-frequency bands, reserves CENELEC-A (3-95 kHz) for utility service providers to exploit for smart grid applications. All the proposed standards

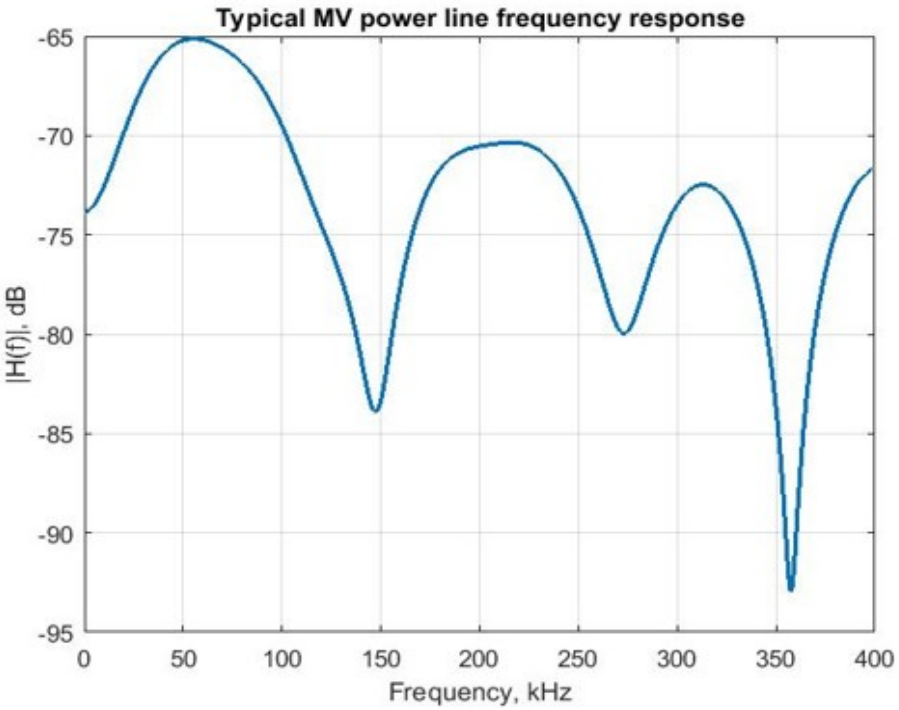


Figure 2.4: Frequency response of a typical power line in MV distribution grid.

adopt orthogonal frequency division multiplexing (OFDM) to efficiently use the available bandwidth and provide high data transmission capabilities. Apart from that, OFDM also provides robustness against the intersymbol interference caused due to the dispersive nature of the channel by using cyclic prefix (CP) and a simple one-tap channel equalization to overcome the channel frequency selectivity.

The PRIME standard uses $f_s = 250$ kHz as sampling frequency and fast Fourier transformation (FFT) size of 512 to modulate the transmission data. Accounting to the FFT size and the bandwidth used for transmission, the PRIME based PLC systems has $\Delta f = 488$ Hz of subcarrier spacing. Among the available subcarriers, exploitation of subcarriers indexed from 86-182 are only used for transmission such that the PRIME signal is located in the frequency range 42-89 kHz. By employing the DPSK modulation scheme with the modulation order ranging from 2 to 8 and switching on or off the convolutional coding, PRIME supports data transmission rates up to 128.6 kbps. An overview of the PHY parameters of PRIME is provided in the first column of Table. 2.1.

Different to PRIME, G3-PLC operates at the sampling frequency of $f_s =$

400 kHz and uses 256 as FFT size. The subcarrier spacing in G3-PLC is $\Delta f = 1.65625$ kHz. Using only the subcarriers indexed from 23 to 58, G3-PLC occupies the frequency range 35.9-90.6 kHz in the CENELEC-A spectrum. Offering three different transmission modes (robust, DBPSK and DQPSK), G3-PLC offers data transmission rate scalable up to 33.4 kbps (maximum using DQPSK). For error correction, a convolutional code in conjunction with Reed Solomon and repetition codes are applied. An overview of the OFDM PHY parameters of G3-PLC is provided in the second column of Table. 2.1.

The standards PRIME and G3-PLC are proposed by the alliance of industries that deploy PLC based devices for metering purposes. To resolve interoperability issues among these standards, a unified standard for next generation NB-PLC technology has been proposed by IEEE. The NB-PLC standard operating in CENELEC-A band proposed by IEEE is called IEEE 1901.2 and is being developed to advance features based on G3-PLC. Operating with a sampling frequency equivalent to $f_s = 400$ kHz, an FFT size of 256, a subcarrier spacing of 1.5625 kHz and employing higher order modulation with error correction coding pertaining to convolutional codes applied in conjunction with Reed Solomon and repetition coding, IEEE 1901.2 offers data transmission rates scalable up to 52.3 kbps. The third column in the Table. 2.1 provides the OFDM parameters of the IEEE 1901.2 standard.

Today, PLC modems based on the PRIME, G3-PLC and IEEE 1901.2 standards are widely deployed for metering applications. As mentioned earlier in section 1.1 of chapter 1, a huge geographical area has already been covered by PLC based modems for energy consumption data collection. However, OFDM based systems are vulnerable to impulsive noise (IN) and narrowband interference (NBI) that are inevitable in PLC environments [40]. To design robust signal processing algorithms to overcome these impairments, a better understanding of these nuisance types is important. Hence, in the upcoming subsections, we present brief description of noise types that make power line a hostile environment for data transmission.

2.3 Noise and Interference in PLC Environment

Different from the conventional radio communication environment, where thermal noise at the receiver front end is considered as the predominant source of

Parameter	PRIME	G3-PLC	IEEE 1901.2
Frequency Range	42-89 kHz	35.9-90.6 kHz	35.9-90.6 kHz
Sampling Frequency	250 kHz	400 kHz	400 kHz
FFT Length	512	256	256
Cyclic Prefix	48	30	30
Subcarrier Spacing	488 kHz	1.5625 kHz	1.5625 kHz
Modulation	DBPSK, DQPSK, D8PSK	DBPSK, DQPSK	DPSK (QAM) M = 2,4,8,16
Forward Error Correction	Conv code	Conv code, RS code, Rep code	Conv code, RS code
Data rate (max)	128.6 kbps	33.4 kbps	52.3 kbps

Table 2.1: PHY layer parameters of OFDM based CENELEC-A band PRIME, G3, IEEE 1901.2 PLC standards. The term "Conv, RS and Rep" in the forward error correction row denote convolutional, Reed Solomon and Repetition codes, respectively.

noise, in the power line environment, the noise is broadly categorized into background noise and IN. Besides, PLC systems are also susceptible to interference from wireless communication systems operating in the close proximity. In the following three subsections, we briefly describe the noise and interference phenomena in the PLC environment.

2.3.1 Background Noise

Due to the thermal noise originating from the front-end amplifiers in PLC systems, background noise with a constant spectral density and constant power typically arises. A different version of the background noise appears if an electrical appliance with an oscillator with rectified power supply, but unsmoothed voltage, is connected to the electrical network. The power of this background noise changes synchronously with the mains absolute voltage with periodicity equivalent to half of the mains cycle [24]. The time-variant background noise is more evident in an in-home scenario where consumers can directly plug their electrical appliances into the electrical network, rather than in the MV distribution and LV access grids. The background noise in MV and LV grids is typically white [41, 42].

2.3.2 Impulsive Noise

In an electrical network, apart from the background noise, noise with very brief time duration arises due to the operation of the power converters connected to the network. Such noise bears higher power spectral density than the background noise and occurs randomly in the time domain and severely impairs communication signals. The IN in PLC environments, depending on the pattern of occurrence, can be classified as [43–45]:

- **Synchronous IN:** A sample of IN falling in this category bears the periodicity of occurrence equivalent to the frequency of alternating current (AC) mains frequency or twice the operational frequency of a typical electrical signal. Operation of the electrical appliances that are connected to the electrical network and have a thyristor-based controller or silicon-controlled rectifier create this type of noise. Apart from this, appliances operating on brush motors also create IN samples when frequent switching of motor brushes takes place [24].

- **Asynchronous IN:** Transients of switching regulators connected to the electrical grid are the main sources of asynchronous IN. A sample of this noise type does not follow a periodicity related to the AC mains frequency. Rather, it occurs randomly and bears higher power spectral density as compared to the background noise [46]. Presence of asynchronous IN is one of the most peculiar characteristics of the MV distribution and LV access grids and make these domains hostile for reliable PLC [47].

The occurrence of asynchronous IN in PLC environment is studied and extensively elaborated in an abundance of literature related to PLC. Over the years, efforts have been made to characterize the occurrence of IN in power lines and models such as Bernoulli-Gaussian (BG), Middleton Class A (MC-A), Markov-Middleton (MM), and Markov-Gaussian (MG) are proposed to depict IN in power line channels [48–51]. The BG model defines the probability density function (PDF) of IN as a mixture of Bernoulli and Gaussian random processes. The MC-A model defines the IN PDF as a sum of different zero mean Gaussian PDFs with different variances weighted by the Poisson distribution. Both BG and MC-A models are widely accepted to model effects of IN in communication systems. Furthermore, to address rare situations where the IN samples occur in burst, the BG model has been extended to MG and MC-A model has been extended to MM.

Among the available models, BG noise model is widely adopted in the literature to represent the effect of IN on PLC systems. BG model ensures a good tradeoff between the mathematical simplicity and fitting the physical phenomenon [52]. Based on this model, a sample of IN is denoted as [53, 54]:

$$i_n = b_n g_n, \tag{2.3}$$

where i_n denotes the n^{th} sample of IN, b_n is an independent and identically distributed (i.i.d.) Bernoulli random variable and g_n is an i.i.d. complex Gaussian random variable with zero mean and variance σ_i^2. The variance σ_i^2 denotes the power of the IN sample.

2.3.3 Narrowband Interference

Apart from the background noise and IN, PLC systems are also susceptible to interferences from wireless communication systems. Typically, if a wireless communication system is operating in close proximity, it creates interference and impairs the operation of PLC systems [55]. Similar to the IN, interference to PLC systems from wireless communication devices is impulsive, but now in the frequency domain. These NBIs act as tone jammers and can be modeled as [55]:

$$e_n = \sum_{k=1}^{c} A_k e^{j(\omega_k n + \phi_k)}, \tag{2.4}$$

where e_n is the n^{th} sample of the NBI, A_k is the amplitude of k^{th} interference, ω_k is the normalized frequency of the k^{th} interference, ϕ_k is the phase of the k^{th} interference, c is the total number of NBIs.

2.4 General Communication Model for PLC

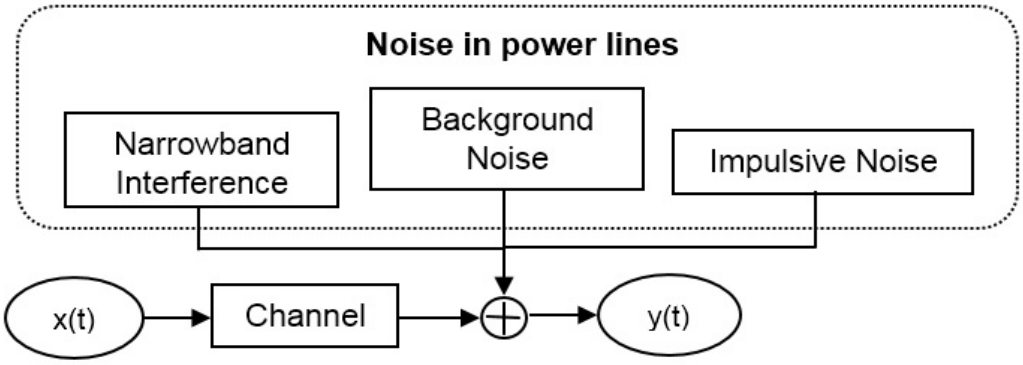


Figure 2.5: A general PLC communication model.

Considering the frequency selective multi path effect and the noise encountered during power line data transmission, a typical PLC model can be graphically represented as in Fig. 2.5 and mathematically expressed as [56]:

$$y(t) = h(t) * x(t) + n(t), \tag{2.5}$$

where y is the received signal, h denotes the channel, x represents the transmitted signal and n refers to the noise occurring during data transmission at a given time t. The notation $*$ in (2.5) refers to the convolution operation between the communication channel and the transmitted signal. Addressing to the theoretical possibility of having an infinite number of reflection paths between a transmitting point and a receiving point in a power line network, the multi path behavior of the channel can hence be expressed as:

$$h(t) = \sum_{i=1}^{\infty} a_i \delta(t - T_i) e^{j\varphi_i}, \tag{2.6}$$

where a_i, T_i and φ_i respectively denote the magnitude, delay and phase of the i^{th} path. Moreover, the time domain system model in (2.5) can be represented in frequency domain as:

$$Y(f) = H(f)X(f) + N(f). \tag{2.7}$$

The variables Y, H, X and N in (2.7) respectively denote the received symbol, channel impulse response, transmitted symbol and the frequency domain noise samples at a given frequency f.

In order to overcome the frequency selective behavior of the power line channel, as mentioned in section 2.2, modern PLC systems exploit multi carrier schemes for data transmission and reception. The communication system model in (2.5) as discrete tone model can be expressed as:

$$\breve{\mathbf{y}} = \mathbf{\Lambda}\breve{\mathbf{h}} + \mathbf{F}\mathbf{n}, \tag{2.8}$$

where $\breve{\mathbf{y}} = [\breve{y}_1, \breve{y}_2, \ldots, \breve{y}_N]^T$ is a vector of received symbols, $\mathbf{\Lambda}$ is diagonal matrix of dimension $N \times N$ with frequency domain transmitted symbols as its entries, $\breve{\mathbf{h}}$ is a vector of dimension $N \times 1$ containing the channel frequency response coefficients corresponding to the discrete frequency points, $\mathbf{F}$ is the $N \times N$ discrete Fourier transform matrix and $\mathbf{n} = [n_1, n_2, \ldots, n_N]^T$ is the vector containing the time domain samples of the noise. The variable N in above notations denotes the number of subcarriers in the system. In the upcoming chapters we will recall this signal model where necessary and adapt it according to the scenario under study.

2.5 Challenges for Reliable Data Transmission using PLC

Summarizing the characteristics of the PLC environment, an active power line network is a complex domain where data transmission is impaired by the channel frequency selectivity and by the occurrence of background noise, IN, and NBI.

In contrast to the traditional signal processing algorithm design done for the conventional communication systems, which are mainly impaired by AWGN efficient and robust signal processing schemes to overcome the effect of IN and NBI, and effective equalization against the channel frequency selectivity are required for the PLC systems. The frequency selective multipath fading of the power lines adds significant importance to an accurate channel estimation strategy to effectively equalize the received symbols against the channel response. In addition, the presence of random IN and NBI demands more sophisticated noise and interference mitigation schemes. In the following two sections, we elaborate on the signal processing algorithms that are usually proposed to mitigate noise and robustly estimate the channel in the presence of IN for the PLC systems.

2.6 Noise and Interference Mitigation

The peculiar behavior of random IN and NBI severely affects the performance of OFDM-based PLC systems [57–59]. In order to overcome such impairments, modern PLC systems use interleaving and forwarding error correction (FEC) coding schemes. Following this approach, FEC codes are first applied to the raw input bits. After generating the symbols to be transmitted by mapping the error correction coded input bits, the symbols are allocated to the subcarriers of the system to generate the OFDM symbols. One such scheme based on LDPC coding for error correction is proposed in [60]. In order to further enhance the performance of PLC systems in the presence of IN, some algorithms propose to use interleaving schemes in conjunction with FEC. In such schemes interleaving is applied either in the time domain or in the frequency domain. To perform frequency domain interleaving (FDI) the sequence of OFDM symbols goes through the FDI block before inverse discrete Fourier transformation (IDFT) is applied at the transmitter [61]. Upon reception, the de-interleaving block at the receiver effectively spreads the IN energy over a large number of OFDM symbols. The errors due to the occurrence of IN and NBI during data transmission are hence distributed among multiple OFDM symbols. Demodulating the received deinterleaved symbols and using a FEC decoder, corrupted data due to noise are recovered at the receiver. By employing such technique, the overall BER performance of the system can be improved. Some other schemes deploy time domain block interleaving (TDBI) in conjunction with FEC. The trade-off between these two approaches is that, at high signal to noise ratio (SNR) regime TDBI performs better than the FDBI, whereas at low SNR regime FDBI outperforms TDBI [39, 62, 63]. Furthermore, IN mitigation schemes exploiting generalized array codes and row and column array codes are proposed in [64]. Some other schemes that exploit Luby transform codes and space-time block codes to overcome the effect of IN are proposed in [42, 65].

Apart from the coding and interleaving schemes, signal processing algorithms that are tailored specifically to overcome IN and NBI can be used. In this section, we briefly elaborate on the noise and interference mitigation schemes that are proposed for OFDM based PLC systems.

2.6.1 Impulsive Noise Mitigation

One of the most intuitive approaches to reliably exchange data between two communicating nodes in an IN environment is to perform retransmission of data until an error-free reception is achieved [66]. However, a scheme like this not only instigates latency in data transmission time hindering real-time communication between the participating nodes, but also makes inefficient use of the available bandwidth [58]. To overcome the need for data retransmission, signal processing algorithms that mitigate the effect of IN on communication signals are proposed. Broadly categorized into passive mitigation and active mitigation, the proposed algorithms are detailed in the following two subsections.

2.6.1.1 Passive Mitigation

In passive mitigation, the power of IN is suppressed by nulling or clipping the IN corrupted samples in the received signal [67, 68]. These schemes compare the amplitude of each sample in the received signal with a reference threshold value to determine the IN corrupted samples. Any sample bearing amplitude higher than the threshold value is considered corrupted by the IN. The IN affected samples are then nulled, following the nulling algorithm, or their amplitudes are clipped to a safe level, following the clipping algorithm to suppress the IN power. Simplified nulling schemes for PLC system are proposed in [69, 70]. Likewise, a simplified clipping algorithm is proposed in [71, 72] for PLC systems. To further improve the system's performance, an algorithm that exploits clipping and nulling schemes in conjunction is proposed in [73,74]. The performance gain by exploiting both schemes in an ad-hoc manner is significant [73, 75]. Enhancing the performance that can be achieved by passive mitigation schemes, a composite algorithm that incorporates sample replacement in addition to nulling and clipping of IN corrupted samples is proposed in [75]. Following this approach, three different threshold values are defined for clipping, replacing and nulling operations. The amplitude of each sample in the received signal is compared with all three reference threshold values and appropriate action is carried out: clipping of sample having amplitude higher than clipping threshold, nulling of the sample that has amplitude higher than the nulling threshold and replacing the samples having amplitudes between clipping and nulling thresholds by predetermined values.

Passive noise mitigation algorithms employ threshold values that are exper-

Operation	Remarks	References
Nulling	Threshold based operation	[68–70, 76]
Clipping	Threshold based operation	[67, 71, 72, 74]
Nulling + Clipping	Threshold based operation	[73, 74]
Clipping + Replacement + Nulling	Threshold based operation	[75]
Nulling/clipping	Adaptive Threshold based operation	[58, 77–79]

Table 2.2: Impulsive noise mitigation schemes based on passive approach.

imentally determined. The identified threshold value is used as a static limit to capture IN affected samples. However, the noise in power lines, in general, is not static and changes depending upon the number of loads connected to the grid, the number of switching operations performed to stabilize the network and ongoing human interaction with the electrical appliances at a specific point of time [74]. Due to the variation in the occurrence and power of the IN, adaptive mitigation schemes that adjust their threshold values according to changes in the IN are required to better segregate the corrupted samples in the received signal. Several such algorithms are reported in [58, 76, 77]. By employing an adaptive determination of the threshold value, the dynamic non-linear operation based algorithm suppress IN power better than the classical non-linear operation based schemes. Moreover, an iterative nulling scheme based on an adaptive threshold to mitigate the effect of IN in the received signal is proposed in [76], where the threshold value for nulling is dynamically determined by exploiting outlier detection theory. Furthermore, an iterative approach comprising of pre-IN mitigation block, dealing with the stronger portion of the IN, and post-IN mitigation block, dealing with cancellation of residual IN in the received signal, is adopted [76]. A similar adaptive clipping based IN mitigation algorithm is proposed in [58]. In addition, some variants of the non-linear IN mitigation schemes are also reported in [78, 79].

2.6.1.2 Active Mitigation

Unlike passive mitigation schemes, active mitigation schemes estimate the IN and cancel it from the received signal. To estimate the IN, active mitigation schemes first identify the location of IN samples. After estimating the IN support, amplitude and phase estimation is performed. Finally the estimated IN is canceled from the received signal to achieve IN mitigated samples for demodulation.

In order to estimate the IN support, active mitigation algorithms exploit the time domain sparsity characteristics of IN. Sparse signal recovery schemes that are based on compressive sensing (CS) and sparse Bayesian learning methods are proposed in [80–84] to estimate the support of IN. After estimating the support, two different approaches are then followed to mitigate the effect of IN from the received signal in [80]. In the first approach, the samples in the received signal corresponding to the estimated locations of the IN samples are blanked. Different to this, in the second approach the amplitude and phase of the IN samples at the selected locations are estimated. Thus, the estimated IN is then subtracted from the received signal. Similar to the noise cancellation approach proposed in [80], [85] and [86] propose IN cancellation schemes that rely on sparse Bayesian learning methods. A further enhancement in IN sample support estimation is proposed in [81], which assumes CS-based estimated supports as a coarse estimation of IN support and further refines it by exploiting a priori information on the IN samples distribution. A similar approach is proposed in [84] where the fine estimation of IN support is done by employing an adaptive threshold based operation. Moreover, the algorithms proposed in [61, 87] exploit time domain interleaving in conjunction with the support detection algorithm to estimate the IN samples occurring in burst. Different to the application of sparse signal recovery techniques, the algorithms proposed in [88, 89] perform replica signal estimation of impulses to mitigate the effect of IN. After estimating it the impulse replica is canceled from the received signal. The resulting samples are then compared against a reference threshold value to perform fine estimation of the IN support. This procedure is iterated until all the IN samples are estimated.

To enhance the precision in IN support detection that can be achieved by the CS-based schemes, a basis pursuit (BP) based algorithm is proposed in [90,91]. This approach performs IN support detection with minimal number of measurements and bears high computational complexity. Reducing the com-

Operation	Remarks	References
Nulling	CS-based support estimation	[80]
Cancellation	CS-based support estimation + least-squares amplitude estimation, Replica Signal Estimation	[61, 80, 85–89]
Spectrum Whitening	CS-based support estimation + minimum mean squared error amplitude estimation	[81, 84]
Greedy Pursuit	IN support detection	[92–98, 100, 101]

Table 2.3: Impulsive noise mitigation schemes based on active approach.

putational complexity bared by the BP scheme, algorithms based on iterative greedy pursuit of IN support are proposed in [92–96]. The proposed schemes exploit orthogonal matching pursuit (OMP) [92, 95], stagewise OMP [93] and regularized OMP for sparse signal reconstruction. Some other algorithms based on subspace pursuit and compressive sampling matching pursuit (CoSaMP) are proposed in [97,98]. These algorithms require a priori information on the sparsity level of the signal. In a practical power line setup, where it is not possible to have this information beforehand, a high quality estimation of IN cannot be guaranteed. To address this situation, a new greedy algorithm has been proposed in [99]. The proposed scheme, called sparsity adaptive matching pursuit (SAMP), performs blind signal recovery as it does not require a priori information on the signal sparsity. The SAMP approach performs stage-by-stage estimation of the sparsity level and the true support set of the target signal. The performance of the SAMP can be further enhanced by the priori-aided SAMP (PA-SAMP) algorithm proposed in [100]. Moreover, a recent work in the literature proposes an improved version of classical CS based schemes and greedy algorithms based on OMP and CoSaMP for accurate IN support estimation [101].

To summarize the IN mitigation schemes in the existing literature, the proposed algorithms for OFDM-based PLC systems are differentiated and reported as active and passive methods in Table.2.2 and Table.2.3.

2.6.2 Narrowband Interference Mitigation

In order to mitigate the effect of NBI, two approaches (namely: Nyquist windowing and modified soft decision decoding) are proposed in [55] for an OFDM-based PLC system. Different to this, a non-linear technique based on subcarrier excision is proposed in [59]. The subcarrier excision algorithm has a simpler implementation in comparison to the ones proposed in [55], where the subcarriers bearing power higher than a reference threshold value are determined contaminated and are nulled to mitigate the effect of NBI. To further enhance the performance that can be achieved by nulling the NBI-corrupted subcarriers, [102,103] propose implementing a time domain notch filter in conjunction with the subcarrier nulling algorithm. By exploiting the time domain notch filter and frequency domain nulling algorithm in an ad-hoc manner, the noise floor problem that is typically encountered while using the nulling algorithm is mitigated. Moreover, when the NBI occurs non-orthogonally with the OFDM symbol, the subcarriers close to the ones that are corrupted also need to be nulled to overcome the spectral leakage of the NBI. In order to mitigate this type of problem, [104] proposes using a prediction error filter to localize the erasures to the subcarriers surrounding the interference, while not affecting the remaining subcarriers. Furthermore, [104] employs a bit-interleaved coded OFDM system to overcome the data loss due to the subcarrier erasure.

Apart from performing subcarrier excision/nulling, [105] proposes employing extreme value theory (EVT) to identify subcarriers that are corrupted by NBI in a Turbo coded OFDM transmission. The EVT is used to define a weight function and the corrupted subcarriers are down weighted to mitigate the effect of NBI. Other approaches that employ low rank parity check codes and an optimized time-frequency interleaving scheme to mitigate the effect of NBI on an OFDM-based system are illustrated respectively in [106,107]. Apart from these schemes, some sophisticated CS based algorithms to estimate the NBI and cancel it from the received signal are proposed in [108–113].

2.7 Channel Estimation in PLC Systems

As mentioned earlier in section 2.1.1, in a typical power line network (due to the branching and impedance mismatches along an electrical path) the communication signals are typically received via multiple paths. As a consequence, the high-frequency communication signals are distorted by a frequency selec-

tive channel. To overcome the frequency selectivity it is necessary to estimate its fading behavior [114]. By accurate tracking of the channel during data transmission, the received symbols can be effectively equalized against the channel impairments [115]. Modern NB-PLC systems use OFDM to combat channel frequency selectivity [116]. To facilitate the receiver estimate the channel, apart from the payload symbols, pilot symbols which are known to the receiver are also transmitted within a transmission frame [117]. Upon reception, the receiver exploits the known pilot symbols to estimate the channel and uses it to equalize the payload symbols against the channel frequency selectivity [115,118,119]. However, the brief duration IN samples that occur randomly during the PLC transmission show wide band characteristics and corrupt pilot symbols resulting in erroneous channel estimation [120].

In conventional OFDM based PLC systems, least squares (LS) channel estimators are deployed in conjunction with passive IN mitigation techniques. Following these approaches, the IN power is first suppressed by employing nulling or clipping based algorithms [76, 121]. After suppressing the IN power, the receiver then performs LS estimation of the channel exploiting the known pilot symbols [122, 123]. An iterative approach to improve the performance of such channel estimators is proposed in [76]. Following this approach, a pre-mitigation of IN followed by LS channel estimation and post-mitigation of IN is done at each iteration. Treating IN samples as outliers in the time domain, channel estimation schemes for PLC systems based on domain transformation techniques are proposed in [124–131]. These schemes suppress the IN affected samples in the received signal and perform LS channel estimation to determine the frequency response of the channel. The estimated channel frequency response (CFR) coefficients are then used to generate the channel impulse response (CIR). The coefficients of the CIR other than the ones having most of the channel energy are set to zero and a refined CFR is generated using the enhanced CIR. Unlike domain transformation based denoising algorithm, [130] proposes wavelet based improvement for channel estimation in power line communication environment that is impaired by the IN. Apart from the domain transformation and non-linear noise mitigation based schemes, some robust channel estimators are proposed in [132, 133]. These approaches exploit outliers search scheme to suppress IN samples and unknown data symbols (symbols other than the known pilot symbols) treating them as noise for channel estimation. After reduction of the noise terms, the scheme in [133] performs iterative search process based on robust support vector machine to determine

the channel. Different to [133], [132] proposes a channel estimator to perform maximum-likelihood estimation of channel coefficients exploiting a cost function tolerant to outliers.

Apart from the conventional approaches, [134] proposes an adaptive channel estimation scheme that is based on variable step size (VSS) sign algorithm. The proposed scheme provides significant gain in terms of channel estimation accuracy in comparison to the conventional PLC channel estimation algorithms. Improving the performance gain of the estimator in [134], [135, 136] proposes VSS affine projection sign algorithm to effectively estimate the channel in the presence of IN. Moreover, sign least squares algorithm proposed in [137] offers an efficient approach to estimate finite impulse response power line channel. Some other channel estimation strategies perform joint channel and IN estimation exploiting the time domain sparsity of both entities [138]. Among the two proposed algorithms, first algorithm performs joint estimation of the CIR and IN based on LS principle. The second algorithm aims to reduce the computational complexity by employing expectation-maximization method.

Chapter 3

Impulsive Noise Mitigation and Narrowband Interference Cancellation Schemes for PLC Systems

In the previous chapter we briefly discussed about the behavior of power lines as communication channels and the impairments that make an active electrical network a hostile environment for data transmission. In the section 2.3 of chapter 2, we explained how impulsive noise (IN) and narrowband interference (NBI) play roles of major impairments in power line communication (PLC) environments and elaborated the state-of-the-art signal processing algorithms to mitigate these impairments in section 2.6 of chapter 2.

In this chapter, we will introduce novel IN mitigation and NBI cancellation schemes that are designed for modern PLC systems based on orthogonal frequency division multiplexing (OFDM). For concise representation of the signal processing algorithm design, we divide this chapter into two sections. In the first section, we outline the IN mitigation scheme and devote the second section to describe the NBI cancellation scheme.

3.1 Impulsive Noise Mitigation

IN is inevitable in power line environment and is one of the prominent sources of error that hinder reliable data transmission through power lines in an active electrical network [139]. Mitigation of IN for PLC systems has been a hot

research topic over several recent past years. As outlined in section 2.5 several algorithms have been proposed to get rid of the effect of the IN. In such conventional schemes, the IN contaminated samples in the received signal are based on a threshold value and are then nulled, clipped or nulled and clipped in conjunction to suppress the power of the IN in the received signal. Since the IN corrupted samples are identified based on the threshold value, lack of precision in the threshold identification leads towards faulty detection of the corrupted samples and creating additional distortion in the received signal [58]. To overcome this problem, we propose an active noise mitigation scheme that first identifies the IN corrupted samples and then mitigates the effect of IN only from those samples. By mitigating the effect of IN only from the corrupted samples, the problem of distortion of uncorrupted samples typically encountered while deploying non-linear noise mitigation schemes is overcome [140].

To estimate the IN present in the received signal, the proposed noise mitigation scheme performs successive estimation of order, support, and amplitude of the IN samples. The order that corresponds to the number of IN samples occurring during the transmission of an OFDM symbol is estimated exploiting the information measure approach known as minimum description length (MDL) criterion. After estimating the order of the IN, multiple signal classification (MUSIC) algorithm is exploited to estimate the locations of the IN samples in the received signal. Finally, after estimating the order and support of the IN, the proposed IN estimator performs minimum mean squared error (MMSE) estimation of the amplitude and phase of the IN samples. Thus estimated IN is canceled from the received signal to provide IN mitigated samples for demodulation.

To concisely elaborate on the proposed IN mitigation algorithm, this section is further divided into four subsections. In subsection 3.1.1, the system model for OFDM-based PLC system is introduced. In subsection 3.1.2, the proposed IN order, support, amplitude and phase estimation algorithm is described. The performance of the proposed algorithm is validated numerically in varied IN scenarios and the simulation results are shown in subsection 3.1.3. Finally, a brief conclusion is drawn at the end of this section.

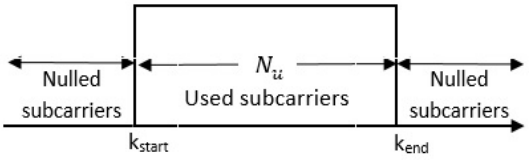


Figure 3.1: Subcarrier structure

3.1.1 System Model

In order to derive the proposed IN mitigation scheme, we begin by considering an OFDM based PLC system with N subcarriers. Among these N subcarriers, only N_u subcarriers, ranging from k_{start} to k_{end}, are used for the data transmission. The remaining $N - N_u$ subcarriers are nulled/unused as shown in Fig. 3.1. The corresponding subcarrier alignment is adopted from the narrowband PLC (NB-PLC) standard [39]. Adapting the general PLC communication system model in (2.8) to the setup shown in Fig. 3.1, the frequency domain received symbols corresponding to an OFDM symbol is expressed as:

$$\breve{\mathbf{y}} = \begin{bmatrix} \mathbf{0} \\ \mathbf{\Lambda}\breve{\mathbf{h}} \\ \mathbf{0} \end{bmatrix} + \mathbf{F}\mathbf{n}, \tag{3.1}$$

where $\mathbf{\Lambda}$ denotes a diagonal matrix of dimension $N_u \times N_u$ whose diagonal entries contain the frequency domain transmitted symbols, $\breve{\mathbf{h}}$ is the channel vector of dimension $N_u \times 1$ containing the coefficients of channel frequency response (CFR) corresponding to the used subcarriers. The matrix $\mathbf{F}$ in (3.1) is the unitary discrete Fourier transform (DFT) matrix having elements as $[\mathbf{F}]_{a,b} = \frac{1}{\sqrt{N}} \exp^{-j2\pi ab/N}$, where the variables a and b respectively denote the row and column index of the DFT matrix. The vector $\mathbf{n} = \mathbf{w} + \mathbf{i}$ in (3.1) is the noise vector containing the time domain samples of background noise and IN where vectors $\mathbf{w}$ and $\mathbf{i}$, each of dimension $N \times 1$, respectively denote background noise and IN. The samples of the background noise are assumed to be independent and identically distributed (i.i.d.) additive white Gaussian noise (AWGN) random variables. The entries of vector $\mathbf{w}$ are therefore Gaussian distributed random variables having zero mean and variance $\bar{\sigma}_w^2$. The IN vector $\mathbf{i}$ is a sparse vector, having only N_{imp} non-zero entries. The IN is modeled based on Bernoulli-Gaussian noise model as introduced in (2.3) in section 2.3.2 of chapter 2 and is defined as:

$$[\mathbf{i}]_n = b_n g_n, \tag{3.2}$$

where $[\mathbf{i}]_n$ is the n^{th} sample of IN, b_n is the Bernoulli distributed random variable and g_n is a random variable which is Gaussian distributed with zero mean and variance $\bar{\sigma_i}^2$. The indexes within the set $\{1, \ldots, N\}$ that are contaminated by IN are denoted by $\mathcal{I} = \left\{n_1, \ldots, n_{N_{imp}}\right\}$ such that $N_{imp} = |\mathcal{I}|$. Furthermore, the signal to noise ratio (SNR) of the system is defined as the ratio $\bar{\sigma}_s^2/\bar{\sigma}_w^2$, where $\bar{\sigma}_s^2$ is the power of the transmitted signal and the IN to background noise power ratio (INR) is defined as the ratio $\bar{\sigma}_i^2/\bar{\sigma}_w^2$, where $\bar{\sigma}_i^2$ is the power of the IN.

Transforming the frequency domain received symbols to time domain samples by inverse discrete Fourier transformation (IDFT), the time domain representation of the signal model in (3.1) is done as:

$$\mathbf{y} = \sqrt{N}\mathbf{F}^H\mathbf{E}\mathbf{\Lambda}\mathbf{G}^H\mathbf{h} + \mathbf{w} + \mathbf{i}, \tag{3.3}$$

where $\mathbf{E}$ is the selection matrix of dimension $N \times N_u$ that has entries as 1s identifying the locations of the subcarriers that are used for the data transmission. The matrix $\mathbf{G}$ in (3.3) is a section of inverse discrete Fourier transformation (IDFT) matrix $\mathbf{F}^H$ that is obtained by selecting the N_u columns of $\mathbf{F}^H$ associated with the locations of used subcarriers and the corresponding L upper rows, L being the length of channel impulse response (CIR). Denoting the CIR by a column vector $\mathbf{h}$ of length $L \times 1$, the CFR $\breve{\mathbf{h}}$ in (3.1) is equivalent to

$$\breve{\mathbf{h}} = \sqrt{N}\mathbf{G}^H\mathbf{h}. \tag{3.4}$$

3.1.1.1 Nulling

Following the conventional non-linear IN mitigation scheme based on nulling, after receiving the samples as in (3.3), the nulling of the IN contaminated samples is done based on the threshold value T_h consists in the following transformation:

$$[\hat{\mathbf{y}}]_{j,\text{null}} = \begin{cases} [\mathbf{y}]_j, & \text{if } |[\mathbf{y}]_j| \leqslant T_h, \\ 0, & \text{otherwise}, \end{cases} \tag{3.5}$$

where T_h is defined as

$$T_h = T\frac{1}{N}\sum_{j=0}^{N-1}|[\mathbf{y}]_j|. \tag{3.6}$$

T in the above equation is the threshold factor, which is designed based on the probability $\mathbb{P}$ that the amplitude of the non-corrupted received samples is higher than T [76]. The value of $\mathbb{P}$ is taken such that IN is reduced

from as many samples as possible. For a reasonably good performance, the value of $\mathbb{P} = 0.1$ is typically taken in the literature [75]. Now, based on the value of $\mathbb{P}$, T can be computed from $\mathbb{P} = \text{Prob}(|[\mathbf{y}]_j| > T) = e^{\frac{-T^2}{2\sigma^2}}$ where $\sigma^2 = \frac{(4-\pi)}{4}(\bar{\sigma}_s^2 + \bar{\sigma}_w^2 + \bar{\sigma}_i^2/P)$ is the variance of the received sample magnitudes which follow Rayleigh distribution assuming that the transmitted samples are complex Gaussian with zero mean and variance $\bar{\sigma}_s^2$ and P denotes the probability of IN sample occurrence. This threshold factor is now multiplied by the average magnitude of the received samples to find the value for nulling threshold T_h as in (3.6).

3.1.1.2 Clipping

Similar to nulling, clipping is also done based on a threshold value that is to be set by the system designer. The threshold value γ for clipping is typically fixed according to the background noise power, transmitted signal power and INR as [58]:

$$\gamma = \sqrt{\frac{2\sigma_1^2\sigma_2^2}{\sigma_2^2 - \sigma_1^2} \ln \frac{\sigma_2}{\sigma_1}}, \tag{3.7}$$

where $\sigma_1^2 = \bar{\sigma}_s^2 + \bar{\sigma}_w^2$ and $\sigma_2^2 = \bar{\sigma}_s^2 + \bar{\sigma}_w^2 + \bar{\sigma}_i^2$. Now, having fixed the threshold, clipping of the received sample is performed based on γ as

$$[\hat{\mathbf{y}}]_{j,\text{clip}} = \begin{cases} [\mathbf{y}]_j, & |[\mathbf{y}]_j| \leqslant \gamma, \\ \gamma \text{csgn}([\mathbf{y}]_j), & \text{otherwise}, \end{cases} \tag{3.8}$$

where $[\hat{y}]_{j,\text{clip}}$ is the received sample after clipping and csgn is the complex signum function.

3.1.1.3 Time Domain Periodogram (TDP) Approach

Unlike non-linear schemes based on nulling and clipping IN contaminated samples, the time domain periodogram (TDP) algorithm performs IN estimation and cancels it from the received signal. To estimate the IN, the TDP scheme follows a two step algorithm. As a first step, the support of IN is estimated using the periodogram of the time domain samples of the frequency domain symbols received in the unused subcarriers of the system. After estimating the support, the amplitude of IN samples is estimated, as a second step, exploiting the conventional minimum mean squared error (MMSE) estimator.

In order to estimate the support of IN, the TDP algorithm extracts the samples from $\breve{\mathbf{y}}$ corresponding to the content of the unused subcarriers after

receiving the frequency domain symbols as in (3.1). To do so, the vector $\breve{\mathbf{y}}$ is multiplied with the transpose of a matrix $\mathbf{S}_x$ that has dimensions $N \times (N - k_{\text{end}})$ as:

$$\bar{\mathbf{y}}' = \mathbf{S}_x^T \breve{\mathbf{y}}. \tag{3.9}$$

The matrix $\mathbf{S}_x$ in (3.9) is a selection matrix that has 1s as its entries at the indexes corresponding to the locations of end $N - k_{\text{end}}$ subcarriers and rest of the other entries are equal to zero. The resulting column vector $\bar{\mathbf{y}}'$ of length $(N - k_{\text{end}}) \times 1$ in (3.9) is now exploited by the TDP algorithm to estimate the IN support. To do so, the frequency domain vector $\bar{\mathbf{y}}'$ is transformed in to time domain using a lower dimensional FFT, producing a low resolution time-domain periodogram of length $N - k_{\text{end}}$ containing the samples only from the IN and background noise, as no data transmission is done in the unused part of the spectrum [141].

The location of samples in the periodogram with large magnitudes are considered as the initial guess for the IN sample support. Taking into account the scaling factor of $\delta l = N/(N - k_{\text{end}})$ the indices of such samples are scaled by δl to estimate the IN sample locations occurring in the OFDM symbol. After determining the support, the IN amplitudes are then estimated via a conventional MMSE algorithm. Finally, the estimated IN samples are subtracted from the time domain signal to mitigate IN, followed by demodulation, symbol de-mapping and decoding [141].

3.1.2 Proposed IN Estimation Scheme

Similar to TDP scheme, the proposed IN mitigation algorithm also follows estimation and cancellation approach to overcome the adversity caused by the occurrence of IN on PLC transmission and takes $\bar{\mathbf{y}}'$ in (3.9) as the observation vector for order and support estimation. The schematic of the proposed algorithm is shown in the block diagram in Fig. 3.2.

The problem of IN support estimation using the samples from the unused subcarriers spectrum can be closely related to the problem of estimating the frequencies of sinusoids that are immersed in the AWGN. The major difference between the two problems is that the support of IN is determined by its location in time domain whereas the frequency of sinusoid is evaluated by its location in frequency domain.

The problem of frequency estimation of sinusoid submerged in AWGN has been widely addressed using the MUSIC algorithm. To estimate the frequen-

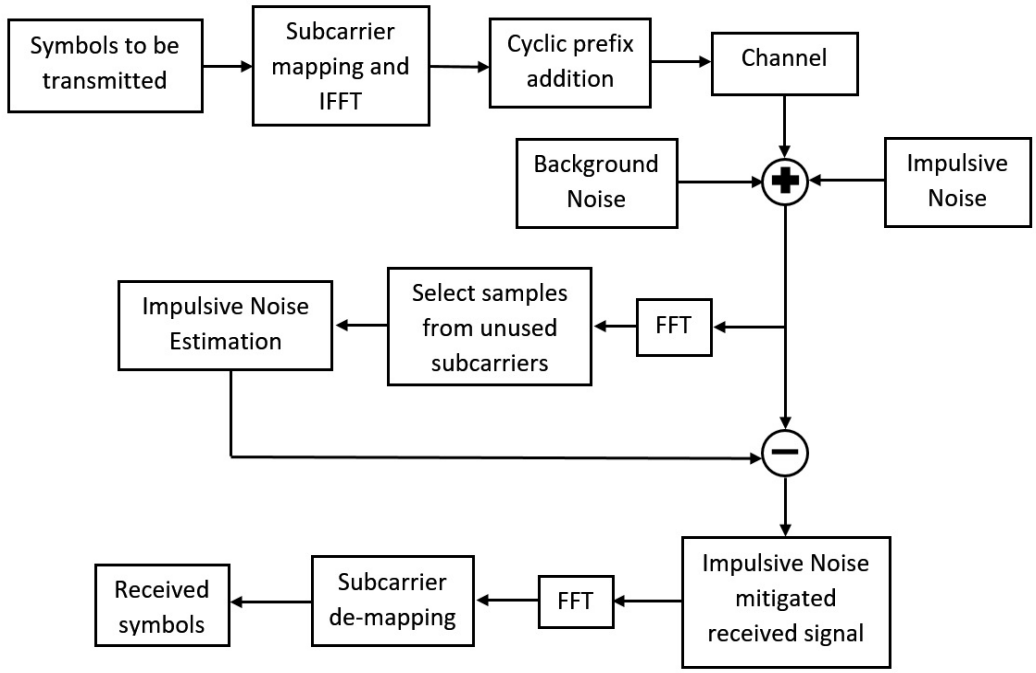


Figure 3.2: Block diagram of the proposed impulsive noise mitigation algorithm.

cies of sinusoids, MUSIC algorithm is typically supplied with the observation vector containing the time domain samples that are mixture of the sinusoid signals and the AWGN. To precisely estimate the sinusoid frequencies, the MUSIC estimator exploits the orthogonality between the signal and noise subspaces [142]. Moreover, being a super-resolution method, the MUSIC estimator resolves spectral lines that are separated in frequency $f = \omega/2\pi$ by less than $1/Z$ cycles per sampling interval, where Z is the total number of samples in an observation vector (in the time domain), and provides precise estimation of the frequencies of the sinusoid signals that are immersed in AWGN [142]. In order to exploit MUSIC estimator for IN support (locations of IN samples in time domain) estimation, the observation vector, however, should bear the frequency domain samples of the signal of interest superimposed by the unwanted signal. Hence, to estimate IN support using MUSIC estimator we resort to exploiting the observation vector in (3.9). Due to the high-resolution characteristic of the proposed scheme, supports of IN samples that are separated by at least $1/N$ sample in the pseudo-spectrum can be well resolved. The improved precision of the support estimation provides better performance than the TDP approach,

where the support is estimated by using a low resolution periodogram.

3.1.2.1 Order Estimation

As mentioned in the previous section, we take the vector $\bar{\mathbf{y}}'$ from (3.9) as the observation vector for IN order and support estimation. From $\bar{\mathbf{y}}'$, a total of $M = (N - k_{\text{end}}) - m + 1$ sample vectors denoted by $\tilde{\bar{\mathbf{y}}}_l$, where $l \in \{1, 2, 3, \ldots, M\}$, are generated, each having length m. Here, m denotes the sample window size used to generate the sample covariance matrix and is chosen such that the necessary condition $m - N_{imp} > N_{imp}$ is satisfied. The sample vector $\tilde{\bar{\mathbf{y}}}_l$ is constructed as

$$\tilde{\bar{\mathbf{y}}}_l = \begin{bmatrix} [\bar{\mathbf{y}}']_{(l+m-1)} \\ [\bar{\mathbf{y}}']_{(l+m-2)} \\ \vdots \\ [\bar{\mathbf{y}}']_l \end{bmatrix}, \tag{3.10}$$

where subindex notation denote the location of sample in vector $\bar{\mathbf{y}}'$. The sample covariance matrix $\hat{\mathbf{R}}$ of size $m \times m$ can now be formed using $\tilde{\bar{\mathbf{y}}}_l$ as

$$\hat{\mathbf{R}} = \frac{1}{M} \sum_{l=1}^{L} \tilde{\bar{\mathbf{y}}}_l \tilde{\bar{\mathbf{y}}}_l^H. \tag{3.11}$$

The eigenvalue decomposition (ED) of $\hat{\mathbf{R}}$ results into m eigenvectors and m eigenvalues. Out of m eigenvalues arranged in decreasing order $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots, \lambda_m$, the N_{imp} largest are associated with the IN samples (signal space in MUSIC terminology), whereas the $m - N_{imp}$ smallest correspond to the background noise samples [142]. By employing the eigenvalues of the sample covariance matrix, we estimate the number of IN impulses by the MDL criterion, according to [143] by

$$\text{MDL}(\hat{o}) = -\log \left(\frac{\prod_{r=\hat{o}+1}^{m} \lambda_r^{1/(m-\hat{o})}}{\frac{1}{m-\hat{o}} \sum_{r=\hat{o}+1}^{m} \lambda_r} \right)^{(m-\hat{o})(N-k_{\text{end}})} + \frac{3\hat{o}}{2} \log \frac{M}{m}. \tag{3.12}$$

The order N_{imp} of IN is estimated as the value of $\hat{o} \in \{0, 1, 2, \ldots, m-1\}$ for which (3.12) is minimum.

3.1.2.2 Support Estimation

After estimating the IN order, the eigenvector set of $\hat{\mathbf{R}}$ can now be split into two subsets. The first subset contains N_{imp} eigenvectors that belong to the

(MUSIC) signal subspace, corresponding to IN. The second subset, defined as the (MUSIC) noise subspace $-\{\hat{\mathbf{g}}_1, \hat{\mathbf{g}}_2, \ldots, \hat{\mathbf{g}}_{m-N_{imp}}\}$, contains $m - N_{imp}$ eigenvectors corresponding to the background noise. We now define the matrix $\hat{\mathbf{G}}$ as the column-wise juxtaposition of the eigenvectors in this second subset.

Having identified the (MUSIC) noise and signal subspaces, the support set of the IN, $\mathcal{I} \in \{n_1, n_2, \ldots, n_{N_{imp}}\}$, is estimated as the sample indexes, as the integers, in the set $q \in [0, N-1]$ for which the rational q/N is closest to the peaks of the function $f(t)$ defined as [142]

$$f(t) = \frac{1}{\mathbf{a}^H(t)\hat{\mathbf{G}}\hat{\mathbf{G}}^H\mathbf{a}(t)}, t = \frac{q}{N} \in [0, 1), \tag{3.13}$$

where $q \in \{0, 1, 2, \ldots, N-1\}$ and $\mathbf{a}(t)$ is the vector of length m defined as

$$\mathbf{a}(t) = [1 \quad e^{jt} \ldots e^{j(m-1)t}]^T. \tag{3.14}$$

Thus identified sample indexes are then arranged in increasing order to appear as entries of $\mathcal{I}$.

As it is well known, when $N \to \infty$ the variance of the MUSIC estimator is asymptotically Gaussian distributed with zero mean and variance V. Analogous to this key observation, the variance of the proposed IN support estimator can also be expressed as [144, 145]

$$V = \frac{6}{Mm^3}\frac{1}{\text{INR}}. \tag{3.15}$$

As seen from (3.15), the variance V is inversely proportional to the cube of the sample window size m and INR. Hence, the variance in support estimation can be reduced by taking wide enough sample window size m or the variance is significantly low when the IN has very high power in comparison to the background noise power. As a rule of thumb, in practice, the value of m should not be considered larger than half of the length of the observation vector. The resolution of support estimator tends to decrease, if the sample window size is taken wider than the half of the length of the observation vector [144].

Due to the variance of the estimator, the true support of IN samples might not be contained in the estimated support set $\mathcal{I}$. Therefore, to overcome this situation, we propose to extend each entry of $\mathcal{I}$ identified by the locations of peaks in (3.13), based on V to ensure that, with high probability, the true supports of IN samples are captured. The extended support set $\tilde{\mathcal{I}}$ can be expressed as $\tilde{\mathcal{I}} \in \{\mathcal{D}_{c1}, \mathcal{D}_{c2} \ldots, \mathcal{D}_{cN_{imp}}\}$, where we assume that each extended subset denoted by $\mathcal{D}_{ci} = \{n_i - r, n_i - r - 1, \ldots, n_i, n_i + 1 \ldots, n_i + r - 1, n_i + r\}$,

$i \in \{1, 2, \ldots, N_{imp}\}$ and $r = (V-1)/2$, is disjoint having no overlapping entry between the consecutive subsets.

3.1.2.3 Amplitude Estimation

After extending the estimated support set to $\tilde{\mathcal{I}}$, the amplitudes of the IN samples at those locations are estimated using the MMSE criterion. This estimation is done separately for each disjoint subset $\mathcal{D}_{ci}$. Then, for each subset, the support of the corresponding IN sample is re-estimated as the sample for which the amplitude is maximum. This procedure is carried out for all subsets and N_{imp} peaks are identified. The indices of the N_{imp} identified peaks, denoted by $\bar{\tilde{\mathcal{I}}}_p, p = 1, \ldots, N_{imp}$, are used to estimate the amplitude and phase of IN samples, denoted by $\hat{A}_p$, via the MMSE estimator as [141]

$$\hat{A}_p = \mathbb{E}([\hat{\mathbf{i}}]_p | \mathbf{y}, \bar{\tilde{\mathcal{I}}}_p) = \frac{\bar{\sigma_i}^2}{\bar{\sigma_i}^2 + \bar{\sigma}_s^2 + \bar{\sigma}_w^2} \mathbf{y}(\bar{\tilde{\mathcal{I}}}_p). \tag{3.16}$$

Finally, the estimate of IN $\hat{\mathbf{i}}$ of length $N \times 1$ having only N_{imp} non-zero entries at the supports of IN samples is generated. This estimate is subtracted from the time domain received signal vector $\mathbf{y}$ in (3.3) providing an IN-mitigated signal for demodulation as shown in Fig. 3.2.

3.1.3 Numerical Validation

In this section, we present the numerical evaluation of the performance of the proposed IN mitigation algorithm. To run the simulations, we consider an OFDM-based PLC system having a total of $N = 1024$ subcarriers. Among the available subcarriers, only $N_u = 768$ subcarriers are used for data transmission. Each used subcarrier conveys complex symbols that are randomly drawn from an uncoded QPSK modulation. The rest $N - N_u = 256$ subcarriers are nulled/unused. The spectral content of these nulled/unused subcarriers is used as the observation vector for IN order and support estimation. The sample window size, to generate sample covariance matrix, is considered as $m = 50$. Furthermore, a frequency flat channel is considered to avoid any contribution coming from the behavior of channel while assessing the performance of the IN mitigation algorithms.

The performance of two optimal informed receivers is also considered, mainly for comparison purposes. The first type of informed receiver is denoted by "Genie aided" in the performance graphs (black curve) and has all

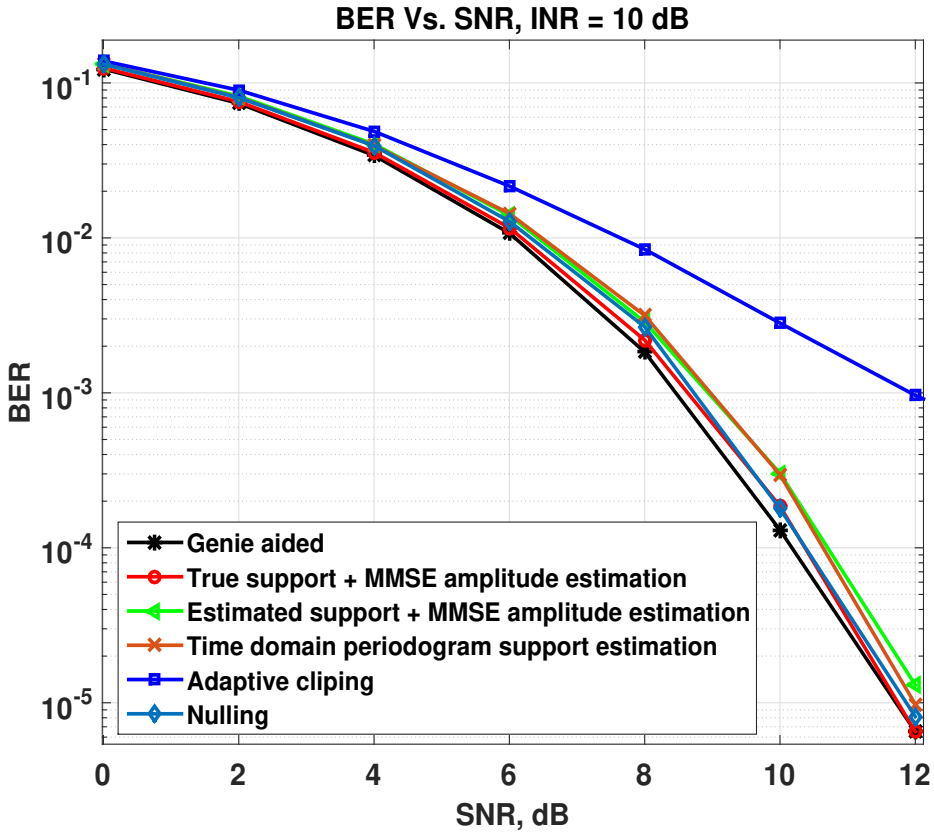


Figure 3.3: BER performance of the proposed scheme along with nulling, clipping and TDP schemes when INR is 10 dB and SNR is changed from 0 dB to 12 dB.

the information about IN, i.e., order, support and amplitude. The second type of informed receiver is denoted by "True support + MMSE amplitude estimation" in the performance graphs (red curve) and has the information about IN sample supports and estimates their amplitudes by the MMSE estimator as in (3.15). Furthermore, we also assess the performance of nulling, clipping and TDP scheme based algorithms introduced in sections 3.1.1.1, 3.1.1.2 and 3.1.1.3, respectively, to show comparison between the performances of the proposed algorithm and the conventional IN mitigation schemes. We consider PLC scenarios where the SNR varies from 0-12 dB and INR ranges from 10-40 dB. We specifically have considered the scenarios where INR $\geq$ 10 dB, as these scenarios define realistic IN environment in an active power line network [44]. Furthermore, the IN sample has probability of occurrence 10^{-2}, where approximately 10 samples of IN occur during transmission of an OFDM symbol.

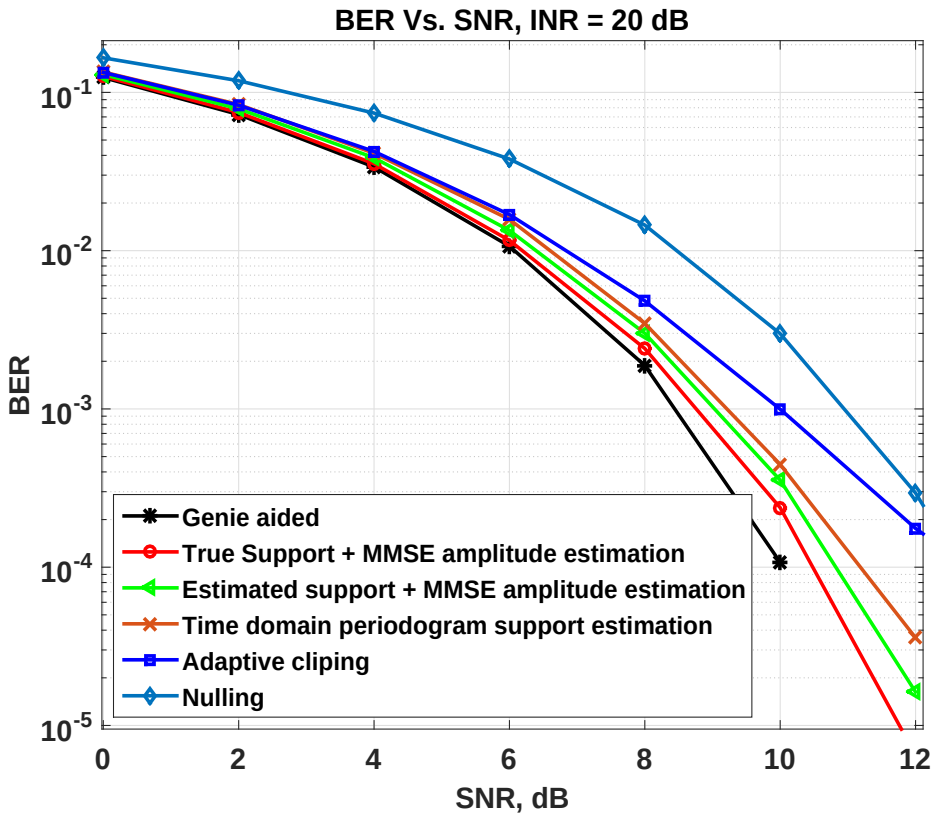


Figure 3.4: BER performance of the proposed scheme along with nulling, clipping and TDP schemes when INR is 20 dB and SNR is changed from 0 dB to 12 dB.

Fig. 3.3 compares the BER of nulling [76], clipping [58], TDP schemes [141] and the proposed algorithm in a typical situation where IN is 10 dB stronger than the background noise. It can be seen from the BER curves that the performance of nulling, TDP and the proposed scheme are very close to each other, and are comparable to the performance of the two informed receivers. The performance of clipping algorithm in this context is bad because some of the useful samples apart from the ones affected by IN are also clipped.

As the IN power increases, it is evident from the simulation result graphs that the performance of the nulling algorithm gradually decreases. Since the threshold value for nulling crucially depends on the mean magnitude of received samples, the thresholding value in this context bear large value due to high mean magnitude, as a result of IN with high INR occurring during transmission and being added to the received samples. Due to the high threshold value, all

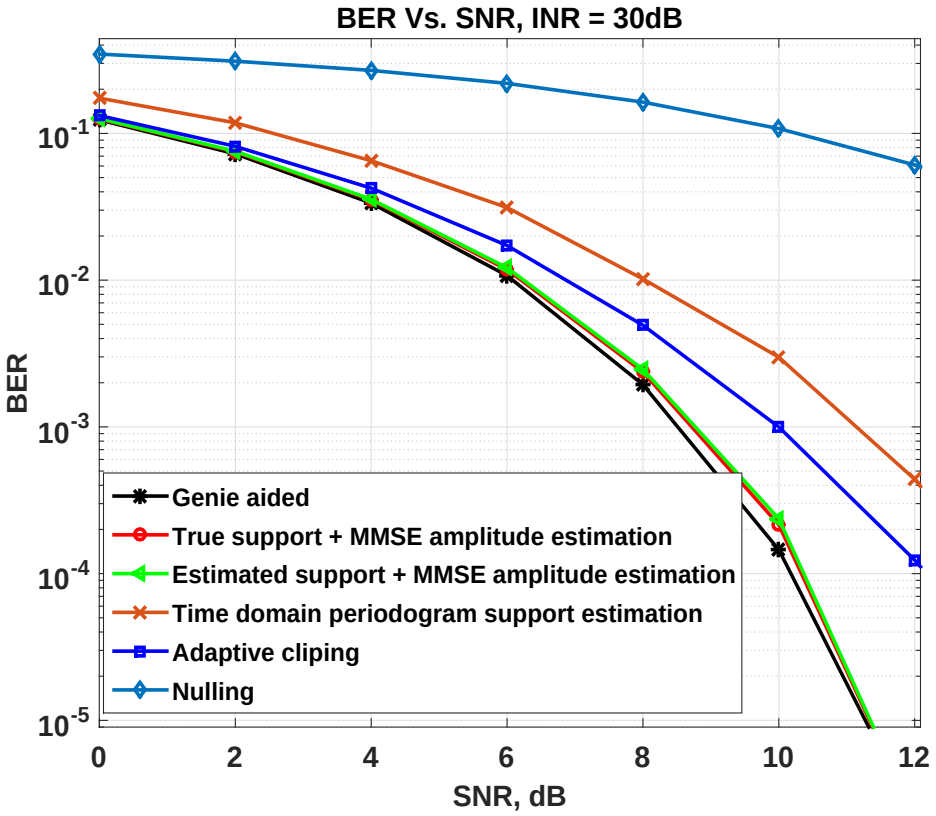


Figure 3.5: BER performance of the proposed scheme along with nulling, clipping and TDP schemes when INR is 30 dB and SNR is changed from 0 dB to 12 dB.

IN samples are not estimated and hence the BER performance of the receiver performing nulling to mitigated IN corrupted samples is degraded.

Similar to the nulling algorithm, the performance of the TDP scheme also tends to decrease as the IN power increases. The major contribution to the performance degradation comes from the fact that low resolution periodogram is typically used for IN support estimation. Due to this, closely located IN samples are not estimated. The requirement of precise support estimation becomes more prominent when the channel is corrupted by even stronger IN. In such a scenario, the effect of any missed IN sample spreads over the whole OFDM symbol while demodulating and the system performance is degraded considerably.

In contrast to the performance of nulling and TDP schemes, the performance of clipping algorithm is more or less consistent in the higher INR re-

gions. Though the clipping algorithm performs better segregation of the IN samples when the power of IN is very high in comparison to the power of the background noise, the performance of clipping algorithm is limited as the clipping magnitude is not enough to truncate the amplitudes of the IN-affected received samples to a safe limit (i.e., even after clipping, the clipped samples still contain a significant amount of IN). As, a result there still remains some room which needs to be covered to improve the BER performance of the PLC system. Moreover, the proposed algorithm outperforms all the algorithms in scenarios with distinct level of INR and provides significant gain in term of BER performance of the system. These observations can readily be made in the simulation results shown in Fig. 3.4, Fig. 3.5 and Fig. 3.6, when INR is 20 dB, 30 dB and 40 dB, respectively.

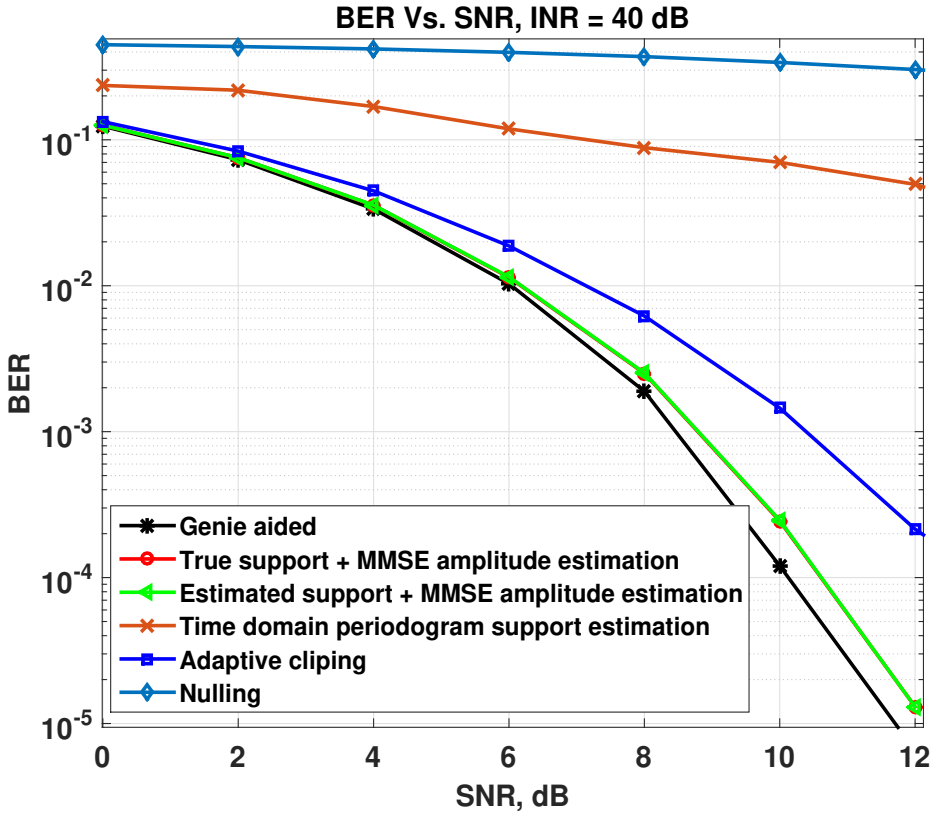


Figure 3.6: BER performance of the proposed scheme along with nulling, clipping and TDP schemes when INR is 40 dB and SNR is changed from 0 dB to 12 dB.

3.2 Successive Impulsive Noise and Narrowband Interference Cancellation

As outlined in the section 2.3.3 of chapter 2, apart from the IN, the interference originating from the wireless communication systems also impair PLC transmission. The NBI mitigation schemes described in section 2.5.2 of this dissertation do not consider the joint effect of IN and NBI on the communication signal. Only few research articles try to address this peculiar noise scenario by resorting to exploitation of hybrid communication (power line + wireless communication) to overcome them [112, 113]. However, in a typical power line environment, the PLC transmission might be affected by the simultaneous occurrence of both impairments. Hence, to achieve reliable data transmission through power lines, it is of utmost importance to design a signal processing algorithm that can effectively mitigate the effect of IN and provide robustness against NBI.

Typical interference from the wireless communication systems to PLC resemble the IN in the sense that, the IN is impulsive in the time domain whereas the NBI is impulsive in the frequency domain. In this section, we propose an algorithm that successively mitigates the IN and cancels NBI from the received signal exploiting the duality of the impairments. To estimate the support of IN and the frequency of NBI, we exploit the subspace-based estimator discussed in section 3.1 of this chapter. Furthermore, the amplitude and phase of the IN samples are estimated using MMSE estimator, and the amplitude and phase of the NBI are estimated using least squares (LS) estimator.

To concisely elaborate on the proposed algorithm, we divide this section into three subsections. In subsection 3.2.1 we revoke the system model presented in section 3.1.1 and introduce the signal model including the effect of NBI occurring during the OFDM transmission. In subsection 3.2.2 we elaborate the proposed algorithm to successively mitigate IN and NBI. In subsection 3.2.3 the numerical validation of the proposed algorithm is done and a brief conclusion is drawn at the end of this section.

3.2.1 System Model

In this section we define the signal model of the OFDM-based PLC system used to elaborate on the proposed successive NBI and IN cancellation algorithm. The subcarrier alignment within an OFDM symbol is considered similar to

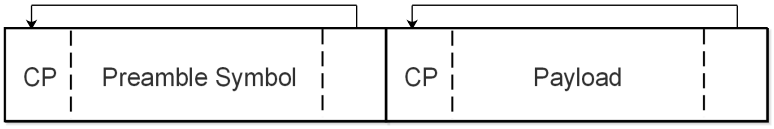


Figure 3.7: Transmission frame structure.

the one shown in Fig. 3.1. In addition, we consider a transmission frame consisting a preamble symbol and a payload symbol as shown in Fig. 3.7. The pilot symbols transmitted in preamble is assumed known to the receiver and are used for transmitter-receiver synchronization and channel estimation.

Recalling the discrete multi tone PLC system model defined in (3.1) and incorporating the occurrence of NBI and IN during PLC reception, the time domain received signal corresponding to the preamble transmission time is expressed as:

$$\mathbf{y}_p = \sqrt{N}\mathbf{F}^H\mathbf{E}\mathbf{\Lambda}_p\mathbf{G}^H\mathbf{h} + \mathbf{n}_p, \tag{3.17}$$

where $\mathbf{F}, \mathbf{E}, \mathbf{G}$ and $\mathbf{h}$ have similar definitions as in (3.3), $\mathbf{\Lambda}_p$ similarly to $\mathbf{\Lambda}$ in (3.3) is a diagonal matrix but of dimensions $N_u \times N_u$ and contains only the pilot symbols that are transmitted in the used subcarriers as its entries, $\mathbf{n}_p = \mathbf{e}_p + \mathbf{i}_p + \mathbf{w}_p$ denotes a column vector of length $N \times 1$ containing the time domain samples of NBI ($\mathbf{e}_p$), IN ($\mathbf{i}_p$) and background noise ($\mathbf{w}_p$). The vectors $\mathbf{i}_p$ and $\mathbf{w}_p$ have similar definitions as $\mathbf{i}$ and $\mathbf{w}$ in (3.3). The NBI is modeled as in section 2.3.3 of chapter 2 and is defined as [55, 108, 146]:

$$[\mathbf{e}_p]_n = \sum_{k=1}^{c} A_k e^{j(\omega_k n + \phi_k)}, \tag{3.18}$$

where $[\mathbf{e}_p]_n$ is the n^{th} sample of NBI, ω_k denotes the frequency of k^{th} NBI, c denotes the total number of NBIs and ϕ_k denotes the phase of k^{th} NBI. Since the adopted model treats each NBI as a tone in frequency domain and characterizes the net effect of multiple NBIs occurring during PLC transmission as a collection of superimposed tones, hence, we rename the vector $\mathbf{e}_p$ as frequency interferer (FI) and will be used to denote the occurrence of more than one NBI in the text to follow.

3.2.2 Proposed NBI and IN Cancellation Scheme

In order to successively cancel the NBI and IN from the received signal, an iterative exploitation of the algorithm in section 3.1 is proposed. The FI esti-

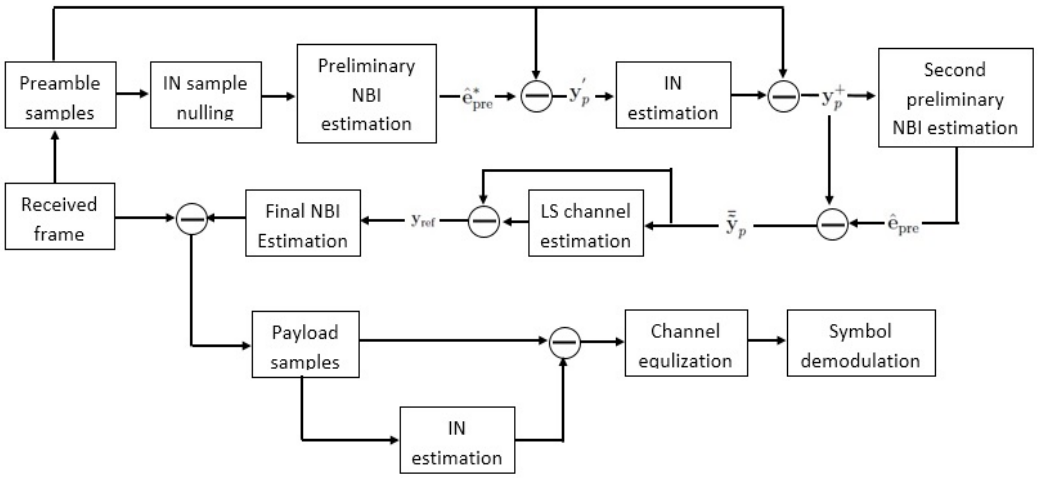


Figure 3.8: Schematic block diagram of the proposed algorithm.

mation is done using the time domain received samples corresponding to the transmitted preamble symbol as an observation vector and the spectrum of the unused subcarriers in the system is exploited to estimate the IN. However, the presence of IN and the channel along with the presence of pilot symbols are undesired signals when it comes to the precise estimation of FI using (3.17) as an observation vector. Therefore, an iterative algorithm is proposed, where we get rid of the IN component followed by the subtraction of the contribution coming from the channel and preamble symbol and obtain a refined observation vector bearing samples corresponding to the FI and background noise only. The schematic diagram of the proposed scheme is shown in Fig 3.8.

To get rid of the IN component, an iterative implementation of the true support estimation (TSE) algorithm proposed in section 3.1 of this chapter is adopted. As a first iteration, the IN contaminated samples in $\mathbf{y}_p$ are identified using the TSE scheme. The IN corrupted samples in the received signal are then nulled. The resulting observation vector, denoted by $\mathbf{y}_p^*$, is now used to perform a preliminary estimation of the FI, denoted by $\hat{\mathbf{e}}_{\text{pre}}^*$. This preliminary estimate of the FI is subtracted from $\mathbf{y}_p$ resulting in

$$\mathbf{y}_p^{'} = \sqrt{N}\mathbf{F}^H\mathbf{E}\mathbf{\Lambda}_p\mathbf{G}^H\mathbf{h} + (\mathbf{e}_p - \hat{\mathbf{e}}_{\text{pre}}^*) + \mathbf{i}_p + \mathbf{w}_p. \tag{3.19}$$

The resulting observation vector $\mathbf{y}_p^{'}$ in (3.19) is now used to estimate the IN occurring during the preamble symbol transmission. We again take advantage of the TSE algorithm to perform the IN estimation at this stage [140]. The

estimated IN noise, $\hat{\mathbf{i}}_p$, is then subtracted from the observation vector $\mathbf{y}_p$, resulting in

$$\mathbf{y}_p^+ = \sqrt{N}\mathbf{F}^H\mathbf{E}\mathbf{\Lambda}_p\mathbf{G}^H\mathbf{h} + \mathbf{e}_p + (\mathbf{i}_p - \hat{\mathbf{i}}_p) + \mathbf{w}_p. \tag{3.20}$$

The resulting vector $\mathbf{y}_p^+$ in (3.20) is now used for the estimation of the FI. Since the components $\mathbf{h}$ and $\mathbf{\Lambda_p}$ in $\mathbf{y}_p^+$ are undesired signals in the observation vector, we now proceed with the iterative estimation of the FI. In order to so, next, we perform a second preliminary estimation of FI and then subtract it from $\mathbf{y}_p^+$. After the second preliminary estimated FI is subtracted from the observation vector, we perform least squares (LS) channel estimation and get rid of $\mathbf{h}$ and Λ_p from $\mathbf{y}_p^+$. This minimization of the undesired signals provides a refined observation vector that contains samples only from the FI and the background noise. The final FI estimation is done using the refined observation vector.

Note: The estimation of $\hat{\mathbf{e}}_{\text{pre}}^*$ is also done following the procedure as mentioned in section 3.2.2.1 using $\mathbf{y}_p^*$ as an observation vector.

3.2.2.1 Frequency Interferer Parameter Estimation

Taking $\mathbf{y}_p^+$ as an observation vector we proceed towards estimating the second preliminary estimate of the FI after the removal of the IN. The estimation of the parameters A_k, ω_k and ϕ_k of the interferer from the equation (3.20) will be carried out in three steps.

- Step 1: Estimation of the number of NBIs (order estimation).
- Step 2: Frequency estimation of each NBI.
- Step 3: Amplitude and phase estimation of each NBI.

Order Estimation of Frequency Interferer

The order of the FI determines the number of the NBIs occurring during a frame transmission. To estimate the number of NBIs, we start by choosing an arbitrary number m, as the sample window size, to generate $M = N - m + 1$ sample vectors from the observation vector in (3.20). Each sample vector of length m denoted by $\tilde{\mathbf{y}}_l$, $l \in \{1, 2, 3, \ldots, M\}$ is constructed as

$$\tilde{\mathbf{y}}_l = \begin{bmatrix} [\mathbf{y}_p^+]_{(l+m-1)} \\ [\mathbf{y}_p^+]_{(l+m-2)} \\ \vdots \\ [\mathbf{y}_p^+]_l \end{bmatrix}, \tag{3.21}$$

where subscripted notation denote the location of sample in vector $\mathbf{y}_p^+$. The window size m to form the sample vector is chosen such that $m - c > c$, as mentioned in section 3.1.2.1 of this chapter. After generating the M sample vectors, a sample covariance matrix $\mathbf{C}$ of size $m \times m$ is formed as:

$$\mathbf{C} = \frac{1}{M} \sum_{l=1}^{L} \tilde{\mathbf{y}}_l \tilde{\mathbf{y}}_l^H. \tag{3.22}$$

The number of NBIs is estimated by evaluating the eigenvalues of the sample covariance matrix, given by the ED of the matrix $\mathbf{C}$. The ED of $\mathbf{C}$ results into m eigenvectors and m eigenvalues. The eigenvalues arranged in decreasing order $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots, \lambda_m$, are then evaluated according to the MDL criterion as in (3.12) to estimate the number of NBIs c [143].

Frequency Estimation of Each NBI

After estimating the number of NBIs, we estimate location of each NBI in the spectrum by using the high resolution frequency estimator MUSIC.

Based on the estimated number of NBIs, the eigenvectors of $\mathbf{C}$ are classified into two subsets. The first subset, denoted by $\hat{\mathbf{S}} = \{\hat{\mathbf{a}}_1, \ldots, \hat{\mathbf{a}}_c\}$, contains the c eigenvectors associated to the c largest eigenvalues of $\mathbf{C}$, also referred to as signal-subspace in MUSIC terminology, and the second subset $\hat{\mathbf{G}} = \{\hat{\mathbf{b}}_1, \ldots, \hat{\mathbf{b}}_{m-c}\}$ contains the remaining $m - c$ eigenvectors of $\mathbf{C}$ namely the noise-subspace in MUSIC terminology. Furthermore, we define the vector $\boldsymbol{\alpha}$ as a function of $\hat{\omega}$ as:

$$\boldsymbol{\alpha}(\hat{\omega}) = [1, e^{-j\hat{\omega}}, \ldots e^{-j(l-1)\hat{\omega}}]^T, \hat{\omega} \in [0, 2\pi]. \tag{3.23}$$

Based on the estimated value of c, the values of $\hat{\omega}$ corresponding to the c largest peaks of the pseudo-periodogram function $\hat{f}(\hat{\omega})$, as defined in (3.24), are the estimated locations of the each NBI in the spectrum [142],

$$\hat{f}(\hat{\omega}) = \frac{1}{\boldsymbol{\alpha}^H(\hat{\omega}) \hat{\mathbf{G}} \hat{\mathbf{G}}^H \boldsymbol{\alpha}(\hat{\omega})}. \tag{3.24}$$

Amplitude and Phase Estimation of Each NBI

After estimating the number and the corresponding frequencies of the NBIs, in this step we perform amplitude and phase estimation of each NBI using the LS estimator.

Without loss of generality, let

$$z_k = A_k e^{j\phi_k}, k \in \{1, 2, \ldots, c\}, \tag{3.25}$$

be the variable containing the values of amplitude and phase of the k^{th} NBI. Consider the column vector

$$\mathbf{z} = [z_1, \ldots, z_c]^T, \tag{3.26}$$

of dimension $c \times 1$, where the variables z_k are defined in (3.25).

Let,

$$\mathbf{Q} = \begin{bmatrix} 1 & \ldots & 1 \\ e^{j\hat{\omega}_1} & \ldots & e^{j\hat{\omega}_c} \\ \vdots & & \vdots \\ e^{j(N-1)\hat{\omega}_1} & \ldots & e^{j(N-1)\hat{\omega}_c} \end{bmatrix} \tag{3.27}$$

be the matrix of size $N \times c$ constructed using the estimated frequencies of NBIs, where $\{\hat{\omega}_1, \ldots, \hat{\omega}_c\}$ are the estimated frequencies.

Using (3.26) and (3.27), (3.20) can now be expressed as:

$$\mathbf{y}_p^+ = \mathbf{Q}\mathbf{z} + \text{noise}. \tag{3.28}$$

where the term "noise" corresponds to the contributions coming from channel, received preamble symbols, IN mitigation error and the background noise in (3.20). From (3.28), the vector $\mathbf{z}$ is estimated by using the LS estimator. The LS estimated vector $\hat{\mathbf{z}}$ of $\mathbf{z}$, is thus given by

$$\hat{\mathbf{z}} = (\mathbf{Q}^H\mathbf{Q})^{-1}\mathbf{Q}^H\mathbf{y}_p^+, \tag{3.29}$$

where $[.]^H$ defines the Hermitian operation.

Exploiting the estimated order, amplitude, frequency and phase, a preliminary estimate of the FI, $\hat{\mathbf{e}}_{\text{pre}}$ is achieved as:

$$[\hat{\mathbf{e}}_{\text{pre}}]_n = \sum_{k=1}^{c} \hat{z}_k e^{j(\hat{\omega}_k n)}, \tag{3.30}$$

where $[\hat{\mathbf{e}}_{\text{pre}}]_n$ denotes the n^{th} sample of $\hat{\mathbf{e}}_{\text{pre}}$.

3.2.2.2 Observation Vector Refinement

The preliminary estimate of the FI, $\hat{\mathbf{e}}_{\text{pre}}$, is subtracted from the observation vector in (3.20),

$$\bar{\bar{\mathbf{y}}}_p = \mathbf{y}_p^+ - \hat{\mathbf{e}}_{\text{pre}} = \sqrt{N}\mathbf{F}^H\mathbf{E}\mathbf{\Lambda}_p\mathbf{G}^H\mathbf{h} + (\mathbf{e}_p - \hat{\mathbf{e}}_{\text{pre}}) + \text{IN}_{\text{residue}} + \mathbf{w}_p, \tag{3.31}$$

where $\text{IN}_{\text{residue}}$ is the error coming from the IN cancellation. To further refine the observation vector for FI estimation, the resulting samples in (3.31) are now used for the channel estimation. The estimated channel along with the known preamble symbols are then used to minimize the contributions coming from the channel and received pilot symbols on the observation vector in (3.20). In order to estimate the channel, the least squares (LS) channel estimator is employed to approximate the coefficients of CFR corresponding to the used subcarriers. To do so, we first transform the samples in (3.31) to frequency domain. The frequency domain transformation is carried out by multiplying (3.31) with the DFT matrix $\mathbf{F}$ as:

$$\breve{\mathbf{y}}_p = \mathbf{F}\bar{\bar{\mathbf{y}}}_p = \begin{bmatrix} \mathbf{0} \\ \mathbf{\Lambda}_p \breve{\mathbf{h}} \\ \mathbf{0} \end{bmatrix} + \text{NBI}_{\text{residue}} + \text{IN}_{\text{residue}} + \mathbf{F}\mathbf{w}_p, \tag{3.32}$$

where $\breve{\mathbf{y}}_p$ is the vector containing the received symbols corresponding to the transmitted preamble symbol and recall that $\breve{\mathbf{h}} = \sqrt{N}\mathbf{G}^H\mathbf{h}$ from (3.4). Exploiting the known pilot symbols the estimated CFR is achieved as:

$$\hat{\breve{\mathbf{h}}} = \mathbf{E}^T \breve{\mathbf{y}}_p ./ \text{diag}(\mathbf{\Lambda}_p), \tag{3.33}$$

where $\text{diag}(\mathbf{\Lambda}_p)$ is a vector of dimension $N_u \times 1$ containing only the diagonal entries of the matrix $\mathbf{\Lambda}_p$ and ./ defines the element wise division operation.

With the estimated channel coefficients and known preamble symbols, the effect of the corresponding quantities from (3.20) is eliminated, providing a refined observation vector for the final FI estimation. To achieve the refined observation vector, the samples in (3.20) are transformed to the frequency domain first and then the subsequent quantities are subtracted as:

$$\bar{\bar{\mathbf{y}}}_{\text{ref}} = \mathbf{F}\mathbf{y}_p^+ - \begin{bmatrix} \mathbf{0} \\ \mathbf{\Lambda}_p \hat{\breve{\mathbf{h}}} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{\Lambda}_p(\breve{\mathbf{h}} - \hat{\breve{\mathbf{h}}}) \\ \mathbf{0} \end{bmatrix} + \mathbf{F}\mathbf{e}_p + \text{IN}_{\text{residue}} + \mathbf{F}\mathbf{w}_p. \tag{3.34}$$

After minimizing the effect of channel, preamble symbols and IN from the received signal corresponding to the preamble transmission time, a refined observation vector in frequency domain is obtained as in (3.34). The time domain samples of vector $\bar{\bar{\mathbf{y}}}_{\text{ref}}$, which is a majority mixture of the FI and background noise, is now used for final FI estimation. The refined observation vector in time domain is hence expressed as:

$$\mathbf{y}_{\text{ref}} = \mathbf{F}^H \bar{\bar{\mathbf{y}}}_{\text{ref}} = \mathbf{e}_p + \mathbf{w}_p + \text{residue}, \tag{3.35}$$

where "residue" defines the net error coming from the IN cancellation and the minimization of channel contribution while obtaining the refined observation vector for FI estimation.

3.2.2.3 Final Estimation of NBI and Cancellation

The final estimation of the FI is carried out by following the steps exactly as mentioned in the subsection 3.2.2.1, using the new observation vector $\mathbf{y}_{\text{ref}}$ from (3.35). The n^{th} sample of FI is hence achieved as:

$$[\hat{\mathbf{e}}_p]_n = \sum_{k=1}^{c} \hat{A}_k e^{j(\hat{\omega}_k n + \hat{\phi}_k)}, \tag{3.36}$$

where $\hat{A}, \hat{\omega}$, and $\hat{\phi}$ are the estimated values of A, ω, and ϕ.

3.2.2.4 Impulsive Noise Estimation and Cancellation

In order to cancel the IN from the payload received signal, the IN estimation and cancellation algorithm proposed in section 3.1.2 is exploited. The frequency domain symbols received corresponding to the payload transmission time is expressed as:

$$\breve{\mathbf{y}} = \begin{bmatrix} \mathbf{0} \\ \mathbf{\Lambda}\breve{\mathbf{h}} \\ \mathbf{0} \end{bmatrix} + \mathbf{F}(\mathbf{e} + \mathbf{i} + \mathbf{w}), \tag{3.37}$$

where $\mathbf{\Lambda}$ denotes a diagonal matrix of dimension $N_u \times N_u$ whose diagonal entries contain the frequency domain transmitted payload symbols, and the vectors $\mathbf{e}, \mathbf{i}$ and $\mathbf{w}$ denote the FI, IN and background noise. Since FI is assumed deterministic for the duration of a frame transmission, the estimated FI, $\hat{\mathbf{e}}_p$, is also canceled from the received signal corresponding to the payload transmission time as:

$$\bar{\breve{\mathbf{y}}} = \begin{bmatrix} \mathbf{0} \\ \mathbf{\Lambda}\breve{\mathbf{h}} \\ \mathbf{0} \end{bmatrix} + \mathbf{F}(\mathbf{e} - \hat{\mathbf{e}}_{\text{pre}} + \mathbf{i} + \mathbf{w}), \tag{3.38}$$

After the cancellation of the FI from (3.37), samples from the unused part of the spectrum is segregated to generate the observation vector for IN estimation as:

$$\bar{\mathbf{y}}' = \mathbf{S}_x^T \bar{\breve{\mathbf{y}}}. \tag{3.39}$$

Here we recall that the matrix $\mathbf{S}_x$ in (3.39) is a selection matrix that has 1s as its entries at the indexes corresponding to the locations of end $N - k_{\text{end}}$ subcarriers and rest of the other entries being zero as defined in (3.9). After the IN is estimated, it is canceled from the received signal in (3.38) to obtain IN mitigated samples for demodulation. Furthermore, the received symbols are equalized against the channel selectivity by exploiting the channel estimated in (3.33).

3.2.3 Numerical Validation

In this section, we evaluate the BER performance of the proposed algorithm. The system parameters, preamble symbol, and channel model are derived from the NB-PLC standard, IEEE 1901.2. The system parameters for the CENELEC-A band OFDM system in consideration are shown in the table below.

Parameter	**Value**
Sampling frequency	400 kHz
Total number of subcarriers	256
Length of preamble symbol	256
Length of cyclic prefix	30
Inter carrier spacing	1.5625 kHz
First used subcarrier location	35.9375 kHz
Last used subcarrier location	90.625 kHz
Modulation	QPSK (uncoded)

Table 3.1: Parameters of OFDM based PLC system derived from IEEE 1901.2 (NB-PLC) for simulation.

The channel considered for the simulations follows a statistical multi-path fading model introduced in (2.2). The channel response at a given frequency (f) is expressed as:

$$H(f) = \sum_{t=1}^{N_{\text{path}}} g_t e^{-(a_0+a_1 f)d_t} e^{-j2\pi f \frac{d_t}{v_o}} \tag{3.40}$$

where N_{path} is the total number of propagation paths between the transmitter and the receiver, g_t is the path gain summarizing the reflection and the transmission along t^{th} propagation path, a_o and a_1 are the attenuation parameters

that depend on the transmission line impedance characteristics, f is the frequency in Hertz, d_t is the length of t^{th} propagation path and v_o is the wave propagation speed. The above channel model is implemented using the realistic parameter values, as given in the standard IEEE 1901.2, where $a_0 = 1 \times 10^{-3}$, $a_1 = 2.5 \times 10^{-3}$, d is the Gaussian random variable having mean 1000 and standard deviation 400, g is also a Gaussian random variable with zero mean and variance 1 that is scaled by 1000, $N_{\text{path}} = 5$ and $v_o = 3/4 \times 10^8$ [39].

To demonstrate the performance of the proposed algorithm, we define two scenarios characterized as:

- **scenario I:** One NBI, whose frequency is uniformly distributed between $[0, F_s]$, occurs during a frame transmission and IN samples occur with the probability of corrupting 12 samples out of 1024 samples in a frame.
- **scenario II:** Two NBIs, whose frequencies are uniformly distributed between $[0, F_s]$, occur during a frame transmission and IN samples occur with the probability of corrupting 20 samples out of 1024 samples in a frame.

The simulation results showing the BER performance of the algorithm in scenario I and scenario II are shown in Fig. 3.9 and Fig. 3.10 respectively. As shown in the simulation results, the proposed scheme has performance close to the AWGN bound and is superior then the nulling and clipping with frequency excision schemes. The AWGN bound in the simulation results defines the scenario when there is no occurrence of NBI and IN and the transmitted signal is only impaired by the background noise. Apart from this, the performance of the non-iterative implementation of the algorithm is also shown. As anticipated, the non-iterative approach reaches the saturation after SNR of 10 dB in both cases. Conversely, the iterative approach converges towards a reasonable SNR value in both scenarios. The significance of iterative estimation becomes more evident at high SNR regions where the preamble symbol with high power acts as strong noise for the FI estimation.

3.3 Conclusion

In the first section of this chapter, we proposed an IN mitigation technique for PLC systems. The proposed algorithm effectively exploits the samples received in the unused part of the spectrum to estimate the IN samples in the received

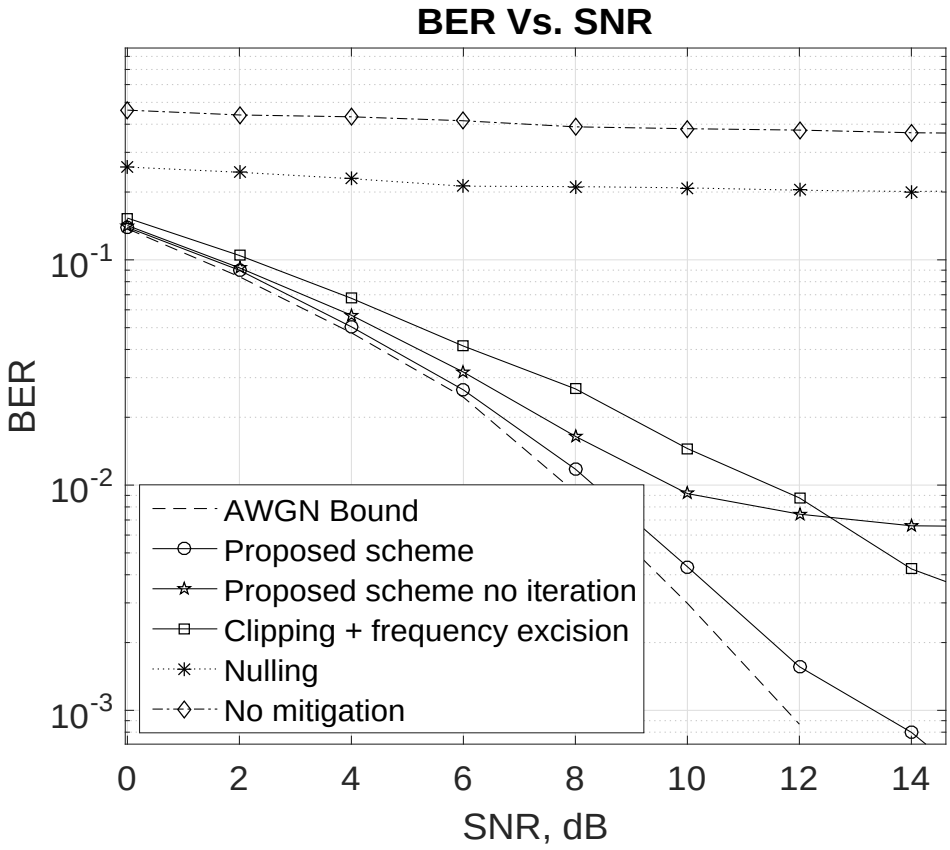


Figure 3.9: BER performance in scenario I, where the impulsive noise power to background noise power ratio is 40 dB and the signal power to interference power is -40 dB.

signal. As shown via numerical validation, the proposed algorithm works accurately in scenarios having IN for a remarkable range of different values of INR. The proposed scheme outperforms the existing non-linear IN mitigation algorithms like nulling and clipping, and also provides better IN mitigation capability than the sophisticated TDP-based IN suppression algorithm. The superiority of the proposed scheme over such techniques comes from the fact that the effect of IN is removed only from the corrupted samples in the received signal, thanks to the higher resolution in the identification of the IN support using the MUSIC algorithm. Furthermore, as shown by the simulation results, the precision in support estimation of the proposed algorithm becomes more prominent when the channel is corrupted by even stronger IN making the proposed algorithm better suited for PLC to support smart grid applications.

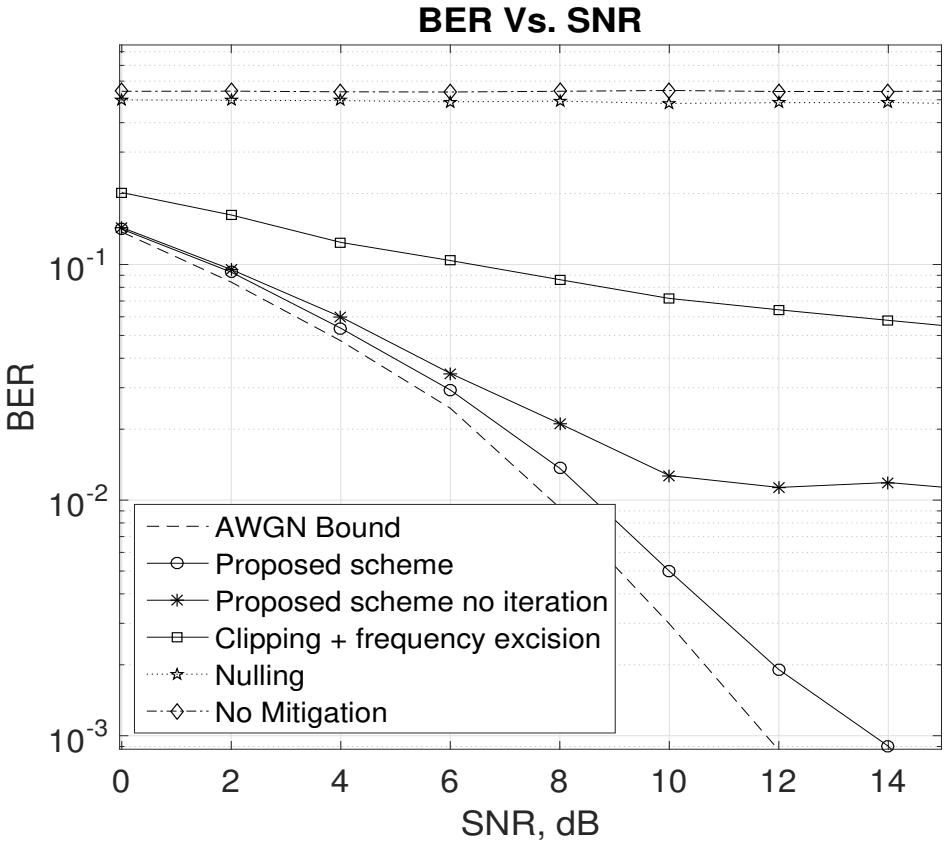


Figure 3.10: BER performance in scenario II, where the impulsive noise power to background noise power ratio is 40 dB and the signal power to interference power is -40 dB.

Addressing to the challenge of performing reliable data transmission in hostile power line environment, in the second section of this chapter we proposed a noise mitigation algorithm that can effectively cancel the effects of IN and concurrently occurring NBI for PLC systems. The proposed scheme bears high precision in parameter estimation and hence the effective cancellation of both impairments from the received signal can be achieved. The performance of the presented scheme is close to the AWGN bound and is consistent over different scenarios having distinct level of NBI and IN power. Moreover, the use of the same algorithm for estimating the IN samples support in the time domain and the localization of NBI in the frequency domain facilitates an efficient PLC receiver architecture for smart grid applications.

As a general conclusion, we can say that the unused part of the OFDM transmission spectrum provides an opportunity to perform effective estimation

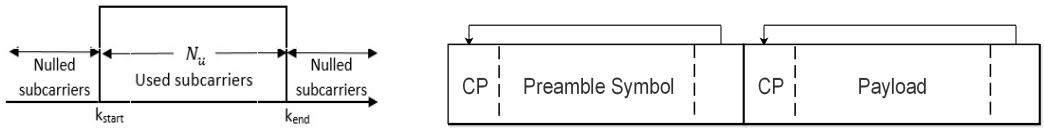
of the inevitable impairments in power line environment. Upon canceling the effect of unavoidable noises like IN and NBI, samples that are fairly free from the error can be obtained for demodulation in PLC systems.

Chapter 4

Channel Estimation for PLC Systems

In section 2.6 of chapter 2 we presented an overview of the channel estimation strategies that have been proposed for PLC systems. Most of the proposed algorithms perform least squares (LS) channel estimation in conjunction with IN power suppression schemes to estimate the coefficients of the channel frequency response (CFR). The LS channel estimator, which typically assumes the noise to be white, looses its optimality in the presence of IN. In a situation when the effect of IN is not eliminated completely from the received signal, the error in channel estimation still persists [132]. In order to enhance the performance of LS estimators in the presence of IN, domain transformation techniques are proposed [124–131]. By exploiting domain transformation techniques, the CFR estimated by the LS estimator can be refined to have a more accurate estimation of the power line channel. However, since the domain transformation techniques exploit threshold based operations to identify the channel taps with significant energy, these schemes show inconsistent behavior in PLC environments. Hence, channel estimation strategies that are not only robust, but also consistent, are required to exploit PLC based technologies to support smart grid applications.

In this chapter, we propose novel channel estimation strategies that exploit estimated parameters of IN to effectively estimate the channel for PLC systems [147]. To concisely elaborate on the proposed channel estimation schemes this chapter is divided into three sections. In the first section, the system model of a typical OFDM-based PLC system defined in (3.1) is recalled and adapted according to the context of the current chapter. In the second section, we



(a) Subcarrier structure (b) Frame structure

Figure 4.1: Illustration of subcarrier and frame structure.

derive two maximum-likelihood (ML) estimators for PLC systems assuming that the samples of IN are estimated and are available to the receiver for further exploitation. Furthermore, in the third section of this chapter, we formalize ML estimators for PLC systems that jointly estimate channel impulse response (CIR) and IN. Moreover, we study the performance of the proposed estimators in typical PLC scenarios and present evaluation of their channel estimation errors.

4.1 System Model

In order to derive ML channel estimators, we consider an OFDM based PLC system with N subcarriers. Among N subcarriers, only N_u subcarriers, ranging from k_{start} to k_{end}, are used for the data transmission. The remaining $N - N_u$ subcarriers are nulled/unused as shown in Fig. 4.1(a). The corresponding subcarrier alignment is derived from the narrowband PLC (NB-PLC) standard IEEE 1901.2 [39]. Furthermore, a transmission frame containing a preamble and the information bearing payload OFDM symbol, as shown in Fig. 4.1(b), is also considered. Symbols in the preamble of the frame are assumed to be known to the receiver and are used for both transmitter-receiver synchronization and channel estimation.

Characterizing the occurrence of IN during transmission, the frequency domain received symbols corresponding to the preamble is expressed as:

$$\breve{\mathbf{y}}_p = \begin{bmatrix} \mathbf{0} \\ \mathbf{\Lambda}_p \breve{\mathbf{h}} \\ \mathbf{0} \end{bmatrix} + \mathbf{F}\mathbf{n}_p. \tag{4.1}$$

The variables in (4.1) follow a similar notation as in (3.1). However, a subscripted variable p is used in (4.1) to specifically denote preamble received symbols and the noise vector $\mathbf{n}_p = \mathbf{w}_p + \mathbf{i}_p$ contains time domain samples of

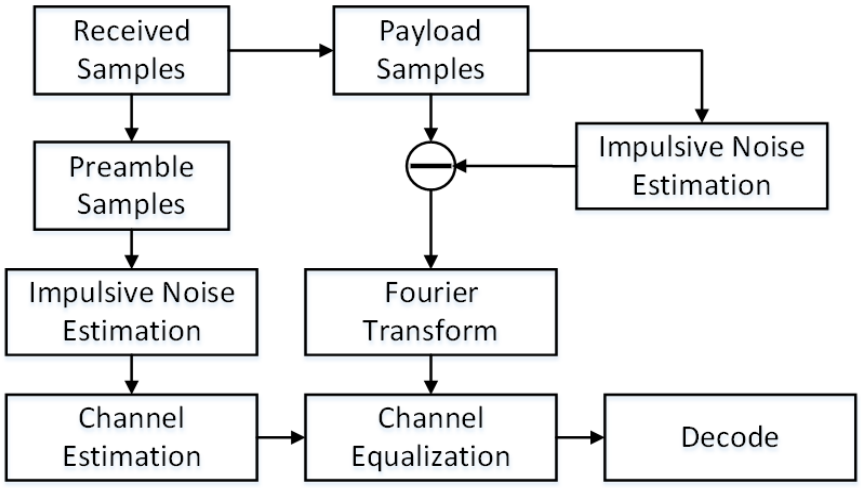


Figure 4.2: Block diagram of the proposed scheme.

the background noise and IN occurring during the preamble transmission time.

4.2 Exploiting Estimated IN Samples for Channel Estimation

After receiving the symbols transmitted in the preamble as in (4.1), the IN can be efficiently estimated exploiting the algorithm that is proposed in section 3.1 of chapter 3. In this section, we will introduce two strategies, namely stochastic and conditional, to exploit the estimated IN for CFR estimation. In the upcoming two subsections, we derive two ML channel estimators that are based on these approaches and follow schematic that is shown in Fig. 4.2.

4.2.1 Stochastic Setting

Under stochastic setting, the time domain samples of estimated IN are considered random and Gaussian distributed with zero mean and variance $|\hat{\mathbf{i}}_p|^2$, where $\hat{\mathbf{i}}_p$ denotes the sparse vector containing the samples of estimated IN. Assuming that the receiver knows the background noise variance $\bar{\sigma}_w^2$ and $\hat{\mathbf{i}}_p$, the log-likelihood function $z_{\text{stochastic}}$ of (4.1) can be written as:

$$z_{\text{stochastic}} = \text{logdet}(\hat{\mathbf{C}}_n) + \bar{\mathbf{A}}^H[\mathbf{F}\hat{\mathbf{C}}_n\mathbf{F}^H]^{-1}\bar{\mathbf{A}}, \tag{4.2}$$

where

$\bar{\mathbf{A}} = \left[\breve{\mathbf{y}}_p - \begin{pmatrix} \mathbf{0} \\ \mathbf{\Lambda}_p \breve{\mathbf{h}} \\ \mathbf{0} \end{pmatrix}\right]$ and $\hat{\mathbf{C}}_n = \bar{\sigma}_w^2 \mathbf{I}_N + \text{diag}(|\hat{\mathbf{i}}_p|^2)$ is the noise covariance matrix.

Taking the derivative of the likelihood function in (4.2) and equating it to zero, the channel estimator under stochastic setting can be written in the closed form as:

$$\hat{\breve{\mathbf{h}}}_{\text{stochastic}} = \mathbf{\Lambda_p}^{-1}[\mathbf{F}_2 \hat{\mathbf{C}}_n^{-1} \mathbf{F}_2^H]^{-1} \mathbf{F}_2 \hat{\mathbf{C}}_n^{-1} \mathbf{F}^H \breve{\mathbf{y}}_p, \tag{4.3}$$

where $\mathbf{F}_2$ is the sub-DFT matrix that contains only the rows of $\mathbf{F}$ that identify the locations of used subcarriers.

4.2.2 Conditional Setting

Under conditional setting, the estimated IN is treated as a deterministic quantity. By replacing the IN vector $\mathbf{i}_p$ in (4.1) by a column vector $\mathbf{i}_c$ that contains only N_{imp} non-zero elements of $\mathbf{i}_p$ (neglecting the zero entries), the frequency domain received symbols vector $\breve{\mathbf{y}}_p$ can be expressed as:

$$\breve{\mathbf{y}}_p = \begin{bmatrix} \breve{\mathbf{y}}_1 \\ \breve{\mathbf{y}}_2 \\ \breve{\mathbf{y}}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{\Lambda}_p \breve{\mathbf{h}} \\ \mathbf{0} \end{bmatrix} + \tilde{\mathbf{F}} \mathbf{i}_c + \mathbf{F} \mathbf{w}_p, \tag{4.4}$$

where the vectors $\breve{\mathbf{y}}_1$, $\breve{\mathbf{y}}_2$ and $\breve{\mathbf{y}}_3$ are decomposition of $\breve{\mathbf{y}}$ such that $\breve{\mathbf{y}}_1$ contains the received symbols in the first $(k_{\text{start}} - 1)$ unused subcarriers, $\breve{\mathbf{y}}_2$ contains the received symbols in N_u used subcarriers and $\breve{\mathbf{y}}_3$ contains the received symbols in the last $N - k_{\text{end}}$ unused subcarriers. The matrix $\tilde{\mathbf{F}}$ in (4.4) is the sub-DFT matrix that contains only the columns of $\mathbf{F}$ that correspond to the locations of IN samples.

Assuming deterministic IN and replacing $\mathbf{i}_c$ by its estimate $\hat{\mathbf{i}}_c$, the log-likelihood function of (4.4) is written as:

$$z_{\text{conditional}} = \text{logdet}(\bar{\sigma}_w^2 \mathbf{I}_N) + \bar{\mathbf{B}}^H \left(\frac{1}{\bar{\sigma}_w^2} \mathbf{I}_N\right) \bar{\mathbf{B}}, \tag{4.5}$$

where $\bar{\mathbf{B}} = \left[\begin{pmatrix} \breve{\mathbf{y}}_1 \\ \breve{\mathbf{y}}_2 \\ \breve{\mathbf{y}}_3 \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \mathbf{\Lambda_p} \breve{\mathbf{h}} \\ \mathbf{0} \end{pmatrix} - \tilde{\mathbf{F}} \hat{\mathbf{i}}_c \right]$.

Differentiating (4.5) and equating to zero, the estimated channel vector can be expressed in the closed form as:

$$\hat{\tilde{\mathbf{h}}}_{\text{conditional}} = \mathbf{\Lambda_p}^{-1}(\breve{\mathbf{y}}_2 - \tilde{\mathbf{F}}_2\hat{\mathbf{i}}_c), \tag{4.6}$$

where $\tilde{\mathbf{F}}_2$ is the sub-DFT matrix containing only the rows of $\tilde{\mathbf{F}}$ that correspond to the locations of used subcarriers.

4.2.3 Numerical Validation

In this subsection, we present numerical evaluation of the performance of the channel estimators derived in (4.3) and (4.6). In order to run the simulations, we consider an OFDM based PLC system operating in the CENELEC-A band and has $N = 256$ subcarriers. Among these 256, only 36 subcarriers indexed from 23^{rd} to 58^{th} are used for data transmission, where each subcarrier conveys a complex symbol that is randomly drawn from QPSK modulation. The subcarriers other than the ones bearing data are nulled/unused. The frequency selective multipath channel model defined in (3.40) is used to simulate realistic effect of the power line channel over the communication signal. Moreover the channel parameters are adopted from the standard IEEE 1901.2 [39].

The IN is defined according to the Bernoulli-Gaussian noise model as in (3.2). The performance of the proposed channel estimators is evaluated in terms of mean squared error (MSE) and BER of the receiver. The performances of two other algorithms that are based on nulling [76] and clipping [58] the IN samples followed by LS channel estimation are also shown in the graphs, to compare with the performances of the proposed estimators. In these two approaches, first the IN samples are identified using threshold reference value and then those samples are nulled/clipped. After suppressing the IN present in the received samples, LS channel estimation is performed using the symbols received in preamble and the known pilot symbols. Apart from these, the performances of two genie aided receivers which have true IN information and perform ML channel estimations are also assessed. Fig. 4.3 shows the MSE when SNR is 10 dB and INR is varied from 20-60 dB. Similarly Fig. 4.4 shows the MSE when SNR is varied from 0-30 dB and INR is fixed to 40 dB. Finally, Fig. 4.5 shows the BER performance of the system where SNR is varied from 0-15 dB and INR is equal to 40 dB.

As evident from the simulation results, the proposed estimators outperform conventional strategies while estimating the channel in typical power line environments with distinct level of INR. The numerically evaluated performances

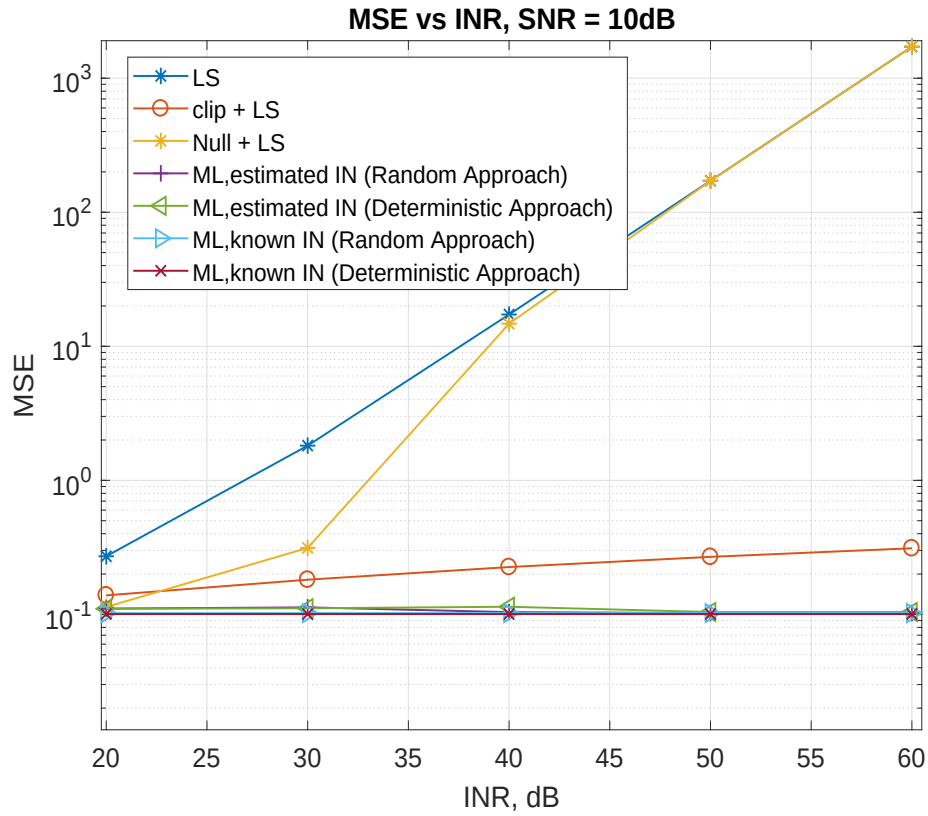


Figure 4.3: MSE of channel estimation of the proposed estimators along with the algorithms based on nulling and clipping impulsive noise corrupted samples followed by the least squares channel estimation.

show that the estimators proposed in (4.3) and (4.6) estimate the channel more accurately than the conventional channel estimation strategies. As shown in the simulation result graphs, the proposed estimators yield lower MSE and BER than the conventional channel estimators. Moreover, the estimator proposed in (4.3), which estimates the CFR by weighting the time domain samples of the received signal by the noise covariance has lower channel estimation error in comparison to the variance of the estimator that is proposed in (4.6), which performs CFR estimation by canceling out the estimated IN from the received signal.

4.3 Joint ML Estimation of CIR and IN

In the previous section we showed that a more accurate channel estimation can be performed, provided the PLC receiver has an estimated information on

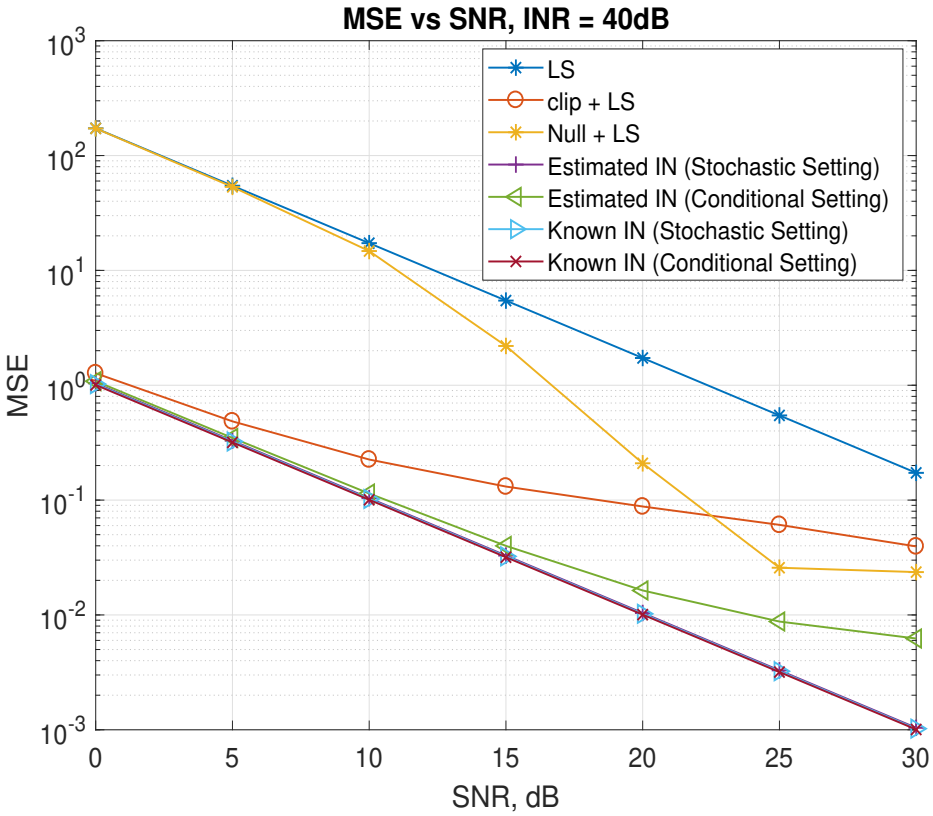


Figure 4.4: MSE of channel estimation of the proposed estimators along with the algorithms based on nulling and clipping impulsive noise corrupted samples followed by the least squares channel estimation.

the IN occurring during the data transmission. After establishing ML channel estimators in section 4.2 that rely on the estimated IN, in this section, we direct our attention towards designing ML estimators that can jointly estimate both the CIR and the IN. The proposed ML estimators, while performing channel estimation, rely on a cost function that is solely a function of CIR. The dependency of the cost function over the IN is removed by first estimating the IN and replacing the IN variable in the cost function by its estimated value. In order to avoid any confusion between the ML estimators proposed in section 4.2 and the estimators that are proposed in this section, we will call the ML estimators in the text to follow as random ML and deterministic ML.

To concisely elaborate on the proposed channel estimators, we divide this section into six subsections. In subsection 4.3.1, we recall the system model of the OFDM-based PLC system model in (4.1). In subsection 4.3.2, we derive the deterministic ML estimator and the random ML estimator. In subsection

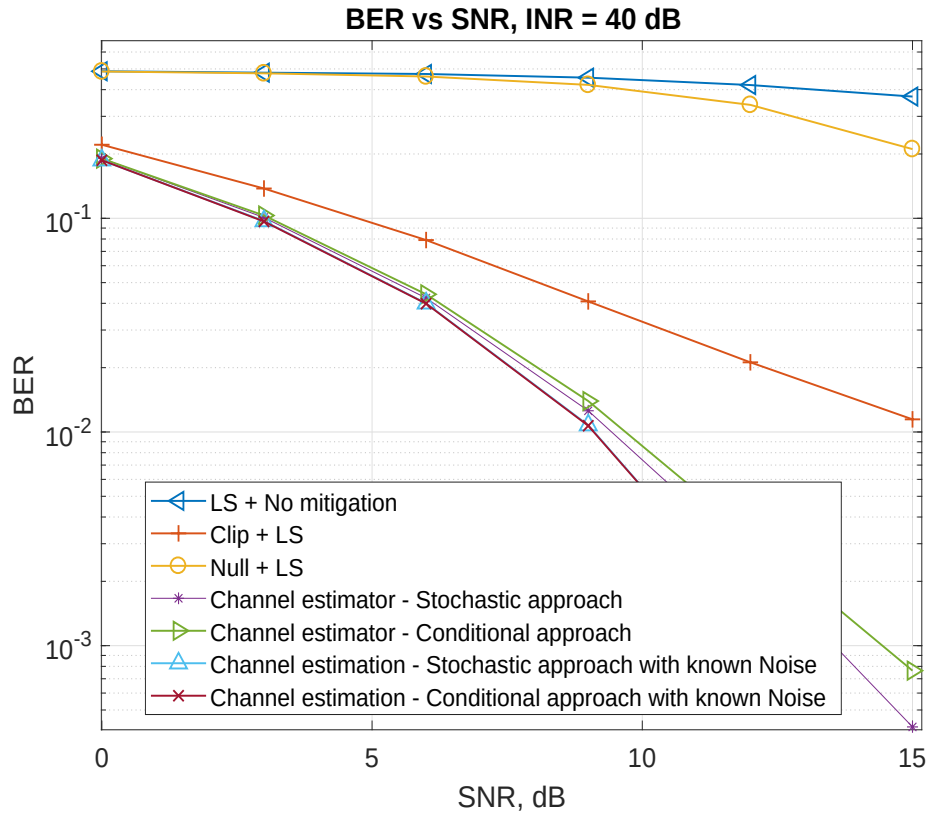


Figure 4.5: BER performance of the proposed estimators along with the algorithms based on nulling and clipping impulsive noise affected samples followed by the least squares channel estimation.

4.3.3, we elaborate on the characterization of the channel estimation errors of both estimators. In subsection 4.3.4, we present all analytical evaluation of the performance of the proposed estimators. Furthermore, we verify the performance of both estimators numerically and provide discussions on the simulation results in subsection 4.3.5. Finally, a brief conclusion is drawn and presented in subsection 4.3.6.

4.3.1 System Model

Recalling from the system model in (3.3), the CIR is expressed as:

$$\breve{\mathbf{h}} = \sqrt{N}\mathbf{G}^H\mathbf{h} \tag{4.7}$$

and the time domain received samples vector is defined as:

$$\mathbf{y}_p = \sqrt{N}\mathbf{F}^H\mathbf{E}\mathbf{\Lambda}_p\mathbf{G}^H\mathbf{h} + \mathbf{i}_p + \mathbf{w}_p. \tag{4.8}$$

4.3.2 Proposed ML Estimators

In this section, we present the derivations of the proposed ML channel estimators. After receiving the samples transmitted during the preamble transmission time, as shown in (4.8), respective cost functions are derived for the deterministic ML and the random ML estimators. The derivation of cost functions is done following similar assumptions as those that were made in sections 4.2.1 and 4.2.2 of this chapter. To concisely elaborate on the proposed ML estimators, we divide this section in to two subsections. In subsection 4.3.2.1, we derive the deterministic ML estimator. In subsection 4.3.2.2, we present the derivation of the random ML estimator.

4.3.2.1 Deterministic ML Estimator

Under the deterministic setting, we express the time domain received signal corresponding to preamble in (4.8) as

$$\mathbf{y}_p = \sqrt{N}\mathbf{F}\mathbf{E}\mathbf{\Lambda}_p\mathbf{G}^H\mathbf{h} + \mathbf{U}\mathbf{i}_c + \mathbf{w}_p, \tag{4.9}$$

where $\mathbf{i}_c$ is a non-sparse vector containing only the non-zero entries of $\mathbf{i}_p$ and

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_{n_1} & \cdots & \mathbf{u}_{n_{N_{imp}}} \end{bmatrix}.$$

The matrix $\mathbf{U}$ in (4.9) is of dimension $N \times N_{imp}$, where each column vector $\mathbf{u}_{\mathcal{I}}$ is of length $N \times 1$ and has a single 1 identifying the location of IN sample as indexed by $\mathcal{I}$. Furthermore, we also define the matrix $\mathbf{P}$ as a diagonal matrix whose entries are ones only at the locations where there are IN samples. In other words, $\mathbf{P} = \mathbf{U}\mathbf{U}^T$.

Assuming deterministic IN samples, the log-likelihood function of (4.9) takes the form

$$\zeta_{DML}\left(\mathbf{h}, \mathbf{i}_c, \sigma_w^2\right) = N \log \sigma_w^2 + \frac{1}{\sigma_w^2}\left\|\mathbf{y}_p - \sqrt{N}\mathbf{\Theta}\mathbf{h} - \mathbf{U}\mathbf{i}_c\right\|^2, \tag{4.10}$$

where $\Theta = \mathbf{F}\mathbf{E}\mathbf{\Lambda}_p\mathbf{G}^H$, which will be used through out the remainder of the chapter. Differentiating (4.10) with respect to IN variable and equating it to zero, the estimated IN takes the form:

$$\hat{\mathbf{i}}_c = \mathbf{U}^H(\mathbf{y}_p - \sqrt{N}\Theta\mathbf{h}). \tag{4.11}$$

Replacing the IN variable $\mathbf{i}_c$ by its estimate, the likelihood function in (4.10) can be equivalently reformulated, only as a function of $\mathbf{h}$, as:

$$\zeta_{DML}\left(\mathbf{h}\right) = \frac{1}{N - N_{imp}}\left(\mathbf{y}_p - \sqrt{N}\mathbf{\Theta}\mathbf{h}\right)^H \mathbf{P}^{\perp}\left(\mathbf{y}_p - \sqrt{N}\mathbf{\Theta}\mathbf{h}\right), \tag{4.12}$$

where $\mathbf{P}^{\perp} = \mathbf{I}_N - \mathbf{P} = \mathbf{I}_N - \sum_{i=1}^{N_{imp}} \mathbf{u}_{n_i}\mathbf{u}_{n_i}^T$ is a diagonal matrix containing ones in its diagonal entries corresponding to the locations of the samples in the received signal that are not impaired by the IN.

Taking the derivative of (4.12) with respect to $\mathbf{h}$ and equating it to zero, assuming that the supports of the IN samples are provided by a previous inference procedure (e.g., the one reported in [140]), the deterministic ML estimator can be expressed as:

$$\hat{\mathbf{h}}_{DML} = \frac{1}{\sqrt{N}} \left(\Theta^H \hat{\mathbf{P}}^{\perp} \Theta\right)^{-1} \Theta^H \hat{\mathbf{P}}^{\perp} \mathbf{y}_p, \tag{4.13}$$

where $\hat{\mathbf{P}}^{\perp} = \mathbf{I}_N - \hat{\mathbf{P}}$ is the estimated value of $\mathbf{P}^{\perp}$ and $\hat{\mathbf{P}}$ is the estimated value of $\mathbf{P}$.

4.3.2.2 Random ML Estimator

Under the random setting, the time domain received signal model in (4.8) is expressed as:

$$\mathbf{y}_p = \sqrt{N}\Theta\mathbf{h} + \mathbf{n}, \tag{4.14}$$

where $\mathbf{n} = \mathbf{i}_p + \mathbf{w}_p$. Furthermore, we assume that the noise follows a zero mean circular Gaussian distribution $\mathbf{n} \sim \mathcal{CN}(0, \mathbf{C}_n)$, where the covariance matrix has the structure

$$\mathbf{C}_n = \sigma^2\mathbf{P} + \sigma_w^2\mathbf{P}^{\perp},$$

where $\sigma^2 = \sigma_i^2 + \sigma_w^2$ denotes the cumulative noise power at the locations of the IN samples. The log-likelihood function of (4.14), assuming random IN, takes the form

$$\zeta_{RML}\left(\mathbf{h}, \sigma^2, \sigma_w^2\right) = \log\det(\mathbf{C}_n) + \bar{\mathbf{y}}^H \mathbf{C}_n^{-1} \bar{\mathbf{y}}, \tag{4.15}$$

where $\bar{\mathbf{y}} = \mathbf{y}_p - \sqrt{N}\Theta\mathbf{h}$.

Assuming that the support of the IN is previously estimated, the resulting noise covariance matrix can be constructed as $\hat{\mathbf{C}}_n = \hat{\sigma}^2\hat{\mathbf{P}} + \hat{\sigma}_w^2\hat{\mathbf{P}}^{\perp}$, where $\mathrm{tr}[\hat{\mathbf{P}}] = \hat{N}_{imp}$ and $\mathrm{tr}[\hat{\mathbf{P}}^{\perp}] = N - \hat{N}_{imp}$. Differentiating (4.15) with respect to the cumulative noise power and the background noise power and equating them to zero, the cumulative noise power and the background noise power in terms of locations of IN samples and its orthogonal locations and the received

signal can be expressed as:

$$\hat{\sigma}^2 = \frac{1}{\hat{N}_{imp}} \bar{\mathbf{y}}^H \hat{\mathbf{P}} \bar{\mathbf{y}}$$
$$\text{and} \tag{4.16}$$
$$\hat{\sigma}_w^2 = \frac{1}{N - \hat{N}_{imp}} \bar{\mathbf{y}}^H \hat{\mathbf{P}}^{\perp} \bar{\mathbf{y}}.$$

Furthermore, plugging the values of estimated cumulative noise power and the background noise power into (4.15), the normalized likelihood function for the random ML estimator, only in terms of $\mathbf{h}$, can be expressed as:

$$\zeta_{RML}(\mathbf{h}) = \frac{\hat{N}_{imp}}{N} \log\left(\frac{1}{\hat{N}_{imp}} \bar{\mathbf{y}}^H \hat{\mathbf{P}} \bar{\mathbf{y}}\right) + \frac{N - \hat{N}_{imp}}{N} \log\left(\frac{1}{N - \hat{N}_{imp}} \bar{\mathbf{y}}^H \hat{\mathbf{P}}^{\perp} \bar{\mathbf{y}}\right). \tag{4.17}$$

The ML estimate of the CIR is now achieved by finding the optimum of the likelihood function in (4.17). In order to do so, we solve $\nabla\zeta_{RML}(\mathbf{h}) = \mathbf{0}$ by applying the Newton-Raphson method [148]. Initiating the solution finding procedure with an initial guess $\mathbf{h}_0$ and exploiting the Hessian and the gradient of the likelihood function, an iterative approximation of the root can be computed as:

$$\mathcal{R}(\mathbf{h}_{\mathrm{k}+1}) = \mathcal{R}(\mathbf{h}_{\mathrm{k}}) - [\mathbb{H}\zeta_{RML}(\mathbf{h}_{\mathrm{k}})]^{-1} \nabla\zeta_{RML}(\mathbf{h}_{\mathrm{k}}), \tag{4.18}$$

where $\mathcal{R}(\mathbf{h}) = \left[\mathrm{Re}^T(\mathbf{h}), \mathrm{Im}^T(\mathbf{h})\right]^T$, $\mathbb{H}\zeta_{RML}$ is the Hessian matrix and $\nabla\zeta_{RML}$ is the gradient of the likelihood function. The final estimate of the CIR is achieved when $\frac{||\mathcal{R}(\mathbf{h}_{\mathrm{k}+1}) - \mathcal{R}(\mathbf{h}_{\mathrm{k}})||^2}{||\mathcal{R}(\mathbf{h}_{\mathrm{k}})||^2} \leq t$, where t is a threshold that defines the stopping criterion.

In order to proceed with the iterative-approximation of the channel estimate, we begin by evaluating the Hessian matrix and the gradient of the cost function with respect to $\mathbf{h}$.

Calculation of Hessian and Gradient of Cost Function

To evaluate the gradient of the likelihood function in (4.17), we begin by evaluating the complex derivatives that establish

$$\nabla\zeta_{RML}(\mathbf{h}) = \frac{\partial\zeta_{RML}(\mathbf{h})}{\partial\mathcal{R}(\mathbf{h})} = \widetilde{\mathbf{\Xi}} \begin{bmatrix} \frac{\partial\zeta_{RML}(\mathbf{h})}{\partial\mathbf{h}^*} \\ \frac{\partial\zeta_{RML}(\mathbf{h})}{\partial\mathbf{h}} \end{bmatrix}, \tag{4.19}$$

where $\frac{\partial\zeta_{RML}(\mathbf{h})}{\partial\mathbf{h}}$ denotes derivative of the cost function with respect to the CIR, $\frac{\partial\zeta_{RML}(\mathbf{h})}{\partial\mathbf{h}^*}$ is the derivative of the cost function with respect to the complex

conjugate of the CIR and $\widetilde{\boldsymbol{\Xi}} = \begin{bmatrix} \mathbf{I}_L & \mathbf{I}_L \\ -\mathrm{j}\mathbf{I}_L & \mathrm{j}\mathbf{I}_L \end{bmatrix}$. Furthermore, the Hessian of the cost function can be calculated by evaluating

$$\mathbb{H}\zeta_{RML}\left(\mathbf{h}\right) = \frac{\partial^2 \zeta_{RML}\left(\mathbf{h}\right)}{\partial \mathcal{R}\left(\mathbf{h}\right) \partial \mathcal{R}^T\left(\mathbf{h}\right)} = \widetilde{\boldsymbol{\Xi}} \begin{bmatrix} \frac{\partial^2 \zeta_{RML}(\mathbf{h})}{\partial \mathbf{h}^* \partial \mathbf{h}^H} & \frac{\partial^2 \zeta_{RML}(\mathbf{h})}{\partial \mathbf{h}^* \partial \mathbf{h}^T} \\ \frac{\partial^2 \zeta_{RML}(\mathbf{h})}{\partial \mathbf{h} \partial \mathbf{h}^H} & \frac{\partial^2 \zeta_{RML}(\mathbf{h})}{\partial \mathbf{h} \partial \mathbf{h}^T} \end{bmatrix} \widetilde{\boldsymbol{\Xi}}^T. \tag{4.20}$$

To simplify the notation, we express the Hessian matrix equivalently as

$$\mathbb{H}\zeta_{RML}\left(\mathbf{h}\right) = \widetilde{\boldsymbol{\Xi}} \begin{bmatrix} \bar{\mathbb{H}}_c\left(\mathbf{h}\right) & \mathbb{H}_c\left(\mathbf{h}\right) \\ \left(\mathbb{H}_c\left(\mathbf{h}\right)\right)^T & \left(\bar{\mathbb{H}}_c\left(\mathbf{h}\right)\right)^* \end{bmatrix} \widetilde{\boldsymbol{\Xi}}^T, \tag{4.21}$$

where $\mathbb{H}_c\left(\mathbf{h}\right) = \frac{\partial^2 \zeta_{RML}(\mathbf{h})}{\partial \mathbf{h}^* \partial \mathbf{h}^T}$ and $\bar{\mathbb{H}}_c\left(\mathbf{h}\right) = \frac{\partial^2 \zeta_{RML}(\mathbf{h})}{\partial \mathbf{h}^* \partial \mathbf{h}^H}$.

Proceeding with the calculations of the complex derivatives, the elements of the gradient vector and the Hessian matrix can be evaluated as:

$$\frac{\partial \zeta_{RML}\left(\mathbf{h}\right)}{\partial \mathbf{h}^*} = -\frac{\hat{N}_{imp}}{\sqrt{N}} \frac{\boldsymbol{\Theta}^H \hat{\mathbf{P}} \bar{\mathbf{y}}}{\bar{\mathbf{y}}^H \hat{\mathbf{P}} \bar{\mathbf{y}}} - \frac{N - \hat{N}_{imp}}{\sqrt{N}} \frac{\boldsymbol{\Theta}^H \hat{\mathbf{P}}^\perp \bar{\mathbf{y}}}{\bar{\mathbf{y}}^H \hat{\mathbf{P}}^\perp \bar{\mathbf{y}}}, \tag{4.22}$$

$$\mathbb{H}_c\left(\mathbf{h}\right) = \frac{\hat{N}_{imp}}{\bar{\mathbf{y}}^H \hat{\mathbf{P}} \bar{\mathbf{y}}} \boldsymbol{\Theta}^H \left(\hat{\mathbf{P}} - \frac{\hat{\mathbf{P}} \bar{\mathbf{y}} \bar{\mathbf{y}}^H \hat{\mathbf{P}}}{\bar{\mathbf{y}}^H \hat{\mathbf{P}} \bar{\mathbf{y}}} \right) \boldsymbol{\Theta} + \frac{N - \hat{N}_{imp}}{\bar{\mathbf{y}}^H \hat{\mathbf{P}}^\perp \bar{\mathbf{y}}} \boldsymbol{\Theta}^H \left(\hat{\mathbf{P}}^\perp - \frac{\hat{\mathbf{P}}^\perp \bar{\mathbf{y}} \bar{\mathbf{y}}^H \hat{\mathbf{P}}^\perp}{\bar{\mathbf{y}}^H \hat{\mathbf{P}}^\perp \bar{\mathbf{y}}} \right) \boldsymbol{\Theta}, \tag{4.23}$$

and

$$\bar{\mathbb{H}}_c\left(\mathbf{h}\right) = -\hat{N}_{imp} \frac{\boldsymbol{\Theta}^H \hat{\mathbf{P}} \bar{\mathbf{y}} \bar{\mathbf{y}}^H \hat{\mathbf{P}} \boldsymbol{\Theta}^*}{\left(\bar{\mathbf{y}}^H \hat{\mathbf{P}} \bar{\mathbf{y}}\right)^2} - \left(N - \hat{N}_{imp}\right) \frac{\boldsymbol{\Theta}^H \hat{\mathbf{P}}^\perp \bar{\mathbf{y}} \bar{\mathbf{y}}^H \hat{\mathbf{P}}^\perp \boldsymbol{\Theta}^*}{\left(\bar{\mathbf{y}}^H \hat{\mathbf{P}}^\perp \bar{\mathbf{y}}\right)^2}, \tag{4.24}$$

where we recall that $\bar{\mathbf{y}} = \mathbf{y}_p - \sqrt{N}\Theta\mathbf{h}$. Upon evaluating (4.22), (4.23) and (4.24), and plugging them into (4.19) and (4.21), the gradient and the Hessian of the cost can be readily obtained.

Observation 1:

In this section we derived the random ML estimator by using gradient search algorithm. In the upcoming section we will show how the same random ML estimator can be computed in a much more efficient way.

4.3.2.3 Simplified Computation of the Random ML Estimator

In order to derive the simplified random ML estimator, we begin by recalling the log-likelihood function in (4.15) that takes the form

$$\zeta_{RML}\left(\mathbf{h}, \sigma^2, \sigma_w^2\right) = \log\det\left(\sigma^2 P + \sigma_w^2 \mathbf{P}^\perp\right) + \bar{\mathbf{y}}^H \left(\frac{1}{\sigma^2}\mathbf{P} + \frac{1}{\sigma_w^2}\mathbf{P}^\perp \right) \bar{\mathbf{y}}, \tag{4.25}$$

where

$$\bar{\mathbf{y}} = \mathbf{y}_p - \sqrt{N}\Theta\mathbf{h}. \tag{4.26}$$

By taking the derivative of the cost function with respect to the channel we can readily see that

$$\frac{\partial \zeta_{RML}\left(\mathbf{h}, \sigma^2, \sigma_w^2\right)}{\partial \mathbf{h}^*} = -\sqrt{N}\Theta^H \left(\frac{1}{\sigma^2}\mathbf{P} + \frac{1}{\sigma_w^2}\mathbf{P}^{\perp}\right)\bar{\mathbf{y}}. \tag{4.27}$$

Forcing (4.27) to be equal to zero we see that the ML channel estimator is a solution to the problem

$$\Theta^H \left[\frac{\mathbf{P}}{\sigma^2} + \frac{\mathbf{P}^{\perp}}{\sigma_w^2}\right]\mathbf{y}_p = \Theta^H \left(\frac{\mathbf{P}}{\sigma^2} + \frac{\mathbf{P}^{\perp}}{\sigma_w^2}\right)\Theta\sqrt{N}\hat{\mathbf{h}} \tag{4.28}$$

such that

$$\bar{\mathbf{y}} = \mathbf{y}_p - \sqrt{N}\Theta\hat{\mathbf{h}} = \left(\mathbf{I}_N - \Theta\left(\Theta^H\left(\frac{\mathbf{P}}{\sigma^2} + \frac{\mathbf{P}^{\perp}}{\sigma_w^2}\right)\Theta\right)^{-1}\Theta^H\left[\frac{\mathbf{P}}{\sigma^2} + \frac{\mathbf{P}^{\perp}}{\sigma_w^2}\right]\right)\mathbf{y}_p. \tag{4.29}$$

Recalling the $N \times N_{imp}$ selection matrix $\mathbf{U}$ in (4.9), such that $\mathbf{P} = \mathbf{U}\mathbf{U}^H$, and using the matrix inverse lemma we can express

$$\begin{aligned}\left(\Theta^H\left(\frac{\mathbf{P}}{\sigma^2} + \frac{\mathbf{P}^{\perp}}{\sigma_w^2}\right)\Theta\right)^{-1} &= \sigma_w^2\left(\left(\frac{\sigma_w^2}{\sigma^2} - 1\right)\Theta^H\mathbf{U}\mathbf{U}^H\Theta + \Theta^H\Theta\right)^{-1} \\ &= \sigma_w^2\left[\left[\Theta^H\Theta\right]^{-1} - \left[\Theta^H\Theta\right]^{-1}\Theta^H\mathbf{U}\left[\left(\frac{\sigma_w^2}{\sigma^2} - 1\right)^{-1}\mathbf{I}_{N_{imp}} + \mathbf{U}^H\mathbf{P}_\Theta\mathbf{U}\right]^{-1}\mathbf{U}^H\Theta\left[\Theta^H\Theta\right]^{-1}\right]\end{aligned} \tag{4.30}$$

where $\mathbf{P}_\Theta = \Theta\left(\Theta^H\Theta\right)^{-1}\Theta^H$. Multiplying (4.30) by Θ and Θ^H on both sides and taking the common factor, we obtain

$$\Theta\left[\Theta^H\left(\frac{\mathbf{P}}{\sigma^2} + \frac{\mathbf{P}^{\perp}}{\sigma_w^2}\right)\Theta\right]^{-1}\Theta^H = \sigma_w^2\mathbf{P}_\Theta\left[\mathbf{I}_N - \mathbf{U}\left[\left(\frac{\sigma_w^2}{\sigma^2} - 1\right)^{-1}\mathbf{I}_{N_{imp}} + \mathbf{U}^H\mathbf{P}_\Theta\mathbf{U}\right]^{-1}\mathbf{U}^H\mathbf{P}_\Theta\right]. \tag{4.31}$$

Upon replacing $\Theta\left[\Theta^H\left(\frac{\mathbf{P}}{\sigma^2} + \frac{\mathbf{P}^{\perp}}{\sigma_w^2}\right)\Theta\right]^{-1}\Theta^H$ in (4.29) by its resulting value obtained in (4.31) we get

$$\bar{\mathbf{y}} = \mathbf{y}_p - \sqrt{N}\Theta\hat{\mathbf{h}} = \mathbf{P}_\Theta^{\perp}\mathbf{y}_p - \mathbf{P}_\Theta\mathbf{U}\left[\left(\frac{\sigma_w^2}{\sigma^2} - 1\right)^{-1}\mathbf{I}_{N_{imp}} + \mathbf{U}^H\mathbf{P}_\Theta\mathbf{U}\right]^{-1}\mathbf{U}^H\mathbf{P}_\Theta^{\perp}\mathbf{y}_p, \tag{4.32}$$

where $\mathbf{P}_\Theta^{\perp} = \mathbf{I}_N - \mathbf{P}_\Theta$. This further implies that

$$\bar{\mathbf{y}}^H\mathbf{P}\bar{\mathbf{y}} = \mathbf{y}_p^H\mathbf{P}_\Theta^{\perp}\mathbf{U}\left[\mathbf{I} + \left(\frac{\sigma_w^2}{\sigma^2} - 1\right)\mathbf{U}^H\mathbf{P}_\Theta\mathbf{U}\right]^{-2}\mathbf{U}^H\mathbf{P}_\Theta^{\perp}\mathbf{y}_p \tag{4.33}$$

and

$$\bar{\mathbf{y}}^H\mathbf{P}^{\perp}\bar{\mathbf{y}} = \mathbf{y}_p^H\mathbf{P}_{\Theta}^{\perp}\mathbf{y}_p + \left(\frac{\sigma_w^2}{\sigma^2}-1\right)\mathbf{y}_p^H\mathbf{P}_{\Theta}^{\perp}\mathbf{U}\left[\mathbf{I}+\left(\frac{\sigma_w^2}{\sigma^2}-1\right)\mathbf{U}^H\mathbf{P}_{\Theta}\mathbf{U}\right]^{-1}\mathbf{U}^H\mathbf{P}_{\Theta}^{\perp}\mathbf{y}_p \tag{4.34}$$
$$-\frac{\sigma_w^2}{\sigma^2}\mathbf{y}_p^H\mathbf{P}_{\Theta}^{\perp}\mathbf{U}\left[\mathbf{I}+\left(\frac{\sigma_w^2}{\sigma^2}-1\right)\mathbf{U}^H\mathbf{P}_{\Theta}\mathbf{U}\right]^{-2}\mathbf{U}^H\mathbf{P}_{\Theta}^{\perp}\mathbf{y}_p.$$

Inserting the resulting values of $\bar{\mathbf{y}}^H\mathbf{P}\bar{\mathbf{y}}$ and $\bar{\mathbf{y}}^H\mathbf{P}^{\perp}\bar{\mathbf{y}}$ back into the random ML cost function in (4.25) we see that

$$\zeta_{RML}\left(\hat{\mathbf{h}},\sigma^2,\sigma_w^2\right) = N_{imp}\log\sigma^2 + (N-N_{imp})\log\sigma_w^2 \tag{4.35}$$
$$+\frac{1}{\sigma_w^2}\mathbf{y}_p^H\mathbf{P}_{\Theta}^{\perp}\mathbf{y}_p - \frac{\sigma^2-\sigma_w^2}{\left(\sigma_w^2\right)^2}\mathbf{y}_p^H\mathbf{P}_{\Theta}^{\perp}\mathbf{U}\left[\mathbf{I}+\left(\frac{\sigma^2-\sigma_w^2}{\sigma_w^2}\right)\mathbf{U}^H\mathbf{P}_{\Theta}^{\perp}\mathbf{U}\right]^{-1}\mathbf{U}^H\mathbf{P}_{\Theta}^{\perp}\mathbf{y}_p.$$

To proceed further, we now estimate the two variances σ^2, σ_w^2 as the minimizers of the cost function in (4.35) in the region $0 \le \sigma_w^2 \le \sigma^2$. For this, we assume that $N_{imp} + L < N$ such that the eigenvalue/eigenvector decomposition of the matrix $\mathbf{U}^H\mathbf{P}_{\Theta}^{\perp}\mathbf{U}$ takes the form

$$\mathbf{U}^H\mathbf{P}_{\Theta}^{\perp}\mathbf{U} = \sum_{i=1}^{N_{imp}}\lambda_i\mathbf{u}_i\mathbf{u}_i^H,$$

where $\lambda_i > 0$ for all i and $\lambda_i = 1$ for $i > \min\{N_{imp}, L\}$. The cost function in (4.35) can therefore be rewritten as

$$\zeta_{RML}\left(\hat{\mathbf{h}},\sigma^2,\sigma_w^2\right) = N_{imp}\log\sigma^2 + (N-N_{imp})\log\sigma_w^2 + \frac{1}{\sigma_w^2}\mathbf{y}_p^H\mathbf{P}_{\Theta}^{\perp}\mathbf{y}_p \tag{4.36}$$
$$-\frac{\sigma^2-\sigma_w^2}{\sigma_w^2}\sum_{i=1}^{N_{imp}}\frac{1}{\sigma_w^2+\left(\sigma^2-\sigma_w^2\right)\lambda_i}\mathbf{y}_p^H\mathbf{P}_{\Theta}^{\perp}\mathbf{U}\mathbf{u}_i\mathbf{u}_i^H\mathbf{U}^H\mathbf{P}_{\Theta}^{\perp}\mathbf{y}_p.$$

To find the minimum of the cost function in the region $0 \le \sigma_w^2 \le \sigma^2$, we first investigate the behavior of (4.36) in the boundary of the feasibility region. We begin by observing that when $\sigma_w^2 \to 0$ for any fixed σ^2 we have

$$\zeta_{RML}\left(\hat{\mathbf{h}},\sigma^2,\sigma_w^2\right) = N_{imp}\log\sigma^2 + (N-N_{imp})\log\sigma_w^2 \tag{4.37}$$
$$+\frac{1}{\sigma_w^2}\mathbf{y}_p^H\left(\mathbf{P}_{\Theta}^{\perp}-\mathbf{P}_{\Theta}^{\perp}\mathbf{U}\left(\mathbf{U}^H\mathbf{P}_{\Theta}^{\perp}\mathbf{U}\right)^{-1}\mathbf{U}^H\mathbf{P}_{\Theta}^{\perp}\right)\mathbf{y}_p$$
$$+\frac{1}{\sigma^2}\sum_{i=1}^{N_{imp}}\frac{1}{\lambda_i^2}\left(1-(1-\lambda_i)\frac{\sigma_w^2}{\sigma^2}\right)\mathbf{y}_p^H\mathbf{P}_{\Theta}^{\perp}\mathbf{U}\mathbf{u}_i\mathbf{u}_i^H\mathbf{U}^H\mathbf{P}_{\Theta}^{\perp}\mathbf{y}_p + o\left(\sigma_w^2\right).$$

As $\zeta_{RML}\left(\hat{\mathbf{h}},\sigma^2,\sigma_w^2\right)\to+\infty$ the minimum is always attained for some $\sigma_w^2>0$. Furthermore, when $\sigma^2\to\infty$ for any fixed σ_w^2 we have

$$\begin{aligned}\zeta_{RML}\left(\hat{\mathbf{h}},\sigma^2,\sigma_w^2\right) &= N_{imp}\log\sigma^2+(N-N_{imp})\log\sigma_w^2 \\ &+\frac{1}{\sigma_w^2}\mathbf{y}_p^H\left(\mathbf{P}_\Theta^\perp-\mathbf{P}_\Theta^\perp\mathbf{U}\left(\mathbf{U}^H\mathbf{P}_\Theta^\perp\mathbf{U}\right)^{-1}\mathbf{U}^H\mathbf{P}_\Theta^\perp\right)\mathbf{y}_p \\ &+\frac{1}{\sigma^2}\mathbf{y}_p^H\mathbf{P}_\Theta^\perp\mathbf{U}\left(\mathbf{U}^H\mathbf{P}_\Theta^\perp\mathbf{U}\right)^{-2}\mathbf{U}^H\mathbf{P}_\Theta^\perp\mathbf{y}_p+o\left(\frac{1}{\sigma^2}\right).\end{aligned} \tag{4.38}$$

As $\zeta_{RML}\left(\hat{\mathbf{h}},\sigma^2,\sigma_w^2\right)\to+\infty$ the minimum is always attained for $\sigma^2<\infty$. Finally when $\sigma_w^2=\sigma^2$ we have

$$\zeta_{RML}\left(\hat{\mathbf{h}},\sigma^2,\sigma^2\right)=N\log\sigma^2+\frac{1}{\sigma^2}\mathbf{y}_p^H\mathbf{P}_\Theta^\perp\mathbf{y}_p. \tag{4.39}$$

The minimum of (4.39) in this boundary is obtained at $\hat{\sigma}^2=\hat{\sigma}_w^2=N^{-1}\mathbf{y}_p^H\mathbf{P}_\Theta^\perp\mathbf{y}_p$, where $\zeta_{RML}=N\log\left(N^{-1}\mathbf{y}_p^H\mathbf{P}_\Theta^\perp\mathbf{y}_p\right)+N$.

Based upon the observations we made above, we can conclude that the minimum of the cost function is either obtained at an inflexion point of the feasible region, or on the boundary line where $\sigma_w^2=\sigma^2$, in which case the optimum is given by $\sigma^2=\sigma_w^2=N^{-1}\mathbf{y}_p^H\mathbf{P}_\Theta^\perp\mathbf{y}_p$. We now investigate the position of the inflexion points for which we begin by taking the partial derivatives of the cost function with respect to both σ^2 and σ_w^2. The derivatives take the form

$$\begin{aligned}\frac{\partial\zeta_{RML}\left(\hat{\mathbf{h}},\sigma^2,\sigma_w^2\right)}{\partial\sigma^2} &= \frac{N_{imp}}{\sigma^2}\Bigg[1-\frac{1}{N_{imp}}\sum_{i=1}^{N_{imp}}\frac{\sigma^2}{\left(\left(1-\lambda_i\right)\sigma_w^2+\lambda_i\sigma^2\right)^2}\mathbf{y}_p^H\mathbf{P}_\Theta^\perp\mathbf{U} \\ &\qquad\times\mathbf{u}_i\mathbf{u}_i^H\mathbf{U}^H\mathbf{P}_\Theta^\perp\mathbf{y}_p\Bigg]\end{aligned} \tag{4.40}$$

and

$$\begin{aligned}\frac{\partial\zeta_{RML}\left(\hat{\mathbf{h}},\sigma^2,\sigma_w^2\right)}{\partial\sigma_w^2} &= \frac{N-N_{imp}}{\sigma_w^2}-\left(\frac{1}{\sigma_w^2}\right)^2\mathbf{y}_p^H\mathbf{P}_\Theta^\perp\mathbf{y}_p \\ &+\frac{1}{\sigma_w^2}\left(\frac{\sigma^2}{\sigma_w^2}\right)\sum_{i=1}^{N_{imp}}\frac{1}{\left(1-\lambda_i\right)\sigma_w^2+\sigma^2\lambda_i}\mathbf{y}_p^H\mathbf{P}_\Theta^\perp\mathbf{U}\mathbf{u}_i\mathbf{u}_i^H\mathbf{U}^H\mathbf{P}_\Theta^\perp\mathbf{y}_p \\ &-\frac{1}{\sigma_w^2}\left(1-\frac{\sigma^2}{\sigma_w^2}\right)\sum_{i=1}^{N_{imp}}\frac{\left(1-\lambda_i\right)\sigma_w^2}{\left(\left(1-\lambda_i\right)\sigma_w^2+\sigma^2\lambda_i\right)^2}\mathbf{y}_p^H\mathbf{P}_\Theta^\perp\mathbf{U}\mathbf{u}_i\mathbf{u}_i^H\mathbf{U}^H\mathbf{P}_\Theta^\perp\mathbf{y}_p.\end{aligned} \tag{4.41}$$

Forcing (4.40) and (4.41) to zero, we can see that any inflexion point corresponds to a $\beta = \sigma^2/\sigma_w^2$ given by a solution to the following equation

$$\mathbf{y}_p^H \mathbf{P}_\Theta^\perp \mathbf{y}_p - \sum_{i=1}^{N_{imp}} \frac{\lambda_i \beta^2 + \beta\left(1 + \frac{N}{N_{\text{imp}}} - 2\lambda_i\right) - (1-\lambda_i)}{(1-\lambda_i+\lambda_i\beta)^2} \mathbf{y}_p^H \mathbf{P}_\Theta^\perp \mathbf{U}\mathbf{u}_i\mathbf{u}_i^H\mathbf{U}^H\mathbf{P}_\Theta^\perp\mathbf{y}_p = 0 \tag{4.42}$$

in the region $\beta \geq 1$. Note that this is a polynomial equation of degree $2N_{imp}$ in β, so we can find the solutions by polynomial rooting in the real axis. Denoting the left hand side expression in (4.42) by $f(\beta)$, the expression can be re-written as $f(\beta) = 0$. Observe that $f(\beta)$ is continuous and differentiable on the positive real axis, and that

$$\lim_{\beta\to\infty} f(\beta) = \mathbf{y}_p^H \left(\mathbf{P}_\Theta^\perp - \mathbf{P}_\Theta^\perp \mathbf{U}\left(\mathbf{U}^H\mathbf{P}_\Theta^\perp\mathbf{U}\right)^{-1}\mathbf{U}^H\mathbf{P}_\Theta^\perp\right)\mathbf{y}_p > 0. \tag{4.43}$$

In a situation when $L < N_{imp}$, there exists some eigenvalues λ_i that are equal to one, such that $\lim_{\beta\to 0^-} f(\beta) = -\infty$. Thus, by continuity of $f(\beta)$ we can say that there exists at least one solution of (4.43) in the region $(0, +\infty)$. However, it may happen that the solution is obtained for $\beta < 1$, which is outside the feasibility region. Moreover, when $L \geq N_{imp}$ all the eigenvalues are positive and therefore

$$\lim_{\beta\to 0^-} f(\beta) = \mathbf{y}_p^H\mathbf{P}_\Theta^\perp\mathbf{y}_p + \mathbf{y}_p^H\mathbf{P}_\Theta^\perp\mathbf{U}\left(\mathbf{U}^H\mathbf{P}_\Theta\mathbf{U}\right)^{-1}\mathbf{U}^H\mathbf{P}_\Theta^\perp\mathbf{y}_p > 0. \tag{4.44}$$

So we cannot generally state that there exists a zero of $f(\beta)$ in the feasibility region.

Assuming that there exists a solution of (4.44) in the region $(1, +\infty)$, one can recover the associated values of $\hat{\sigma}^2$ and $\hat{\sigma}_w^2$ in the inflexion point as

$$\hat{\sigma}^2 = \frac{1}{N_{imp}} \sum_{i=1}^{N_{imp}} \frac{\beta^2}{(1-\lambda_i+\lambda_i\beta)^2} \mathbf{y}_p^H\mathbf{P}_\Theta^\perp\mathbf{U}\mathbf{u}_i\mathbf{u}_i^H\mathbf{U}^H\mathbf{P}_\Theta^\perp\mathbf{y}_p \tag{4.45}$$

and

$$\hat{\sigma}_w^2 = \frac{1}{N-N_{imp}}\mathbf{y}_p^H\mathbf{P}_\Theta^\perp\mathbf{y}_p - \frac{1}{N-N_{imp}}\sum_{i=1}^{N_{imp}} \frac{\beta}{1-\lambda_i+\lambda_i\beta}\mathbf{y}_p^H\mathbf{P}_\Theta^\perp\mathbf{U}\mathbf{u}_i\mathbf{u}_i^H\mathbf{U}^H\mathbf{P}_\Theta^\perp\mathbf{y}_p \tag{4.46}$$

$$+ \frac{1}{N-N_{imp}}\sum_{i=1}^{N_{imp}} \frac{(1-\beta)(1-\lambda_i)}{(1-\lambda_i+\lambda_i\beta)^2}\mathbf{y}_p^H\mathbf{P}_\Theta^\perp\mathbf{U}\mathbf{u}_i\mathbf{u}_i^H\mathbf{U}^H\mathbf{P}_\Theta^\perp\mathbf{y}_p.$$

In summary, we have shown that the random ML channel estimate can be computed as follows:

1. Find the roots of $f(\beta)$ in (4.42) on the positive real axis.
   - If there exits at least one root on the positive real axis such that $\beta > 1$ do the following: compute (4.45) and (4.46) for each root, and evaluate (4.38) accordingly. Select the root for which (4.38) is minimum.
   - Otherwise, take $\beta = 1$ and $\hat{\sigma}^2 = \hat{\sigma}_w^2 = N^{-1}\mathbf{y}_p^H \mathbf{P}_\Theta^\perp \mathbf{y}_p$.

2. Find $\hat{\mathbf{h}}_{RML}$ by inverting (4.28) with $\hat{\sigma}^2$ and $\hat{\sigma}_w^2$ fixed to the above values.

Based on above steps an algorithm as in 4.1 is formulated to estimate the random ML CIR.

4.3.3 Asymptotic Performance Characterization

In this section we evaluate the error in the channel estimation of the proposed methods. We divide this section into two subsections, one for each approach, and outline in detail the characterizations of the variances of the estimation errors of both estimators. In the first subsection, we characterize the channel estimation error of the deterministic ML estimator. In the second subsection, the variance of the channel estimation error of the random ML estimator is evaluated taking an asymptotic approach, when $N \to \infty$.

4.3.3.1 Variance of the Deterministic ML Estimator

To calculate the variance of the deterministic ML channel estimator, we recall (4.13) and replace $\mathbf{y}_p = \sqrt{N}\mathbf{F}\mathbf{E}\mathbf{\Lambda}_p\mathbf{G}^H\bar{\mathbf{h}} + \bar{\mathbf{n}}$, where $\bar{\mathbf{h}}$ denotes the true channel and $\bar{\mathbf{n}}$ denotes the true noise samples, we observe that

$$\hat{\mathbf{h}}_{DML} = \bar{\mathbf{h}} + \frac{1}{\sqrt{N}} \left(\mathbf{\Theta}^H \hat{\mathbf{P}}^\perp \mathbf{\Theta} \right)^{-1} \mathbf{\Theta}^H \hat{\mathbf{P}}^\perp \bar{\mathbf{n}}. \tag{4.47}$$

The true noise in (4.47) is assumed to follow $\bar{\mathbf{n}} \sim \mathcal{CN}\left(\mathbf{0}, \bar{\mathbf{C}}_n\right)$ distribution. Similar to $\mathbf{C}_n$, the true noise covariance, $\bar{\mathbf{C}}_n$, is defined as $\bar{\mathbf{C}}_n = \bar{\sigma}^2\bar{\mathbf{P}} + \bar{\sigma}_w^2\bar{\mathbf{P}}^\perp$, where $\bar{\sigma}^2$ is the true variance of the IN, $\bar{\mathbf{P}}$ is the diagonal matrix with ones as its entries at the locations identifying the true support of the IN samples, $\bar{\sigma}_w^2$ is the true variance of the background noise and $\bar{\mathbf{P}}^\perp = \mathbf{I}_N - \bar{\mathbf{P}}$ is the diagonal matrix with ones at its entries identifying the samples in the received signal that are not corrupted by the IN. Note that $\hat{\mathbf{P}} = \bar{\mathbf{P}}$ and $\hat{\mathbf{P}}^\perp = \bar{\mathbf{P}}^\perp$, only if the estimated support of the IN aligns with the true one.

Trivially from (4.47), we see that $\hat{\mathbf{h}}_{DML}$ is unbiased regardless of the IN behavior and $\sqrt{N}\left(\hat{\mathbf{h}}_{DML} - \bar{\mathbf{h}}\right)$ is Gaussian distributed with zero mean and

Algorithm 4.1 Algorithm for simplified random ML estimator

Input: observation $\mathbf{y}_p$, set of samples contaminated with impulse noise $\mathcal{A}$, matrix Θ.

1. Build the selection matrix $\mathbf{U}$, the projection matrix $\mathbf{P}_{\Theta}^{\perp} = \mathbf{I}_N - \Theta\left(\Theta^H\Theta\right)^{-1}\Theta^H$, and compute the eigendecomposition in (5.7).

2. Compute the solutions to the equation in (5.7) on the region $\beta \geq 1$. Let them be denoted by $\beta_1, \ldots, \beta_R$.

3. If there are no solutions to the equation in (5.7), take $\hat{\sigma}^2 = \hat{\sigma}_w^2 = N^{-1}\mathbf{y}_p^H\mathbf{P}_{\Theta}^{\perp}\mathbf{y}_p$. Otherwise, for each solution β_r, compute $\hat{\sigma}_r^2$ and $\hat{\sigma}_{w,r}^2$ as (4.45)-(4.46) with $\beta = \beta_r$, $r = 1, \ldots, R$. Then, take

$$r_{opt} = \arg\min_r \zeta_{RML}\left(\hat{\mathbf{h}}, \hat{\sigma}_r^2, \hat{\sigma}_{w,r}^2\right).$$

   If $\zeta_{GML}\left(\hat{\mathbf{h}}, \hat{\sigma}_{r_{opt}}^2, \hat{\sigma}_{w,r_{opt}}^2\right) \leq N\log\left(N^{-1}\mathbf{y}_p^H\mathbf{P}_{\Theta}^{\perp}\mathbf{y}_p\right) + N$, take

$$\hat{\sigma}^2 = \hat{\sigma}_{r_{opt}}^2 \text{ and } \hat{\sigma}_w^2 = \hat{\sigma}_{w,r_{opt}}^2.$$

   Otherwise, take $\hat{\sigma}^2 = \hat{\sigma}_w^2 = N^{-1}\mathbf{y}_p^H\mathbf{P}_{\Theta}^{\perp}\mathbf{y}_p$.

4. Compute the channel estimate as

$$\hat{\mathbf{h}}_{RML} = \frac{1}{\sqrt{N}}\left[\Theta^H\left(\frac{\hat{\mathbf{P}}}{\hat{\sigma}^2} + \frac{\hat{\mathbf{P}}^{\perp}}{\hat{\sigma}_w^2}\right)\Theta\right]^{-1}\Theta^H\left[\frac{\hat{\mathbf{P}}}{\hat{\sigma}^2} + \frac{\hat{\mathbf{P}}^{\perp}}{\hat{\sigma}_w^2}\right]\mathbf{y}_p,$$

   where the matrices $\hat{\mathbf{P}}$ and $\hat{\mathbf{P}}^{\perp}$ respectively denote the diagonal matrices with ones at the estimated locations where IN samples are present and where IN samples are not present in the received signal.

covariance

$$\mathbf{C}_{DML} = \left(\boldsymbol{\Theta}^H\hat{\mathbf{P}}^{\perp}\boldsymbol{\Theta}\right)^{-1}\boldsymbol{\Theta}^H\hat{\mathbf{P}}^{\perp}\bar{\mathbf{C}}_n\hat{\mathbf{P}}^{\perp}\boldsymbol{\Theta}\left(\boldsymbol{\Theta}^H\hat{\mathbf{P}}^{\perp}\boldsymbol{\Theta}\right)^{-1}. \tag{4.48}$$

4.3.3.2 Variance of the Random ML Estimator

To evaluate the variance of the random ML estimator, we take an asymptotic approach and characterize the behavior of the estimation error by assuming large symbol size N. In order to do so, we recall the cost function $\zeta_{RML}(\mathbf{h})$ derived in (4.17) and evaluate its gradient around the true channel $\bar{\mathbf{h}}$. The

Taylor series expansion of the gradient of the cost function can be expressed as

$$\nabla\zeta_{RML}\left(\mathbf{h}\right)=\left[\nabla\zeta_{RML}\left(\mathbf{h}\right)\right]_{\mathbf{h}=\bar{\mathbf{h}}}+\left[\mathbb{H}\zeta_{RML}\left(\mathbf{h}\right)\right]_{\mathbf{h}=\bar{\mathbf{h}}}\mathcal{R}\left(\mathbf{h}-\bar{\mathbf{h}}\right)+\xi(\mathbf{h},\bar{\mathbf{h}}),\quad(4.49)$$

where $\xi(\mathbf{h},\bar{\mathbf{h}})$ is a column vector. Evaluating (4.49) at $\mathbf{h}=\hat{\mathbf{h}}_{RML}$, assuming that the Hessian in invertible at this point and using the fact that $\nabla\zeta_{RML}\left(\hat{\mathbf{h}}\right)$ by definition, we obtain

$$\mathcal{R}\left(\hat{\mathbf{h}}_{RML}-\bar{\mathbf{h}}\right)=-\left[\mathbb{H}\zeta_{RML}\left(\mathbf{h}\right)\right]_{\mathbf{h}=\bar{\mathbf{h}}}^{-1}\left[\nabla\zeta_{RML}\left(\mathbf{h}\right)\right]_{\mathbf{h}=\bar{\mathbf{h}}}-\left[\mathbb{H}\zeta_{RML}\left(\mathbf{h}\right)\right]_{\mathbf{h}=\bar{\mathbf{h}}}^{-1}\xi(\hat{\mathbf{h}},\bar{\mathbf{h}}),\quad(4.50)$$

or equivalently,

$$\hat{\mathbf{h}}_{RML}-\bar{\mathbf{h}}=-\boldsymbol{\Xi}\left[\mathbb{H}\zeta_{RML}\left(\mathbf{h}\right)\right]_{\mathbf{h}=\bar{\mathbf{h}}}^{-1}\left[\nabla\zeta_{RML}\left(\mathbf{h}\right)\right]_{\mathbf{h}=\bar{\mathbf{h}}}-\boldsymbol{\Xi}\left[\mathbb{H}\zeta_{RML}\left(\mathbf{h}\right)\right]_{\mathbf{h}=\bar{\mathbf{h}}}^{-1}\xi(\hat{\mathbf{h}},\bar{\mathbf{h}}),\quad(4.51)$$

where we have used $\boldsymbol{\Xi}=[\mathbf{I}_L,\mathrm{j}\mathbf{I}_N]$. Owing to the consistency of the likelihood function in (4.17) (proven in the appendix), the quantity $\boldsymbol{\Xi}[\mathbb{H}\zeta_{RML}(\mathbf{h})]_{\mathbf{h}=\bar{\mathbf{h}}}^{-1}\xi(\hat{\mathbf{h}},\bar{\mathbf{h}})\rightarrow 0$ in probability at a sufficiently high rate. Hence, we evaluate the asymptotic distribution of $-\boldsymbol{\Xi}\left[\mathbb{H}\zeta_{RML}\left(\mathbf{h}\right)\right]_{\mathbf{h}=\bar{\mathbf{h}}}^{-1}\left[\nabla\zeta_{RML}\left(\mathbf{h}\right)\right]_{\mathbf{h}=\bar{\mathbf{h}}}$ to characterize large symbol size error of the random ML estimator. To do so, as a first step we define the convergence of the Hessian matrix and then proceed with the characterization of the asymptotic distribution of the cost function gradient.

Convergence of Hessian Matrix

To determine the convergence of the Hessian matrix derived in (4.21), we evaluate its entries by studying the bilinear form

$$\mathcal{R}^{T}\left(\mathbf{a}\right)\mathbb{H}\zeta_{RML}\left(\mathbf{h}\right)\mathcal{R}\left(\mathbf{b}\right)=2\,\mathrm{Re}\left[\mathbf{a}^{H}\mathbb{H}_{c}\left(\mathbf{h}\right)\mathbf{b}+\mathbf{a}^{H}\bar{\mathbb{H}}_{c}\left(\mathbf{h}\right)\mathbf{b}^{*}\right],$$

where $\mathbf{a}$ and $\mathbf{b}$ are vectors of appropriate dimensions. At true channel and applying the weak law of large numbers, we can see that [149]

$$\begin{aligned}\mathbf{a}^{H}\mathbb{H}_{c}\left(\bar{\mathbf{h}}\right)\mathbf{b}-\left(\mathbf{a}^{H}\boldsymbol{\Theta}^{H}\left(\frac{1}{\widetilde{\sigma}^{2}}\hat{\mathbf{P}}+\frac{1}{\widetilde{\sigma}_{w}^{2}}\hat{\mathbf{P}}^{\perp}\right)\boldsymbol{\Theta}\mathbf{b}\right)&\rightarrow 0\\ \mathbf{a}^{H}\bar{\mathbb{H}}_{c}\left(\bar{\mathbf{h}}\right)\mathbf{b}&\rightarrow 0\end{aligned}\quad(4.52)$$

with probability one, where $\widetilde{\sigma}^{2}=\frac{1}{\hat{N}_{imp}}\mathrm{tr}\left[\hat{\mathbf{P}}\bar{\mathbf{C}}_{n}\right]$ and $\widetilde{\sigma}_{w}^{2}=\frac{1}{N-\hat{N}_{imp}}\mathrm{tr}\left[\hat{\mathbf{P}}^{\perp}\bar{\mathbf{C}}_{n}\right]$.

Exploiting (4.52) and (4.20) we can conclude that

$$\mathbb{H}\zeta_{RML}\left(\bar{\mathbf{h}}\right)-2\begin{bmatrix}\mathrm{Re}\left[\boldsymbol{\Theta}^{H}\widetilde{\mathbf{C}}_{n}^{-1}\boldsymbol{\Theta}\right] & -\mathrm{Im}\left[\boldsymbol{\Theta}^{H}\widetilde{\mathbf{C}}_{n}^{-1}\boldsymbol{\Theta}\right]\\ \mathrm{Im}\left[\boldsymbol{\Theta}^{H}\widetilde{\mathbf{C}}_{n}^{-1}\boldsymbol{\Theta}\right] & \mathrm{Re}\left[\boldsymbol{\Theta}^{H}\widetilde{\mathbf{C}}_{n}^{-1}\boldsymbol{\Theta}\right]\end{bmatrix}\rightarrow\mathbf{0},\quad(4.53)$$

where $\widetilde{\mathbf{C}}_{n}^{-1}=\frac{1}{\widetilde{\sigma}^{2}}\hat{\mathbf{P}}+\frac{1}{\widetilde{\sigma}_{w}^{2}}\hat{\mathbf{P}}^{\perp}$, almost surely.

Asymptotic Distribution of the Gradient of Cost Function

Similar to the Hessian, for the evaluation of the asymptotic distribution of the cost function gradient, we exploit the identity that establishes

$$\mathcal{R}^T(\mathbf{a})\,\nabla\zeta_{RML}(\mathbf{h}) = 2\,\mathrm{Re}\left[\mathbf{a}^H \frac{\partial\zeta_{RML}(\mathbf{h})}{\partial\mathbf{h}^*}\right], \tag{4.54}$$

for a complex vector $\mathbf{a}$ of appropriate dimension. Furthermore, from (4.22), upon applying strong law of large numbers [149]

$$\left.\frac{\bar{\mathbf{y}}^H\hat{\mathbf{P}}\bar{\mathbf{y}}}{\hat{N}_{imp}}\right|_{\mathbf{h}=\bar{\mathbf{h}}} \to \widetilde{\sigma}^2 \quad \text{and} \quad \left.\frac{\bar{\mathbf{y}}^H\hat{\mathbf{P}}^{\perp}\bar{\mathbf{y}}}{N-\hat{N}_{imp}}\right|_{\mathbf{h}=\bar{\mathbf{h}}} \to \widetilde{\sigma}_w^2$$

almost surely. This further implies that

$$\sqrt{N}\mathbf{a}^H\left.\frac{\partial\zeta_{RML}(\mathbf{h})}{\partial\mathbf{h}^*}\right|_{\mathbf{h}=\bar{\mathbf{h}}} = -\frac{1}{\widetilde{\sigma}^2}\mathbf{a}^H\boldsymbol{\Theta}^H\hat{\mathbf{P}}\bar{\mathbf{y}} - \frac{1}{\widetilde{\sigma}_w^2}\mathbf{a}^H\boldsymbol{\Theta}^H\hat{\mathbf{P}}^{\perp}\bar{\mathbf{y}} + o_p(1). \tag{4.55}$$

Asymptotic Variance of Random ML Estimator

Plugging the values of the Hessian and the gradient of the cost function evaluated at $\mathbf{h} = \bar{\mathbf{h}}$ into (4.51), the channel estimation error of the random ML estimator can hence be expressed as

$$\sqrt{N}\left(\hat{\mathbf{h}}_{RML} - \bar{\mathbf{h}}\right) = \left(\boldsymbol{\Theta}^H\widetilde{\mathbf{C}}_n^{-1}\boldsymbol{\Theta}\right)^{-1}\boldsymbol{\Theta}^H\widetilde{\mathbf{C}}_n^{-1}\bar{\mathbf{n}}. \tag{4.56}$$

Furthermore, the estimation error is asymptotically equivalent in law to a Gaussian random variable with zero mean and covariance being equal to

$$\mathbf{C}_{RML} = \left(\boldsymbol{\Theta}^H\widetilde{\mathbf{C}}_n^{-1}\boldsymbol{\Theta}\right)^{-1}\boldsymbol{\Theta}^H\widetilde{\mathbf{C}}_n^{-1}\bar{\mathbf{C}}_n\widetilde{\mathbf{C}}_n^{-1}\boldsymbol{\Theta}\left(\boldsymbol{\Theta}^H\widetilde{\mathbf{C}}_n^{-1}\boldsymbol{\Theta}\right)^{-1}. \tag{4.57}$$

4.3.4 Performance Assessment

In this section we analyze and compare the performance of both estimators in varied scenarios by exploiting the derived covariances in (4.48) and (4.57). As evident, both estimators rely on estimated support of IN to perform channel estimation. Hence, to assess the performance of proposed estimators we define two different scenarios. To begin with, we first consider an ideal scenario where the support of the IN is perfectly estimated and characterize the variance of both channel estimators. In the second assessment, we consider scenarios where the support of the IN is estimated with error and determine the variance of the channel estimation error for the proposed channel estimators.

4.3.4.1 Correct Estimation of the IN Support

In an ideal situation, when the estimated support and the true support of the IN match, then $\bar{\mathbf{P}} = \hat{\mathbf{P}}$ and $\bar{\mathbf{P}}^{\perp} = \hat{\mathbf{P}}^{\perp}$. In such a situation $\widetilde{\sigma}^2 = \bar{\sigma}^2$, $\widetilde{\sigma}_w^2 = \bar{\sigma}_w^2$ and $\widetilde{\mathbf{C}}_n = \bar{\mathbf{C}}_n = \hat{\mathbf{C}}_n$.

Hence, the variances of the channel estimation errors of the two estimators when the support of the IN is accurately estimated, can be expressed as:

$$\mathbf{C}_{DML} = \bar{\sigma}_w^2 \left(\boldsymbol{\Theta}^H \bar{\mathbf{P}}^{\perp} \boldsymbol{\Theta}\right)^{-1} \tag{4.58}$$

and

$$\mathbf{C}_{RML} = \left(\boldsymbol{\Theta}^H \bar{\mathbf{C}}_n^{-1} \boldsymbol{\Theta}\right)^{-1}. \tag{4.59}$$

Recalling that $\bar{\mathbf{C}}_n^{-1} = \bar{\sigma}_w^{-2}\bar{\mathbf{P}}^{\perp} + \bar{\sigma}^{-2}\bar{\mathbf{P}}$, we see that $\bar{\mathbf{C}}_n^{-1} \geq \bar{\sigma}_w^{-2}\bar{\mathbf{P}}^{\perp}$, implying

$$\mathbf{C}_{RML} \leq \mathbf{C}_{DML}. \tag{4.60}$$

From (4.60) we can conclude that, provided that the support of the IN samples are correctly estimated, the random ML estimator asymptotically outperforms the performance of the deterministic ML estimator, as its asymptotic variance is lower. The superior performance of the random ML estimator is reasonable as it exploits all samples in the received signal to determine the CIR coefficients, whereas the deterministic ML estimator only uses those samples that are not contaminated by the IN. However, when the IN power is very high than the background noise power, then $\bar{\mathbf{C}}_n^{-1} \approx \bar{\sigma}_w^{-2}\bar{\mathbf{P}}^{\perp}$ and, hence, the variances of both estimators converge to approximately similar values such that $\mathbf{C}_{RML} \approx \mathbf{C}_{DML}$.

Furthermore,

$$\mathbf{C}_{DML} = \bar{\sigma}_w^2(\boldsymbol{\Theta}^H \bar{\mathbf{P}}^{\perp} \boldsymbol{\Theta})^{-1} = \mathbf{CRB}_{DML}$$

and

$$\mathbf{C}_{RML} = (\boldsymbol{\Theta}^H \bar{\mathbf{C}}_n^{-1} \boldsymbol{\Theta})^{-1} = \mathbf{CRB}_{RML}$$

are the CRBs of the deterministic and the random ML estimators respectively. *(Refer to the Appendix section for the proof that inverse of* $(\boldsymbol{\Theta}^H \bar{\mathbf{P}}^{\perp} \boldsymbol{\Theta})$ *exists.)*

4.3.4.2 Incorrect Estimation of the IN Support

In practice, it might happen that the support of the IN is estimated with error. In such a situation $\hat{\mathbf{P}} \neq \bar{\mathbf{P}}$ and $\hat{\mathbf{P}}^{\perp} \neq \bar{\mathbf{P}}^{\perp}$ further implying that $\hat{\mathbf{C}}_n \neq \bar{\mathbf{C}}_n$. In a realistic scenario, a mismatch between the estimated and the true noise

covariance matrices typically takes place when the support of the IN samples is either underestimated or overestimated.

In order to investigate the performance of both estimators in underestimated and overestimated scenarios, first we consider a situation when the IN support is underestimated and formulate the estimation error variances for both approaches. Following the underestimated support scenario, we then characterize the channel estimation errors of both estimators considering a situation when the supports of the IN are overestimated.

Underestimated IN Support

In this scenario we assume that, among all the IN affected samples, only a subset is identified. The estimated support of the IN is therefore a smaller subset of all the truly contaminated samples. Under this assumption, we see that $\bar{N}_{imp} > \hat{N}_{imp}$, where $\bar{N}_{imp}$ is the true number of IN samples, and $\bar{\mathbf{P}} - \hat{\mathbf{P}} > 0$, such that $\hat{\mathbf{P}}\bar{\mathbf{P}} = \hat{\mathbf{P}}$, $\hat{\mathbf{P}}\bar{\mathbf{P}}^{\perp} = 0$, $\widetilde{\sigma}^2 = \bar{\sigma}^2$ and

$$\widetilde{\sigma}_w^2 = \bar{\sigma}^2 \frac{\bar{N}_{imp} - \hat{N}_{imp}}{N - \hat{N}_{imp}} + \bar{\sigma}_w^2 \frac{N - \bar{N}_{imp}}{N - \hat{N}_{imp}}.$$

The variance of the deterministic ML estimator, under above assumptions, can hence be expressed as:

$$\begin{aligned} \mathbf{C}_{DML} &= \bar{\sigma}_w^2 \left(\mathbf{\Theta}^H \hat{\mathbf{P}}^{\perp} \mathbf{\Theta} \right)^{-1} \\ &+ \left(\bar{\sigma}^2 - \bar{\sigma}_w^2 \right) \left(\mathbf{\Theta}^H \hat{\mathbf{P}}^{\perp} \mathbf{\Theta} \right)^{-1} \mathbf{\Theta}^H \left(\bar{\mathbf{P}} - \hat{\mathbf{P}} \right) \mathbf{\Theta} \left(\mathbf{\Theta}^H \hat{\mathbf{P}}^{\perp} \mathbf{\Theta} \right)^{-1}. \end{aligned} \tag{4.61}$$

Similarly, the asymptotic variance of the random ML estimator takes the form:

$$\begin{aligned} \mathbf{C}_{RML} = \left(\mathbf{\Theta}^H \widetilde{\mathbf{C}}_n^{-1} \mathbf{\Theta} \right)^{-1} \mathbf{\Theta}^H \times \\ \left(\frac{1}{\widetilde{\sigma}^2} \hat{\mathbf{P}} + \frac{\bar{\sigma}^2}{\widetilde{\sigma}_w^4} (\bar{\mathbf{P}} - \hat{\mathbf{P}}) + \frac{\bar{\sigma}_w^2}{\widetilde{\sigma}_w^4} \bar{\mathbf{P}}^{\perp} \right) \times \\ \mathbf{\Theta} \left(\mathbf{\Theta}^H \widetilde{\mathbf{C}}_n^{-1} \mathbf{\Theta} \right)^{-1}, \end{aligned} \tag{4.62}$$

where we have used the fact that $\widetilde{\mathbf{C}}_n^{-1} \bar{\mathbf{C}}_n \widetilde{\mathbf{C}}_n^{-1}$ in (4.57) is equivalent to $(\frac{1}{\bar{\sigma}^2} \hat{\mathbf{P}} + \frac{\bar{\sigma}^2}{\widetilde{\sigma}_w^4} (\bar{\mathbf{P}} - \hat{\mathbf{P}}) + \frac{\bar{\sigma}_w^2}{\widetilde{\sigma}_w^4} \bar{\mathbf{P}}^{\perp})$ when the support of the IN is underestimated.

Overestimated IN Support

In this scenario, we assume that the estimated set of IN contaminated samples is larger than the true one and also that all the truly contaminated samples are retained in the estimated support set.

Under these two assumptions, we see that $\hat{N}_{imp} > \bar{N}_{imp}$, $\hat{\mathbf{P}} = \bar{\mathbf{P}} + \mathbf{V}\mathbf{V}^H$ for some selection matrix $\mathbf{V}$ with dimensions $N \times (\hat{N}_{imp} - \bar{N}_{imp})$, resulting into $\bar{\mathbf{P}}\mathbf{V} = \mathbf{0}$ (such that $\bar{\mathbf{P}}^{\perp}\mathbf{V} = \mathbf{V}$) implying $\hat{\mathbf{P}}\bar{\mathbf{P}} = \bar{\mathbf{P}}$, $\hat{\mathbf{P}}^{\perp}\bar{\mathbf{P}} = \mathbf{0}$, $\widetilde{\sigma}_w^2 = \bar{\sigma}_w^2$ and

$$\widetilde{\sigma}^2 = \bar{\sigma}^2 \frac{\bar{N}_{imp}}{\hat{N}_{imp}} + \bar{\sigma}_w^2 \frac{\hat{N}_{imp} - \bar{N}_{imp}}{\hat{N}_{imp}}.$$

Upon evaluating the variance of the deterministic ML estimator, the resulting variance can be readily expressed as:

$$\mathbf{C}_{DML} = \bar{\sigma}_w^2 \left(\mathbf{\Theta}^H \hat{\mathbf{P}}^{\perp} \mathbf{\Theta} \right)^{-1}. \tag{4.63}$$

Similarly, the asymptotic variance of the random ML estimator can be expressed as:

$$\begin{aligned} \mathbf{C}_{RML} = & \left(\mathbf{\Theta}^H \widetilde{\mathbf{C}}_n^{-1} \mathbf{\Theta} \right)^{-1} \mathbf{\Theta}^H \times \\ & \left(\frac{\bar{\sigma}^2}{\widetilde{\sigma}^4} \bar{\mathbf{P}} + \frac{\bar{\sigma}_w^2}{\widetilde{\sigma}^4} (\hat{\mathbf{P}} - \bar{\mathbf{P}}) + \frac{1}{\widetilde{\sigma}_w^2} \hat{\mathbf{P}}^{\perp} \right) \times \\ & \mathbf{\Theta} \left(\mathbf{\Theta}^H \widetilde{\mathbf{C}}_n^{-1} \mathbf{\Theta} \right)^{-1}, \end{aligned} \tag{4.64}$$

where we have used the fact that taking into account when the IN support is over estimated, $\widetilde{\mathbf{C}}_n^{-1} \bar{\mathbf{C}}_n \widetilde{\mathbf{C}}_n^{-1}$ in (4.57) is equivalent to $(\frac{\bar{\sigma}^2}{\widetilde{\sigma}^4} \bar{\mathbf{P}} + \frac{\bar{\sigma}_w^2}{\widetilde{\sigma}^4} (\hat{\mathbf{P}} - \bar{\mathbf{P}}) + \frac{1}{\widetilde{\sigma}_w^2} \hat{\mathbf{P}}^{\perp})$.

Observation 2:

After deriving the variances of the proposed estimators by segregating the scenarios when the estimated noise covariance does not match the true one, it is legitimate to say that the derived expressions for the estimation error variances allow us to study individual scenarios on a case by case basis. Therefore, in order to validate the performances of the proposed estimators in IN support underestimated and overestimated scenarios, we resort to channel estimation simulations for both estimators and validate their performances numerically.

4.3.5 Numerical Validation

In this section, we numerically evaluate the performance of the proposed channel estimators in typical PLC scenarios. For the purpose of simulation, we consider an OFDM based PLC system with $N = 256$ subcarriers. Among the 256 subcarriers, only $N_u = 252$ subcarriers are used for data transmission. Similar to the simulation setup that we defined in section 4.2.3, we adopt

the frequency selective multipath channel defined in (3.40). The values of all the channel parameters are adopted as proposed in the standard [39]. Furthermore, the occurrence of IN is defined according to the Bernoulli-Gaussian noise model as defined in (3.2).

To assess the performance of random ML and deterministic ML estimators, we resort to different PLC scenarios. Among the simulated scenarios, Fig. 4.6 shows the comparison of the performance of the proposed estimators when INR is fixed to 10 dB and SNR is varied from 0 dB to 30 dB. Fig. 4.7 shows the performance of proposed estimators when SNR is fixed to 20 dB and INR is varied from 10 dB to 40 dB. Fig. 4.8 reports the evaluation of the channel estimation errors of the proposed estimators when the mistakes are made in IN support estimation. Finally, Fig. 4.9 shows evaluation of the computational efficiency of the proposed estimators.

In an ideal situation when the estimated IN support matches the true one, the channel estimator that exploits the estimated noise treating it as a random quantity outperforms the performance of the deterministic ML estimator. The superior performance of the random ML estimator holds true at different SNR values and provides significant gain in terms of low mean-squared error (MSE) for the channel estimation. The better performance of the random ML estimator comes from the fact that it exploits all the samples in the received signal to approximate the coefficients of CIR. In contrast, the deterministic ML estimator, which exploits only the non-contaminated samples in the received signal, results in a higher variance than the random approach. Nonetheless, both estimators, under accurate IN support estimation, fairly reach their corresponding CRBs as shown in Fig. 4.6. However, when the IN bears very high power in comparison to the background noise, the performance of both estimators tends to converge to a saturated variance as reported in Fig. 4.7.

Referring to the scenarios defined in section 4.3.4.2, numerical validation of the performance of both approaches is done considering situations when the support of the IN is estimated with error. Regardless of whether the support of the IN is underestimated and overestimated, the random ML estimator is found to outperform the deterministic ML channel estimator, as shown in Fig. 4.8. In a situation where the support of the IN is underestimated, the channel estimator following the deterministic approach cannot distinguish all the IN corrupted samples. As a consequence, the performance of the deterministic ML estimator degrades resulting into having higher channel estimation error variance as compared to the random ML estimator. Similarly when the

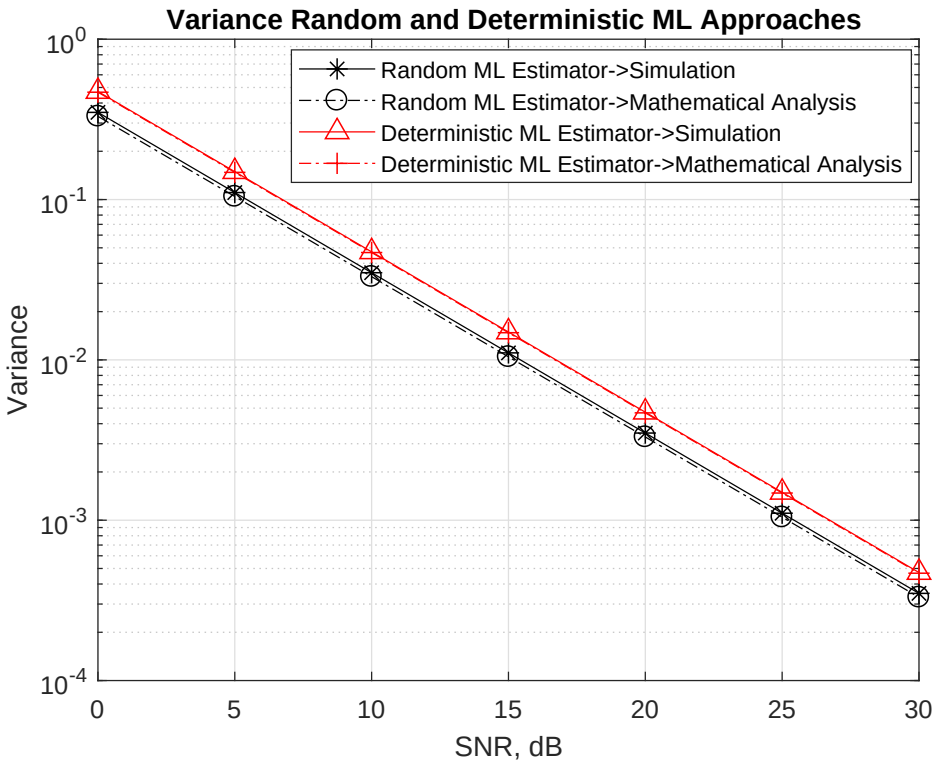


Figure 4.6: Performance of both estimators when impulsive to background noise power ratio is fixed to 10 dB and signal to noise power ratio is varied from 0 dB to 30 dB.

IN samples supports are over estimated, the number of samples available for channel estimation to the deterministic ML estimator is reduced and hence higher variance is perceived by the estimator. In contrast, the random ML estimator still has the opportunity to eliminate all the IN contaminated samples and approximate the impulse response of the channel, resulting into lower variance than the deterministic ML channel estimator. Nevertheless, both estimators reach their respective CRBs when the estimated support and true support coincide. These effects are well demonstrated in the simulation result reported in Fig. 4.8. The x-axis in Fig. 4.8 is scaled to denote subsets of true IN support set when they are underestimated *(extreme at the left)* and ranging up to the situation where the noise support is overestimated *(extreme at the right)*. The mid-point, denoted by covariance matching point in the figure, defines the situation when the estimated support and the true support of IN sample coincide.

The conventional random ML estimator derived in section 4.3.2.2 performs

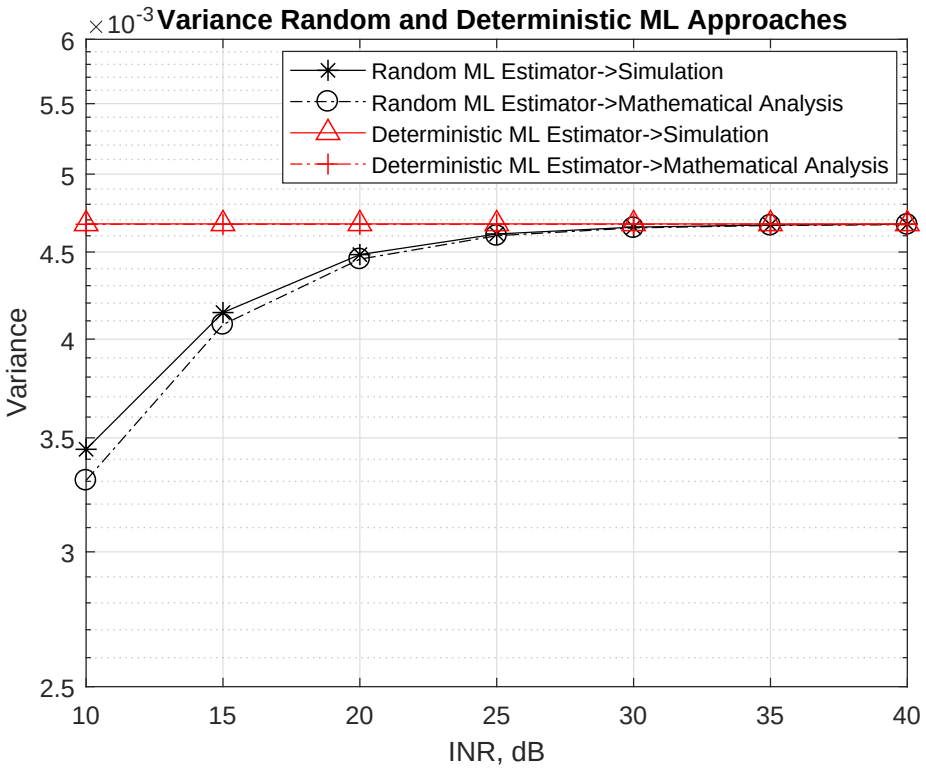


Figure 4.7: Performance of both estimators when the signal to background noise power ratio is fixed to 20 dB and the IN to background noise power is varied from 10 dB to 40 dB.

iterative-approximation of the channel estimate, where in each iteration the Hessian matrix and the gradient vector needs to be calculated along with the matrix inversion and matrix multiplication operations. By exploiting the simplified estimator derived in 4.3.2.3, the computational burden of the random ML estimator can be significantly reduced. As shown in Fig. 4.9, where we evaluate the computational time taken by random ML, simplified random ML and the deterministic ML estimators to estimate the channel, the deterministic ML and simplified random ML estimators are computationally inexpensive and are timely consistent on estimating the channel on varied INR scenarios (fixed SNR). On the other hand the conventional random ML estimator shows efficient channel estimation capability in high INR regions and computational inefficiency in low INR regions.

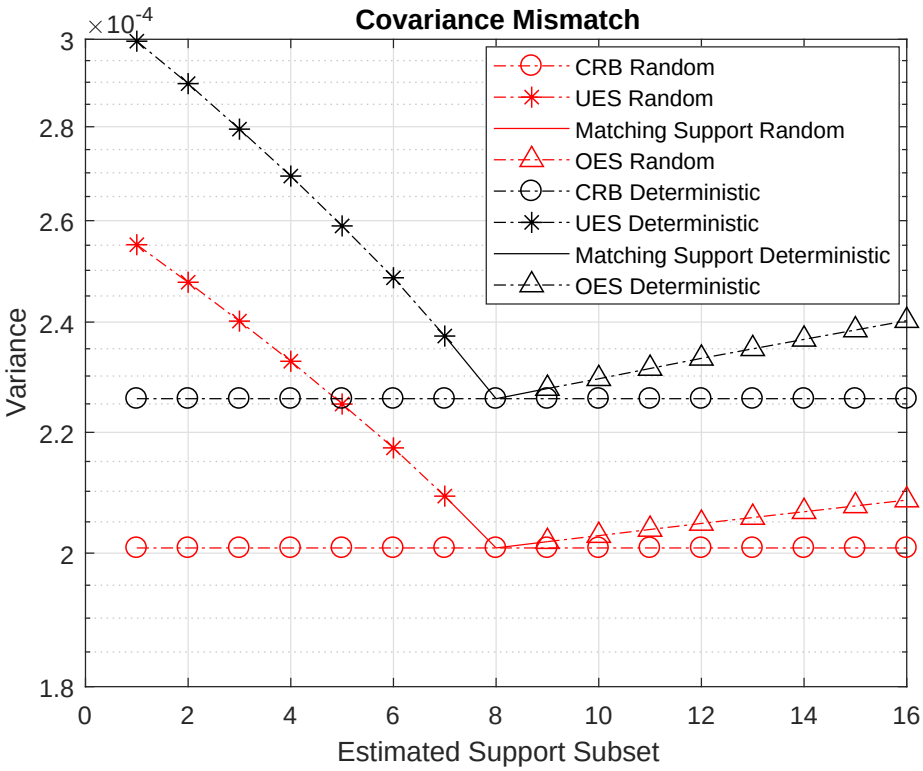


Figure 4.8: Performance of proposed estimators when the impulsive to background noise power ratio is fixed to 10 dB and the signal to background noise power is fixed to 20 dB under noise covariance mismatch scenarios. UES and OES in the legend respectively denote IN support underestimated and overestimated scenarios.

4.4 Conclusion

In this chapter we proposed robust channel estimators for PLC systems. The proposed channel estimators exploit the estimated IN samples, rather than suppressing them, while estimating the channel.

The channel estimation error of both estimators was evaluated and analytically characterized. It was observed that irrespective of the behavior of IN both estimators perform consistent channel estimation and outperform the performance of the classical channel estimators for the PLC systems. Both estimators reach their respective Cramer-Rao bounds when the support of IN is estimated accurately. Moreover, the random ML estimator is found to outperform the performance of deterministic ML estimator in all typical PLC scenarios.

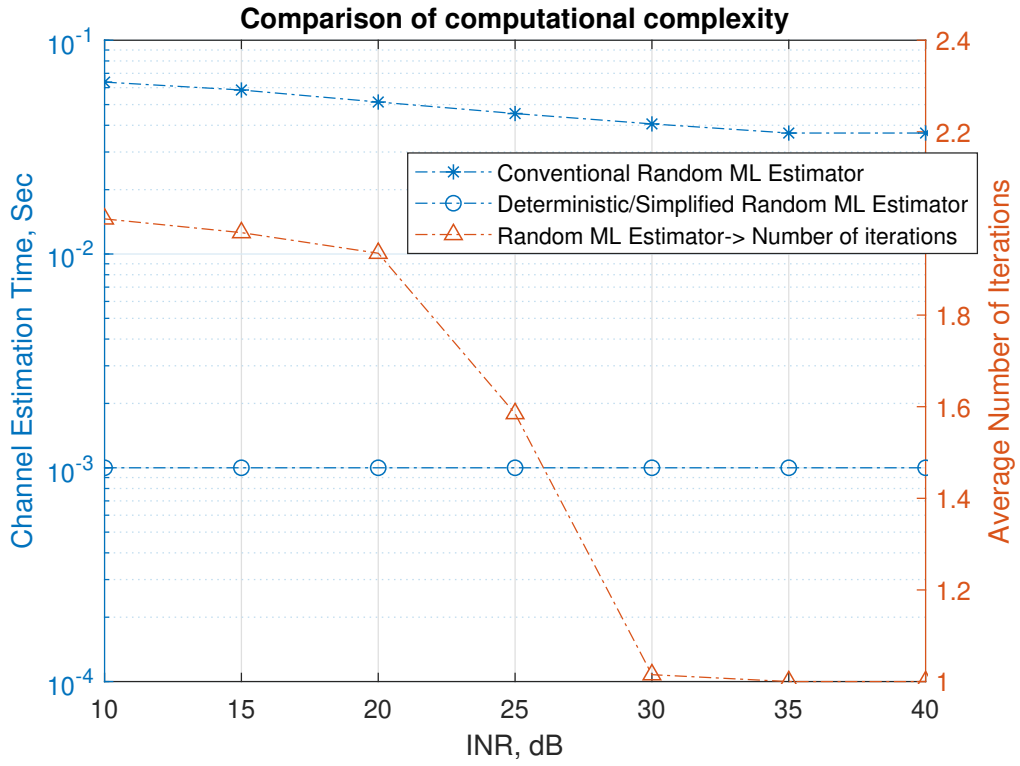


Figure 4.9: Computational efficiency in terms of time taken by the proposed estimators to estimate the channel in scenarios where the signal to background noise power ratio is fixed at 20 dB and the impulsive to background noise power ratio is varied from 10 dB to 40 dB.

4.5 Appendix: Consistency of the likelihood function in (4.17) for the random approach based estimator

In order to check the consistency of the random ML estimator, we assume that the symbol size N grows to infinity and the number of subcarriers bearing pilot symbols N_u along with the number of IN samples N_{imp} are a function of N such that

$$0 < \liminf_N \frac{N_{imp}}{N} \leq \limsup_N \frac{N_{imp}}{N} < +\infty,$$
$$0 < \liminf_N \frac{N_u}{N} \leq \limsup_N \frac{N_u}{N} < 1.$$

Furthermore, when N grows large, we assume that the diagonal values of $\mathbf{\Lambda}_p^H \mathbf{\Lambda}_p$ are contained in a fixed interval of the positive real axis.

Recalling from (4.17),

$$\zeta_{RML}(\mathbf{h}) = \frac{N_{imp}}{N} \log\left(\frac{1}{N_{imp}}\left(\mathbf{y}_p - \sqrt{N}\mathbf{\Theta}\mathbf{h}\right)^H \mathbf{P}\left(\mathbf{y}_p - \sqrt{N}\mathbf{\Theta}\mathbf{h}\right)\right)$$
$$+ \frac{N - N_{imp}}{N} \log\left(\frac{1}{N - N_{imp}}\left(\mathbf{y}_p - \sqrt{N}\mathbf{\Theta}\mathbf{h}\right)^H \mathbf{P}^{\perp}\left(\mathbf{y}_p - \sqrt{N}\mathbf{\Theta}\mathbf{h}\right)\right),$$

or, equivalently,

$$\zeta_{RML}(\mathbf{h}) = \frac{N_{imp}}{N} \log\left[\frac{N}{N_{imp}}\left(\mathbf{\Theta}\left(\bar{\mathbf{h}} - \mathbf{h}\right) + \frac{\mathbf{n}}{\sqrt{N}}\right)^H \times \right.$$
$$\left. \mathbf{P}\left(\mathbf{\Theta}\left(\bar{\mathbf{h}} - \mathbf{h}\right) + \frac{\mathbf{n}}{\sqrt{N}}\right)\right]$$
$$+ \frac{N - N_{imp}}{N} \log\left[\frac{N}{N - N_{imp}}\left(\mathbf{\Theta}\left(\bar{\mathbf{h}} - \mathbf{h}\right) + \frac{\mathbf{n}}{\sqrt{N}}\right)^H \times\right.$$
$$\left. \mathbf{P}^{\perp}\left(\mathbf{\Theta}\left(\bar{\mathbf{h}} - \mathbf{h}\right) + \frac{\mathbf{n}}{\sqrt{N}}\right)\right].$$

In order to prove point wise consistency of the random approach based cost function, we establish the following proposition.

Proposition 1. *Under above conditions, we have that*

$$\zeta_{RML}(\mathbf{h}) - \bar{\zeta}_{RML}(\mathbf{h}) \to 0$$

almost surely, where

$$\bar{\zeta}_{RML}(\mathbf{h}) = \frac{N_{imp}}{N} \log\left(\frac{N}{N_{imp}}\left(\bar{\mathbf{h}} - \mathbf{h}\right)^H \mathbf{\Theta}^H \mathbf{P} \mathbf{\Theta}\left(\bar{\mathbf{h}} - \mathbf{h}\right) + \widetilde{\sigma}^2\right)$$
$$+ \frac{N - N_{imp}}{N} \log\left(\frac{N}{N - N_{imp}}\left(\bar{\mathbf{h}} - \mathbf{h}\right)^H \mathbf{\Theta}^H \mathbf{P}^{\perp} \mathbf{\Theta}\left(\bar{\mathbf{h}} - \mathbf{h}\right) + \widetilde{\sigma}_w^2\right),$$

taking into account

$$\widetilde{\sigma}^2 = \frac{1}{N_{imp}} \mathrm{tr}\left[\mathbf{P}\bar{\mathbf{C}}_n\right] \quad \textit{and} \quad \widetilde{\sigma}_w^2 = \frac{1}{N - N_{imp}} \mathrm{tr}\left[\mathbf{P}^{\perp}\bar{\mathbf{C}}_n\right].$$

Before sketching a proof of the above result, it is interesting to note that the random method appears to be consistent even if the true noise covariance $\bar{\mathbf{C}}_n$ is completely different from the the one assumed in the model (note that this is a property that is shared with the deterministic ML approach).

Proof. Observe that we can write

$$\zeta_{RML}(\mathbf{h}) - \bar{\zeta}_{RML}(\mathbf{h}) = \frac{N_{imp}}{N} \log\left(\frac{\left(\mathbf{m} + \frac{\mathbf{n}}{\sqrt{N}}\right)^H \mathbf{P}\left(\mathbf{m} + \frac{\mathbf{n}}{\sqrt{N}}\right)}{\mathbf{m}^H\mathbf{P}\mathbf{m} + \frac{N_{imp}}{N}\widetilde{\sigma}^2}\right)$$
$$+ \frac{N - N_{imp}}{N} \log\left(\frac{\left(\mathbf{m} + \frac{\mathbf{n}}{\sqrt{N}}\right)^H \mathbf{P}^{\perp}\left(\mathbf{m} + \frac{\mathbf{n}}{\sqrt{N}}\right)}{\mathbf{m}^H\mathbf{P}^{\perp}\mathbf{m} + \frac{N-N_{imp}}{N}\widetilde{\sigma}_w^2}\right).$$

where we have denoted $\mathbf{m} = \mathbf{\Theta}\left(\bar{\mathbf{h}} - \mathbf{h}\right)$ for simplicity.

Applying the strong law of large numbers, it is trivial to observe that the two arguments of the logarithms converge to one, such that the preposition is verified. ∎

After establishing the point wise convergence in probability of the cost function, we now assess the consistency of the estimate of the channel impulse response. This does not follow from point wise convergence and must be extended to uniform convergence over a certain compact subset.

Let us fix a positive constant $C > 0$ large enough, and denote by $\mathcal{K} \subset \mathbb{C}^L$ the compact subset

$$\mathcal{K} = \left\{\mathbf{h} \in \mathbb{C}^L : \limsup_N \left\|\bar{\mathbf{h}} - \mathbf{h}\right\|^2 \leq C\right\}.$$

The following proposition generalizes the consistency of the random ML cost function under the above assumptions, uniformly in $\mathcal{K}$.

Proposition 2. *Under the above assumptions,*

$$\sup_{\mathbf{h}\in\mathcal{K}} \left|\zeta_{RML}(\mathbf{h}) - \bar{\zeta}_{RML}(\mathbf{h})\right| \to 0$$

almost surely.

Proof. Using the fact that $\log(1+x) < x$ for $x > -1$, the triangular inequality and the fact that $\mathbf{m}^H\mathbf{P}\mathbf{m} \geq \mathbf{0}$, $\mathbf{m}^H\mathbf{P}^\perp\mathbf{m} \geq \mathbf{0}$, we see that

$$\left|\zeta_{RML}(\mathbf{h}) - \bar{\zeta}_{RML}(\mathbf{h})\right| \leq \frac{|\epsilon(\mathbf{m})|}{\widetilde{\sigma}^2} + \frac{\left|\epsilon^\perp(\mathbf{m})\right|}{\widetilde{\sigma}_w^2},$$

where

$$\epsilon(\mathbf{m}) = \left(\mathbf{m} + \frac{\mathbf{n}}{\sqrt{N}}\right)^H \mathbf{P}\left(\mathbf{m} + \frac{\mathbf{n}}{\sqrt{N}}\right) - \left(\mathbf{m}^H\mathbf{P}\mathbf{m} + \frac{N_{imp}}{N}\widetilde{\sigma}^2\right)$$

and where $\epsilon^\perp(\mathbf{m})$ is equivalently defined by replacing $\mathbf{P}$ with $\mathbf{P}^\perp$. Note that

$$\widetilde{\sigma}^2 = \frac{1}{N_{imp}}\mathrm{tr}\left[\mathbf{P}\bar{\mathbf{C}}_n\right] \geq \lambda_{\min}\left(\bar{\mathbf{C}}_n\right) > 0,$$

which implies that $\inf_N \widetilde{\sigma}^2 > 0$ because the eigenvalues of $\bar{\mathbf{C}}_n$ are bounded away from zero. A similar conclusion holds true for $\widetilde{\sigma}_w^2$. We therefore only need to show that $\sup_{\mathbf{h}\in\mathcal{K}}|\epsilon(\mathbf{m})| \to 0$ and $\sup_{\mathbf{h}\in\mathcal{K}}\left|\epsilon^\perp(\mathbf{m})\right| \to 0$ almost surely. Let us prove this for $\epsilon(\mathbf{m})$, the proof for $\epsilon^\perp(\mathbf{m})$ following the same arguments. Note that, by the triangular inequality,

$$|\epsilon(\mathbf{m})| \leq 2\left|\frac{\mathbf{m}^H\mathbf{P}\mathbf{n}}{\sqrt{N}}\right| + \left|\frac{\mathbf{n}^H\mathbf{P}\mathbf{n}}{N} - \frac{N_{imp}}{N}\widetilde{\sigma}^2\right|.$$

The second term does not depend on $\mathbf{m}$ and converges almost surely to zero by the strong law of the large numbers. It remains to show that the first term above, which will be denoted by $\epsilon_1(\mathbf{m})$ converges almost surely to zero. To see this, consider a network of vectors $\mathcal{H} = \{\mathbf{h}_k \in \mathcal{K}, k \in \mathcal{J}\}$ such that

$$\sup_{\mathbf{h}\in\mathcal{K}} \min_{k\in\mathcal{J}} \|\mathbf{h} - \mathbf{h}_k\| < \frac{1}{N}.$$

This can clearly be achieved for $|\mathcal{J}| = cN^{2L}$ with $c > 0$ being a positive constant independent of N. For a given $\mathbf{h}$, let k be an index (that depends on $\mathbf{h}$) such that $\|\mathbf{h} - \mathbf{h}_k\| < N^{-1}$, and denote $\mathbf{m}_k = \boldsymbol{\Theta}\left(\bar{\mathbf{h}} - \mathbf{h}_k\right)$. We can decompose

$$\epsilon_1(\mathbf{m}) = \epsilon_1(\mathbf{m}) - \epsilon_1(\mathbf{m}_k) + \epsilon_1(\mathbf{m}_k)$$

and observe that, by Cauchy-Schwarz,

$$\begin{aligned}
|\epsilon_1(\mathbf{m}) - \epsilon_1(\mathbf{m}_k)|^2 &= \left|\frac{(\mathbf{m} - \mathbf{m}_k)^H \mathbf{P}\mathbf{n}}{\sqrt{N}}\right|^2 \\
&\leq \|\mathbf{m} - \mathbf{m}_k\|^2 \frac{\mathbf{n}^H\mathbf{P}\mathbf{n}}{N} \\
&\leq \lambda_{\max}\left(\boldsymbol{\Theta}^H\boldsymbol{\Theta}\right)\|\mathbf{h} - \mathbf{h}_k\|^2 \frac{\mathbf{n}^H\mathbf{P}\mathbf{n}}{N} \\
&\leq \frac{\lambda_{\max}\left(\boldsymbol{\Theta}^H\boldsymbol{\Theta}\right)}{N^2}\frac{\mathbf{n}^H\mathbf{P}\mathbf{n}}{N},
\end{aligned}$$

from where it follows that $\sup_{\mathbf{h}\in\mathcal{K}} |\epsilon_1(\mathbf{m}) - \epsilon_1(\mathbf{m}_k)| \to 0$ almost surely by the Borell-Cantelli lemma (note that $\sup_N \lambda_{\max}\left(\mathbf{\Theta}^H\mathbf{\Theta}\right) < \infty$ by assumption). On the other hand, we can also see that for any $\epsilon > 0$ and $r > 0$, we have

$$\begin{aligned}
\mathbb{P}\left(\sup_{k\in\mathcal{J}} |\epsilon_1(\mathbf{m}_k)| > \epsilon\right) &\leq cN^{2L} \sup_{k\in\mathcal{J}} \mathbb{P}\left(|\epsilon_1(\mathbf{m}_k)| > \epsilon\right) \\
&\leq cN^{2L} \sup_{k\in\mathcal{J}} \frac{\mathbb{E}\left[|\epsilon_1(\mathbf{m}_k)|^{2r}\right]}{\epsilon^{2r}} \\
&\leq cN^{2L} \sup_{k\in\mathcal{J}} \frac{\mathbb{E}\left[\left|(\mathbf{m}_k)^H \mathbf{P}\mathbf{n}\right|^{2r}\right]}{\epsilon^{2r}N^r} \\
&= cN^{2L} \left(\sup_{k\in\mathcal{J}} (\mathbf{m}_k)^H \mathbf{P}\mathbf{m}_k\right)^r \frac{\mathbb{E}\left[|n|^{2r}\right]}{\epsilon^{2r}N^r},
\end{aligned}$$

where n is a standard complex Gaussian random variable. By choosing $r > 2L + 1$ and applying the Borel-Cantelli lemma we obtain the desired result. ∎

Having established uniform convergence of the random ML cost function, consistency of the random ML channel estimates follows from the standard argument.

4.6 Appendix: Inverse of $\left(\mathbf{\Theta}^H\bar{\mathbf{P}}^{\perp}\mathbf{\Theta}\right)$ in (4.58) exists

Here we prove that the inverse of $\left(\mathbf{\Theta}^H\bar{\mathbf{P}}^{\perp}\mathbf{\Theta}\right)$ in (4.58) exists.

Proof. To see that the inverse exists, recalling $\mathbf{\Theta} = \mathbf{F}\mathbf{E}\mathbf{\Lambda}_p\mathbf{G}^H$ we note that

$$\left(\mathbf{\Theta}^H\mathbf{P}^{\perp}\mathbf{\Theta}\right)^{-1} = \left(\mathbf{G}\mathbf{\Lambda}_p^H\mathbf{E}^H\mathbf{F}^H\mathbf{P}^{\perp}\mathbf{F}\mathbf{E}\mathbf{\Lambda}_p\mathbf{G}^H\right)^{-1}.$$

Since $\mathbf{\Lambda}_p^H\mathbf{\Lambda}_p > 0$ and $\mathbf{G}$ is a flat full row rank matrix, when $L < N_u$, then $\mathbf{G}\mathbf{\Lambda}_p^H\mathbf{\Lambda}_p\mathbf{G}^H > 0$.

On the other hand,

$$\begin{aligned}
\det\left[\mathbf{G}\mathbf{\Lambda}_p^H\mathbf{E}^H\mathbf{F}^{\mathbf{H}}\mathbf{P}^{\perp}\mathbf{F}\mathbf{E}\mathbf{\Lambda}_p\mathbf{G}^H\right] &= \det\left[\mathbf{G}\mathbf{\Lambda}_p^H\mathbf{\Lambda}_p\mathbf{G}^H\right] \\
&\times \det\left[\mathbf{I}_{N_{imp}} - \mathbf{U}^H\mathbf{F}\mathbf{E}\left(\mathbf{\Lambda}_p\mathbf{G}^H\left(\mathbf{G}\mathbf{\Lambda}_p^H\mathbf{\Lambda}_p\mathbf{G}^H\right)^{-1}\mathbf{G}\mathbf{\Lambda}_p^H\right)\mathbf{E}^H\mathbf{F}^{\mathbf{H}}\mathbf{U}\right].
\end{aligned}$$

Now, since $L < N_u$ by assumption, $\mathbf{G}\mathbf{\Lambda}_p^H$ is flat and full row rank, and

$$\mathbf{\Lambda}_p\mathbf{G}^H\left(\mathbf{G}\mathbf{\Lambda}_p^H\mathbf{\Lambda}_p\mathbf{G}^H\right)^{-1}\mathbf{G}\mathbf{\Lambda}_p^H < \mathbf{I}_{N_u}.$$

Therefore,

$$\mathbf{U}^H\mathbf{F}\mathbf{E}\left(\mathbf{\Lambda}_p\mathbf{G}^H\left(\mathbf{G}\mathbf{\Lambda}_p^H\mathbf{\Lambda}_p\mathbf{G}^H\right)^{-1}\mathbf{G}\mathbf{\Lambda}_p^H\right)\mathbf{E}^H\mathbf{F}^{\mathbf{H}}\mathbf{U} < \mathbf{U}^H\mathbf{F}\mathbf{E}\mathbf{E}^H\mathbf{F}^{\mathbf{H}}\mathbf{U} < \mathbf{I}_{N_{imp}}$$

where we have used the fact that $\mathbf{U}\mathbf{U}^H < \mathbf{I}_N$ because $N_u < N$ by assumption. ∎

Chapter 5

Maximum-Likelihood Detection of IN Support

In this chapter we propose maximum-likelihood (ML) impulsive noise (IN) support detectors. Unlike the IN estimator proposed in chapter 3, which relied on the measurement data of dimension smaller than the dimension of the sparse signal for IN support estimation, the proposed schemes perform incremental greedy search to identify the most likely IN support in the received signal. In this chapter, we recall the statistical definitions introduced in chapter 4, used for the derivations of the ML channel estimators, to derive the deterministic ML (DML) and random ML (RML) IN support detectors.

In order to detect the IN support, the proposed schemes perform multiple hypothesis tests on each sample of the received signal. Depending on the output of the hypothesis test, the contamination or not of the IN on the sample is determined. If the sample is identified as contaminated by the IN, its location is incorporated in the estimated support set. If the sample is identified as not corrupted by the IN, its location is not added in the estimated support set. This greedy procedure is continued until the most likely IN support is identified. We build multiple hypothesis tests under the Bonferroni and Benjamini and Hochberg criteria and numerically assess their performance in terms of false missed target, false detection, false alarm and missed target rates.

To concisely elaborate on the proposed schemes, we divide this chapter into 5 sections. In section 2, we present the adopted system model. In section 3, we derive the DML and RML support detection schemes and evaluate them under the Bonferroni and Benjamini and Hochberg criteria. In section 4, we present the numerical validation of the performance of the proposed estimators.

Finally, we provide some conclusions in section 5 and conclude the chapter by presenting important derivations in the appendix.

5.1 System Model

We consider the same system model that has been defined in section 4.1 of chapter 4. Similar notations and variable definitions are also adopted. A proper definition will be presented when introducing new variables and notations that were not present in chapter 4.

5.2 ML Approach (Deterministic Setting)

To derive the DML IN support estimator we begin by defining $\mathcal{A}$ as the set of active impulses and assume that $N - |\mathcal{A}| - L > 0$. We recall here that the DML cost function defined in chapter 4 can be written as

$$\begin{aligned}\zeta_{DML}\left(\mathbf{h}, \mathbf{i}_{\mathcal{A}}, \sigma_w^2, \mathcal{A}\right) = &\frac{1}{N\sigma_w^2}\sum_{n\in\mathcal{A}}\left|y_p(n) - \sqrt{N}\Theta^H(n)\,\mathbf{h} - i(n)\right|^2 \\ &+ \frac{1}{N\sigma_w^2}\left(\mathbf{y}_p - \sqrt{N}\Theta\mathbf{h}\right)^H \mathbf{P}_{\mathcal{A}}^{\perp}\left(\mathbf{y}_p - \sqrt{N}\Theta\mathbf{h}\right) \\ &+ \log\left(\sigma_w^2\pi\right).\end{aligned}$$

where $\mathbf{i}_{\mathcal{A}} = [i(n)]_{n\in\mathcal{A}}$ and

$$\mathbf{P}_{\mathcal{A}} = \sum_{n\in\mathcal{A}}\mathbf{u}_n\mathbf{u}_n^H, \quad \mathbf{P}_{\mathcal{A}}^{\perp} = \mathbf{I}_N - \mathbf{P}_{\mathcal{A}}.$$

We will also define the $N \times |\mathcal{A}|$ selection matrix $\mathbf{U}_{\mathcal{A}}$ such that $\mathbf{P}_{\mathcal{A}} = \mathbf{U}_{\mathcal{A}}\mathbf{U}_{\mathcal{A}}^H$. Let us denote by $\hat{\mathbf{h}}_{\mathcal{A}}$, $\hat{\imath}_{\mathcal{A}}$ and $\hat{\sigma}^2_{w,\mathcal{A}}$ the ML estimators[1] of the channel impulse response, the IN amplitude and the noise power respectively, namely

$$\begin{aligned}\hat{\mathbf{h}}_{\mathcal{A}} &= \frac{1}{\sqrt{N}}\left(\Theta^H\mathbf{P}_{\mathcal{A}}^{\perp}\Theta\right)^{-1}\Theta^H\mathbf{P}_{\mathcal{A}}^{\perp}\mathbf{y}_p \\ \hat{\imath}_{\mathcal{A}} &= \mathbf{U}_{\mathcal{A}}^H\left(\mathbf{y}_p - \sqrt{N}\Theta^H\mathbf{h}\right) \\ \hat{\sigma}^2_{w,\mathcal{A}} &= \frac{1}{N}\mathbf{y}_p^H\left(\mathbf{P}_{\mathcal{A}}^{\perp} - \mathbf{P}_{\mathcal{A}}^{\perp}\Theta\left(\Theta^H\mathbf{P}_{\mathcal{A}}^{\perp}\Theta\right)^{-1}\Theta^H\mathbf{P}_{\mathcal{A}}^{\perp}\right)\mathbf{y}_p.\end{aligned}$$

[1]Note that we make explicit in the notation the fact that these estimators depend on the assumed support for the impulses, namely $\mathcal{A}$.

We define by $\zeta_{DML}^{(\mathcal{A})}$ the statistic that is obtained by replacing all these estimators back into the DML cost function, that is

$$\zeta_{DML}^{(\mathcal{A})} = \zeta_{DML}\left(\hat{\mathbf{h}}_{\mathcal{A}}, \hat{\imath}_{\mathcal{A}}, \hat{\sigma}^2_{w,\mathcal{A}}, \mathcal{A}\right) = \log\left(\mathbf{y}_p^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{y}_p\right) + ct.,$$

where $ct.$ denotes a constant independent of the support $\mathcal{A}$ and where we have defined the matrix

$$\mathcal{R}_{\mathcal{A}}^{\perp} = \mathbf{P}_{\mathcal{A}}^{\perp} - \mathbf{P}_{\mathcal{A}}^{\perp}\Theta\left(\Theta^H \mathbf{P}_{\mathcal{A}}^{\perp}\Theta\right)^{-1}\Theta^H \mathbf{P}_{\mathcal{A}}^{\perp}. \tag{5.1}$$

Remark 1. *Our approach for detecting the support of the IN will follow a greedy procedure that will compare the statistic $\zeta_{DML}^{(\mathcal{A})}$ with the statistic $\zeta_{DML}^{(\mathcal{A}\cup\{j\})}$ for every $j \notin \mathcal{A}$. Recalling that these values are negative log-likelihoods, a value of $\zeta_{DML}^{(\mathcal{A}\cup\{j\})}$ that is much lower than $\zeta_{DML}^{(\mathcal{A})}$ will indicate that the support $\mathcal{A}\cup\{j\}$ is much more probable than the support $\mathcal{A}$. When this is the case, the algorithm will incorporate the j^{th} sample in the support, and continue with the greedy procedure. We will describe this procedure in a much more formal way in the following subsection.*

The main problem to this greedy approach comes from the fact that, at each step, the algorithm must compute the values $\zeta_{DML}^{(\mathcal{A}\cup\{j\})}$ for every $j \notin \mathcal{A}$ in order to decide whether the j^{th} sample must be included in the IN support. This is computationally complex, since it implies doing matrix inversions in the computation of $\mathcal{R}_{\mathcal{A}\cup\{j\}}^{\perp}$ as defined in (5.1). We can avoid this problem by noting that $\mathbf{P}_{\mathcal{A}\cup\{j\}}^{\perp} = \mathbf{P}_{\mathcal{A}}^{\perp} - \mathbf{u}_j\mathbf{u}_j^H$, so that a direct application of the matrix inverse lemma leads to

$$\begin{aligned}
\mathbf{y}_p^H \mathcal{R}_{\mathcal{A}\cup\{j\}}^{\perp} \mathbf{y}_p &= \mathbf{y}_p^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{y}_p - \frac{\mathbf{y}_p^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{u}_j \mathbf{u}_j^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{y}_p}{1 - \mathbf{u}_j^H \Theta^H \left(\Theta^H \mathbf{P}_{\mathcal{A}}^{\perp}\Theta\right)^{-1}\Theta\mathbf{u}_j} \\
&= \mathbf{y}_p^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{y}_p - \frac{\mathbf{y}_p^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{u}_j \mathbf{u}_j^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{y}_p}{\mathbf{u}_j^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{u}_j},
\end{aligned}$$

where in the last step we have used the fact that $\mathbf{u}_j^H \mathbf{P}_{\mathcal{A}} \mathbf{u}_j = 0$ and $\mathbf{P}_{\mathcal{A}}^{\perp}\mathbf{u}_j = \mathbf{u}_j$ because $j \notin \mathcal{A}$. Using the above identity, we readily see that we can express

$$\zeta_{DML}^{(\mathcal{A})} - \zeta_{DML}^{(\mathcal{A}\cup\{j\})} = \log\left(\frac{\mathbf{y}_p^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{y}_p}{\mathbf{y}_p^H \mathcal{R}_{\mathcal{A}\cup\{j\}}^{\perp} \mathbf{y}_p}\right) = -\log\left(1 - \frac{\mathbf{y}_p^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{u}_j \mathbf{u}_j^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{y}_p}{\mathbf{y}_p^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{y}_p \mathbf{u}_j^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{u}_j}\right).$$

Note that a direct application of the Cauchy-Schwarz inequality shows that

$$\frac{\mathbf{y}_p^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{u}_j \mathbf{u}_j^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{y}_p}{\mathbf{y}_p^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{y}_p \mathbf{u}_j^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{u}_j} < 1$$

and therefore the quantity inside the logarithm is always well defined.

The proposed algorithm will incorporate the j^{th} sample into the support $\mathcal{A}$ when the difference between negative log-likelihoods $\zeta_{DML}^{(\mathcal{A})} - \zeta_{DML}^{(\mathcal{A}\cup\{j\})}$ is sufficiently high. Noting that the function $-\log(1-x)$ is monotonically increasing for $x \in (0,1)$, this is equivalent to a sufficiently large value of the statistic

$$\mathcal{T}_{\mathcal{A}}(j) = \frac{\mathbf{y}_p^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{u}_j \mathbf{u}_j^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{y}_p}{\mathbf{y}_p^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{y}_p \mathbf{u}_j^H \mathcal{R}_{\mathcal{A}}^{\perp} \mathbf{u}_j}. \tag{5.2}$$

Observe that we can compute $\mathcal{T}_{\mathcal{A}}(j)$ for all $j \notin \mathcal{A}$ without doing any matrix inversions, simply by using the value of $\mathcal{R}_{\mathcal{A}}^{\perp}$ computed once with the assumed support $\mathcal{A}$. This significantly simplifies the computational complexity of the proposed algorithm for support detection.

Now, the first question that we need to answer is, how large must $\mathcal{T}_{\mathcal{A}}(j)$ be in order to provide good confidence that the j^{th} sample is contaminated by IN? In other words, how do we choose a threshold value α_j so that we decide that the j^{th} sample is contaminated by IN if $\mathcal{T}_{\mathcal{A}}(j) > \alpha_j$? Clearly, in order to best fix this threshold, we should take into account all the statistics $\{\mathcal{T}_{\mathcal{A}}(j), j \notin \mathcal{A}\}$ and not only the j^{th} statistic alone. In other words, we should consider the problem as a multiple hypothesis test.

5.2.1 Multiple Hypothesis Test

For a given value of the support $\mathcal{A}$, we ask ourselves whether there exist other samples outside $\mathcal{A}$ that are contaminated by IN. To solve this, we consider the multiple hypothesis test composed of $N - |\mathcal{A}|$ binary decisions corresponding to the samples not included in $\mathcal{A}$. The j^{th} binary test is defined for every $j \notin \mathcal{A}$ as

$$\begin{aligned} H_0(j) &: \left(\mathbf{y}_p - \sqrt{N}\Theta\mathbf{h}\right)_j \sim \mathcal{CN}\left(0, \sigma_w^2\right) \\ H_1(j) &: \left(\mathbf{y}_p - \sqrt{N}\Theta\mathbf{h}\right)_j \sim \mathcal{CN}\left(i(j), \sigma_w^2\right). \end{aligned}$$

The null hypothesis $H_0(j)$ assumes that the j^{th} sample is free from IN, whereas the alternative one $H_1(j)$ considers IN contamination with complex amplitude $i(j)$. For each binary test ($H_0(j)$ versus $H_1(j)$) we may construct the generalized likelihood ratio statistic in (5.2) and we decide that the alternative is correct if

$$\mathcal{T}_{\mathcal{A}}(j) > \alpha_j$$

for a given threshold α_j. We will now show how the threshold values α_j for each $j \notin \mathcal{A}$ can be selected.

We begin by considering a single binary hypothesis test, namely the j^{th} one. Conventional binary hypothesis tests select the threshold value α_j in order to have a certain probability of false alarm, defined as the probability of incorrectly deciding for $H_1(j)$ when the correct hypothesis is $H_0(j)$, namely

$$\mathrm{FA}_j = \mathbb{P}_{H_0(j)}\left(\mathcal{T}_{\mathcal{A}}(j) > \alpha_j\right),$$

where $\mathbb{P}_{H_0(j)}(\cdot)$ is the probability of a certain event under the hypothesis $H_0(j)$. To fix this probability, we must investigate the statistical behavior of the statistic $\mathcal{T}_{\mathcal{A}}(j)$ under $H_0(j)$, meaning that $\mathbf{y}_p - \sqrt{N}\Theta\mathbf{h} \sim \mathcal{CN}(0, \sigma_w^2\mathbf{I}_N)$. To investigate the statistical behavior of $\mathcal{T}_{\mathcal{A}}(j)$, observe that we can express this matrix as $\mathcal{R}_{\mathcal{A}}^{\perp} = \mathbf{Q}_{\mathcal{A}}\mathbf{Q}_{\mathcal{A}}^H$ where $\mathbf{Q}_{\mathcal{A}}$ is an $N \times (N - |\mathcal{A}| - L)$ matrix of orthogonal columns. Since linear transformations of Gaussian random variables are Gaussian random variables, we see that

$$\mathbf{w}_{\mathcal{A}} \triangleq \mathbf{Q}_{\mathcal{A}}^H\mathbf{y}_p \sim \mathcal{CN}\left(0, \sigma_w^2\mathbf{I}_{N-|\mathcal{A}|-L}\right)$$

where we have used the fact that $\mathcal{R}_{\mathcal{A}}^{\perp}\Theta = \mathbf{0}$. Observe that we can express $\mathcal{T}_{\mathcal{A}}(j)$ as

$$\mathcal{T}_{\mathcal{A}}(j) = \frac{|v_{\mathcal{A}}(j)|^2}{\|\mathbf{v}_{\mathcal{A}}(j)\|^2 + |v_{\mathcal{A}}(j)|^2},$$

where

$$v_{\mathcal{A}}(j) = \frac{1}{\sqrt{\sigma_w^2\mathbf{u}_j^H\mathcal{R}_{\mathcal{A}}^{\perp}\mathbf{u}_j}}\mathbf{u}_j^H\mathbf{Q}_{\mathcal{A}}\mathbf{w}_{\mathcal{A}}$$

and

$$\mathbf{v}_{\mathcal{A}}(j) = \frac{1}{\sqrt{\sigma_w^2}}\left(\mathbf{I}_{N-|\mathcal{A}|-L} - \frac{1}{\mathbf{u}_j^H\mathcal{R}_{\mathcal{A}}^{\perp}\mathbf{u}_j}\mathbf{Q}_{\mathcal{A}}^H\mathbf{u}_j\mathbf{u}_j^H\mathbf{Q}_{\mathcal{A}}\right)\mathbf{w}_{\mathcal{A}}.$$

It can readily be seen that $v_{\mathcal{A}}(j)$ and $\mathbf{v}_{\mathcal{A}}(j)$ are independent standardized Gaussian random vectors, so that $|v_{\mathcal{A}}(j)|^2 \sim \frac{1}{2}\chi_2^2$ and $\|\mathbf{v}_{\mathcal{A}}(j)\|^2 \sim \frac{1}{2}\chi_{2(N-|\mathcal{A}|-L-1)}^2$. This implies that

$$\begin{aligned}\frac{\mathcal{T}_{\mathcal{A}}(j)}{1-\mathcal{T}_{\mathcal{A}}(j)} = \frac{|v_{\mathcal{A}}(j)|^2}{\|\mathbf{v}_{\mathcal{A}}(j)\|^2} &\sim \frac{\chi_2^2}{\chi_{2(N-|\mathcal{A}|-L-1)}^2} \\ &= \frac{1}{N-|\mathcal{A}|-L-1}F\left(2, 2\left(N-|\mathcal{A}|-L-1\right)\right),\end{aligned} \tag{5.3}$$

where $F(d_1, d_2)$ is the Snedecor F-distribution. In particular, we know that

$$\mathbb{P}\left(F\left(2, 2d\right) \leq x\right) = 1 - \left(\frac{d}{x+d}\right)^{d}.$$

We can therefore compute the false alarm probability for the j^{th} binary test (denoted as FA_j) in closed form as

$$\begin{aligned}\mathrm{FA}_j = \mathbb{P}_{H_0(j)}\left(\mathcal{T}_{\mathcal{A}}\left(j\right) \geq \alpha_j\right) &= \mathbb{P}_{H_0(j)}\left(\frac{\mathcal{T}_{\mathcal{A}}\left(j\right)}{1-\mathcal{T}_{\mathcal{A}}\left(j\right)} \geq \frac{\alpha_j}{1-\alpha_j}\right) \\ &= \left(1-\alpha_j\right)^{N-|\mathcal{A}|-L-1}.\end{aligned}$$

Therefore, we can fix the threshold α_j that guarantees a certain false alarm probability FA_j as

$$\alpha_j = 1 - \mathrm{FA}_j^{1/(N-|\mathcal{A}|-L-1)}.$$

This procedure ensures a certain false alarm probability for each of the binary hypothesis test. We next ask ourselves about the performance of the whole multiple hypothesis test procedure. We propose two different performance criteria, which will lead to two different methodologies for the multiple hypothesis test definition.

5.2.1.1 Family-wise False Alarm Rate (Bonferroni Criterion)

We define the family-wise false alarm rate as the probability that *any* of the binary hypothesis tests results in false alarm, so that there exists a certain $j \notin \mathcal{A}$ such that the j^{th} test declares $H_1(j)$ provided that $H_0(j)$ is true. This means that the family wise false alarm rate is given by

$$\begin{aligned}\overline{\mathrm{FA}} &= \mathbb{P}_{H_0}\left(\bigcup\nolimits_{j\notin\mathcal{A}}\left\{\mathcal{T}_{\mathcal{A}}\left(j\right) > \alpha_j\right\}\right) \leq \sum\nolimits_{j\notin\mathcal{A}}\mathbb{P}_{H_0(j)}\left(\mathcal{T}_{\mathcal{A}}\left(j\right) > \alpha_j\right) \\ &\leq \left(N - |\mathcal{A}|\right)\max_{j\notin\mathcal{A}}\left(\mathrm{FA}_j\right),\end{aligned}$$

where the first inequality follows from the conventional union bound. One can ensure that the family wise error rate is below a certain value $\overline{\mathrm{FA}}$ by forcing all the tests to have the same threshold α, which is fixed to guarantee a pairwise false alarm rate $\mathrm{FA}_j = \overline{\mathrm{FA}}/\left(N - |\mathcal{A}|\right)$. This is sometimes referred to as the Bonferroni criterion in multiple hypothesis testing [150].

5.2.1.2 False Discovery Rate (Benjamini and Hochberg Criterion)

It is well known that the Bonferroni method provides a very strict control of the false alarm probabilities but is not very powerful at all. An alternative criterion is based on the false discovery rate (FDR), which measures the expected proportion of errors among the set of rejected hypotheses.

To better understand the definition of FDR, consider the explanation in Table 5.1. The two central columns represent the true hypothesis of a binary hypothesis test, whereas the two central rows represent the acceptance/rejection of the null hypothesis. The quantities x and v represent the number of true negatives and true positives respectively, whereas z gives the number of false alarms (false positives) and y the number of detection errors (false negatives).

- As we have seen above, the false alarm is defined as the percentage of false positives among all the realizations of the test under $H_{\mathbf{0}}$, that is $\mathbb{E}\left[\frac{z}{x+z}\right]$.
- Likewise, we could define the missed target (MT) probability as the percentage of false negatives among all the realizations of the test under H_1, namely $\mathbb{E}\left[\frac{y}{y+v}\right]$.
- The FDR measures the percentage of false positives (errors) among the total number of realizations for which $H_{\mathbf{0}}$ is rejected, that is $\mathbb{E}\left[\frac{z}{z+v}\right]$.
- Finally, we could define the false missed target rate (FMR) as the percentage of false negatives (errors) among all the realizations for which $H_{\mathbf{0}}$ is accepted.

Benjamini and Hochberg [151] suggested the FDR criterion as an extremely powerful procedure as compared to the family-wise error rate in multiple hypothesis testing. This method is implemented as follows. Consider the p-value of the j^{th} test, defined as the probability, under the null hypothesis, of obtaining a result equal to or more extreme than what was actually observed.

$$p_j = \mathbb{P}\left(\frac{F\left(2, 2\left(N - |\mathcal{A}| - L - 1\right)\right)}{N - |\mathcal{A}| - L - 1} > \frac{\mathcal{T}_{\mathcal{A}}\left(j\right)}{1 - \mathcal{T}_{\mathcal{A}}\left(j\right)}\right) = \left(1 - \mathcal{T}_{\mathcal{A}}\left(j\right)\right)^{(N-|\mathcal{A}|-L-1)}. \tag{5.4}$$

Let $p_{(j)}$ denote the sequence of p-values in non-decreasing order. Then, we select

$$k = \max\left\{j : p_{(j)} \leq \frac{j}{N - |\mathcal{A}|}\text{FDR}\right\},$$

Table 5.1: Definition of the different probabilities in a binary hypothesis test.

	True H_0	True H_1	
H_0 accepted	x	y	$\mathrm{FMR} = \mathbb{E}\left[\frac{y}{x+y}\right]$
H_0 rejected	z	v	$\mathrm{FDR} = \mathbb{E}\left[\frac{z}{z+v}\right]$
	$\mathrm{FA} = \mathbb{E}\left[\frac{z}{x+z}\right]$	$\mathrm{MT} = \mathbb{E}\left[\frac{y}{y+v}\right]$	

Algorithm 5.1 Deterministic Bonferroni

Set $\mathcal{A} = \varnothing$ and fix a value for the family wise false alarm rate $\overline{\mathrm{FA}}$.

1. For $j \in [N] \backslash \mathcal{A}$, compute $\mathcal{T}_{\mathcal{A}}(j)$ according to the formula in (5.2).
2. Let $\mathcal{J}$ contain the set of values j for which
$$\mathcal{T}_{\mathcal{A}}(j) > \alpha_j = 1 - \left(\frac{\overline{\mathrm{FA}}}{N - |\mathcal{A}|}\right)^{1/(N-|\mathcal{A}|-L-1)}.$$
3. If $\mathcal{J} = \varnothing$, stop and return $\mathcal{A}$. Otherwise, replace $\mathcal{A}$ with $\mathcal{A} \cup \mathcal{J}$ and return to (1).

where FDR is the selected false discovery rate, and reject the null hypothesis corresponding to the tests with the lowest p-values $p_{(1)}, \ldots, p_{(k)}$.

5.2.2 Proposed Algorithms (Incremental Greedy Procedure)

To finalize this section, we provide an explicit description of the two proposed greedy algorithms for IN support detection. Depending on whether the target is the $\overline{\mathrm{FA}}$ or the FDR, we obtain two different algorithms 5.1 and 5.2 which are described in what follows.

5.3 ML Approach (Random Setting)

The above greedy algorithm for the detection of IN support can also be formulated according to the random (Gaussian) ML principle. In this section, we will investigate how this can be effectively implemented in practice and we will provide some guidelines on how to choose the thresholds.

Algorithm 5.2 Deterministic Benjamini and Hochberg

Set $\mathcal{A} = \varnothing$ and fix a value for the family wise false discovery ratio FDR.

1. For $j \in [N] \backslash \mathcal{A}$, compute $\mathcal{T}_{\mathcal{A}}(j)$ according to the formula in (5.2) and the corresponding p-values according to the formula in (5.4). Let $p_{i(j)}$ denote the sequence of p-values in non-decreasing order, namely $p_{i(1)} \leq p_{i(2)} \leq \ldots \leq p_{i(N-|\mathcal{A}|)}$.

2. Select k such that

$$k = \max \left\{ j : p_{(j)} \leq \frac{j}{N - |\mathcal{A}|} \mathrm{FDR} \right\}$$

and let $\mathcal{J} = \{i(1), \ldots, i(k)\}$.

3. If $\mathcal{J} = \varnothing$, stop and return $\mathcal{A}$. Otherwise, replace $\mathcal{A}$ with $\mathcal{A} \cup \mathcal{J}$ and return to (1).

As we have seen in chapter 4, we can write the normalized negative log-likelihood for the RML approach as

$$\zeta_{RML}\left(\mathbf{h}, \sigma^2, \sigma_w^2, \mathcal{A}\right) = \frac{1}{N} \log\det \mathbf{C}_n + \frac{1}{N} \left(\mathbf{y}_p - \sqrt{N}\Theta\mathbf{h}\right)^H \mathbf{C}_n^{-1} \left(\mathbf{y}_p - \sqrt{N}\Theta\mathbf{h}\right), \tag{5.5}$$

where $\mathbf{C}_n = \sigma^2 \mathbf{P}_{\mathcal{A}} + \sigma_w^2 \mathbf{P}_{\mathcal{A}}^{\perp}$ and $\sigma^2 = \sigma_i^2 + \sigma_w^2$. By taking the RML channel estimate (denoted here $\hat{\mathbf{h}}_{\mathcal{A}}$) and inserting it back into the cost function we obtain

$$\zeta_{RML}\left(\hat{\mathbf{h}}_{\mathcal{A}}, \sigma^2, \sigma_w^2, \mathcal{A}\right) = \frac{|\mathcal{A}|}{N} \log \sigma^2 + \left(1 - \frac{|\mathcal{A}|}{N}\right) \log \sigma_w^2 + \frac{1}{N\sigma_w^2} \mathbf{y}_p^H \mathbf{P}_{\Theta}^{\perp} \mathbf{y}_p \tag{5.6}$$

$$- \left(\frac{\sigma^2}{\sigma_w^2} - 1\right) \frac{1}{N} \sum_{i=1}^{|\mathcal{A}|} \frac{\mathbf{y}_p^H \mathbf{P}_{\Theta}^{\perp} \mathbf{U}_{\mathcal{A}} \mathbf{u}_{\mathcal{A}}(i) \mathbf{u}_{\mathcal{A}}^H(i) \mathbf{U}_{\mathcal{A}}^H \mathbf{P}_{\Theta}^{\perp} \mathbf{y}_p}{\sigma_w^2 + (\sigma^2 - \sigma_w^2) \lambda_{\mathcal{A}}(i)},$$

where $\mathbf{U}_{\mathcal{A}}$ is an $N \times |\mathcal{A}|$ selection matrix that points at the positions of the set $\mathcal{A}$, and we have used the eigendecomposition

$$\mathbf{U}_{\mathcal{A}}^H \mathbf{P}_{\Theta}^{\perp} \mathbf{U}_{\mathcal{A}} = \sum_{i=1}^{|\mathcal{A}|} \lambda_{\mathcal{A}}(i) \mathbf{u}_{\mathcal{A}}(i) \mathbf{u}_{\mathcal{A}}^H(i). \tag{5.7}$$

The minimum of this cost function can be established by the methodology established in chapter 4, which essentially simplifies to a conventional line

search. We will denote by $\hat{\sigma}^2_{\mathcal{A}}, \hat{\sigma}^2_{w,\mathcal{A}}$ the corresponding minima, and let

$$\zeta^{(\mathcal{A})}_{RML} = \zeta_{RML}\left(\hat{\mathbf{h}}_{\mathcal{A}}, \hat{\sigma}^2_{\mathcal{A}}, \hat{\sigma}^2_{w,\mathcal{A}}, \mathcal{A}\right).$$

The proposed estimator for the IN support will follow the same greedy approach as the one for the DML criterion, based on

$$\zeta^{(\mathcal{A})}_{RML} - \zeta^{(\mathcal{A}\cup\{j\})}_{RML}. \tag{5.8}$$

The main idea is to compute $\zeta^{(\mathcal{A}\cup\{j\})}_{RML}$ for all $j \notin \mathcal{A}$. Assuming that the value of the negative log-likelihood $\zeta^{(\mathcal{A}\cup\{j\})}_{RML}$ is sufficiently lower than the value of $\zeta^{(\mathcal{A})}_{RML}$, we will have evidence that the model that incorporates the j^{th} sample into the IN support is more likely. This will lead us to the conclusion that the j^{th} sample should be incorporated into the model.

The main problem of the above procedure comes from the fact that we need to compute the statistic $\zeta^{(\mathcal{A}\cup\{j\})}_{RML}$ for every $j \notin \mathcal{A}$. This implies that, for every $j \notin \mathcal{A}$, we need to generate the eigenvalue decomposition of the matrix $\mathbf{U}^H_{\mathcal{A}\cup\{j\}}\mathbf{P}^{\perp}_{\Theta}\mathbf{U}_{\mathcal{A}\cup\{j\}}$ and optimize the corresponding cost function in (5.6) with respect to the two variables $\sigma^2 \geq \sigma^2_w$. We have seen in chapter 4 that the optimization of the cost function with respect to these two variables can be carried out in a very efficient way by considering the Karush-Kuhn-Tucker conditions with respect to the variable $\beta = \sigma^2/\sigma^2_w$ and conducting a simple line search. Hence, the main computational difficulty of the greedy algorithm comes from the computation of the eigenvalue decomposition of $\mathbf{U}^H_{\mathcal{A}\cup\{j\}}\mathbf{P}^{\perp}_{\Theta}\mathbf{U}_{\mathcal{A}\cup\{j\}}$ for every $j \notin \mathcal{A}$. Fortunately, we can rely on the matrix inverse lemma to avoid this complicated step, thereby significantly simplifying the implementation of the algorithm.

To see this point, observe that we can express the above cost function as

$$\zeta_{RML}\left(\hat{\mathbf{h}}_{\mathcal{A}}, \sigma^2, \sigma^2_w, \mathcal{A}\right) = \frac{|\mathcal{A}|}{N}\log\sigma^2 + \left(1 - \frac{|\mathcal{A}|}{N}\right)\log\sigma^2_w + \frac{1}{N\sigma^2_w}\mathbf{y}_p^H\mathcal{R}_{\mathcal{A}}\left(\frac{\sigma^2_w}{\sigma^2 - \sigma^2_w}\right)\mathbf{y}_p,$$

where we have defined the matrix function

$$\mathcal{R}_{\mathcal{A}}(x) = \mathbf{P}^{\perp}_{\Theta} - \mathbf{P}^{\perp}_{\Theta}\mathbf{U}_{\mathcal{A}}\left(x\mathbf{I}_{|\mathcal{A}|} + \mathbf{U}^H_{\mathcal{A}}\mathbf{P}^{\perp}_{\Theta}\mathbf{U}_{\mathcal{A}}\right)^{-1}\mathbf{U}^H_{\mathcal{A}}\mathbf{P}^{\perp}_{\Theta}.$$

Note that, according to the definition in (5.1), we have $\mathcal{R}^{\perp}_{\mathcal{A}} = \mathcal{R}_{\mathcal{A}}(0)$. Furthermore, using the eigendecomposition in (5.7) we can also express $\mathcal{R}_{\mathcal{A}}(x)$ as

$$\mathcal{R}_{\mathcal{A}}(x) = \mathbf{P}^{\perp}_{\Theta} - \sum_{i=1}^{|\mathcal{A}|}\frac{1}{x + \lambda_{\mathcal{A}}(i)}\mathbf{P}^{\perp}_{\Theta}\mathbf{U}_{\mathcal{A}}\mathbf{u}_{\mathcal{A}}(i)\,\mathbf{u}^H_{\mathcal{A}}(i)\,\mathbf{U}^H_{\mathcal{A}}\mathbf{P}^{\perp}_{\Theta}. \tag{5.9}$$

A direct application of the matrix inversion lemma shows that

$$\mathbf{P}_{\Theta}^{\perp}\mathbf{U}_{\mathcal{A}}\left(x\mathbf{I}_{|\mathcal{A}|}+\mathbf{U}_{\mathcal{A}}^{H}\mathbf{P}_{\Theta}^{\perp}\mathbf{U}_{\mathcal{A}}\right)^{-1}\mathbf{U}_{\mathcal{A}}^{H}\mathbf{P}_{\Theta}^{\perp}=\mathbf{I}_{N}-x\left(x\mathbf{I}_{N}+\mathbf{P}_{\Theta}^{\perp}\mathbf{U}_{\mathcal{A}}\mathbf{U}_{\mathcal{A}}^{H}\mathbf{P}_{\Theta}^{\perp}\right)^{-1}$$

and this implies that

$$\mathcal{R}_{\mathcal{A}\cup\{j\}}\left(x\right)=\mathcal{R}_{\mathcal{A}}\left(x\right)-\frac{\mathcal{R}_{\mathcal{A}}\left(x\right)\mathbf{u}_{j}\mathbf{u}_{j}^{H}\mathcal{R}_{\mathcal{A}}\left(x\right)}{x+\mathbf{u}_{j}^{H}\mathcal{R}_{\mathcal{A}}\left(x\right)\mathbf{u}_{j}}. \tag{5.10}$$

We obtain

$$\begin{aligned}\zeta_{RML}\left(\hat{\mathbf{h}}_{\mathcal{A}\cup\{j\}},\sigma^{2},\sigma_{w}^{2},\mathcal{A}\cup\{j\}\right)&=\frac{|\mathcal{A}|+1}{N}\log\sigma^{2}+\left(1-\frac{|\mathcal{A}|+1}{N}\right)\log\sigma_{w}^{2}\\&+\frac{1}{N\sigma_{w}^{2}}\mathbf{y}_{p}^{H}\mathcal{R}_{\mathcal{A}}\left(\alpha\right)\mathbf{y}_{p}-\frac{1}{N\sigma_{w}^{2}}\frac{\mathbf{y}_{p}^{H}\mathcal{R}_{\mathcal{A}}\left(\alpha\right)\mathbf{u}_{j}\mathbf{u}_{j}^{H}\mathcal{R}_{\mathcal{A}}\left(\alpha\right)\mathbf{y}_{p}}{\alpha+\mathbf{u}_{j}^{H}\mathcal{R}_{\mathcal{A}}\left(\alpha\right)\mathbf{u}_{j}},\end{aligned}$$

where we have defined

$$\alpha=\frac{\sigma_{w}^{2}}{\sigma^{2}-\sigma_{w}^{2}}.$$

Observe that $\alpha=(\beta-1)^{-1}$ with $\beta=\sigma^{2}/\sigma_{w}^{2}$ as defined in chapter 4. Clearly one must have $\beta\geq 1$ so that in practice $\alpha\in(0,\infty)$.

Remark 2. *The identity in (5.10) allows to convert the term* $\mathbf{y}_{p}^{H}\mathcal{R}_{\mathcal{A}\cup\{j\}}\left(x\right)\mathbf{y}_{p}$ *into products and sums of quantities of the type* $\mathbf{y}_{p}^{H}\mathcal{R}_{\mathcal{A}}\left(x\right)\mathbf{y}_{p}$, $\mathbf{y}_{p}^{H}\mathcal{R}_{\mathcal{A}}\left(x\right)\mathbf{u}_{j}$ *and* $\mathbf{u}_{j}^{H}\mathcal{R}_{\mathcal{A}}\left(x\right)\mathbf{u}_{j}$. *Using the eigendecomposition of* $\mathcal{R}_{\mathcal{A}}\left(x\right)$ *in (5.9) we can therefore find an expression of the cost function* $\zeta_{RML}\left(\hat{\mathbf{h}}_{\mathcal{A}\cup\{j\}},\sigma^{2},\sigma_{w}^{2},\mathcal{A}\cup\{j\}\right)$ *without having to compute a novel eigenvalue decomposition of* $\mathcal{R}_{\mathcal{A}\cup\{j\}}$ *for each* $j\notin\mathcal{A}$. *This leads to a significant reduction of the computational complexity of the algorithm.*

The optimization of the cost function $\zeta_{RML}\left(\hat{\mathbf{h}}_{\mathcal{A}\cup\{j\}},\sigma^{2},\sigma_{w}^{2},\mathcal{A}\cup\{j\}\right)$ with respect to the pair $(\sigma^{2},\sigma_{w}^{2})$ follows the approach that was described in chapter 4. Taking derivatives with respect to these two quantities, we obtain

$$\frac{\partial\zeta_{RML}\left(\hat{\mathbf{h}}_{\mathcal{A}\cup\{j\}},\sigma^{2},\sigma_{w}^{2},\mathcal{A}\cup\{j\}\right)}{\partial\sigma^{2}}=\frac{|\mathcal{A}|+1}{N\sigma^{2}}-\frac{\mathbf{y}_{p}^{H}\mathcal{R}_{\mathcal{A}\cup\{j\}}\left(\alpha\right)\left(\mathbf{I}_{N}-\mathcal{R}_{\mathcal{A}\cup\{j\}}\left(\alpha\right)\right)\mathbf{y}_{p}}{N\sigma_{w}^{2}\left(\sigma^{2}-\sigma_{w}^{2}\right)}$$

$$\begin{aligned}\frac{\partial\zeta_{RML}\left(\hat{\mathbf{h}}_{\mathcal{A}\cup\{j\}},\sigma^{2},\sigma_{w}^{2},\mathcal{A}\cup\{j\}\right)}{\partial\sigma_{w}^{2}}&=\frac{N-|\mathcal{A}|-1}{N\sigma_{w}^{2}}\\&+\frac{\mathbf{y}_{p}^{H}\mathcal{R}_{\mathcal{A}\cup\{j\}}\left(\alpha\right)\left(\mathbf{I}_{N}-\beta\mathcal{R}_{\mathcal{A}\cup\{j\}}\left(\alpha\right)\right)\mathbf{y}_{p}}{N\sigma_{w}^{2}\left(\sigma^{2}-\sigma_{w}^{2}\right)}\end{aligned}$$

with $\beta = \sigma^2/\sigma_w^2$. Therefore, the minimum will be attained at the pair of points[2]

$$\hat{\sigma}^2_{\mathcal{A}\cup\{j\}} = \frac{\beta^2_{\mathcal{A}\cup\{j\}}}{\left(\beta_{\mathcal{A}\cup\{j\}} - 1\right)\left(|\mathcal{A}| + 1\right)} \mathbf{y}_p^H \mathcal{R}_{\mathcal{A}\cup\{j\}}\left(\alpha_{\mathcal{A}\cup\{j\}}\right)\left(\mathbf{I}_N - \mathcal{R}_{\mathcal{A}\cup\{j\}}\left(\alpha_{\mathcal{A}\cup\{j\}}\right)\right)\mathbf{y}_p$$

$$\hat{\sigma}^2_{w,\mathcal{A}\cup\{j\}} = \frac{1}{\left(\beta_{\mathcal{A}\cup\{j\}} - 1\right)\left(N - |\mathcal{A}| - 1\right)} \mathbf{y}_p^H \mathcal{R}_{\mathcal{A}\cup\{j\}}\left(\alpha_{\mathcal{A}\cup\{j\}}\right) \times \left(\mathbf{I}_N - \beta_{\mathcal{A}\cup\{j\}} \mathcal{R}_{\mathcal{A}\cup\{j\}}\left(\alpha_{\mathcal{A}\cup\{j\}}\right)\right)\mathbf{y}_p,$$

where $\alpha_{\mathcal{A}} = \left(\beta_{\mathcal{A}} - 1\right)^{-1}$. $\beta_{\mathcal{A}}$ is a solution to the following equation in $\beta \geq 1$

$$\frac{1}{N - |\mathcal{A}| - 1} \mathbf{y}_p^H \mathcal{R}_{\mathcal{A}\cup\{j\}}(\alpha)\left(\mathbf{I}_N - \beta \mathcal{R}_{\mathcal{A}\cup\{j\}}(\alpha)\right)\mathbf{y}_p = \frac{\beta}{|\mathcal{A}| + 1} \mathbf{y}_p^H \mathcal{R}_{\mathcal{A}\cup\{j\}}(\alpha) \times \left(\mathbf{I}_N - \mathcal{R}_{\mathcal{A}\cup\{j\}}(\alpha)\right)\mathbf{y}_p$$

if there exists one, or otherwise $\hat{\sigma}^2_{\mathcal{A}\cup\{j\}} = \hat{\sigma}^2_{w,\mathcal{A}\cup\{j\}} = \frac{1}{N}\mathbf{y}_p^H \mathbf{P}^{\perp}_{\Theta}\mathbf{y}_p$. Note that all these equations can be formulated in terms of the eigendecomposition of $\mathcal{R}_{\mathcal{A}}(x)$ by using the identity in (5.10).

5.3.1 Multiple Hypothesis Test

Following the same approach as in the DML method, one can now formulate a greedy algorithm for the detection of the IN support. As before, given a support set $\mathcal{A}$, we ask ourselves whether there exist other samples outside $\mathcal{A}$ that are contaminated by IN. To solve this, we consider the multiple hypothesis test composed of $N - |\mathcal{A}|$ binary decisions corresponding to each of the j^{th} samples not included in $\mathcal{A}$. The j^{th} test is defined for every $j \notin \mathcal{A}$ as

$$H_0(j) : \left(\mathbf{y}_p - \sqrt{N}\Theta\mathbf{h}\right) \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2\mathbf{P}_{\mathcal{A}} + \sigma_w^2\mathbf{P}^{\perp}_{\mathcal{A}}\right)$$
$$H_1(j) : \left(\mathbf{y}_p - \sqrt{N}\Theta\mathbf{h}\right) \sim \mathcal{CN}\left(\mathbf{0}, \sigma^2\mathbf{P}_{\mathcal{A}\cup\{j\}} + \sigma_w^2\mathbf{P}^{\perp}_{\mathcal{A}\cup\{j\}}\right).$$

Note that, as in the case of the DML approach, the null hypothesis $H_0(j)$ assumes that the j^{th} sample is free from IN, whereas the alternative one $H_1(j)$ considers that the j^{th} sample has larger variance σ^2. For each binary test ($H_0(j)$ versus $H_1(j)$) we decide that the alternative is correct if

$$\zeta^{(\mathcal{A})}_{RML} - \zeta^{(\mathcal{A}\cup\{j\})}_{RML} > \alpha_j$$

[2]Of course, the estimator of σ^2 only makes sense when $|\mathcal{A}| > 0$.

for a given threshold α_j, where we recall that $\zeta_{RML}^{(\mathcal{A})}$ is the value of the negative log-likelihood function after replacing all parameters with their RML estimates, namely

$$\zeta_{RML}^{(\mathcal{A})} = \zeta_{RML}\left(\hat{\mathbf{h}}_{\mathcal{A}}, \hat{\sigma}_{\mathcal{A}}^2, \hat{\sigma}_{w,\mathcal{A}}^2, \mathcal{A}\right).$$

As we have seen before, the selection of the threshold values α_j for each $j \notin \mathcal{A}$ can be done by imposing a certain value of the family-wise false alarm probability (Bonferroni approach) or by forcing a certain false discovery rate (Benjamini and Hochberg approach). Either way, we need to investigate the statistical behavior of the test $\zeta_{RML}^{(\mathcal{A})} - \zeta_{RML}^{(\mathcal{A}\cup\{j\})}$ under the null hypothesis.

Given the complex form of the RML estimator and the fact that it does not accept a closed form expression, it is in general extremely difficult to characterize the statistical law of $\zeta_{RML}^{(\mathcal{A})}$. In order to solve this issue, we will take here an asymptotic approach and analyze the behavior of this statistic when $N \to \infty$. To that effect, we will be using the result in chapter 4 that establishes the asymptotic behavior of the RML channel estimator. We summarize here the main statistical assumptions that are made in our asymptotic analysis.

(**As1**) The noise term is zero mean, circularly symmetric and Gaussian distributed with covariance $\bar{\mathbf{C}}_n$, that is $\mathbf{n} \sim \mathcal{CN}\left(0, \bar{\mathbf{C}}_n\right)$.

(**As2**) Both N and $|\mathcal{A}|$ converge to infinity at the same rate, so that

$$0 < \liminf_N \frac{|\mathcal{A}|}{N} \leq \limsup_N \frac{|\mathcal{A}|}{N} < 1.$$

(**As3**) The eigenvalues of the matrix $\Theta^H\Theta$ are contained in a compact interval of the positive real axis for all N, that is

$$0 < \inf_N \Theta^H\Theta < \sup_N \Theta^H\Theta < \infty.$$

Furthermore, if $\mathbf{P}_{\mathcal{A}}$ is the $N \times N$ selection matrix corresponding to the set[3] $\mathcal{A} \subset [N]$ we have

$$\inf_N \Theta^H\mathbf{P}_{\mathcal{A}}\Theta > 0 \text{ and } \inf_N \Theta^H\mathbf{P}_{\mathcal{A}}^{\perp}\Theta > 0.$$

(**As4**) The norm of the rows of Θ decays uniformly to zero as $O\left(N^{-1}\right)$ or faster, that is

$$\sup_N \max_{j=1,\ldots,N} N\left\|\mathbf{u}_j^H\Theta\right\| < \infty.$$

[3]Note that both matrix Θ and the set $\mathcal{A}$ depend on the system dimension N, although we obviate this in the notation for the sake of clarity.

Remark 3. *We point out that* (**As1**) – (**As3**) *have been used before in order to establish the asymptotic behavior of* $\hat{\mathbf{h}}_{\mathcal{A}}$ *in chapter 4. Assumption* (**As4**) *is new, and therefore deserves some explanation. Observe that* $\Theta^H\Theta$ *is an* $L \times L$ *matrix with trace equal to the sum of the norms of the rows of* Θ*, that is*

$$\mathrm{tr}\left[\Theta^H\Theta\right] = \sum\nolimits_{j=1}^{N} \mathrm{tr}\left[\Theta^H\mathbf{u}_j\mathbf{u}_j^T\Theta\right] = \sum\nolimits_{j=1}^{N} \left\|\mathbf{u}_j^H\Theta\right\|^2.$$

Now, according to (**As3**), $\mathrm{tr}\left[\Theta^H\Theta\right]$ *must be uniformly bounded for all* N*, which implies that the above series must be convergent and therefore* $\left\|\mathbf{u}_j^H\Theta\right\|^2 \to 0$ *as* $j \to \infty$*. However, this does not imply convergence to zero of each particular row, that is* $\left\|\mathbf{u}_j^H\Theta\right\|^2 \to 0$ *for* $N \to \infty$*, a fact that is needed in some steps of the following analysis. This new assumption is imposed by* (**As4**).

The following assumption guarantees that the minimum of the cost function is achieved, with probability one and for large enough N, in the interior of the feasibility region.

(**As5**) It holds that

$$\limsup_{N}\left(\frac{1}{N}\mathrm{tr}\left[\bar{\mathbf{C}}_n\right] - \frac{1}{|\mathcal{A}|}\mathrm{tr}\left[\bar{\mathbf{C}}_n\mathbf{P}_{\mathcal{A}}\right]\right) < 0$$

Remark 4. *The assumption could well be dropped by separately considering the realizations for which this does not hold. However, we prefer to impose it for the sake of clarity in the proofs. Note, in any case, that the assumption holds when* $\mathcal{A}$ *is included in the true support of the IN. Indeed, when this is the case, we have*

$$\frac{1}{|\mathcal{A}|}\mathrm{tr}\left[\bar{\mathbf{C}}_n\mathbf{P}_{\mathcal{A}}\right] = \sigma^2$$

$$\frac{1}{N}\mathrm{tr}\left[\bar{\mathbf{C}}_n\right] = \frac{N_{imp}}{N}\sigma^2 + \left(1 - \frac{N_{imp}}{N}\right)\sigma_w^2$$

and therefore

$$\frac{1}{N}\mathrm{tr}\left[\bar{\mathbf{C}}_n\right] - \frac{1}{|\mathcal{A}|}\mathrm{tr}\left[\bar{\mathbf{C}}_n\mathbf{P}_{\mathcal{A}}\right] = \left(1 - \frac{N_{imp}}{N}\right)\left(\sigma_w^2 - \sigma^2\right) < 0.$$

However, this may not be the case for general values of $\mathcal{A}$*.*

Let us first recall the main asymptotic result of the RML channel estimator as established in chapter 4. If $\hat{\mathbf{h}}_{\mathcal{A}}$ denotes the RML channel estimator when the support $\mathcal{A}$ is assumed and denoting as $\bar{\mathbf{h}}$ the true channel impulse response, under (**As1**) – (**As3**) we have

$$\hat{\mathbf{h}}_{\mathcal{A}} = \bar{\mathbf{h}} + \frac{1}{\sqrt{N}}\left(\Theta^H\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{-1}\Theta\right)^{-1}\Theta^H\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{-1}\mathbf{n} + o_p\left(\frac{1}{\sqrt{N}}\right), \tag{5.11}$$

where

$$\widetilde{\mathbf{C}}_{n,\mathcal{A}} = \widetilde{\sigma}^2_{\mathcal{A}}\mathbf{P}_{\mathcal{A}} + \widetilde{\sigma}^2_{w,\mathcal{A}}\mathbf{P}^{\perp}_{\mathcal{A}} \tag{5.12}$$

$$\widetilde{\sigma}^2_{\mathcal{A}} = \frac{1}{|\mathcal{A}|}\mathrm{tr}\left[\mathbf{P}_{\mathcal{A}}\bar{\mathbf{C}}_n\right] \text{ and } \quad \widetilde{\sigma}^2_{w,\mathcal{A}} = \frac{1}{N-|\mathcal{A}|}\mathrm{tr}\left[\mathbf{P}^{\perp}_{\mathcal{A}}\bar{\mathbf{C}}_n\right]. \tag{5.13}$$

This result can be used to establish the following proposition, which provides an asymptotic description of the statistic $\zeta^{(\mathcal{A})}_{RML} - \zeta^{(\mathcal{A}\cup\{j\})}_{RML}$.

Proposition 1. *Under* (**As1**) – (**As5**), *we have*

$$\zeta^{(\mathcal{A})}_{RML} - \zeta^{(\mathcal{A}\cup\{j\})}_{RML} = \frac{1}{N}\log\left(\frac{\widetilde{\sigma}^2_{w,\mathcal{A}}}{\widetilde{\sigma}^2_{\mathcal{A}}}\right) + \frac{1}{N}\left\{\bar{\mathbf{C}}_n\right\}_{jj}\left(\frac{\widetilde{\sigma}^2_{\mathcal{A}} - \widetilde{\sigma}^2_{w,\mathcal{A}}}{\widetilde{\sigma}^2_{w,\mathcal{A}}\widetilde{\sigma}^2_{\mathcal{A}}}\right) + \frac{1}{N}\mathbf{n}^H\mathcal{Q}^{\perp}_{\mathcal{A}}\Psi_{\mathcal{A}}(j)\,\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{n} + o_p\left(\frac{1}{N}\right)$$

where we have defined

$$\mathcal{Q}^{\perp}_{\mathcal{A}} = \widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}} - \widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}}\Theta\left(\Theta^H\widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}}\Theta\right)^{-1}\Theta^H\widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}} \tag{5.14}$$

and

$$\Psi_{\mathcal{A}}(j) = \frac{\left\{\bar{\mathbf{C}}_n\right\}_{jj} - \widetilde{\sigma}^2_{\mathcal{A}}}{|\mathcal{A}|}\mathbf{P}_{\mathcal{A}} + \frac{\widetilde{\sigma}^2_{w,\mathcal{A}} - \left\{\bar{\mathbf{C}}_n\right\}_{jj}}{N-|\mathcal{A}|}\mathbf{P}^{\perp}_{\mathcal{A}} + \left(\widetilde{\sigma}^2_{\mathcal{A}} - \widetilde{\sigma}^2_{w,\mathcal{A}}\right)\frac{\widetilde{\sigma}^2_{w,\mathcal{A}}}{\widetilde{\sigma}^2_{\mathcal{A}}}\mathbf{u}_j\mathbf{u}_j^T.$$

Proof. See Appendix 5.6. ∎

The above proposition provides a means to fix the threshold levels α_j, $j \notin \mathcal{A}$, to guarantee a certain asymptotic false alarm probability for each binary hypothesis test. To ensure that, observe that under the null hypothesis $H_0(j)$ we have

$$\mathbf{y}_p \sim \mathcal{CN}\left(\sqrt{N}\Theta\mathbf{h}, \sigma^2\mathbf{P}_{\mathcal{A}} + \sigma^2_w\mathbf{P}^{\perp}_{\mathcal{A}}\right)$$

which means that $\widetilde{\sigma}^2_{\mathcal{A}} = \sigma^2$, $\widetilde{\sigma}^2_{w,\mathcal{A}} = \left\{\bar{\mathbf{C}}_n\right\}_{jj} = \sigma^2_w$ so that

$$\zeta^{(\mathcal{A})}_{RML} - \zeta^{(\mathcal{A}\cup\{j\})}_{RML} = \frac{1}{N}\log\left(\frac{\sigma^2_w}{\sigma^2}\right) + \frac{1}{N}\left(\frac{\sigma^2 - \sigma^2_w}{\sigma^2}\right) + \frac{1}{N}\mathbf{n}^H\mathcal{Q}^{\perp}_{\mathcal{A}}\Psi_{\mathcal{A}}(j)\,\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{n} + o_p\left(\frac{1}{N}\right)$$

where

$$\Psi_{\mathcal{A}}(j) = \frac{\sigma^2_w - \sigma^2}{|\mathcal{A}|}\mathbf{P}_{\mathcal{A}} + \left(\sigma^2 - \sigma^2_w\right)\frac{\sigma^2_w}{\sigma^2}\mathbf{u}_j\mathbf{u}_j^T.$$

Now, it can be shown from (**As3**) – (**As4**) that

$$\frac{1}{N}\mathbf{n}^H\mathcal{Q}^{\perp}_{\mathcal{A}}\frac{\mathbf{P}_{\mathcal{A}}}{|\mathcal{A}|}\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{n} = \frac{1}{(\sigma^2)^2}\frac{1}{N}\mathbf{n}^H\frac{\mathbf{P}_{\mathcal{A}}}{|\mathcal{A}|}\mathbf{n} + o_p\left(\frac{1}{N}\right)$$

$$\frac{1}{N}\mathbf{n}^H\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{u}_j\mathbf{u}_j^T\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{n} = \frac{1}{(\sigma^2_w)^2}\frac{1}{N}\mathbf{n}^H\mathbf{u}_j\mathbf{u}_j^T\mathbf{n} + o_p\left(\frac{1}{N}\right)$$

and this shows that

$$\zeta_{RML}^{(\mathcal{A})} - \zeta_{RML}^{(\mathcal{A}\cup\{j\})} = \frac{1}{N}\log\left(\frac{\sigma_w^2}{\sigma^2}\right) + \frac{1}{N}\left(1-\frac{\sigma_w^2}{\sigma^2}\right) - \frac{1}{N}\left(1-\frac{\sigma_w^2}{\sigma^2}\right)\left[\frac{\mathbf{n}^H\mathbf{P}_{\mathcal{A}}\mathbf{n}}{\sigma^2|\mathcal{A}|} - \frac{\mathbf{n}^H\mathbf{u}_j\mathbf{u}_j^T\mathbf{n}}{\sigma_w^2}\right] + o_p\left(\frac{1}{N}\right).$$

Now, clearly $\mathbf{u}_j^T\mathbf{n} \sim \mathcal{CN}(0,\sigma_w^2)$ and $\mathbf{U}_{\mathcal{A}}^H\mathbf{n} \sim \mathcal{CN}\left(0,\sigma^2\mathbf{I}_{|\mathcal{A}|}\right)$ and these two random variables are independent, so that it follows that

$$\frac{\mathbf{n}^H\mathbf{P}_{\mathcal{A}}\mathbf{n}}{\sigma^2|\mathcal{A}|} - \frac{\mathbf{n}^H\mathbf{u}_j\mathbf{u}_j^T\mathbf{n}}{\sigma_w^2} \sim \frac{1}{2}\left(\frac{1}{|\mathcal{A}|}\chi^2_{2|\mathcal{A}|} - \chi^2_2\right),$$

where χ_2^2 and $\chi^2_{2|\mathcal{A}|}$ are two independent Chi-Square variables with 2 and $2|\mathcal{A}|$ degrees of freedom respectively. Let us write $\chi = \frac{1}{|\mathcal{A}|}\chi^2_{2|\mathcal{A}|} - \chi^2_2$. The cumulative density function of this random variable is well known to be [152, 153]:

$$F_\chi(y) = \mathbb{P}(\chi \le y) = \\ = \begin{cases} \left(\frac{|\mathcal{A}|}{|\mathcal{A}|+1}\right)^{|\mathcal{A}|}\mathrm{e}^{\frac{y}{2}} & y<0 \\ 1 - \frac{\mathrm{e}^{-\frac{y|\mathcal{A}|}{2}}}{|\mathcal{A}|+1}\sum_{i=0}^{|\mathcal{A}|-1}\sum_{\ell=0}^{i}\frac{1}{(i-l)!}\left(\frac{|\mathcal{A}|}{|\mathcal{A}|+1}\right)^{|\mathcal{A}|-1-i}\left(\frac{y|\mathcal{A}|}{2}\right)^{i-\ell} & y\ge 0. \end{cases} \tag{5.15}$$

In order to achieve a certain false alarm probability we should choose α_j such that

$$\mathrm{FA}_j = \mathbb{P}_{H_0(j)}\left(\zeta_{RML}^{(\mathcal{A})} - \zeta_{RML}^{(\mathcal{A}\cup\{j\})} \ge \alpha_j\right).$$

This probability can be asymptotically approximated by

$$\mathrm{FA}_j = \mathbb{P}\left(\chi \le \left(\frac{2\sigma^2}{\sigma^2-\sigma_w^2}\right)\log\left(\frac{\sigma_w^2}{\sigma^2}\right) + 2 - N\alpha_j\left(\frac{2\sigma^2}{\sigma^2-\sigma_w^2}\right)\right).$$

and therefore we should fix α_j such that

$$\alpha_j = \frac{1}{N}\log\left(\frac{\sigma_w^2}{\sigma^2}\right) + \frac{\sigma^2-\sigma_w^2}{2N\sigma^2}\left(2 - F_\chi^{-1}(\mathrm{FA}_j)\right)$$

where $F_\chi^{-1}(p)$ is the functional inverse of $F_\chi(y)$.

With this, we have all the ingredients to formulate the two incremental greedy support estimation algorithms based on RML approach.

5.3.2 Proposed Algorithms (Incremental Greedy Procedure)

The following two algorithms are the unconditional (random) counterparts of the DML procedures. Contrary to what happens in the DML algorithm,

we cannot simply initialize the support set to $\mathcal{A} = \varnothing$ and run the greedy algorithm, because our statistical description of $\zeta_{RML}^{(\mathcal{A})} - \zeta_{RML}^{(\mathcal{A}\cup\{j\})}$ is only valid asymptotically when $N, |\mathcal{A}| \to \infty$. Consequently, the case $|\mathcal{A}| = 0$ needs to be handled with care. It can be shown that, under $H_0(j)$, $\zeta_{RML}^{(\varnothing)} - \zeta_{RML}^{(\{j\})}$ is always equal to zero for all N sufficiently large. Therefore, we propose to initialize the algorithm by selecting the sample index that provides the smallest value of $\zeta_{RML}^{(\{j\})}$. If all the samples provide the same value $\zeta_{RML}^{(\{j\})} = \zeta_{RML}^{(\varnothing)}$, the algorithm declares $\mathcal{A} = \varnothing$ and finishes. Otherwise, a greedy estimation procedure is carried out as described in algorithms 5.3 and 5.4.

Algorithm 5.3 Random Bonferroni

Let $k^* = \arg\min_k \zeta_{RML}^{(\{k\})}$. If $\zeta_{ML}^{(\varnothing)} - \zeta_{RML}^{(\{k^*\})} > 0$ initialize $\mathcal{A} = \{k\}$, otherwise decide that there are no impulses. Fix a value for the family wise false alarm rate $\overline{\text{FA}}$.

1. For $j \in [N]\backslash\mathcal{A}$, compute $\zeta_{RML}^{(\mathcal{A})} - \zeta_{RML}^{(\mathcal{A}\cup\{j\})}$.
2. Let $\mathcal{J}$ contain the set of values j for which $\zeta_{RML}^{(\mathcal{A})} - \zeta_{RML}^{(\mathcal{A}\cup\{j\})} > \alpha_j$, where
$$\alpha_j = \frac{1}{N}\log\left(\frac{\sigma_w^2}{\sigma^2}\right) + \frac{\sigma^2 - \sigma_w^2}{2N\sigma^2}\left(2 - F_\chi^{-1}\left(\frac{\overline{\text{FA}}}{N - |\mathcal{A}|}\right)\right).$$
3. If $\mathcal{J} = \varnothing$, stop and return $\mathcal{A}$. Otherwise, replace $\mathcal{A}$ with $\mathcal{A} \cup \mathcal{J}$ and return to (1).

The most complex part of the proposed two algorithms is clearly the computation of the statistic $\zeta_{RML}^{(\mathcal{A}\cup\{j\})}$ for all $j \in [N]\backslash\mathcal{A}$, which corresponds to step 1 in the algorithms. As mentioned before, the computational complexity of this step can be considerably diminished by exploiting the identity in (5.10) together with the eigendecomposition of the matrix $\mathcal{R}_{\mathcal{A}}(x)$ in (5.9), which is valid for all $j \in [N]\backslash\mathcal{A}$.

In order to further simplify the computational complexity of the multiple hypothesis test, we consider suboptimal versions of the two algorithms above, obtained by replacing

$$\zeta_{RML}^{(\mathcal{A}\cup\{j\})} = \zeta_{RML}\left(\hat{\mathbf{h}}_{\mathcal{A}\cup\{j\}}, \hat{\sigma}^2_{\mathcal{A}\cup\{j\}}, \hat{\sigma}^2_{w,\mathcal{A}\cup\{j\}}, \mathcal{A}\cup\{j\}\right)$$

with the approximation

$$\zeta_{RML}^{(\mathcal{A}\cup\{j\})} \simeq \zeta_{RML}\left(\hat{\mathbf{h}}_{\mathcal{A}\cup\{j\}}, \hat{\sigma}^2_{\mathcal{A}}, \hat{\sigma}^2_{w,\mathcal{A}}, \mathcal{A}\cup\{j\}\right). \tag{5.16}$$

Algorithm 5.4 Random Benjamini and Hochberg

Let $k^* = \arg\min_k \zeta_{RML}^{(\{k\})}$. If $\zeta_{RML}^{(\varnothing)} - \zeta_{RML}^{(\{k^*\})} > 0$ initialize $\mathcal{A} = \{k\}$, otherwise decide that there are no impulses. Fix a value for the family wise false discovery ratio FDR.

1. For $j \in [N] \backslash \mathcal{A}$, compute $\zeta_{RML}^{(\mathcal{A})} - \zeta_{RML}^{(\mathcal{A} \cup \{j\})}$ and the corresponding p-values according to

$$p_j = F_\chi \left(\left(\frac{2\sigma^2}{\sigma^2 - \sigma_w^2} \right) \log \left(\frac{\sigma_w^2}{\sigma^2} \right) + 2 - N \left(\zeta_{RML}^{(\mathcal{A})} - \zeta_{RML}^{(\mathcal{A} \cup \{j\})} \right) \left(\frac{2\sigma^2}{\sigma^2 - \sigma_w^2} \right) \right)$$

   where $F_\chi(\cdot)$ is defined in (5.15). Let $p_{i(j)}$ denote the sequence of p-values in non-decreasing order, namely $p_{i(1)} \leq p_{i(2)} \leq \ldots \leq p_{i(N-|\mathcal{A}|)}$.

2. Select k such that

$$k = \max \left\{ j : p_{(j)} \leq \frac{j}{N - |\mathcal{A}|} \mathrm{FDR} \right\}$$

   and let $\mathcal{J} = \{i(1), \ldots, i(k)\}$.

3. If $\mathcal{J} = \varnothing$, stop and return $\mathcal{A}$. Otherwise, replace $\mathcal{A}$ with $\mathcal{A} \cup \mathcal{J}$ and return to (1).

These suboptimal versions of the algorithms will avoid the optimization step at each new evaluation of $j \in [N]\backslash\mathcal{A}$, replacing the true optimizers $\hat{\sigma}^2_{\mathcal{A}\cup\{j\}}, \hat{\sigma}^2_{w,\mathcal{A}\cup\{j\}}$ by the approximations $\hat{\sigma}^2_{\mathcal{A}}, \hat{\sigma}^2_{w,\mathcal{A}}$.

5.4 Numerical Validation

In this section we numerically validate the theoretical results derived above. First, we evaluate the accuracy of the statistical description of the two statistics derived under the deterministic and the random models. Then, we evaluate the performance of the multi-hypothesis tests for the two models under the Bonferroni and the Benjamini and Hochberg criteria.

5.4.1 Validation of the Detection Statistics

We consider here a scenario with a variable number of impulses, where the SNR is fixed to 10 dB and the INR to 20 dB. The channel impulse response is randomly selected with an exponentially decaying power delay profile of average duration equal to 20 samples, and the receiver assumed a total channel length of $L = 30$ samples. In order to study the accuracy of the statistical description provided above, we simulate different instances of the number of subcarriers N and the number of impulses N_{imp}, the position of which is randomly fixed according to a uniform law. For each scenario, a total of 100 realizations of the input signal is generated, and for each such realization we compute the $N - N_{imp}$ statistics corresponding to (5.2) and (5.8) for all j outside the support of the IN. The resulting values are then conveniently transformed and compared to the two densities in (5.3) –exact DML– and (5.15) –asymptotic RML– respectively.

In Figures 5.1 and 5.2 we represent the histograms (solid blue lines) and probability density functions (red dotted line) corresponding to the DML statistic for different values of N and N_{imp}. Observe that there is a perfect match between the histograms and the corresponding density functions, which in this case are exact. In Figures 5.3 and 5.4 we represent the same quantities for the RML statistic, where now the red lines represent the asymptotic approximation of the corresponding law. Given the fact that this asymptotic law is derived under the assumption that $N, N_{imp} \to \infty$ at the same rate, it seems reasonable that a higher accuracy is observed for large values of these two parameters. In any case, it is worth pointing out that the asymptotic law

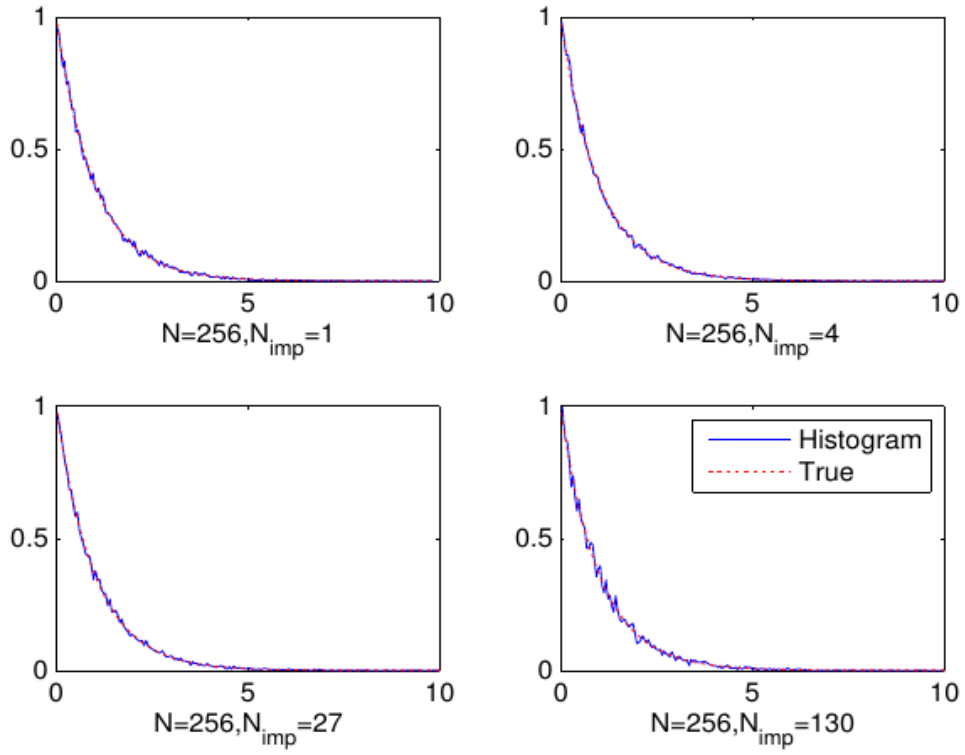


Figure 5.1: Histogram of the DML statistic versus true density for different values of N_{imp} when $N = 256$.

provides a reasonable approximation for values of N_{imp} as low as $N_{imp} = 1$. Observe also that for low values of N_{imp}, which corresponds to less accurate approximations of the asymptotic density, a mass point is observe in the simulated histograms. This mass point corresponds to realizations for which no inflexion point is found on the feasible region of the RML cost function, and therefore the optimum value is selected in the boundary line where $\sigma^2 = \sigma_w^2$.

5.4.2 Evaluation of the Support Detection Algorithms

In this section we provide an evaluation of the support detection algorithms presented above. We consider a scenario with $N = 256$ subcarriers with a random number of impulses, where the activation of the IN is modeled according to a Bernoulli random variable with activation probability equal to 0.1 [58]. The SNR is fixed to 10 dB and INR is fixed to 20 dB. The CIR is randomly selected as in the above subsection and each algorithm is run for 1000 realiza-

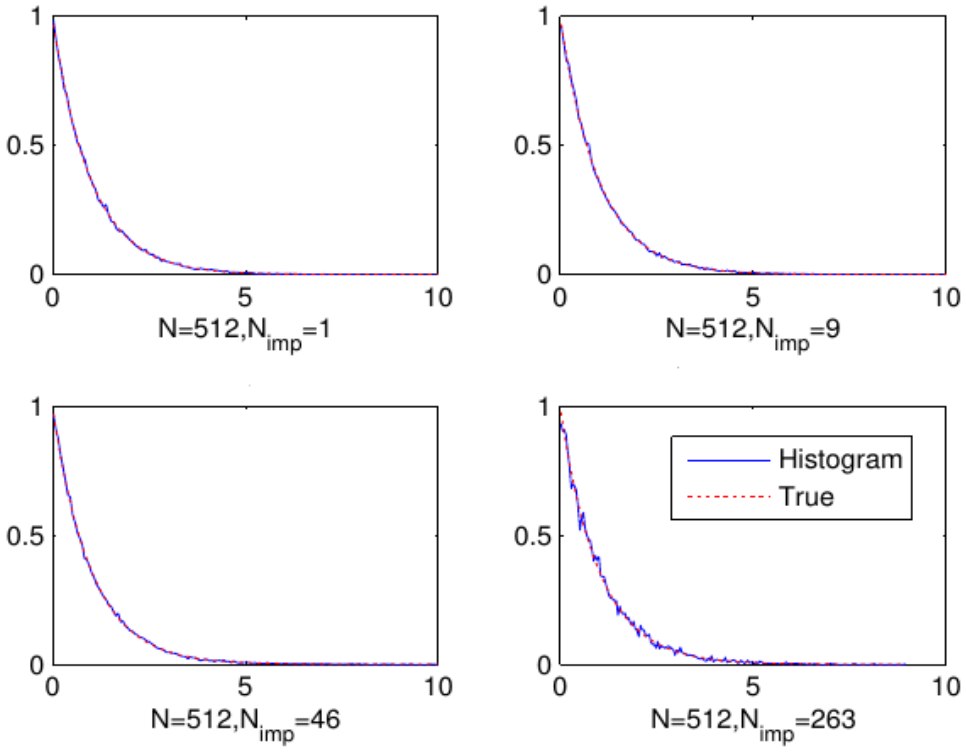


Figure 5.2: Histogram of the DML statistic versus true density for different values of N_{imp} when $N = 512$.

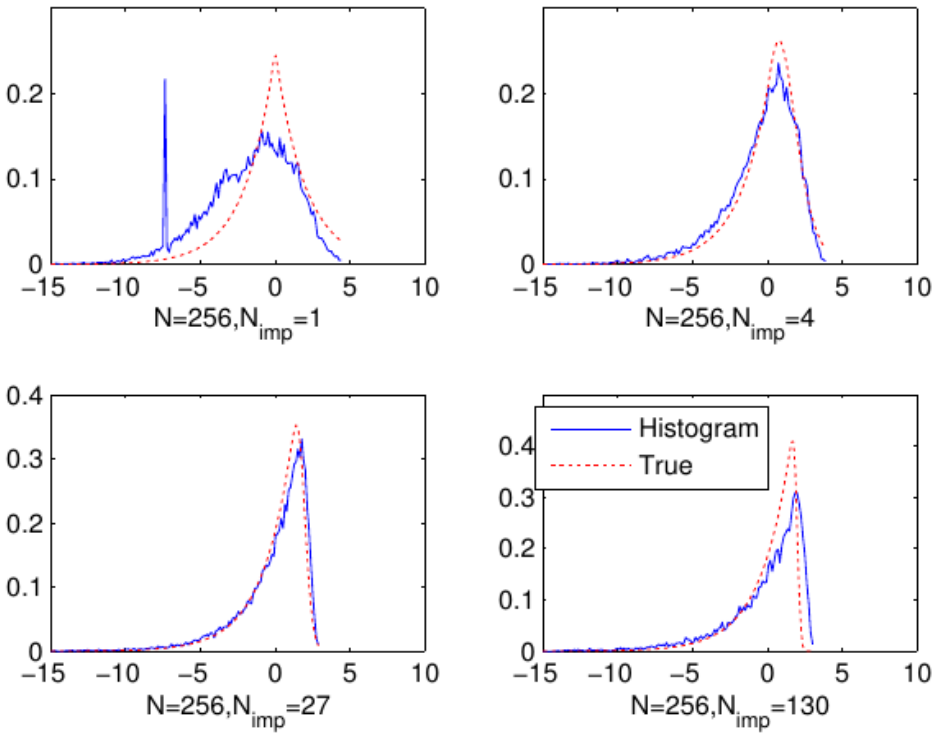


Figure 5.3: Histogram of the RML statistic versus asymptotic density for different values of N_{imp} when $N = 256$.

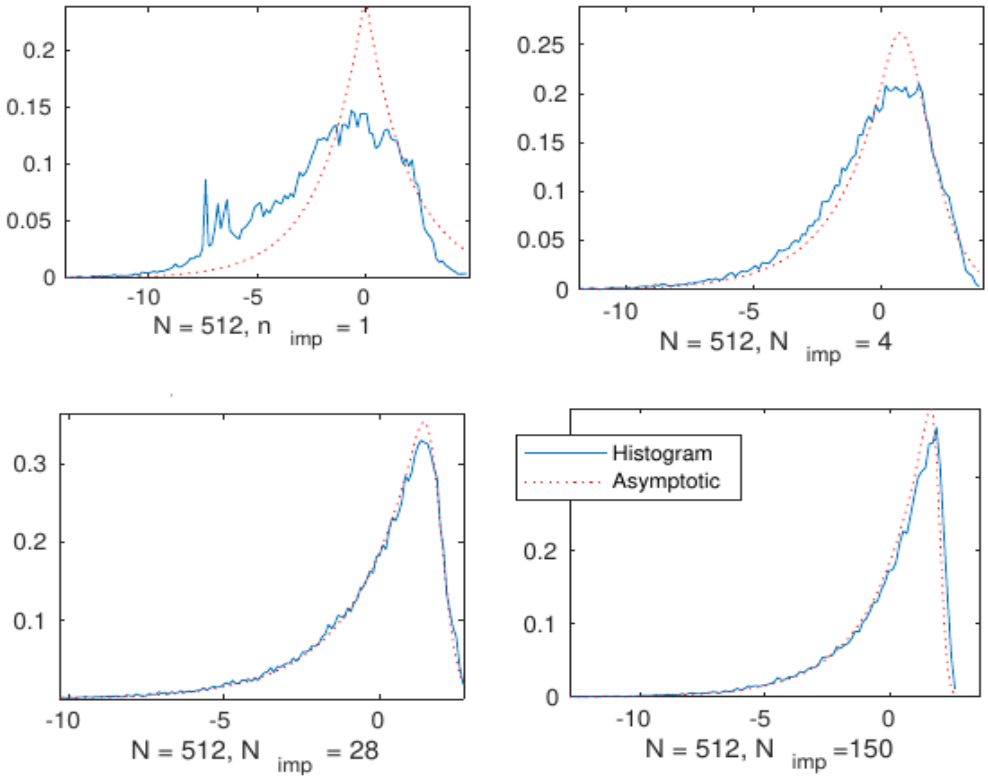


Figure 5.4: Histogram of the RML statistic versus asymptotic density for different values of N_{imp} when $N = 512$.

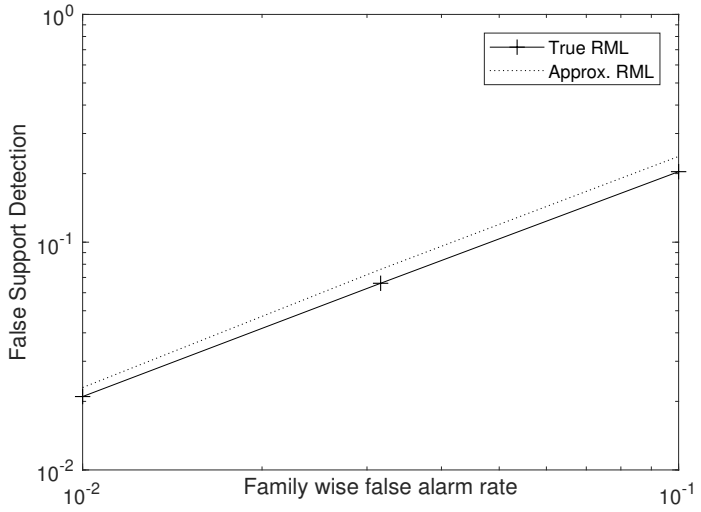


Figure 5.5: False support detection performance of approximated RML and true RML.

tions of signal, channel and noise. The detection thresholds are selected using the Bonferroni and Benjamini and Hochberg approaches according to some target values that took values in the range $10^{-2}, \ldots, 10^{-1}$. The performance of RML under Bonferroni and Benjamini and Hochberg criterion is respectively denoted by RBF and RBH in the graphs. Similarly, the performance of DML under Bonferroni and Benjamini and Hochberg criterion is denoted respectively by DBF and DBH in the graphs.

Figures 5.5 and 5.6 show the comparison of true RML and the approximated RML derived in (5.16) in terms of false support detection and missed support. We define the number of falsely detected IN support as false detection and number of missed IN support as missed support. To simulate the approximated RML algorithm the background and IN powers are frozen for the whole sequence of tests and updated every time a new set of samples is included in the support set. As shown, the approximated RML shows comparable false support detection and missed support performance as shown by the true RML algorithm. It should be noted here that the performance curves of both estimators do not properly align with each other because the approximated RML is the suboptimal characterization of the true RML algorithm. In the numerical validations to follow, we adopt approximated version of RML for computational efficiency and compare its performance with the DML-based

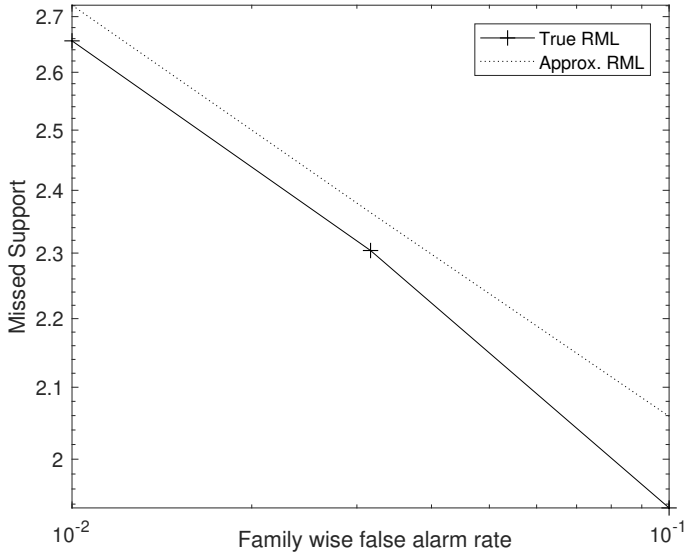


Figure 5.6: Missed support comparison of approximated RML and true RML.

schemes. Apart from the DML-based and RML-based algorithms, we also simulate a greedy naive (traditional) approach of IN support detection, which is derived without assuming anything about the structure of the received signal. This naive approach therefore assumes that the signal consists of a Gaussian signal with unknown variance (containing the desired signal plus noise) contaminated by impulsive samples of unknown magnitude. The resulting method is implemented by particularizing the DML estimation algorithm to the case $\Theta = \mathbf{0}$ and $L = 0$.

Figures 5.7 and 5.8 show the performance of IN support detection algorithms, in terms of FDR and FMR. As can be seen in Fig. 5.7, RBH shows zero FDR when a relaxed false alarm target is chosen. As the false alarm target value is strict, an FDR value is attained by the RBH which is comparable with the one that is achieved by the RBF scheme. Comparing the performance of the other algorithms, it should be noted that the traditional algorithms (denoted by TBF and TBH in the graphs) bear lower FDR than the DML and RBF schemes in all tested false alarm probability values. However, analyzing the FMR performance of the IN support detection algorithms in Fig. 5.8, we see that the TBF and TBH bear high FMRs than the other schemes. Similar to what has been observed during FDR analysis, RBH shows zero FMR when implemented with relaxed false alarm target probability and an FMR value that is comparable with the one achieved by RBF is attained when tested un-

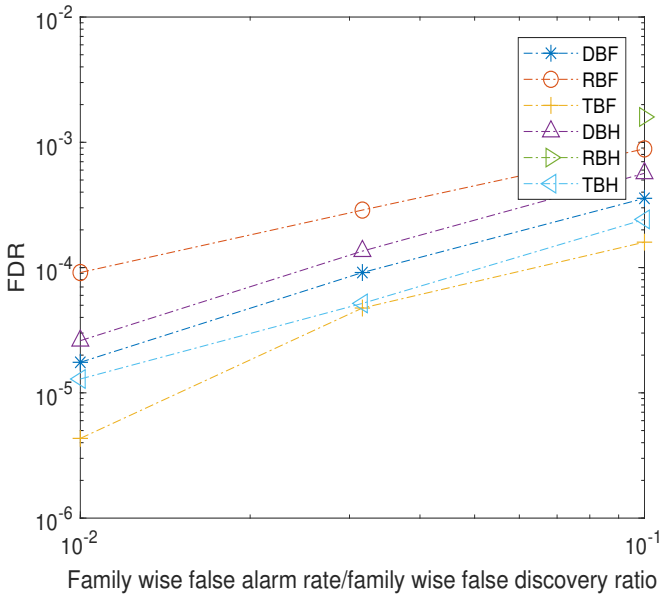


Figure 5.7: FDR performance of the support detection algorithms.

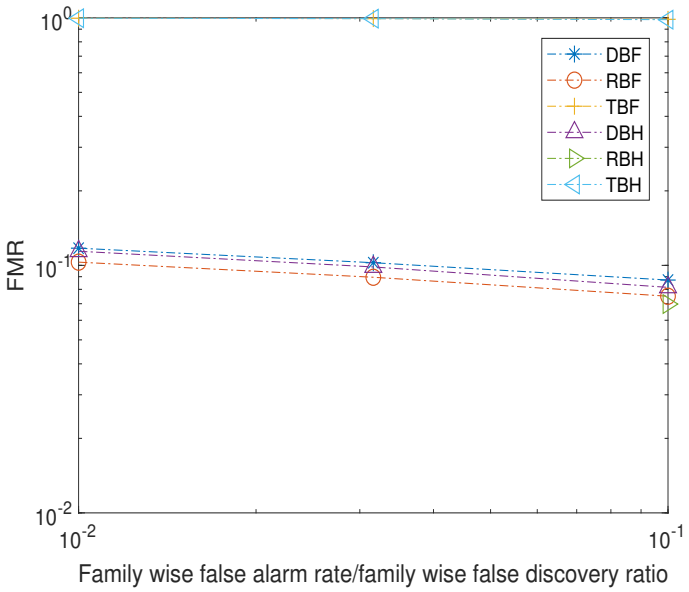


Figure 5.8: FMR performance of the support detection algorithms.

der high false alarm target probability. Moreover, the DML-based algorithms bear higher FMR than the RBF scheme.

In figures 5.9 and 5.10 the FA and MT performance of the IN support detec-

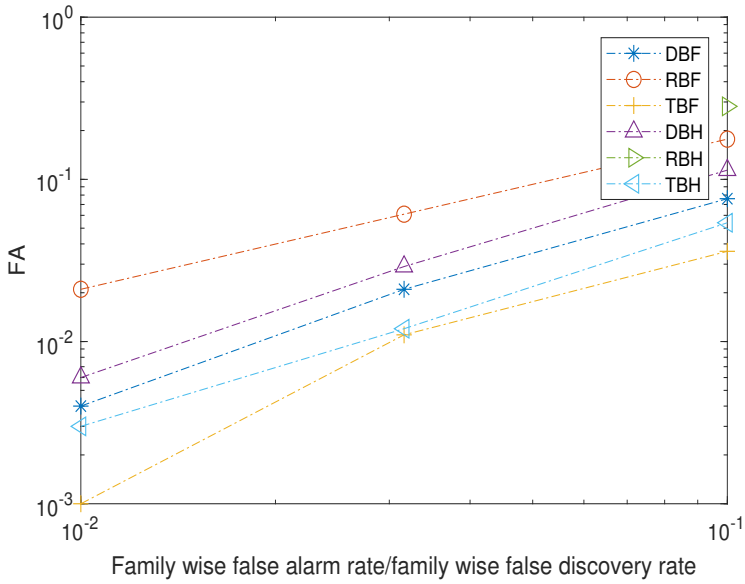


Figure 5.9: FA performance of the support detection algorithms.

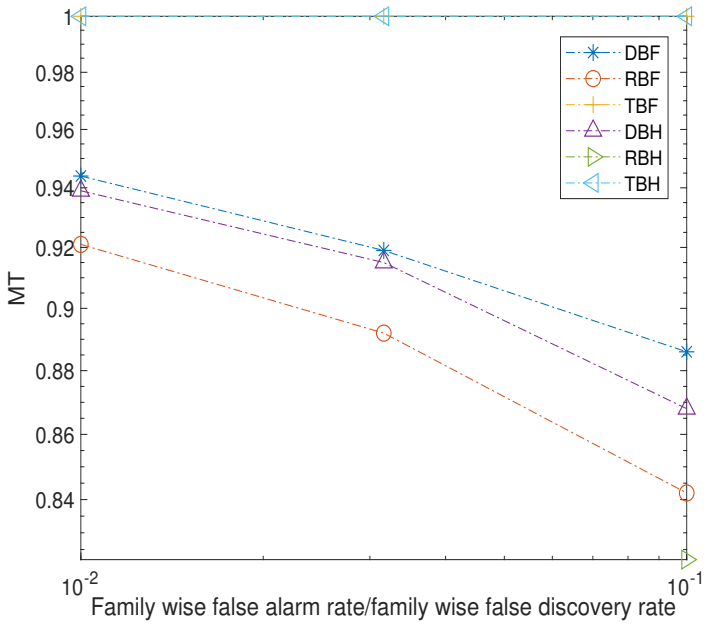


Figure 5.10: MT performance of the support detection algorithms.

tion algorithms is reported. As can be seen, the traditional greedy algorithms TBF and TBH show lower FA and higher MT rates in comparison to the other algorithms. Among the RBF and DML-based algorithms, DML-based algo-

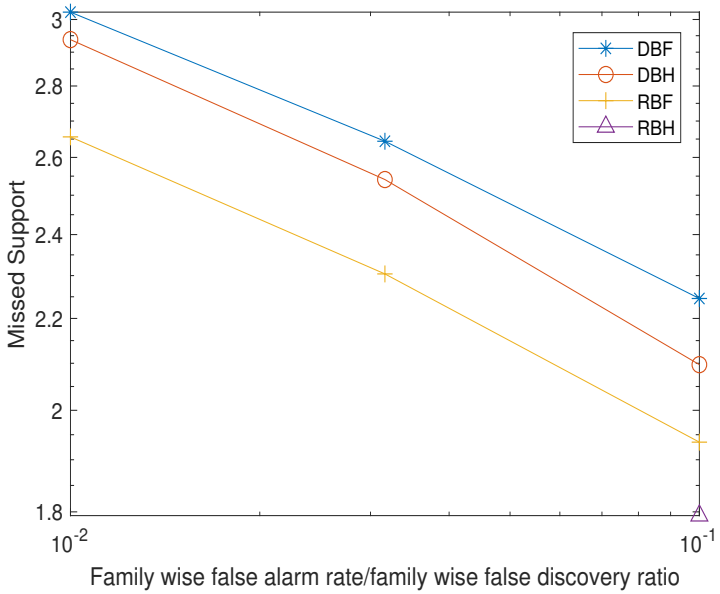


Figure 5.11: Missed support comparison of the support detection algorithms.

rithms show lower FA but bear higher MT rates than the RBF scheme. The RBH, similar to its performance during FDR and FMR analysis, shows zero FA when false alarm probability is relaxed and shows lower MT in comparison to RBF, DML-based algorithms and traditional algorithms. The numerical evaluation showing the missed support and false support detection rates of the RML-based and DML-based approaches is reported in figures 5.11 and 5.12. As can been seen, RML approach implemented under Benjamini and Hochberg criterion does not misses any support to be detected when a relaxed false alarm probability is imposed. Upon increasing the target false alarm probability value, it should be noted that the RML-based approaches still outperform other IN support detection algorithms. In a similar fashion, RML under Benjamini and Hochberg criterion does not report any false support detection when a relaxed target false alarm probability is chosen. However, under stricter target false alarm probability, RML-based approaches bear higher false support detection rates than the DML-based approaches.

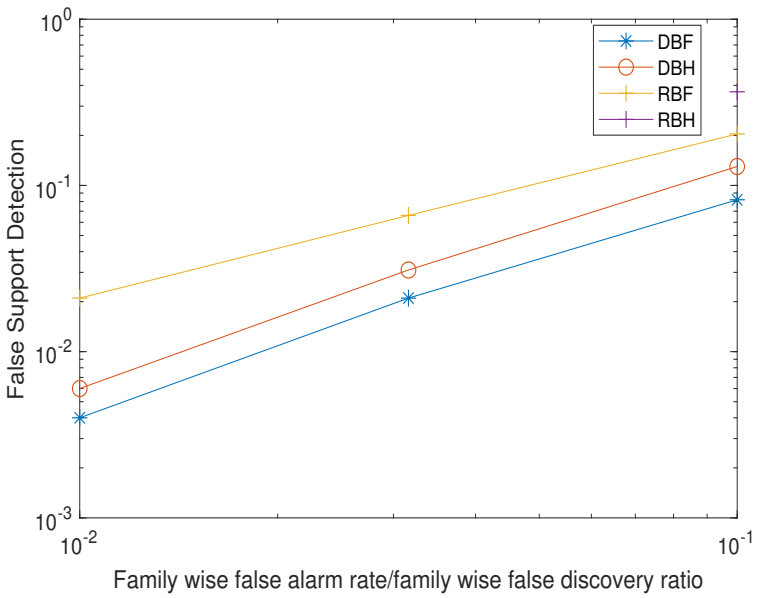


Figure 5.12: False support detection performance of the support detection algorithms.

5.5 Conclusions

In this chapter we proposed two ML IN support detection algorithms. Both schemes perform incremental greedy search of the probable IN samples location in the received signal. Two multiple hypothesis tests were built according to the Bonferroni and Benjamini and Hochberg criteria. Among the proposed IN detection algorithms, the scheme that is derived under random setting segregates the IN samples better than the one derived under deterministic setting. Moreover, we verified numerically that implementing the RML algorithm under Benjamini and Hochberg criterion performs an accurate determination of IN support rather than implementing it under Bonferroni criterion. We also evaluated the performance of naive algorithms and verified that a poor IN support detection performance is achieved if the received signal structure is not considered while estimating the IN samples location.

5.6 Appendix: Proof of Proposition 1

We begin by a lemma that allows to convert the quantity $\zeta_{RML}^{(\mathcal{A})}$ into a much simpler expression that depends only quadratically on the noise vector $\mathbf{n}$. The proof is based on the asymptotic behavior of the RML channel estimate as established in (5.11).

Lemma 2. *Under the hypothesis of Proposition 1, we can express*

$$\zeta_{RML}^{(\mathcal{A})} = \frac{1}{N}\log\det\left(\widetilde{\mathbf{C}}_{n,\mathcal{A}}\right) + \frac{1}{N}\mathbf{n}^H\mathcal{Q}_{\mathcal{A}}^{\perp}\mathbf{n} + o_p\left(\frac{1}{N}\right) \tag{5.17}$$

where $\bar{\mathbf{C}}_n$ is the true covariance matrix of $\mathbf{n}$,$\mathcal{Q}_{\mathcal{A}}^{\perp}$ is defined in (5.14) and $\widetilde{\mathbf{C}}_{n,\mathcal{A}}$ is defined in (5.12).

Proof. We consider first the RML cost function in (5.5). We claim that under the assumptions of Proposition 1, with probability one the minimum is always achieved at an interior point of the achievable region. This can be shown by studying the function $f(\beta)$ in (4.42) and noting that

$$\frac{1}{N}f(1) = \frac{1}{N}\mathbf{y}_p^H\mathbf{P}_{\Theta}^{\perp}\mathbf{y}_p - \frac{1}{|\mathcal{A}|}\mathbf{y}_p^H\mathbf{P}_{\Theta}^{\perp}\mathbf{P}_{\mathcal{A}}\mathbf{P}_{\Theta}^{\perp}\mathbf{y}_p \to \frac{1}{N}\mathrm{tr}\left[\bar{\mathbf{C}}_n\right] - \frac{1}{|\mathcal{A}|}\mathrm{tr}\left[\bar{\mathbf{C}}_n\mathbf{P}_{\mathcal{A}}\right]$$

almost surely, which is a negative quantity by assumption. This guarantees (see further chapter 4) that $f(\beta) = 0$ in the region $\beta \geq 1$, which in turn implies that the minimum of the RML cost function is in the interior of the feasible region.

Assume from now on that N is large enough so that the above holds. By optimizing first with respect to the two noise powers we obtain the estimates

$$\hat{\sigma}_{\mathcal{A}}^2 = \frac{1}{|\mathcal{A}|}\bar{\mathbf{y}}^H(\mathbf{h})\,\mathbf{P}_{\mathcal{A}}\bar{\mathbf{y}}(\mathbf{h})$$
$$\hat{\sigma}_{w,\mathcal{A}}^2 = \frac{1}{N-|\mathcal{A}|}\bar{\mathbf{y}}^H(\mathbf{h})\,\mathbf{P}_{\mathcal{A}}^{\perp}\bar{\mathbf{y}}(\mathbf{h})$$

where $\bar{\mathbf{y}}(\mathbf{h}) = \mathbf{y}_p - \sqrt{N}\Theta\mathbf{h}$. Therefore, for all N sufficiently large, the concentrated log-likelihood function takes the form

$$\zeta_{RML}\left(\mathbf{h},\hat{\sigma}_{\mathcal{A}}^2,\hat{\sigma}_{w,\mathcal{A}}^2,\mathcal{A}\right) = \frac{|\mathcal{A}|}{N}\log\left(\frac{1}{|\mathcal{A}|}\bar{\mathbf{y}}^H(\mathbf{h})\,\mathbf{P}_{\mathcal{A}}\bar{\mathbf{y}}(\mathbf{h})\right)$$
$$+\left(1-\frac{|\mathcal{A}|}{N}\right)\times\log\left(\frac{1}{N-|\mathcal{A}|}\bar{\mathbf{y}}^H(\mathbf{h})\,\mathbf{P}_{\mathcal{A}}^{\perp}\bar{\mathbf{y}}(\mathbf{h})\right)+1.$$

Now, let $\hat{\mathbf{h}}_{\mathcal{A}}$ denote the minimum of the above cost function with respect to $\mathbf{h}$. We can write

$$\zeta_{RML}^{(\mathcal{A})} = \zeta_{RML}\left(\hat{\mathbf{h}}_{\mathcal{A}}, \hat{\sigma}^2_{\mathcal{A}}, \hat{\sigma}^2_{w,\mathcal{A}}, \mathcal{A}\right) = \frac{|\mathcal{A}|}{N}\log\left(\widetilde{\sigma}^2_{\mathcal{A}}\right) + \left(1 - \frac{|\mathcal{A}|}{N}\right)\log\left(\widetilde{\sigma}^2_{w,\mathcal{A}}\right) + 1 + \delta_{\mathcal{A}}$$

where the error term $\delta_{\mathcal{A}}$ can be expressed as

$$\delta_{\mathcal{A}} = \frac{|\mathcal{A}|}{N}\log\left(\frac{1}{|\mathcal{A}|}\frac{\bar{\mathbf{y}}^H\left(\hat{\mathbf{h}}_{\mathcal{A}}\right)\mathbf{P}_{\mathcal{A}}\bar{\mathbf{y}}\left(\hat{\mathbf{h}}_{\mathcal{A}}\right)}{\widetilde{\sigma}^2_{\mathcal{A}}}\right) + \left(1 - \frac{|\mathcal{A}|}{N}\right) \times \log\left(\frac{1}{N - |\mathcal{A}|}\frac{\bar{\mathbf{y}}^H\left(\hat{\mathbf{h}}_{\mathcal{A}}\right)\mathbf{P}^{\perp}_{\mathcal{A}}\bar{\mathbf{y}}\left(\hat{\mathbf{h}}_{\mathcal{A}}\right)}{\widetilde{\sigma}^2_{w,\mathcal{A}}}\right).$$

Next, we make use of the asymptotic characterization of the channel estimator $\hat{\mathbf{h}}_{\mathcal{A}}$ in (5.11). We begin by noticing that we can express the quadratic form

$$\frac{1}{|\mathcal{A}|}\bar{\mathbf{y}}^H\left(\hat{\mathbf{h}}_{\mathcal{A}}\right)\mathbf{P}_{\mathcal{A}}\bar{\mathbf{y}}\left(\hat{\mathbf{h}}_{\mathcal{A}}\right) = \frac{N}{|\mathcal{A}|}\left(\bar{\mathbf{h}} - \hat{\mathbf{h}}_{\mathcal{A}}\right)^H\Theta^H\mathbf{P}_{\mathcal{A}}\Theta\left(\bar{\mathbf{h}} - \hat{\mathbf{h}}_{\mathcal{A}}\right) + + 2\frac{\sqrt{N}}{|\mathcal{A}|}\mathrm{Re}\left[\left(\bar{\mathbf{h}} - \hat{\mathbf{h}}_{\mathcal{A}}\right)^H\Theta^H\mathbf{P}_{\mathcal{A}}\mathbf{n}\right] + \frac{1}{|\mathcal{A}|}\mathrm{tr}\left[\mathbf{P}_{\mathcal{A}}\mathbf{n}\mathbf{n}^H\right]$$

By recalling (5.11), we can easily establish that

$$\left(\bar{\mathbf{h}} - \hat{\mathbf{h}}_{\mathcal{A}}\right)^H\Theta^H\mathbf{P}_{\mathcal{A}}\Theta\left(\bar{\mathbf{h}} - \hat{\mathbf{h}}_{\mathcal{A}}\right) = \widetilde{\sigma}^4_{\mathcal{A}}\frac{1}{N}\mathbf{n}^H\mathcal{Q}_{\mathcal{A}}\mathbf{P}_{\mathcal{A}}\mathcal{Q}_{\mathcal{A}}\mathbf{n} + o_p\left(\frac{1}{N}\right)$$

$$\frac{1}{\sqrt{N}}\left(\bar{\mathbf{h}} - \hat{\mathbf{h}}_{\mathcal{A}}\right)^H\Theta^H\mathbf{P}_{\mathcal{A}}\mathbf{n} = -\widetilde{\sigma}^2_{\mathcal{A}}\frac{1}{N}\mathbf{n}^H\mathcal{Q}_{\mathcal{A}}\mathbf{P}_{\mathcal{A}}\mathbf{n} + o_p\left(\frac{1}{N}\right)$$

where we have introduced the quantity

$$\mathcal{Q}_{\mathcal{A}} = \widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}}\Theta\left(\Theta^H\widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}}\Theta\right)^{-1}\Theta^H\widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}}. \tag{5.18}$$

After some manipulation, we see that

$$\frac{1}{|\mathcal{A}|}\bar{\mathbf{y}}^H\left(\hat{\mathbf{h}}_{\mathcal{A}}\right)\mathbf{P}_{\mathcal{A}}\bar{\mathbf{y}}\left(\hat{\mathbf{h}}_{\mathcal{A}}\right) = \left(\widetilde{\sigma}^2_{\mathcal{A}}\right)^2\frac{N}{|\mathcal{A}|}\frac{1}{N}\mathbf{n}^H\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{P}_{\mathcal{A}}\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{n} + o_p\left(\frac{1}{N}\right)$$

and, following the same methodology,

$$\frac{1}{N - |\mathcal{A}|}\bar{\mathbf{y}}^H\left(\hat{\mathbf{h}}_{\mathcal{A}}\right)\mathbf{P}^{\perp}_{\mathcal{A}}\bar{\mathbf{y}}\left(\hat{\mathbf{h}}_{\mathcal{A}}\right) = \left(\widetilde{\sigma}^2_{w,\mathcal{A}}\right)^2\frac{N}{N - |\mathcal{A}|}\frac{1}{N}\mathbf{n}^H\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{P}^{\perp}_{\mathcal{A}}\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{n} + o_p\left(\frac{1}{N}\right).$$

Furthermore, one can easily see that

$$\frac{\bar{\mathbf{y}}^H\left(\hat{\mathbf{h}}\right)\mathbf{P}_{\mathcal{A}}\bar{\mathbf{y}}\left(\hat{\mathbf{h}}\right)}{|\mathcal{A}|} = \widetilde{\sigma}^2_{\mathcal{A}} + O_p\left(\frac{1}{N}\right) \quad \text{and} \quad \frac{\bar{\mathbf{y}}^H\left(\hat{\mathbf{h}}\right)\mathbf{P}^{\perp}_{\mathcal{A}}\bar{\mathbf{y}}\left(\hat{\mathbf{h}}\right)}{N - |\mathcal{A}|} = \widetilde{\sigma}^2_{w,\mathcal{A}} + O_p\left(\frac{1}{N}\right).$$

Now, using a Taylor expansion of the logarithm we obtain

$$\begin{aligned}
\delta_{\mathcal{A}} &= \frac{|\mathcal{A}|}{N}\frac{1}{\widetilde{\sigma}^2_{\mathcal{A}}}\left(\frac{\bar{\mathbf{y}}^H\left(\hat{\mathbf{h}}_{\mathcal{A}}\right)\mathbf{P}_{\mathcal{A}}\bar{\mathbf{y}}\left(\hat{\mathbf{h}}_{\mathcal{A}}\right)}{|\mathcal{A}|} - \widetilde{\sigma}^2_{\mathcal{A}}\right) \\
&+ \left(1-\frac{|\mathcal{A}|}{N}\right)\frac{1}{\widetilde{\sigma}^2_{w,\mathcal{A}}}\left(\frac{\bar{\mathbf{y}}^H\left(\hat{\mathbf{h}}_{\mathcal{A}}\right)\mathbf{P}^{\perp}_{\mathcal{A}}\bar{\mathbf{y}}\left(\hat{\mathbf{h}}_{\mathcal{A}}\right)}{N-|\mathcal{A}|} - \widetilde{\sigma}^2_{w,\mathcal{A}}\right) + O_p\left(\frac{1}{N^2}\right) \\
&= \frac{|\mathcal{A}|}{N}\left(\widetilde{\sigma}^2_{\mathcal{A}}\frac{N}{|\mathcal{A}|}\frac{1}{N}\mathbf{n}^H\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{P}_{\mathcal{A}}\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{n} - 1\right) \\
&+ \left(1-\frac{|\mathcal{A}|}{N}\right)\left(\widetilde{\sigma}^2_{w,\mathcal{A}}\frac{N}{N-|\mathcal{A}|}\frac{1}{N}\mathbf{n}^H\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{P}^{\perp}_{\mathcal{A}}\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{n} - 1\right) + o_p\left(\frac{1}{N}\right) \\
&= \frac{1}{N}\mathbf{n}^H\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{n} - 1 + o_p\left(\frac{1}{N}\right)
\end{aligned}$$

which directly leads to (5.17). ∎

Observe that we can write,if $j \notin \mathcal{A}$,

$$\zeta^{(\mathcal{A})}_{RML} - \zeta^{(\mathcal{A}\cup\{j\})}_{RML} = \mathcal{V}_1 + \mathcal{V}_2 + \mathcal{V}_3 + o_p\left(\frac{1}{N}\right)$$

where we have defined

$$\begin{aligned}
\mathcal{V}_1 &= \frac{1}{N}\log\det\left(\widetilde{\mathbf{C}}_{n,\mathcal{A}}\right) - \frac{1}{N}\log\det\left(\widetilde{\mathbf{C}}_{n,\mathcal{A}\cup\{j\}}\right) \\
\mathcal{V}_2 &= \frac{1}{N}\mathbf{n}^H\left(\widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}} - \widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}\cup\{j\}}\right)\mathbf{n} \\
\mathcal{V}_3 &= \frac{1}{N}\mathbf{n}^H\left(\mathcal{Q}_{\mathcal{A}\cup\{j\}} - \mathcal{Q}_{\mathcal{A}}\right)\mathbf{n}
\end{aligned}$$

and where $\mathcal{Q}_{\mathcal{A}}$ is defined in (5.18). We analyze these three terms separately in what follows. We will use the fact that

$$\begin{aligned}
\widetilde{\mathbf{C}}_{n,\mathcal{A}\cup\{j\}} &= \widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}}\mathbf{P}_{\mathcal{A}} + \widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}}\mathbf{e}_j\mathbf{e}_j^T + \widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}\mathbf{P}^{\perp}_{\mathcal{A}\cup\{j\}} \\
\widetilde{\mathbf{C}}_{n,\mathcal{A}} &= \widetilde{\sigma}^2_{\mathcal{A}}\mathbf{P}_{\mathcal{A}} + \widetilde{\sigma}^2_{w,\mathcal{A}}\mathbf{e}_j\mathbf{e}_j^T + \widetilde{\sigma}^2_{w,\mathcal{A}}\mathbf{P}^{\perp}_{\mathcal{A}\cup\{j\}}.
\end{aligned}$$

Next, we proceed to study the asymptotic behavior of these three terms, up to error terms that decay to zero in probability faster than N^{-1}.

5.6.1 Study of $\mathcal{V}_1$

Using the structure of matrices $\widetilde{\mathbf{C}}_{n,\mathcal{A}\cup\{j\}}$ and $\widetilde{\mathbf{C}}_{n,\mathcal{A}}$ above, it readily follows that

$$\begin{aligned}\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}} &= \frac{|\mathcal{A}|}{|\mathcal{A}|+1}\widetilde{\sigma}^2_{\mathcal{A}} + \frac{1}{|\mathcal{A}|+1}\left\{\bar{\mathbf{C}}_n\right\}_{jj},\\ \widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}} &= \frac{N-|\mathcal{A}|}{N-|\mathcal{A}|-1}\widetilde{\sigma}^2_{w,\mathcal{A}} - \frac{1}{N-|\mathcal{A}|-1}\left\{\bar{\mathbf{C}}_n\right\}_{jj}.\end{aligned} \tag{5.19}$$

Therefore, using the Taylor expansion of the logarithm, we obtain

$$\begin{aligned}\frac{1}{N}\log\det\widetilde{\mathbf{C}}_{n,\mathcal{A}\cup\{j\}} &= \frac{|\mathcal{A}|+1}{N}\log\left(\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}}\right) + \frac{N-|\mathcal{A}|-1}{N}\log\left(\widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}\right)\\ &= \frac{1}{N}\log\det\widetilde{\mathbf{C}}_{n,\mathcal{A}} + \frac{1}{N}\log\left(\widetilde{\sigma}^2_{\mathcal{A}}\right) - \frac{1}{N}\log\left(\widetilde{\sigma}^2_{w,\mathcal{A}}\right)\\ &\quad + \frac{1}{N}\left(\frac{\left\{\bar{\mathbf{C}}_n\right\}_{jj}}{\widetilde{\sigma}^2_{\mathcal{A}}} - 1\right) + \frac{1}{N}\left(1 - \frac{\left\{\bar{\mathbf{C}}_n\right\}_{jj}}{\widetilde{\sigma}^2_{w,\mathcal{A}}}\right) + o\left(\frac{1}{N}\right).\end{aligned}$$

This readily shows that

$$\begin{aligned}\mathcal{V}_1 = &-\frac{1}{N}\log\left(\widetilde{\sigma}^2_{\mathcal{A}}\right) + \frac{1}{N}\log\left(\widetilde{\sigma}^2_{w,\mathcal{A}}\right)\\ &- \frac{1}{N}\left(\frac{\left\{\bar{\mathbf{C}}_n\right\}_{jj}}{\widetilde{\sigma}^2_{\mathcal{A}}} - 1\right) - \frac{1}{N}\left(1 - \frac{\left\{\bar{\mathbf{C}}_n\right\}_{jj}}{\widetilde{\sigma}^2_{w,\mathcal{A}}}\right) + o\left(\frac{1}{N}\right)\end{aligned}$$

5.6.2 Study of $\mathcal{V}_2$

Using the expression of the inverses of $\widetilde{\mathbf{C}}_{n,\mathcal{A}}$ and $\widetilde{\mathbf{C}}_{n,\mathcal{A}\cup\{j\}}$ together with the identities in (5.19) we obtain

$$\begin{aligned}\widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}} - \widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}\cup\{j\}} &= \frac{1}{\widetilde{\sigma}^2_{\mathcal{A}}}\left(1 - \frac{\widetilde{\sigma}^2_{\mathcal{A}}}{\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}}}\right)\mathbf{P}_{\mathcal{A}} + \frac{1}{\widetilde{\sigma}^2_{w,\mathcal{A}}}\left(1 - \frac{\widetilde{\sigma}^2_{w,\mathcal{A}}}{\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}}}\right)\mathbf{e}_j\mathbf{e}_j^T\\ &\quad + \frac{1}{\widetilde{\sigma}^2_{w,\mathcal{A}}}\left(1 - \frac{\widetilde{\sigma}^2_{w,\mathcal{A}}}{\widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}}\right)\mathbf{P}^{\perp}_{\mathcal{A}\cup\{j\}}\\ &= \left(\frac{\left\{\bar{\mathbf{C}}_n\right\}_{jj}}{\widetilde{\sigma}^2_{\mathcal{A}}} - 1\right)\frac{\mathbf{P}_{\mathcal{A}}}{|\mathcal{A}|\,\widetilde{\sigma}^2_{\mathcal{A}} + \left\{\bar{\mathbf{C}}_n\right\}_{jj}}\\ &\quad + \left(1 - \frac{\left\{\bar{\mathbf{C}}_n\right\}_{jj}}{\widetilde{\sigma}^2_{w,\mathcal{A}}}\right)\frac{\mathbf{P}^{\perp}_{\mathcal{A}}}{(N-|\mathcal{A}|)\,\widetilde{\sigma}^2_{w,\mathcal{A}} - \left\{\bar{\mathbf{C}}_n\right\}_{jj}}\\ &\quad + \frac{1}{\widetilde{\sigma}^2_{w,\mathcal{A}}}\left(\frac{(N-|\mathcal{A}|-1)\,\widetilde{\sigma}^2_{w,\mathcal{A}}}{(N-|\mathcal{A}|)\,\widetilde{\sigma}^2_{w,\mathcal{A}} - \left\{\bar{\mathbf{C}}_n\right\}_{jj}} - \frac{(|\mathcal{A}|+1)\,\widetilde{\sigma}^2_{w,\mathcal{A}}}{|\mathcal{A}|\,\widetilde{\sigma}^2_{\mathcal{A}} + \left\{\bar{\mathbf{C}}_n\right\}_{jj}}\right)\mathbf{e}_j\mathbf{e}_j^T.\end{aligned} \tag{5.20}$$

Therefore, multiplying on both sides by the noise vector we obtain

$$\mathcal{V}_2 = \frac{1}{N}\left(\frac{\left\{\bar{\mathbf{C}}_n\right\}_{jj}}{\widetilde{\sigma}^2_{\mathcal{A}}} - 1\right)\frac{\mathbf{n}^H\mathbf{P}_{\mathcal{A}}\mathbf{n}}{|\mathcal{A}|\,\widetilde{\sigma}^2_{\mathcal{A}}} + \frac{1}{N}\left(1 - \frac{\left\{\bar{\mathbf{C}}_n\right\}_{jj}}{\widetilde{\sigma}^2_{w,\mathcal{A}}}\right)\frac{\mathbf{n}^H\mathbf{P}^{\perp}_{\mathcal{A}}\mathbf{n}}{(N-|\mathcal{A}|)\,\widetilde{\sigma}^2_{w,\mathcal{A}}} + \frac{1}{N}\frac{\widetilde{\sigma}^2_{\mathcal{A}} - \widetilde{\sigma}^2_{w,\mathcal{A}}}{\widetilde{\sigma}^2_{\mathcal{A}}\widetilde{\sigma}^2_{w,\mathcal{A}}}\mathbf{n}^H\mathbf{e}_j\mathbf{e}_j^T\mathbf{n} + o_p\left(\frac{1}{N}\right).$$

5.6.3 Study of $\mathcal{V}_3$

The main idea of the proof is based on the Cauchy integral description of the projection matrix. Recall that, for a full column rank tall matrix $\mathbf{A}$, we denote $\mathcal{P}_{\mathbf{A}} = \mathbf{A}\left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H$. This projection matrix $\mathcal{P}_{\mathbf{A}}$ admits a very convenient integral expression given by

$$\mathcal{P}_{\mathbf{A}} = \frac{1}{2\pi\mathrm{j}}\oint_{C^-}\left(\mathbf{A}\mathbf{A}^H - z\mathbf{I}_N\right)^{-1}dz$$

where C^- is a negatively oriented simple contour enclosing the eigenvalues of $\mathbf{A}^H\mathbf{A}$ and not zero (this can be shown by conventional Cauchy residue calculus). Using this result, we can easily express

$$\mathcal{Q}_{\mathcal{A}} = \frac{1}{2\pi\mathrm{j}}\oint_{C^-}\mathbf{Q}_{\mathcal{A}}(z)dz$$

where $\mathbf{Q}_{\mathcal{A}}(z)$ denotes the resolvent matrix defined as

$$\mathbf{Q}_{\mathcal{A}}(z) = \left(\Theta\Theta^H - z\widetilde{\mathbf{C}}_{n,\mathcal{A}}\right)^{-1}.$$

In order to simplify the following derivations, it becomes convenient to define a new matrix $\widetilde{\mathbf{C}}^{(j)}_{n,\mathcal{A}}$ as

$$\widetilde{\mathbf{C}}^{(j)}_{n,\mathcal{A}} = \widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}}\mathbf{P}_{\mathcal{A}} + \widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}\mathbf{P}^{\perp}_{\mathcal{A}}.$$

This matrix can be seen as a very small perturbation between $\widetilde{\mathbf{C}}_{n,\mathcal{A}}$ and $\widetilde{\mathbf{C}}_{n,\mathcal{A}\cup\{j\}}$.

Using this auxiliary matrix and its corresponding resolvent, namely

$$\mathbf{Q}^{(j)}_{\mathcal{A}}(z) = \left(\Theta\Theta^H - z\widetilde{\mathbf{C}}^{(j)}_{n,\mathcal{A}}\right)^{-1}$$

we can write $\mathcal{V}_3 = \mathcal{V}_{31} + \mathcal{V}_{32}$ where

$$\mathcal{V}_{31} = \frac{1}{2\pi\mathrm{j}}\frac{1}{N}\oint_{C^-}\mathbf{n}^H\left[\mathbf{Q}_{\mathcal{A}\cup\{j\}}(z) - \mathbf{Q}^{(j)}_{\mathcal{A}}(z)\right]\mathbf{n}dz \tag{5.21}$$

$$\mathcal{V}_{32} = \frac{1}{2\pi\mathrm{j}}\frac{1}{N}\oint_{C^-}\mathbf{n}^H\left[\mathbf{Q}^{(j)}_{\mathcal{A}}(z) - \mathbf{Q}_{\mathcal{A}}(z)\right]\mathbf{n}dz \tag{5.22}$$

where now C^- is a negatively oriented simple contour that encloses all the eigenvalues of $\Theta^H \widetilde{\mathbf{C}}_{n,\mathcal{A}}^{-1}\Theta$, $\Theta^H \widetilde{\mathbf{C}}_{n,\mathcal{A}\cup\{j\}}^{-1}\Theta$ and $\Theta^H \left(\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{(j)}\right)^{-1}\Theta$, and not zero. Observe that these three matrices are positive definite ($\Theta^H\Theta > 0$ by assumption) and therefore do not have zero as eigenvalues.

We will next deal with $\mathcal{V}_{31}$ and $\mathcal{V}_{32}$ separately. As we will see, the techniques used to expand these two quantities are somewhat different.

5.6.3.1 Study of $\mathcal{V}_{31}$

In order to study the quantity $\mathcal{V}_{31}$ we first make use of the resolvent identity[4] and express

$$\mathbf{Q}_{\mathcal{A}\cup\{j\}}(z) - \mathbf{Q}_{\mathcal{A}}^{(j)}(z) = z\mathbf{Q}_{\mathcal{A}\cup\{j\}}(z)\left[\widetilde{\mathbf{C}}_{n,\mathcal{A}\cup\{j\}} - \widetilde{\mathbf{C}}_{n,\mathcal{A}}^{(j)}\right]\mathbf{Q}_{\mathcal{A}}^{(j)}(z).$$

Now, using the expression of $\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{(j)}$ and $\widetilde{\mathbf{C}}_{n,\mathcal{A}\cup\{j\}}$, we obtain

$$\widetilde{\mathbf{C}}_{n,\mathcal{A}\cup\{j\}} - \widetilde{\mathbf{C}}_{n,\mathcal{A}}^{(j)} = \left(\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}\right)\mathbf{e}_j\mathbf{e}_j^T.$$

The following proposition provides an easy-to-use tool to integrate several quantities that will appear in the rest of the proof.

Proposition 3. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *denote two* $N \times K$ *matrices such that* $\mathbf{A}^H\mathbf{A} > \mathbf{0}$ *and* $\mathbf{B}^H\mathbf{B} > \mathbf{0}$. *Consider the two resolvents* $\mathbf{Q}_A(z) = \left(\mathbf{A}\mathbf{A}^H - z\mathbf{I}_N\right)^{-1}$ *and* $\mathbf{Q}_B(z) = \left(\mathbf{B}\mathbf{B}^H - z\mathbf{I}_N\right)^{-1}$ *and let* C^- *denote a simple negatively oriented contour enclosing the eigenvalues of* $\mathbf{A}^H\mathbf{A}$ *and* $\mathbf{B}^H\mathbf{B}$ *but not zero. Then, for any* $N \times N$ *matrix* $\mathbf{C}$ *we have*

$$\frac{1}{2\pi\mathrm{j}}\oint_{C^-} z\mathbf{Q}_A(z)\mathbf{C}\mathbf{Q}_B(z)dz = \mathcal{P}_{\mathbf{A}}^{\perp}\mathbf{C}\mathcal{P}_{\mathbf{B}}^{\perp} - \mathbf{C} = \mathcal{P}_{\mathbf{A}}\mathbf{C}\mathcal{P}_{\mathbf{B}} - \mathcal{P}_{\mathbf{A}}\mathbf{C} - \mathbf{C}\mathcal{P}_{\mathbf{B}}$$

where $\mathcal{P}_{\mathbf{A}}^{\perp} = \mathbf{I}_N - \mathbf{A}\left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H$ *and equivalently for* $\mathcal{P}_{\mathbf{B}}^{\perp}$. *On the other hand, if* C_0^- *is a simple negatively oriented contour enclosing zero but no eigenvalues of either* $\mathbf{A}^H\mathbf{A}$ *or* $\mathbf{B}^H\mathbf{B}$, *we have*

$$\frac{1}{2\pi\mathrm{j}}\oint_{C_0^-} z\mathbf{Q}_A(z)\mathbf{C}\mathbf{Q}_B(z)dz = -\mathcal{P}_{\mathbf{A}}^{\perp}\mathbf{C}\mathcal{P}_{\mathbf{B}}^{\perp}.$$

Proof. Consider the eigendecompositions of $\mathbf{A}\mathbf{A}^H$ and $\mathbf{B}\mathbf{B}^H$ as

$$\mathbf{A}\mathbf{A}^H = \sum_{n=1}^{K}\lambda_n\mathbf{u}_n\mathbf{u}_n^H + 0\times\mathcal{P}_{\mathbf{A}}^{\perp} \quad \text{and} \quad \mathbf{B}\mathbf{B}^H = \sum_{n=1}^{K}\gamma_n\mathbf{v}_n\mathbf{v}_n^H + 0\times\mathcal{P}_{\mathbf{B}}^{\perp}$$

[4] $\mathbf{Q}_1 - \mathbf{Q}_2 = \mathbf{Q}_1\left[\mathbf{Q}_2^{-1} - \mathbf{Q}_1^{-1}\right]\mathbf{Q}_2$

where $\mathcal{P}^{\perp}_{\mathbf{A}}$ and $\mathcal{P}^{\perp}_{\mathbf{B}}$ are defined as above and $0\times$ symbolizes the fact that the associated eigenvalue is zero. We can write

$$\mathbf{Q}_A(z) = \sum_{n=1}^{K} \frac{1}{\lambda_n - z}\mathbf{u}_n\mathbf{u}_n^H - \frac{1}{z}\mathcal{P}^{\perp}_{\mathbf{A}} \tag{5.23}$$

$$\mathbf{Q}_B(z) = \sum_{m=1}^{K} \frac{1}{\gamma_m - z}\mathbf{v}_m\mathbf{v}_m^H - \frac{1}{z}\mathcal{P}^{\perp}_{\mathbf{B}}. \tag{5.24}$$

Inserting these expressions into the integral and using the fact that C does not enclose $z = 0$, we see that

$$\begin{aligned}\frac{1}{2\pi\mathrm{j}}\oint_{C^-} z\mathbf{Q}_A(z)\mathbf{C}\mathbf{Q}_B(z)dz = &\sum_{m=1}^{K}\sum_{n=1}^{K}\mathbf{u}_n\mathbf{u}_n^H\mathbf{C}\mathbf{v}_m\mathbf{v}_m^H\frac{1}{2\pi\mathrm{j}}\oint_{C^-} z\frac{1}{\lambda_n - z}\frac{1}{\gamma_m - z}dz \\ &- \mathcal{P}_{\mathbf{A}}\mathbf{C}\mathcal{P}^{\perp}_{\mathbf{B}} - \mathcal{P}^{\perp}_{\mathbf{A}}\mathbf{C}\mathcal{P}_{\mathbf{B}}\end{aligned}$$

where we have applied the Cauchy residue theorem, which implies that

$$\frac{1}{2\pi\mathrm{j}}\oint_{C^-}\left(\frac{1}{\gamma_m - z}\right)dz = 1.$$

The result of the proposition follows by showing that

$$\frac{1}{2\pi\mathrm{j}}\oint_{C^-} z\frac{1}{\lambda_n - z}\frac{1}{\gamma_m - z}dz = -1$$

Indeed, if we first assume that $\gamma_m \neq \lambda_n$ we can write the above integral as

$$\frac{1}{\gamma_m - \lambda_n}\frac{1}{2\pi\mathrm{j}}\oint_{C^-} z\left(\frac{1}{\lambda_n - z} - \frac{1}{\gamma_m - z}\right)dz = \frac{\lambda_n - \gamma_m}{\gamma_m - \lambda_n} = -1.$$

In the same way, if we assume that $\lambda_n = \gamma_m$, we have

$$\frac{1}{2\pi\mathrm{j}}\oint_{C^-}\frac{z}{(\lambda_n - z)^2}dz = -1$$

and the first result is proven. Regarding the second result, it follows from the same reasoning and the fact that $\mathbf{A}^H\mathbf{A} > \mathbf{0}$ and $\mathbf{B}^H\mathbf{B} > \mathbf{0}$. ∎

A direct application of the above proposition shows that

$$\begin{aligned}\frac{1}{2\pi\mathrm{j}}\oint_{C^-}\left(\mathbf{Q}_{\mathcal{A}\cup\{j\}}(z) - \mathbf{Q}^{(j)}_{\mathcal{A}}(z)\right)dz = &\left(\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}\right)\times \\ &\left[\mathcal{Q}^{\perp}_{\mathcal{A}\cup\{j\}}\mathbf{e}_j\mathbf{e}_j^T\mathcal{Q}^{\perp}_{\mathcal{A}^{(j)}} - \widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}\cup\{j\}}\mathbf{e}_j\mathbf{e}_j^T\left(\widetilde{\mathbf{C}}^{(j)}_{n,\mathcal{A}}\right)^{-1}\right]\end{aligned}$$

where $\mathcal{Q}^{\perp}_{\mathcal{A}\cup\{j\}}$ is as defined in (5.14) and where we have defined

$$\begin{aligned}\mathcal{Q}^{\perp}_{\mathcal{A}^{(j)}} &= \left(\widetilde{\mathbf{C}}^{(j)}_{n,\mathcal{A}}\right)^{-1} - \mathcal{Q}_{\mathcal{A}^{(j)}} \\ &= \left(\widetilde{\mathbf{C}}^{(j)}_{n,\mathcal{A}}\right)^{-1} - \left(\widetilde{\mathbf{C}}^{(j)}_{n,\mathcal{A}}\right)^{-1}\Theta\left(\Theta^H\left(\widetilde{\mathbf{C}}^{(j)}_{n,\mathcal{A}}\right)^{-1}\Theta\right)^{-1}\Theta^H \times \left(\widetilde{\mathbf{C}}^{(j)}_{n,\mathcal{A}}\right)^{-1}.\end{aligned}$$

Next, observe that, since $j \notin \mathcal{A}$,

$$\widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}\cup\{j\}} = \left(\widetilde{\mathbf{C}}^{(j)}_{n,\mathcal{A}}\right)^{-1} + \frac{\widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}}}{\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}}\widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}}\mathbf{e}_j\mathbf{e}_j^T$$

and this implies that (by the matrix inversion lemma)

$$\mathcal{Q}^{\perp}_{\mathcal{A}\cup\{j\}} = \mathcal{Q}^{\perp}_{\mathcal{A}^{(j)}} - \frac{\left(\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}\right)}{1 + \left(\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}\right)\mathbf{e}_j^T\mathcal{Q}^{\perp}_{\mathcal{A}^{(j)}}\mathbf{e}_j}\mathcal{Q}^{\perp}_{\mathcal{A}^{(j)}}\mathbf{e}_j\mathbf{e}_j^T\mathcal{Q}^{\perp}_{\mathcal{A}^{(j)}}$$

Using this, we can re-write the above expression as

$$\begin{aligned}&\frac{1}{2\pi\mathrm{j}}\oint_{C^-}\left(\mathbf{Q}_{\mathcal{A}\cup\{j\}}(z) - \mathbf{Q}^{(j)}_{\mathcal{A}}(z)\right)dz = \\ &= \frac{\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}}{1 + \left(\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}\right)\mathbf{e}_j^T\mathcal{Q}^{\perp}_{\mathcal{A}^{(j)}}\mathbf{e}_j}\mathcal{Q}^{\perp}_{\mathcal{A}^{(j)}}\mathbf{e}_j\mathbf{e}_j^T\mathcal{Q}^{\perp}_{\mathcal{A}^{(j)}} - \frac{\left(\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}\right)}{\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}}\widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}}\mathbf{e}_j\mathbf{e}_j^T\end{aligned}$$

and we can therefore conclude that

$$\begin{aligned}\mathcal{V}_{31} = &\frac{\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}}{\frac{\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}}}{\widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}} - \left(\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}\right)\mathbf{e}_j^T\mathcal{Q}_{\mathcal{A}^{(j)}}\mathbf{e}_j}\times \\ &\times\left[\frac{1}{N}\mathbf{n}^H\mathcal{Q}_{\mathcal{A}^{(j)}}\mathbf{e}_j\mathbf{e}_j^T\mathcal{Q}_{\mathcal{A}^{(j)}}\mathbf{n} - \frac{1}{\widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}}\frac{1}{N}\mathbf{n}^H\mathbf{e}_j\mathbf{e}_j^T\mathcal{Q}_{\mathcal{A}^{(j)}}\mathbf{n}\right. \\ &\left. - \frac{1}{\widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}}\frac{1}{N}\mathbf{n}^H\mathcal{Q}_{\mathcal{A}^{(j)}}\mathbf{e}_j\mathbf{e}_j^T\mathbf{n} + \mathbf{e}_j^T\mathcal{Q}_{\mathcal{A}^{(j)}}\mathbf{e}_j\frac{\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}}{\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}}\widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}}\frac{1}{N}\mathbf{n}^H\mathbf{e}_j\mathbf{e}_j^T\mathbf{n}\right].\end{aligned}$$

5.6.3.2 Study of $\mathcal{V}_{32}$

We begin again by employing the resolvent identity to show that

$$\mathbf{Q}^{(j)}_{\mathcal{A}}(z) - \mathbf{Q}_{\mathcal{A}}(z) = z\mathbf{Q}^{(j)}_{\mathcal{A}}(z)\left(\widetilde{\mathbf{C}}^{(j)}_{n,\mathcal{A}} - \widetilde{\mathbf{C}}_{n,\mathcal{A}}\right)\mathbf{Q}_{\mathcal{A}}(z) \tag{5.25}$$

where we recall that

$$\widetilde{\mathbf{C}}^{(j)}_{n,\mathcal{A}} - \widetilde{\mathbf{C}}_{n,\mathcal{A}} = \left(\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{\mathcal{A}}\right)\mathbf{P}_{\mathcal{A}} + \left(\widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{w,\mathcal{A}}\right)\mathbf{P}^{\perp}_{\mathcal{A}}.$$

Now, a direct application of Proposition 3 shows that

$$\begin{aligned}
&\frac{1}{2\pi \mathrm{j}} \oint_{C^-} \left(\mathbf{Q}_{\mathcal{A}}^{(j)}(z) - \mathbf{Q}_{\mathcal{A}}(z) \right) dz \\
&= \mathcal{Q}_{\mathcal{A}^{(j)}}^{\perp} \left(\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{(j)} - \widetilde{\mathbf{C}}_{n,\mathcal{A}} \right) \mathcal{Q}_{\mathcal{A}}^{\perp} - \left(\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{(j)} \right)^{-1} \left(\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{(j)} - \widetilde{\mathbf{C}}_{n,\mathcal{A}} \right) \widetilde{\mathbf{C}}_{n,\mathcal{A}}^{-1} \\
&= \left(\widetilde{\sigma}_{\mathcal{A}\cup\{j\}}^2 - \widetilde{\sigma}_{\mathcal{A}}^2 \right) \left[\mathcal{Q}_{\mathcal{A}^{(j)}}^{\perp} \mathbf{P}_{\mathcal{A}} \mathcal{Q}_{\mathcal{A}}^{\perp} - \left(\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{(j)} \right)^{-1} \mathbf{P}_{\mathcal{A}} \widetilde{\mathbf{C}}_{n,\mathcal{A}}^{-1} \right] \\
&+ \left(\widetilde{\sigma}_{w,\mathcal{A}\cup\{j\}}^2 - \widetilde{\sigma}_{w,\mathcal{A}}^2 \right) \left[\mathcal{Q}_{\mathcal{A}^{(j)}}^{\perp} \mathbf{P}_{\mathcal{A}}^{\perp} \mathcal{Q}_{\mathcal{A}}^{\perp} - \left(\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{(j)} \right)^{-1} \mathbf{P}_{\mathcal{A}}^{\perp} \widetilde{\mathbf{C}}_{n,\mathcal{A}}^{-1} \right]
\end{aligned}$$

This allows to conclude that

$$\begin{aligned}
\mathcal{V}_{32} = &\left(\widetilde{\sigma}_{\mathcal{A}\cup\{j\}}^2 - \widetilde{\sigma}_{\mathcal{A}}^2 \right) \frac{1}{N} \mathbf{n}^H \left(\mathcal{Q}_{\mathcal{A}^{(j)}}^{\perp} \mathbf{P}_{\mathcal{A}} \mathcal{Q}_{\mathcal{A}}^{\perp} - \left(\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{(j)} \right)^{-1} \mathbf{P}_{\mathcal{A}} \widetilde{\mathbf{C}}_{n,\mathcal{A}}^{-1} \right) \mathbf{n} \\
&+ \left(\widetilde{\sigma}_{w,\mathcal{A}\cup\{j\}}^2 - \widetilde{\sigma}_{w,\mathcal{A}}^2 \right) \frac{1}{N} \mathbf{n}^H \left(\mathcal{Q}_{\mathcal{A}^{(j)}}^{\perp} \mathbf{P}_{\mathcal{A}}^{\perp} \mathcal{Q}_{\mathcal{A}}^{\perp} - \left(\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{(j)} \right)^{-1} \mathbf{P}_{\mathcal{A}}^{\perp} \widetilde{\mathbf{C}}_{n,\mathcal{A}}^{-1} \right) \mathbf{n}
\end{aligned}$$

5.6.4 Completing the study of $\mathcal{V}_3$

To complete the proof, we now show that we can asymptotically replace $\mathcal{Q}_{\mathcal{A}^{(j)}}^{\perp}$ by $\mathcal{Q}_{\mathcal{A}}^{\perp}$ in the above expressions. This can be ensured by noting that

$$\mathcal{Q}_{\mathcal{A}^{(j)}}^{\perp} = \frac{1}{2\pi \mathrm{j}} \oint_{C_0^-} \mathbf{Q}_{\mathcal{A}}^{(j)}(z) dz$$

where C_0^- is a negatively oriented contour that encloses zero and no eigenvalue of either $\Theta^H \widetilde{\mathbf{C}}_{n,\mathcal{A}}^{-1} \Theta$ or $\Theta^H \left(\widetilde{\mathbf{C}}_{n,\mathcal{A}}^{(j)} \right)^{-1} \Theta$. A very similar expression holds for $\mathcal{Q}_{\mathcal{A}}^{\perp}$. Now, the resolvent identity in (5.25) and the corresponding result in Proposition 3 shows that

$$\mathcal{Q}_{\mathcal{A}^{(j)}}^{\perp} - \mathcal{Q}_{\mathcal{A}}^{\perp} = -\mathcal{Q}_{\mathcal{A}^{(j)}}^{\perp} \Delta \mathcal{Q}_{\mathcal{A}}^{\perp}$$

with

$$\Delta = \left(\widetilde{\sigma}_{\mathcal{A}\cup\{j\}}^2 - \widetilde{\sigma}_{\mathcal{A}}^2 \right) \mathbf{P}_{\mathcal{A}} + \left(\widetilde{\sigma}_{w,\mathcal{A}\cup\{j\}}^2 - \widetilde{\sigma}_{w,\mathcal{A}}^2 \right) \mathbf{P}_{\mathcal{A}}^{\perp}.$$

Therefore, for N large enough we can write

$$\mathcal{Q}_{\mathcal{A}^{(j)}}^{\perp} = \mathcal{Q}_{\mathcal{A}}^{\perp} \left(\mathbf{I} + \Delta \mathcal{Q}_{\mathcal{A}}^{\perp} \right)^{-1} = \mathcal{Q}_{\mathcal{A}}^{\perp} + \sum_{n=1}^{\infty} \left(-\Delta \mathcal{Q}_{\mathcal{A}}^{\perp} \right)^n$$

where the last series converges when $\left\|\Delta\mathcal{Q}_{\mathcal{A}}^{\perp}\right\| < 1$. This always holds for N sufficiently high, because

$$\begin{aligned}\left\|\Delta\mathcal{Q}_{\mathcal{A}}^{\perp}\right\| &\leq \left(\left|\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{\mathcal{A}}\right| + \left|\widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{w,\mathcal{A}}\right|\right)\left\|\mathcal{Q}_{\mathcal{A}}^{\perp}\right\| \\ &\leq \left(\left|\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{\mathcal{A}}\right| + \left|\widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{w,\mathcal{A}}\right|\right)\left\|\left(\widetilde{\mathbf{C}}^{(j)}_{n,\mathcal{A}}\right)^{-1}\right\| \\ &= \frac{\frac{1}{|\mathcal{A}|+1}\left|\left\{\bar{\mathbf{C}}_n\right\}_{jj} - \widetilde{\sigma}^2_{\mathcal{A}}\right| + \frac{1}{N-|\mathcal{A}|-1}\left|\widetilde{\sigma}^2_{w,\mathcal{A}} - \left\{\bar{\mathbf{C}}_n\right\}_{jj}\right|}{\min\left(\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}}, \widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}\right)} = O\left(N^{-1}\right)\end{aligned}$$

where we have used the identity in (5.19). Now, this also implies that

$$\left\|\mathcal{Q}^{\perp}_{\mathcal{A}^{(j)}} - \mathcal{Q}^{\perp}_{\mathcal{A}}\right\| = \left\|\mathcal{Q}^{\perp}_{\mathcal{A}}\Delta\mathcal{Q}^{\perp}_{\mathcal{A}}\left(\mathbf{I} + \Delta\mathcal{Q}^{\perp}_{\mathcal{A}}\right)^{-1}\right\| = O\left(N^{-1}\right)$$

and consequently

$$\begin{aligned}\frac{1}{N}\mathbf{n}^H\mathcal{Q}^{\perp}_{\mathcal{A}^{(j)}}\mathbf{P}_{\mathcal{A}}\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{n} &= \frac{1}{N}\mathbf{n}^H\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{P}_{\mathcal{A}}\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{n} + O\left(N^{-1}\right) \\ \frac{1}{N}\mathbf{n}^H\mathcal{Q}^{\perp}_{\mathcal{A}^{(j)}}\mathbf{P}^{\perp}_{\mathcal{A}}\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{n} &= \frac{1}{N}\mathbf{n}^H\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{P}^{\perp}_{\mathcal{A}}\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{n} + O\left(N^{-1}\right).\end{aligned}$$

This directly shows that

$$\begin{aligned}\mathcal{V}_{32} = &\left(\frac{\left\{\bar{\mathbf{C}}_n\right\}_{jj} - \widetilde{\sigma}^2_{\mathcal{A}}}{|\mathcal{A}|}\right)\frac{1}{N}\mathbf{n}^H\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{P}_{\mathcal{A}}\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{n} - \frac{1}{\widetilde{\sigma}^2_{\mathcal{A}}}\frac{1}{|\mathcal{A}|}\left(\frac{\left\{\bar{\mathbf{C}}_n\right\}_{jj}}{\widetilde{\sigma}^2_{\mathcal{A}}} - 1\right)\frac{1}{N}\mathbf{n}^H\mathbf{P}_{\mathcal{A}}\mathbf{n} \\ &+ \left(\frac{\widetilde{\sigma}^2_{w,\mathcal{A}} - \left\{\bar{\mathbf{C}}_n\right\}_{jj}}{N-|\mathcal{A}|}\right)\frac{1}{N}\mathbf{n}^H\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{P}^{\perp}_{\mathcal{A}}\mathcal{Q}^{\perp}_{\mathcal{A}}\mathbf{n} \\ &- \frac{1}{\widetilde{\sigma}^2_{w,\mathcal{A}}}\frac{1}{N-|\mathcal{A}|}\left(1 - \frac{\left\{\bar{\mathbf{C}}_n\right\}_{jj}}{\widetilde{\sigma}^2_{w,\mathcal{A}}}\right)\frac{1}{N}\mathbf{n}^H\mathbf{P}^{\perp}_{\mathcal{A}}\mathbf{n} + o_p(N^{-1})\end{aligned}$$

On the other hand, we obviously have

$$\mathbf{e}_j^T\mathcal{Q}_{\mathcal{A}^{(j)}}\mathbf{e}_j = \frac{\widetilde{\sigma}^2_{w,\mathcal{A}}}{\widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}}\mathbf{e}_j^T\widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}}\Theta\left(\Theta^H\widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}}\Theta + \epsilon^{(j)}_{\mathcal{A}}\Theta^H_{\mathcal{A}}\Theta_{\mathcal{A}}\right)^{-1}\Theta^H\widetilde{\mathbf{C}}^{-1}_{n,\mathcal{A}}\mathbf{e}_j$$

where $\Theta_{\mathcal{A}}$ is an $|\mathcal{A}| \times L$ matrix containing the rows of Θ as indexed by $\mathcal{A}$ and where

$$\epsilon^{(j)}_{\mathcal{A}} = \frac{\widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}}\left(\widetilde{\sigma}^2_{\mathcal{A}} - \widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}}\right) + \widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}}\left(\widetilde{\sigma}^2_{w,\mathcal{A}\cup\{j\}} - \widetilde{\sigma}^2_{w,\mathcal{A}}\right)}{\widetilde{\sigma}^2_{w,\mathcal{A}}\widetilde{\sigma}^2_{\mathcal{A}\cup\{j\}}\widetilde{\sigma}^2_{\mathcal{A}}}.$$

Note that, by assumption $\left\|\Theta^H\mathbf{e}_j\right\| = O\left(N^{-1/2}\right)$, and, since $\epsilon^{(j)}_{\mathcal{A}} = O\left(N^{-1}\right)$ we see that

$$\mathbf{e}_j^T\mathcal{Q}_{\mathcal{A}^{(j)}}\mathbf{e}_j = \mathbf{e}_j^T\mathcal{Q}_{\mathcal{A}}\mathbf{e}_j + O\left(N^{-2}\right) = O\left(N^{-1}\right).$$

This implies that

$$\mathcal{V}_{31} = \left(\widetilde{\sigma}^2_{\mathcal{A}} - \widetilde{\sigma}^2_{w,\mathcal{A}}\right) \frac{\widetilde{\sigma}^2_{w,\mathcal{A}}}{\widetilde{\sigma}^2_{\mathcal{A}}} \frac{1}{N} \Bigg[\mathbf{n}^H \mathcal{Q}_{\mathcal{A}} \mathbf{e}_j \mathbf{e}_j^T \mathcal{Q}_{\mathcal{A}} \mathbf{n} - \frac{1}{\widetilde{\sigma}^2_{w,\mathcal{A}}} \mathbf{n}^H \mathbf{e}_j \mathbf{e}_j^T \mathcal{Q}_{\mathcal{A}} \mathbf{n} - \frac{1}{\widetilde{\sigma}^2_{w,\mathcal{A}}} \mathbf{n}^H \mathcal{Q}_{\mathcal{A}} \mathbf{e}_j \mathbf{e}_j^T \mathbf{n} \Bigg] + o_p\left(N^{-1}\right).$$

Combining the expressions, we obtain the result.

Chapter 6

Conclusions and Future Research Directions

Exploiting already existing power lines in the electrical grid also for communication purposes makes power line communication (PLC) the most economical enabler of the smart grid. The almost ubiquitous electrical power network provides extensive outreach that is exclusive to PLC-based technologies, making PLC suitable for smart grid communication in all domains of the electrical power distribution network. However, as it has been conveyed throughout this thesis, power lines, which were designed only to transfer electrical energy from the generators to the end consumers, are one of the harshest environments for data transmission. Communication signals with frequencies higher than the fundamental operating frequency of electrical signal are frequency selective attenuated when passing through the power line channel. Besides, the operation of the power line network components generates random noise that is impulsive in nature and corrupts data transmitted through the power lines. Furthermore, PLC transmissions are also affected by narrowband interferences (NBIs) originating from the wireless communication systems operating in the close proximity. Therefore, reliable data exchange using power lines is not trivial. To address this problem, in this thesis we studied signal processing algorithm design for PLC systems and proposed robust and efficient schemes to overcome the inevitable impairments in power line environment.

In the first section of chapter 3 of this thesis, we dealt with the problem of impulsive noise (IN) mitigation for PLC systems. To overcome the effect of IN, we proposed an active noise mitigation scheme based on IN estimation and cancellation. To effectively estimate the IN present in the received signal

we proposed to exploit the samples received in the unused subcarriers of the orthogonal frequency division multiplexing (OFDM) PLC systems. We verified the performance of the proposed algorithm numerically and concluded that, in the power line environment, the effect of IN can be effectively overcome when the location of IN in the received signal is properly estimated and the effect of IN is canceled only from the corrupted samples in the received signal. We then paid our attention towards designing an NBI interference cancelation scheme that is tailored specifically for the PLC systems. Exploiting the sparsity of the IN and NBI, where IN is sparse in time domain and NBI is sparse in frequency domain, in the second section of chapter 3 we extended the IN estimation algorithm proposed in the first section also to successively cancel the NBI for PLC systems. We verified the performance of the proposed noise cancellation scheme by simulating its performance in typical power line environment and concluded that mitigating the effect of IN and NBI only from the samples that are corrupted, significantly improves the BER performance of the OFDM based PLC systems.

In chapter 4 of this thesis, we proposed maximum-likelihood (ML) channel estimators for PLC systems. The proposed estimators exploit the estimated IN samples to estimate the ML coefficient of channel in the presence of IN. We proposed two different IN exploitation methods and evaluated, both numerically and analytically, the performance of both estimators in typical power line environments. We verified and concluded that better channel estimation performance is achieved if rather than suppressing the power of IN present in the received signal and performing least squares estimation of the channel, the IN is estimated and exploited during channel estimation. Furthermore, we also verified that among the proposed estimators the one that treats the IN as a random quantity outperforms the other estimator that considers IN samples as a deterministic quantity while exploiting them during channel estimation process.

The numerical validation of the ML estimators in chapter 4 showed that the channel estimation accuracy depends on the precision of IN support detection. An optimum channel estimation accuracy was observed when the IN support was precisely estimated. In a situation when the IN is either overestimated or underestimated, the channel estimation accuracy degrades. The more IN support is missed, the less accurately channel is estimated. To address this situation, in chapter 5 we proposed deterministic ML (DML) and random ML (RML) IN support detectors. Both estimators were derived using the same

statistics that were established in chapter 4 to derive the ML channel estimators. The proposed schemes identify IN support by employing an incremental greedy search algorithm. The accuracy of the statistical description used to design the support detectors were verified numerically. Simulation results showed that the RML algorithm performs better segregation of the IN support than the DML algorithms. The performance of the proposed algorithm was also compared with the performance of naive greedy approach. It was further concluded that a better detection of IN can be done by proper exploitation of the received signal structure.

Contributions

The research results and findings of this dissertation are published in international scientific conferences and journals. The signal processing schemes to mitigate the IN and NBI for PLC systems proposed in chapter 3 are published in:

- **D. Shrestha**, X. Mestre and M. Payaró,"Asynchronous Impulsive Noise Mitigation Based on Subspace Support Estimation for PLC Systems", *IEEE International Symposium on Power Line Communications and its Applications*, Bottrop, Germany, 2016.
- **D. Shrestha**, A. Tonello, X. Mestre and M. Payaró, "Simultaneous Cancellation of Narrow Band Interference and Impulsive Noise in PLC Systems", *IEEE International Conference on Smart Grid Communications*, Sydney, Australia, 2016.

The channel estimation strategies proposed in section 4.2 of chapter 4 is published in:

- **D. Shrestha**, X. Mestre and M. Payaró,"Maximum-Likelihood Channel Estimation in Presence of Impulsive Noise for PLC Systems", *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Washington DC, USA, 2016.

and the joint ML CIR and IN estimator proposed in section 4.2 of chapter 4 is published in:

- **D. Shrestha**, X. Mestre and M. Payaró,"On Channel Estimation for Power Line Communication Systems in the Presence of Impulsive Noise",

Journal of Computers and Electrical Engineering, Vol. 72, Pages 406-419, 2018.

The algorithm proposed in chapter 5 has been submitted to:

- X. Mestre, M. Payaró and **D. Shrestha**,"Maximum-Likelihood Detection of Impulsive Noise Support with Application to Channel Parameter Estimation", IEEE International Conference on Acoustics, Speech, and Signal Processing, 2019.
- **D. Shrestha**, X. Mestre and M. Payaró,"On Maximum-Likelihood Detection of the Impulsive Noise Support", IEEE Transactions on Signal Processing.

for review. Moreover, following article has been derived from the research collaboration conducted during this Ph.D. and is not included in this thesis:

- **D. Shrestha**, Z. Pang and D. Dzung,"Precise Clock Synchronization in High Performance Wireless Communication for Time Sensitive Networking", IEEE Access, vol. 6, pp. 8944-8953, 2018.

Future Research Directions

There is no doubt that PLC is the most viable enabler of the smart grid. The data transmission capacity that modern PLC based technologies provide has proved its suitability for basic smart grid application like automatic meter reading. As immense interest in using PLC for the smart grid is rising, the next generation of PLC systems should be able to provide a platform for critical smart grid applications that demand higher data transmission capabilities. Today, as PLC is evolving as a mature technology, implementation of multiple-input-multiple-output (MIMO) in in-home PLC scenarios has been proposed to enhance the data rates of conventional PLC systems [154]. The presence of multiple conductors (namely: line, neutral and protective earth) in an electrical cable provides multiple transmission and reception ports. Exploiting the available conductors three different ports can be exploited for MIMO communication and significantly increase the data rate and communication reliability without the need for higher bandwidth [155]. To this end, a top-down approach based statistical MIMO channel model and analysis and modeling of noise for in-home MIMO PLC are presented in [156–160].

Similar as in the in-home PLC case, application of MIMO schemes can become a possible solution to achieve high data rates in the NB-PLC systems for the smart grid [161]. Some recent works propose exploitation of multi-conductor power lines in the distribution and access grids for MIMO NB-PLC for smart grid communications [161–164]. These works report data rate enhancements in PLC systems and advocate the use of MIMO PLC to support communication requirements of smart grid applications. However, these studies are limited to analyzing the performance of MIMO PLC systems in the presence of background noise only. Moreover, a deterministic channel model that is specific to a medium voltage (MV) distribution network is used. Therefore, to better exploit MIMO for smart grid applications the following topics can be considered as future research directions.

- **Statistical characterization of the MIMO channels:** At present, the MIMO NB-PLC literature refers to bottom-up frequency dependent channel models to assess the system performance. These channel models typically refer to behaviors of specific power line networks. Therefore, the development of signal processing algorithms for MIMO PLC systems and their performance assessment is limited to certain scenarios. In order to develop signal processing algorithms that are tailored for MIMO PLC and validate their performance in multitude of scenarios, a statistical channel model that can characterize the MIMO power line channel behavior in varied power line networks is required.

- **Modeling and characterization of IN:** The presence of IN is considered as one of the specific behaviors of PLC environment. Though the modeling and characterization of IN in single-input-single-output (SISO) transmission modes is richly discussed in the PLC literature, the analysis and the effect of IN on the MIMO PLC transmission is yet to be explored. Hence, to exploit MIMO transmissions using the multiconductor power lines in the low voltage access and MV distribution grids, proper modeling and characterization of IN in MIMO setting are required.

- **Effect of NBI on multiple data transmission and receiving channels:** As extensively argued during this dissertation, it is widely accepted that PLC systems are susceptible to the interference originating from the operation of the closely located wireless communication systems. In this sense, characterization and study of NBI and its effect on MIMO PLC communication needs to be carried out.

- **Suppression and mitigation of IN and NBI:** Since modern PLC systems offer only SISO transmission modes, the noise and interference mitigation schemes are developed taking this into consideration. However, to exploit MIMO PLC, signal processing algorithms that are capable of mitigating the effect of IN and NBI from the signals received via multiple ports are of utmost importance.

Bibliography

[1] S. Howell, Y. Rezgui, J.-L. Hippolyte, B. Jayan, and H. Li, "Towards the next generation of smart grids: Semantic and holonic multi-agent management of distributed energy resources," *Renewable and Sustainable Energy Reviews*, vol. 77, pp. 193 – 214, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032117304392

[2] V. C. Gungor, D. Sahin, T. Kocak, S. Ergut, C. Buccella, C. Cecati, and G. P. Hancke, "A survey on smart grid potential applications and communication requirements," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 1, pp. 28–42, Feb 2013.

[3] T. F. Garrity, "Getting smart," *IEEE Power and Energy Magazine*, vol. 6, no. 2, pp. 38–45, March 2008.

[4] A. Ipakchi and F. Albuyeh, "Grid of the future," *IEEE Power and Energy Magazine*, vol. 7, no. 2, pp. 52–62, March 2009.

[5] G. T. Heydt, C. C. Liu, A. G. Phadke, and V. Vittal, "Solution for the crisis in electric power supply," *IEEE Computer Applications in Power*, vol. 14, no. 3, pp. 22–30, Jul 2001.

[6] H. Farhangi, "The path of the smart grid," *IEEE Power and Energy Magazine*, vol. 8, no. 1, pp. 18–28, January 2010.

[7] "NIST framework and roadmap for smart grid interoperability standards," *National Institute of Standards and Technology*, vol. 1.0, January 2010.

[8] E. Ancillotti, R. Bruno, and M. Conti, "The role of communication systems in smart grids: Architectures, technical

solutions and research challenges," *Computer Communications*, vol. 36, no. 17, pp. 1665 – 1697, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0140366413002090

[9] W. Wang, Y. Xu, and M. Khanna, "A survey on the communication architectures in smart grid," *Computer Networks*, vol. 55, no. 15, pp. 3604 – 3629, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S138912861100260X

[10] J. Gao, Y. Xiao, J. Liu, W. Liang, and C. P. Chen, "A survey of communication/networking in smart grids," *Future Generation Computer Systems*, vol. 28, no. 2, pp. 391 – 404, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X11000653

[11] M. Chenine, L. Nordstrom, and P. Johnson, "Factors in assessing performance of wide area communication networks for distributed control of power systems," in *2007 IEEE Lausanne Power Tech*, July 2007, pp. 1682–1687.

[12] J. Zhou, R. Q. Hu, and Y. Qian, "Scalable distributed communication architectures to support advanced metering infrastructure in smart grid," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1632–1642, Sept 2012.

[13] H. Khurana, M. Hadley, N. Lu, and D. A. Frincke, "Smart-grid security issues," *IEEE Security Privacy*, vol. 8, no. 1, pp. 81–85, Jan 2010.

[14] D. Shrestha, Z. Pang, and D. Dzung, "Precise clock synchronization in high performance wireless communication for time sensitive networking," *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2018.

[15] A. A. Amarsingh, H. A. Latchman, and D. Yang, "Narrowband power line communications: Enabling the smart grid," *IEEE Potentials*, vol. 33, no. 1, pp. 16–21, Jan 2014.

[16] A. M. Tonello, J. Song, S. Weiss, and F. Yang, "PLC for smart grid: state-of-the-art and challenges," in *Proceedings of CMC*, 2012.

[17] 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, June 2011.

[18] A. Haidine, A. Tabone, and J. Müller, "Deployment of power line communication by European utilities in advanced metering infrastructure," in *2013 IEEE 17th International Symposium on Power Line Communications and Its Applications*, March 2013, pp. 126–130.

[19] M. Gotz, M. Rapp, and K. Dostert, "Power line channel characteristics and their effect on communication system design," *IEEE Communications Magazine*, vol. 42, no. 4, pp. 78–86, Apr 2004.

[20] D. E. Nordell, "Communication systems for distribution automation," in *2008 IEEE/PES Transmission and Distribution Conference and Exposition*, April 2008, pp. 1–14.

[21] S. T. Mak and T. G. Moore, "TWACS, a new viable two-way automatic communication system for distribution networks. part II: Inbound communication," *IEEE Power Engineering Review*, vol. PER-4, no. 8, pp. 51–51, Aug 1984.

[22] M. Nassar, J. Lin, Y. Mortazavi, A. Dabak, I. H. Kim, and B. L. Evans, "Local utility power line communications in the 3-500 kHz band: Channel impairments, noise, and standards," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 116–127, Sept 2012.

[23] C. Cano, A. Pittolo, D. Malone, L. Lampe, A. M. Tonello, and A. G. Dabak, "State of the art in power line communications: From the applications to the medium," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 7, pp. 1935–1952, July 2016.

[24] H. C. Ferreira, L. Lampe, J. Newburry, and T. G. Swart, *Power Line Communications: Theory and Applications for Narrowband and Broadband Communications over Power Lines.* Wiley, 2010.

[25] L. T. Berger, A. Schwager, and J. J. Escudero-Garzás, "Power line communications for smart grid applications," *Journal of Electrical and Computer Engineering*, 2013.

[26] P. Amirshahi and M. Kavehrad, "High-frequency characteristics of overhead multiconductor power lines for broadband communications," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 7, pp. 1292–1303, July 2006.

[27] A. Sendin, J. Simon, I. Urrutia, and I. Berganza, "PLC deployment and architecture for smart grid applications in Iberdrola," in *18th IEEE International Symposium on Power Line Communications and Its Applications*, March 2014, pp. 173–178.

[28] A. M. Tonello, F. Versolatto, B. Bejar, and S. Zazo, "A fitting algorithm for random modeling the PLC channel," *IEEE Transactions on Power Delivery*, vol. 27, no. 3, pp. 1477–1484, July 2012.

[29] H. Philips, "Modelling of powerline communication channels," *IEEE ISPLC*, pp. 14–21, 1999.

[30] M. Zimmermann and K. Dostert, "A multipath model for the powerline channel," *IEEE Transactions on Communications*, vol. 50, no. 4, pp. 553–559, Apr 2002.

[31] J. Cairo, J. Pegueroles, F. Martin-Soto, and M. Hormigo-Gonzalez, "Modeling PLC channel using load impedance mismatch and line parameter determination," *IEEE Transactions on Communications*, 2013.

[32] J. S. Barnes, "A physical multi-path model for power distribution network propagation," *IEEE ISPLC*, pp. 76–89, 1998.

[33] A. B. Dalby, "Signal transmission in powerlines - analysis of powerline circuits," *IEEE ISPLC*, pp. 37–44, 1997.

[34] A. Cataliotti, A. Daidone, and G. Tine, "Power line communication in medium voltage systems: Characterization of MV cables," *IEEE Transactions on Power Delivery*, vol. 23, no. 4, pp. 1896–1902, Oct 2008.

[35] H. Philips, "Performance measurements of powerline channels at high frequencies," *IEEE ISPLC*, 1998.

[36] Y. Xiaoxian, Z. Tao, Z. Baohui, H. Zonghong, C. Jian, and G. Zhiqiang, "Channel model and measurement methods for 10-kV medium-voltage power lines," *IEEE Transactions on Power Delivery*, vol. 22, no. 1, pp. 129–134, Jan 2007.

[37] A. M. Tonello and F. Versolatto, "Bottom up statistical PLC channel modeling part I: Random topology model and efficient transfer function computation," *IEEE Transactions on Power Delivery*, vol. 26, no. 2, pp. 891–898, April 2011.

[38] A. G. Lazaropoulos and P. G. Cottis, "Transmission characteristics of overhead medium-voltage power-line communication channels," *IEEE Transactions on Power Delivery*, vol. 24, no. 3, pp. 1164–1173, July 2009.

[39] "IEEE standard for low-frequency (less than 500 kHz) narrowband power line communications for smart grid applications," *IEEE Std 1901.2-2013*, pp. 1–269, Dec 2013.

[40] H. Gassara, F. Rouissi, and A. Ghazel, "Ofdm communication chain performance assessment under realistic measured narrowband plc channels and noise," in *2014 IEEE Fourth International Conference on Consumer Electronics Berlin (ICCE-Berlin)*, Sept 2014, pp. 101–105.

[41] P. Torio and M. G. Sanchez, "Method to cancel impulsive noise from power-line communication systems by processing the information in the idle carriers," *IEEE Transactions on Power Delivery*, vol. 27, no. 4, pp. 2421–2422, Oct 2012.

[42] N. Andreadou and A. M. Tonello, "On the mitigation of impulsive noise in power-line communications with LT codes," *IEEE Transactions on Power Delivery*, vol. 28, no. 3, pp. 1483–1490, July 2013.

[43] 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, April 2005.

[44] 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.

[45] M. Nassar, K. Gulati, Y. Mortazavi, and B. L. Evans, "Statistical modeling of asynchronous impulsive noise in powerline communication networks," in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, Dec 2011, pp. 1–6.

[46] A. Mehboob, L. Zhang, and J. Khangosstar, "Adaptive impulsive noise mitigation using multi mode compressive sensing for powerline communications," in *2012 IEEE International Symposium on Power Line Communications and Its Applications*, March 2012, pp. 368–373.

[47] M. Korki, N. Hosseinzadeh, H. L. Vu, T. Moazzeni, and C. H. Foh, "Impulsive noise reduction of a narrowband power line communication using optimal nonlinearity technique," in *2011 Australasian Telecommunication Networks and Applications Conference (ATNAC)*, Nov 2011, pp. 1–4.

[48] D. Middleton, *An introduction to statistical communication theory.* McGraw-Hill New York, 1960.

[49] R. Pighi, M. Franceschini, G. Ferrari, and R. Raheli, "Fundamental performance limits of communications systems impaired by impulse noise," *IEEE Transactions on Communications*, vol. 57, no. 1, pp. 171–182, January 2009.

[50] 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.

[51] D. Fertonani and G. Colavolpe, "On reliable communications over channels impaired by bursty impulse noise," in *2008 IEEE International Symposium on Power Line Communications and Its Applications*, April 2008, pp. 357–362.

[52] M. Ghosh, "Analysis of the effect of impulse noise on multicarrier and single carrier QAM systems," *IEEE Transactions on Communications*, vol. 44, no. 2, pp. 145–147, Feb 1996.

[53] S. P. Herath, N. H. Tran, and T. Le-Ngoc, "On optimal input distribution and capacity limit of Bernoulli-Gaussian impulsive noise channels," in *2012 IEEE International Conference on Communications (ICC)*, June 2012, pp. 3429–3433.

[54] F. Abdelkefi, P. Duhamel, and F. Alberge, "Impulsive noise cancellation in multicarrier transmission," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 94–106, Jan 2005.

[55] D. Galda and H. Rohling, "Narrowband interference reduction in OFDM based power line communication systems," *IEEE ISPLC*, pp. 345–351, Apr 2001.

[56] B. Tan and J. S. Thompson, "Powerline communications channel modelling methodology based on statistical features," *CoRR*, vol. abs/1203.3879, 2012.

[57] M. Hoch, "Comparison of PLC G3 and PRIME," in *2011 IEEE International Symposium on Power Line Communications and Its Applications*, April 2011, pp. 165–169.

[58] G. Ndo, P. Siohan, and M. H. Hamon, "Adaptive noise mitigation in impulsive environment: Application to power-line communications," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 647–656, April 2010.

[59] T. Pande, I. H. Kim, and A. Batra, "A method for narrowband interference mitigation in OFDM by minimizing spectral leakage," in *2015 IEEE International Symposium on Power Line Communications and Its Applications (ISPLC)*, March 2015, pp. 19–23.

[60] Y. Hou, R. Liu, B. Dai, and L. Zhao, "Joint channel estimation and ldpc decoding over time-varying impulsive noise channels," *IEEE Transactions on Communications*, vol. 66, no. 6, pp. 2376–2383, June 2018.

[61] L. Bai, T. Zheng, C. Liu, and W. Ba, "Impulsive noise recovery based interleaving in power line communication," in *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*, Oct 2016, pp. 294–299.

[62] K. Razazian, M. Umari, and A. Kamalizad, "Error correction mechanism in the new G3-PLC specification for powerline communication," in *ISPLC2010*, March 2010, pp. 50–55.

[63] J. Lin, T. Pande, I. H. Kim, A. Batra, and B. L. Evans, "Robust transceiver to combat periodic impulsive noise in narrowband powerline communications," in *2015 IEEE International Conference on Communications (ICC)*, June 2015, pp. 752–757.

[64] N. Andreadou and F. N. Pavlidou, "PLC channel: Impulsive noise modelling and its performance evaluation under different array coding schemes," *IEEE Transactions on Power Delivery*, vol. 24, no. 2, pp. 585–595, April 2009.

[65] P. Gao and C. Tepedelenlioglu, "Space-time coding over fading channels with impulsive noise," *IEEE Transactions on Wireless Communications*, vol. 6, no. 1, pp. 220–229, Jan 2007.

[66] S. Lin and J. D. J. Costello, *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, NJ:Prentice-Hall, 1983.

[67] D. F. Tseng, R. B. Yang, T. R. Tsai, Y. S. Han, and W. H. Mow, "Efficient clipping for broadband power line systems in impulsive noise environment," in *2012 IEEE International Symposium on Power Line Communications and Its Applications*, March 2012, pp. 362–367.

[68] S. V. Zhidkov, "Performance analysis and optimization of OFDM receiver with blanking nonlinearity in impulsive noise environment," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 1, pp. 234–242, Jan 2006.

[69] N. P. Cowley, A. Payne, and M. Dawkins, "COFDM tuner with impulse noise reduction," *European Patent Application EP1043874*, Feb 2002.

[70] T. N. Zogakis, P. S. Chow, J. T. Aslanis, and J. M. Cioffi, "Impulse noise mitigation strategies for multicarrier modulation," in *Communications, 1993. ICC '93 Geneva. Technical Program, Conference Record, IEEE International Conference on*, vol. 2, May 1993, pp. 784–788 vol.2.

[71] H. A. Suraweera, C. Chai, J. Shentu, and J. Armstrong, "Analysis of impulse noise mitigation techniques for digital television systems," in *8th International OFDM Workshop*, Sept. 2003, pp. 172–176.

[72] Y.-H. Kim, K.-H. Kim, H.-M. Oh, K.-H. Kim, and S.-C. Kim, "Mitigation of effect of impulsive noise for OFDM systems over power line channels," in *2008 IEEE International Symposium on Power Line Communications and Its Applications*, April 2008, pp. 386–390.

[73] S. V. Zhidkov, "Analysis and comparison of several simple impulsive noise mitigation schemes for OFDM receivers," *IEEE Transactions on Communications*, vol. 56, no. 1, pp. 5–9, January 2008.

[74] F. H. Juwono, Q. Guo, D. Huang, and K. P. Wong, "Deep clipping for impulsive noise mitigation in OFDM-based power-line communications," *IEEE Transactions on Power Delivery*, vol. 29, no. 3, pp. 1335–1343, June 2014.

[75] V. N. Papilaya and A. J. H. Vinck, "Investigation on a new combined impulsive noise mitigation scheme for OFDM transmission," in *2013 IEEE 17th International Symposium on Power Line Communications and Its Applications*, March 2013, pp. 86–91.

[76] Y. R. Chien, "Iterative channel estimation and impulsive noise mitigation algorithm for OFDM-based receivers with application to power-line communications," *IEEE Transactions on Power Delivery*, vol. 30, no. 6, pp. 2435–2442, Dec 2015.

[77] B. Adebisi, K. Anoh, K. M. Rabie, A. Ikpehai, M. Fernando, and A. Wells, "A new approach to peak threshold estimation for impulsive noise reduction over power line fading channels," *IEEE Systems Journal*, pp. 1–12, 2018.

[78] T. Bai, C. Xu, R. Zhang, A. F. A. Rawi, and L. Hanzo, "Joint impulsive noise estimation and data detection conceived for ldpc-coded dmt-based dsl systems," *IEEE Access*, vol. 5, pp. 23 133–23 145, 2017.

[79] M. Mao, G. Xie, J. Gao, G. Wu, and K. Liu, "Adaptive in mitigation in ofdm-based power-line communication system with jointed blanking and robust channel estimation," in *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, Oct 2016, pp. 2573–2576.

[80] L. Lampe, "Bursty impulse noise detection by compressed sensing," in *2011 IEEE International Symposium on Power Line Communications and Its Applications*, April 2011, pp. 29–34.

[81] T. Y. Al-Naffouri, F. F. Al-Shaalan, A. A. Quadeer, and H. Hmida, "Impulsive noise estimation and cancellation in DSL using compressive sampling," in *2011 IEEE International Symposium of Circuits and Systems (ISCAS)*, May 2011, pp. 2133–2136.

[82] J. Zhang, Z. He, P. Chen, and Y. Rong, "A compressive sensing based iterative algorithm for channel and impulsive noise estimation in underwater acoustic ofdm systems," in *OCEANS 2017 - Anchorage*, Sept 2017, pp. 1–5.

[83] H. Zhang, L. L. Yang, and L. Hanzo, "Compressed impairment sensing-assisted and interleaved-double-fft-aided modulation improves broad-

band power line communications subjected to asynchronous impulsive noise," *IEEE Access*, vol. 4, pp. 81–96, 2016.

[84] G. Ren, S. Qiao, and Y. Hei, "Asynchronous impulsive noise mitigation in ofdm using adaptive threshold compressive sensing," in *WAMICON 2014*, June 2014, pp. 1–5.

[85] J. Lin, M. Nassar, and B. L. Evans, "Non-parametric impulsive noise mitigation in OFDM systems using sparse bayesian learning," in *2011 IEEE Global Telecommunications Conference - GLOBECOM 2011*, Dec 2011, pp. 1–5.

[86] J. Lin and B. L. Evans, "Non-parametric mitigation of periodic impulsive noise in narrowband powerline communications," in *2013 IEEE Global Communications Conference (GLOBECOM)*, Dec 2013, pp. 2981–2986.

[87] L. Bai, M. Tucci, and M. Raugi, "Impulsive noise mitigation with interleaving based on music in power line communication," *IEEE Transactions on Smart Grid*, pp. 1–1, 2018.

[88] T. Kitamura, K. Ohno, and M. Itami, "Improving of performance of ofdm reception under class-a impulsive channel by replica signal estimation of impulse," in *2012 IEEE International Conference on Consumer Electronics (ICCE)*, Jan 2012, pp. 620–621.

[89] K. Ogura, A. Nakamura, and M. Itami, "A study on class a impulsive noise cancellation and channel estimation under rayleigh fading environment," in *2018 IEEE International Conference on Consumer Electronics (ICCE)*, Jan 2018, pp. 1–2.

[90] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[91] A. B. Ramirez, R. E. Carrillo, G. Arce, K. E. Barner, and B. Sadler, "An overview of robust compressive sensing of sparse signals in impulsive noise," in *2015 23rd European Signal Processing Conference (EUSIPCO)*, Aug 2015, pp. 2859–2863.

[92] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, Dec 2007.

[93] D. L. Donoho, Y. Tsaig, I. Drori, and J. L. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, Feb 2012.

[94] D. Needell and R. Vershynin, "Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 310–316, April 2010.

[95] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec 1993.

[96] H. Liu, Y. Li, Y. Zhou, and T. K. Truong, "Greedy pursuit algorithms for sparse signal reconstruction in the case of impulsive noise," in *2016 IEEE International Conference on Digital Signal Processing (DSP)*, Oct 2016, pp. 705–709.

[97] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing: Closing the gap between performance and complexity," *Preprint*, 2008.

[98] Needell and J. A. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Preprint*, 2008.

[99] T. T. Do, L. Gan, N. Nguyen, and T. D. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *2008 42nd Asilomar Conference on Signals, Systems and Computers*, Oct 2008, pp. 581–587.

[100] S. Liu, F. Yang, W. Ding, and J. Song, "A priori aided compressive sensing approach for impulsive noise reconstruction," in *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*, Aug 2015, pp. 205–209.

[101] F. Abdelkefi and J. Ayadi, "Efficient techniques for impulsive noise cancellation in cgu/sd systems," *IEEE Transactions on Signal Processing*, vol. 65, no. 14, pp. 3749–3760, July 2017.

[102] K. Shi, Y. Zhou, B. Kelleci, T. W. Fischer, E. Serpedin, and A. l. Karsilayan, "Impacts of narrowband interference on OFDM-UWB receivers:

Analysis and mitigation," *IEEE Transactions on Signal Processing*, vol. 55, no. 3, pp. 1118–1128, March 2007.

[103] B. Kelleci, T. W. Fischer, K. Shi, Y. Zhou, A. I. Karsilayan, and E. Serpedin, "Narrowband interference suppression in multi-band OFDM ultra wideband communication systems: A mixed-mode approach," in *2006 IEEE 12th Digital Signal Processing Workshop 4th IEEE Signal Processing Education Workshop*, Sept 2006, pp. 55–59.

[104] A. Batra and J. R. Zeidler, "Narrowband interference mitigation in bicm ofdm systems," in *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2009, pp. 2605–2608.

[105] S. Kalyani, V. Raj, and K. Giridhar, "Narrowband interference mitigation in Turbo-coded OFDM systems," in *2007 IEEE International Conference on Communications*, June 2007, pp. 1059–1064.

[106] A. K. Yazbek, J. P. Cances, and V. Meghdadi, "Narrowband interference mitigation with LRPC code and OFDM for smart grid applications," in *2016 10th International Conference on Signal Processing and Communication Systems (ICSPCS)*, Dec 2016, pp. 1–6.

[107] S. Liu, F. Yang, J. Song, J. Li, and W. Liu, "An optimized time-frequency interleaving scheme for OFDM-based power line communication systems," in *2014 IEEE International Conference on Communications (ICC)*, June 2014, pp. 4137–4142.

[108] S. Liu, F. Yang, W. Ding, and J. Song, "Double kill: Compressive-sensing-based narrow-band interference and impulsive noise mitigation for vehicular communications," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5099–5109, July 2016.

[109] ——, "Structured compressive sensing based narrowband interference mitigation for vehicular communications," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, June 2015, pp. 2375–2380.

[110] R. M. Devi and G. Prema, "Compressive sensing algorithm for narrowband interference mitigation in MIMO-OFDM," in *2012 IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, Aug 2012, pp. 22–27.

[111] A. Gomaa, K. M. Z. Islam, and N. Al-Dhahir, "Two novel compressed-sensing algorithms for NBI detection in OFDM systems," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2010, pp. 3294–3297.

[112] M. Mokhtar, W. U. Bajwa, M. Elgenedy, and N. Al-Dhahir, "Exploiting block sparsity for joint mitigation of asynchronous nbi and in in hybrid powerline-wireless communications," in *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Nov 2015, pp. 362–367.

[113] M. Elgenedy, M. M. Awadin, R. Hamila, W. U. Bajwa, A. S. Ibrahim, and N. Al-Dhahir, "Sparsity-based joint nbi and impulse noise mitigation in hybrid plc-wireless transmissions," *IEEE Access*, vol. 6, pp. 30 280–30 295, 2018.

[114] Z. W. Zheng, "Channel estimation and channel equalization for the OFDM-based WLAN systems," in *2010 International Conference on E-Business and E-Government*, May 2010, pp. 1691–1694.

[115] J. J. van de Beek, O. Edfors, M. Sandell, S. K. Wilson, and P. O. Borjesson, "On channel estimation in OFDM systems," in *1995 IEEE 45th Vehicular Technology Conference. Countdown to the Wireless Twenty-First Century*, vol. 2, Jul 1995, pp. 815–819 vol.2.

[116] Z. Sadowski, "Comparison of PLC-PRIME and PLC-G3 protocols," in *2015 International School on Nonsinusoidal Currents and Compensation (ISNCC)*, June 2015, pp. 1–6.

[117] "IEEE standard for low-frequency (less than 500 kHz) narrowband power line communications for smart grid applications," *IEEE Std 1901.2-2013*, pp. 1–269, Dec 2013.

[118] J. Rinne, A. Hazmi, and M. Renfors, "Impulse burst position detection and channel estimation schemes for OFDM systems," *IEEE Transactions on Consumer Electronics*, vol. 49, no. 3, pp. 539–545, Aug 2003.

[119] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM system," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 3, pp. 1122–1128, Aug 1998.

[120] Y. H. Ma, P. L. So, and E. Gunawan, "Performance analysis of OFDM systems for broadband power line communications under impulsive noise and multipath effects," *IEEE Transactions on Power Delivery*, vol. 20, no. 2, pp. 674–682, April 2005.

[121] X. Yu, P. Lin, Z. He, and W. Wu, "OFDM channel estimation with impulse noise cancellation," in *2007 International Conference on Wireless Communications, Networking and Mobile Computing*, Sept 2007, pp. 330–333.

[122] H. Ando, A. Nakamura, K. Ohno, and M. Itami, "A study on channel estimation under class A impulsive PLC channel," in *2013 IEEE 17th International Symposium on Power Line Communications and Its Applications*, March 2013, pp. 69–74.

[123] H. Yasui, K. Ohno, and M. Itami, "A study on channel estimation using pilot symbols under impulsive PLC channel," in *18th IEEE International Symposium on Power Line Communications and Its Applications*, March 2014, pp. 322–327.

[124] H. Minn and V. K. Bhargava, "An investigation into time-domain approach for OFDM channel estimation," *IEEE Transactions on Broadcasting*, vol. 46, no. 4, pp. 240–248, Dec 2000.

[125] L. Yang, G. Ren, and Z. Qiu, "Novel noise reduction algorithm for LS channel estimation in OFDM system with frequency selective channels," in *2010 IEEE International Conference on Communication Systems*, Nov 2010, pp. 478–482.

[126] C. Xu, X. Chen, and S. Yu, "A modified time-domain channel estimation method for OFDM systems in impulsive noise environment," in *2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP)*, June 2013, pp. 88–93.

[127] H. Zamiri-Jafarian, M. J. Omidi, and S. Pasupathy, "Improved channel estimation using noise reduction for OFDM systems," in *The 57th IEEE Semiannual Vehicular Technology Conference, 2003. VTC 2003-Spring.*, vol. 2, April 2003, pp. 1308–1312 vol.2.

[128] A. . M. Picorone, L. R. Amado, and M. V. Ribeiro, "Linear and periodically time-varying PLC channels estimation in the presence of impulsive noise," in *ISPLC2010*, March 2010, pp. 255–260.

[129] Y. Zhao and A. Huang, "A novel channel estimation method for OFDM mobile communication systems based on pilot signals and transform-domain processing," in *1997 IEEE 47th Vehicular Technology Conference. Technology in Motion*, vol. 3, May 1997, pp. 2089–2093 vol.3.

[130] G. A. Laguna-Sanchez, R. Barron-Fernandez, and A. Prieto-Guerrero, "Wavelet-based improvement for channel estimation in a power-line communication environment impaired by impulsive noise," in *2010 IEEE Electronics, Robotics and Automotive Mechanics Conference*, Sept 2010, pp. 523–528.

[131] Y. Kang, K. Kim, and H. Park, "Efficient DFT-based channel estimation for OFDM systems on multipath channels," *IET Communications*, vol. 1, no. 2, pp. 197–202, April 2007.

[132] J. Huang, P. Wang, and Q. Wan, "Robust approach for channel estimation in power line communication," *Journal of Communications and Networks*, vol. 14, no. 3, pp. 237–242, June 2012.

[133] M. J. F. G. Garcia, J. L. Rojo-Alvarez, F. Alonso-Atienza, and M. Martinez-Ramon, "Support vector machines for robust channel estimation in OFDM," *IEEE Signal Processing Letters*, vol. 13, no. 7, pp. 397–400, July 2006.

[134] T. Shao, Y. R. Zheng, and J. Benesty, "An affine projection sign algorithm robust against impulsive interferences," *IEEE Signal Processing Letters*, vol. 17, no. 4, pp. 327–330, April 2010.

[135] J. Yoo, J. Shin, and P. Park, "Variable step-size affine projection sign algorithm," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 61, no. 4, pp. 274–278, April 2014.

[136] L. Dai, Z. Wang, Z. Yang, G. Gui, and F. Adachi, "Reliable and energy-efficient ofdm based on structured compressive sensing," in *2014 IEEE International Conference on Communications (ICC)*, June 2014, pp. 3963–3968.

[137] G. Gui, L. Xu, W. Ma, and B. Chen, "Robust adaptive sparse channel estimation in the presence of impulsive noises," in *2015 IEEE International Conference on Digital Signal Processing (DSP)*, July 2015, pp. 628–632.

[138] P. Chen, Y. Rong, S. Nordholm, and Z. He, "Joint channel and impulsive noise estimation in underwater acoustic ofdm systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10 567–10 571, Nov 2017.

[139] S. Moaveninejad, A. Kumar, D. Scazzoli, A. Piti, M. Magarini, S. Bregni, and G. Verticale, "Ber evaluation of post-meter plc services in cenelec-c band," in *2017 IEEE 9th Latin-American Conference on Communications (LATINCOM)*, Nov 2017, pp. 1–6.

[140] D. Shrestha, X. Mestre, and M. Payaró, "Asynchronous impulsive noise mitigation based on subspace support estimation for PLC systems," in *2016 International Symposium on Power Line Communications and its Applications (ISPLC)*, March 2016, pp. 1–6.

[141] T. Y. Al-Naffouri, A. A. Quadeer, and G. Caire, "Impulse noise estimation and removal for OFDM systems," *IEEE Transactions on Communications*, vol. 62, no. 3, pp. 976–989, March 2014.

[142] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, New Jersey, USA: Prentice Hall, 1997.

[143] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 2, pp. 387–392, Apr 1985.

[144] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound: further results and comparisons," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 12, pp. 2140–2150, Dec 1990.

[145] A. Eriksson, P. Stoica, and T. Soderstrom, "Second-order properties of MUSIC and ESPRIT estimates of sinusoidal frequencies in high SNR scenarios," *IEE Proceedings F - Radar and Signal Processing*, vol. 140, no. 4, pp. 266–272, Aug 1993.

[146] A. Batra and J. R. Zeidler, "Narrowband interference mitigation in OFDM systems," in *MILCOM 2008 - 2008 IEEE Military Communications Conference*, Nov 2008, pp. 1–7.

[147] K. Yan, H. Zhang, and H. C. Wu, "Robust multipath channel estimation in the presence of impulsive noise," *IET Communications*, vol. 12, no. 2, pp. 228–235, 2018.

[148] A. S. Ackleh, E. J. Allen, R. B. Kearfott, and P. Seshaiyer, *Classical and Modern Numerical Analysis: Theory, Methods, and Practice*. CRC PRESS, 2009.

[149] W. Feller, *An Introduction to Probability Theory and Its Applications. Vol. 2. 2nd. Edition*. John Wiley & Sons, 1970.

[150] C. E. Bonferroni, "Teoria statistica delle classi e calcolo delle probabilità," *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, 1936.

[151] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society*, vol. Series B. 57 (1), pp. 289–300, 1995.

[152] J. Omura and T. Kailath, "Some useful probability functions," *Tech. Rep*, vol. 7050-6, 1965.

[153] M. Simon, *Probability Distributions Involving Gaussian Random Variables: A Handbook for Engineers and Scientists*. Springer, 2006.

[154] L. Stadelmeier, D. Schill, A. Schwager, D. Schneider, and J. Speidel, "MIMO for inhome power line communications," in *7th International ITG Conference on Source and Channel Coding*, Jan 2008, pp. 1–6.

[155] R. Hashmat, P. Pagani, A. Zeddam, and T. Chonavel, "Analysis and modeling of background noise for inhome MIMO PLC channels," in *IEEE International Symposium on Power Line Communications and Its Applications*, 2010.

[156] R. Hashmat, P. Pagani, T. Chonavel, and A. Zeddam, "A time-domain model of background noise for in-home MIMO PLC networks," *IEEE Transactions on Power Delivery*, vol. 27, no. 4, pp. 2082–2089, Oct 2012.

[157] ——, "Analysis and modeling of background noise for inhome MIMO PLC channels," in *2012 IEEE International Symposium on Power Line Communications and Its Applications*, March 2012, pp. 316–321.

[158] D. Rende, A. Nayagam, K. Afkhamie, L. Yonge, R. Riva, D. Veronesi, F. Osnato, and P. Bisaglia, "Noise correlation and its effect on capacity of inhome MIMO power line channels," in *2011 IEEE International Symposium on Power Line Communications and Its Applications*, April 2011, pp. 60–65.

[159] A. Pittolo and A. M. Tonello, "A synthetic statistical MIMO PLC channel model applied to an in-home scenario," *IEEE Transactions on Communications*, vol. 65, no. 6, pp. 2543–2553, June 2017.

[160] N. Pine and S. Choe, "Modified multipath model for broadband MIMO power line communications," in *2012 IEEE International Symposium on Power Line Communications and Its Applications*, March 2012, pp. 292–297.

[161] A. Elsamadouny, A. E. Shafie, M. Abdallah, and N. Al-Dhahir, "Secure sum-rate-optimal MIMO multicasting over medium-voltage NB-PLC networks," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2017.

[162] T. A. Papadopoulos, A. I. Chrysochos, G. K. Papagiannis, A. El-Samadouny, and N. Al-Dhahir, "MIMO modeling and data rates for medium-voltage narrowband-PLC in smart distribution grids," in *IEEE PES Innovative Smart Grid Technologies, Europe*, Oct 2014, pp. 1–6.

[163] A. ElSamadouny, N. Al-Dhahir, M. Abdallah, A. I. Chrysochos, T. A. Papadopoulos, and G. K. Papagiannis, "Multi-user MIMO broadcasting/multicasting for medium-voltage narrowband-PLC networks," in *2015 IEEE International Symposium on Power Line Communications and Its Applications (ISPLC)*, March 2015, pp. 77–82.

[164] A. ElSamadouny and N. Al-Dhahir, "Sum-rate-optimal dual-QoS MIMO multicasting over medium-voltage NB-PLC networks," in *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Nov 2015, pp. 683–688.