ALMA MATER STUDIORUM
UNIVERSITÁ DI BOLOGNA

SCUOLA DI INGEGNERIA E ARCHITETTURA

Dipartimento Informatica - Scienza e Ingegneria - DISI

Corso di Laurea Magistrale in Ingegneria Informatica

TESI DI LAUREA MAGISTRALE

in

Embedded System

Machine Learning Aided Methods for Resilient Industrial Wireless Sensor Network

Candidato:

Andrea Bombino

Relatore:

Chiar. mo Prof. **Stefano Mattoccia**

Correlatore:

Prof. **Mikael Gidlund**

M.Sc. **Simone Grimaldi**

Ph.D. **Aamir Mahmood**

Anno Accademico 2017/2018

Sessione III

Abstract

A Wireless Sensor Network (WSN) is an infrastructure comprised of sensing, computing, and communication devices, that obtains and processes data to understand the behavior of the monitored environment, and to react to events and phenomena that occur within that. The use of WSN in industry settings is extremely appealing, nevertheless most industrial environments are obstile to reliable radio communication, showing pronounced effects of multipath fading, strong attenuation and radio interference. This motivates a huge effort in research activities, standardization process and industrial investments on this field since the last decade. In our work, we propose mechanisms based on machine learning algorithms that allows us to classify the channel propagation condition (Line-of-Sight and Non-Line-of-Sight) of radio links. The investigated methods provide an useful diagnostic tool in the context of adaptive transmission strategies for improving the quality and reliability of wireless communication.

In particular, the first mechanism is based on the analysis of I/Q data, while the second method is based on bit-error pattern distribution in received packets. Both the solutions have been implemented on real hardware and tested in a number of environment with heterogeneous characteristics, showing promising results.

Prefazione

Una Wireless Sensor Network (WSN) può essere definita come un' infrastruttura composta da sensori/dispositivi in grado di calcolare, comunicare e effettuare sensing dell'ambiente circostante processando e analizzando i dati in modo da reagire a eventi e fenomeni che possono occorrere durante la comunicazione. L'utilizzo delle WSN nel settore industriale è estremamente allettante, nonostante la maggior parte du questi ambienti risulti ostile alle comunicazioni radio affidabili, mostrando effetti sensibili di cammini multipli, forte attenuazione e interferenze radio. Questo motiva un enorme effort nella ricerca, standardizzazione e investimento industriale in questo campo, nell'ultimo decennio. L'uso delle WSN nell'ambiente industriale è soggetto a diverse problematiche, dovuto all'ostilità del canale radio in ambito industriale, come rumore, Shadwoing, cammini multipli e interferenze. Nel nostro progetto, proponiamo meccanismi basati sulle condizioni di propagazione del canale e algoritmi di machine learning che ci permettono di classificare lo stato del canale (Line-of-Sight o Non-Line-of-Sight). I metodi analizzati forniscono un utile strumento diagnostico nel contesto delle strategie di trasmissione adattiva per migliorare la qualità e l'affidabilità della comunicazione wireles.

In particolare, il primo meccanimo è basato sull'analisi degli I/Q data, mentre il secondo metodo è basato sull'analisi della distribuzione del bit-error pattern nel pacchetto ricevuto. Entrambe le soluzione sono state implementate su hardware e testate in differenti ambienti con differenti caratteristiche, mostrando risulati promettenti.

Contents

Acronyms

The following list represents the acronyms used in the thesis.

BEP Bit Error Pattern

BPSK Binary Phase Shift Keying

DSSS Direct Sequence Spread Spectrum

FHSS Frequency Hopping Spread Spectrum

ISM Industry Science Medical

LOS Line-of-Sight

MAC Media Access Control

NLOS Non-Line-of-Sight

NN Neural Network

OQPSK Offset Quadrature Phase Shift Keying

PDF Probability Density Function

PHY Physical Layer

QoS Quality of Service

RF Radio Frequency

RSSI Received Signal Strength Indicator

SDR Software Defined Radio

TSCH Time-slotted Channel Hopping

USRP Universal Software Radio Peripheral

WLAN Wireless Local Area Network

WSN Wireless Sensor Network

Chapter 1

Introduction

Since the beginning of the third Millennium, Wireless Sensor Networks (WSNs) have generated an increasing interest from industrial and research perspectives [1]. A wireless sensor network can be generally described as a multi-hop self-organizing network composed of multiple sensor devices, which communicate with each other wirelessly. It is composed of many sensor nodes with the ability to sense and the possibility to control the environment enabling interaction between persons or computers and the surrounding environment, collecting and processing information. They have resource constraints, with low processing power and, in some cases, restrictions in power consumption.

1.1 Motivation

The use of WSN in industrial systems presents some challenges[2]. Wireless networks use an inherently unreliable communication medium which can be aggravated due to noise and interference in the spectrum band used for communication. The majority of wireless sensor networks work in the 2.4 GHz unlicensed frequency band, reserved for Industrial, Scientific and Medical (ISM) applications. This means that the WSN have to deal with unwanted RF interference, in addition to the typical adverse effects of multipath fading and signal attenuation. The WSN can be used in domains such as agriculture, energy, industrial automation, medical health care, smart building.

In the industry environment, the characteristics of the wireless channel are different in comparison to other WSN environments, such as home and office environments. The sensors are deployed to monitor critical parameters such as vibration, temperature, pressure and motor efficiency. In addition, the wireless channel in many industrial environments is non-stationary on a long time-scale, which can cause abrupt changes in the characteristics of the channel over time. These shortcomings can seriously affect the communication link quality of Wireless Sensor Networks. A set of standards, such as WirelessHART, ISA100.11a, was developed for industrial WSN to partially overcome these limitations.

1.2 An Overview on Availaible Approaches

In the related literature, it is possible to find a number of works targeting the topic of radio-link state estimation in WSN. A common approach for the physical layer measurement is to rely on the Received Signal Strength Indicator (RSSI) values provided by the chipset. Because, the measuraments and calculation involved with RSSI are less complicated, and RSSI values are easily available [3]. Anyway, the RSSI data analysis present some drawbacks due to the low-resolution of the RSSI data available on most WSN plattforms. Therefore, as the computation could be more heavy the I/Q data analysis give us more information about the channel properties and studying the distribution, by means of theoretical model like Rician and Rayleigh distribution [4], so that it is possiblie, it is possible to infer the signal properties. On top of this, machine learning methods are built and trained to classify the radio link propagation-state both offline and online way, with a variable level of complexity and performance.

1.3 Problem Statement

In this thesis, we focus on the radio link state. In particular, we want to distinguish between **Line-of-Sight** (LOS) and **No-line-of-Sight** (NLOS)

conditions. Starting with the analysis of received signals, our goals are analyze, detect and classify the sources of link disruption during the packet transmission using two approaches: first, based on I/Q data analysis and the second, based on analyzing the Bit Error Pattern distribution [5].

1.4 Thesis Contribution

Although, as mentioned in Section 1.2, there are independent efforts in the literature on these topics, we envision the development of a unified framework for channel state detection and classification framework that moves beyond the existing approaches by incorporating link quality prediction and link/network adaption in the framework. To this end, we use machine learning tools along with the spectrum sensing, signal processing and wireless communication principles to enhance the resilience of industrial WSNs. For this purpose, we introduce the classification procedure; combining the results of signals analysis methods [3] and [6] to categorize the radio channel state. We use a Software Defined Radio as transceiver and receiver. Capturing signal with SDR gives us possibility to analyze the statistical properties of the received signal, enabling the use of various machine learning method.

1.5 Thesis Outline

The rest of the thesis is organized as follows. In Chapter 2 we give a description of wireless sensor networks, in particular industrial wireless sensor networks and their main challenges. We also provide a brief description of machine learning solutions with specific interest to supervised learning-based methods. We then explore other related works about this topic in Chapter 3. Then, we describe in detail the Radio Link Sate in Chapter 4, focusing in particular on the techniques to model the issue. In Chapter 5 we elaborate the problem and how we tackle it. The experiments results on real datasets are discussed in Chapter 6. The thesis is concluded with Chapter 7 where future work is discussed as well.

Chapter 2

Background

In this chapter we introduce the fundamental concepts within the scope of this thesis. We begin by describing the wireless sensor network and the main techniques to address the problems. We then outline a number of machine learning techniques commonly used in signal processing.

2.1 Wireless Sensor Network

A WSN is a network formed by a large number of sensor nodes where each node is equipped with a sensor to detect physical phenomena such as light, heat, pressure. WSNs are regarded as a revolutionary information gathering method to build the information and communication system which will greatly improve the reliability and efficiency of infrastructure systems. Compared with the wired solution, WSNs feature easier deployment and better flexibility of devices. The main features of WSNs are: scalability with respect to the number of nodes in the network, self-organization, self-healing, energy efficiency, a sufficient degree of connectivity among nodes, low-complexity, low cost and size of nodes. With the rapid technological development of sensors, WSNs will become the key technology for IoT.

History

Research on WSNs dates back to the early 1980s when the United States Defense Advanced Research Projects Agency (DARPA) carried out the distributed sensor networks (DSNs) programme for the US military [7].
At that time, the Advanced Research Projects Agency Network (ARPANET) had been in operation for a number of years, with about 200 hosts at universities and research institutes. DSNs were assumed to have many spatially distributed low-cost sensing nodes, collaborating with each other but operated autonomously, with information being routed to whichever node not that can use the Information most effectively. Even though early researchers on sensor networks had the vision of a DSN in mind, the technology was not quite ready. The sensors were rather large (i.e. the size of a shoe box and bigger), and the number of potential applications was thus limited. As related technologies mature, the cost of WSN equipment has dropped dramatically, and their applications are gradually expanding from the military areas to the industrial and commercial field.

2.1.1 Architecture

Modern WSN usually include sensor nodes, actuator nodes, gateways and clients. A number of sensor nodes is deployed in the designed area with the scope of monitoring certain physical properties of the environment. The nodes can be organized in star or mesh topologies, while the WSN protocols usually provide self-organizing capabilities by means of multi-hopping communication.
The ***sensor node*** is one of the main parts of a WSN. The hardware of a sensor node generally includes four parts: the power and power management module, a sensor, a micro-controller, and a wireless transceiver. A sensor is in charge of collecting and transforming the signals, such as light, vibration and chemical signals into electrical signals and then transferring them to the micro-controller. The micro-controller receives the data from the sensor and processes them accordingly.

2.1.2 Network Topologies

WSN nodes are typically organized in one of three types of network topologies:

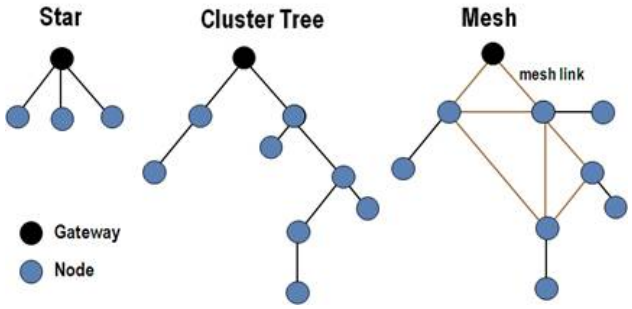

Figure 2.1: WSN topologies

1. **Star**: In a star topology, each node connects directly to a gateway. A single gateway can send messages to as well as receive them from a number of remote nodes. The nodes are not permitted to send messages to each other, which allows low-latency communications between the remote node and the gateway (base station).

2. **Cluster Tree**: It is also called as cascaded star topology. In cluster tree topologies, each node connects to a node that is placed higher in the tree and then to the gateway. The main advantage of the cluster tree topology is that the expansion of a network can be easily done.

3. **Mesh**: It allows transmission of data from one node to another, within its radio transmission range. If a node wants to send a message to a node located out of its radio communication range, it needs an intermediate node to forward the message to the desired node.

2.2 Industrial Wireless Sensor Network

The indoor radio channel has active area of research [8]. Due to increasing use of indoor wireless communications. While wireless communication standards, such as IEEE 802.11, are already largely adopted in office environments for non-critical applications, during the last few years many manufacturing companies have been interested to incorporate wireless communication in their production processes. This implies a certain number of technical challenges due to the highly dynamic changes of the workplace layout overtime, the presence of machinery and highly reflective materials. Hence, the use of WSN in these environments, is subject to typical problems of wireless communications, such as noise, shadowing and multipath fading. The lack of reliability makes it difficult to provide Quality of Service (QoS) guarantees. The characterization of the industrial environments with respect to interference sources and propagation characteristics, is an important step in the development reliable wireless networks and to improve the current IWSN standard, namely WirelessHART, ISA100.11a, WISA, ZigBee [9]. All the aforementioned standards are based on the IEEE 802.15.4 physical layer, but have defined different mechanisms for the upper layers. In addition, we also mention IWLAN [9] that is based on the IEEE 802.11.

2.2.1 IWSN Requirements

Simple deployment, significant cost savings in installations, lack of cabling, high mobility, easy rearrangements related to device configuration and sensor locations make the wireless network technologies also appealing for industrial application. The adaption of wireless technologies appealing in industrial environment possesses additional challenges since the factory environments are typically harsh for wireless communication in terms of interferences, noise and physical obstacles. The following list includes some of the common requirements found in industrial applications:

- **Standardized Solutions**: In order to provide flexibility and freedom to choose among a broad set of suppliers with guaranteed interoper-

ability, standardized and open communication protocols should be used instead of specific proprietary protocols. Besides facilitating repairment and replacement procedures in case of equipment faults, the use of standardized solutions usually extends the lifespan of the network, since there is not a direct dependence on a specific vendor success or failure.

- **Reliable Performance**: For most applications, it is desirable to have network reliability close to 100%. In other words, the sensor data loss should be minimal. Employing redundant paths, self-healing algorithms, and retransmission schemes are some example of possible ways through which highly reliable performance can be achieved, even under the harsh and rapidly changing conditions that characterize industrial environments.

- **Energy efficiency**: In wireless systems, the power needed to feed the devices must come from local sources, normally small batteries. For this reason, efficient energy consumption is a greatly desired quality and wireless sensors must be able to provide a battery life of several months and even years. Considering that self-organizing and self-healing procedures, as well as the data relay, demand different energy levels, it can be said that power consumption is a non-deterministic metric. Energy harvesting from thermal sources, for instance, is one of the most used options to extend the battery life of nodes.

- **Friendly coexistence**: Most wireless technologies operate in the 2.4 GHz ISM frequency band. To enable the deployment of IWSN and Wireless Local Area Networks (WLAN) in the same facilities, which is a remarkably relevant scenario, the involved technologies must be able to coexist without degradation in their performances. The implementation of an opportune frequency channel mapping, channel sensing techniques and the use of frequency hopping algorithms, are some of the widely used solutions to avoid or reduce interference.

- **Operation in Harsh Environments**: Environmental conditions that an IWSN may encounter are strongly dependent on the specific location in which it is deployed. Hence, operational temperatures, humidity, noise and hazard characteristics will greatly vary from case to case. Before deploying a network in a plant, strict verification should be performed to ensure that the equipment used is capable of withstanding the given conditions and, more importantly, if it complies with the legal regulations of the country.

2.2.2 IEEE 802.15.4

The Institute of Electrical and Electronics Engineers (IEEE) supports many working groups to develop and maintain wireless and wired communications standards. For example, 802.3 is wired Ethernet and 802.11 is for wireless LANs (WLANs), also known as Wi-Fi. The 802.15 group of standards specifies a variety of wireless personal area networks (WPANs) for different applications. For instance, 802.15.1 is Bluetooth, 802.15.3 is a high-data-rate category for ultra-wideband (UWB) technologies, and 802.15.6 is for body area networks (BAN). There are several others [10]. The 802.15.4 standard defines the physical layer (PHY) and media access control (MAC) layer of the Open Systems Interconnection (OSI) model of network operation. The PHY defines frequency, power, modulation, and other wireless conditions of the link. The MAC defines the format of the data handling. The remaining layers define other measures for handing the data and related protocol enhancements including the final application. The goal of the standard is to provide a base format to which other protocols and features could be added by way of the upper layers. While three frequency assignments are available, the 2.4-GHz band is by far the most widely used 2.1. Most available chips and modules use this popular ISM band.

Several society, such as ISA, have been actively pushing the applications of wireless technologies in industrial automation [11]. Moreover different standard are realized like ZigBee, WirelessHart and ISA100 that use the same physical layer of IEEE 802.15.4, but they differ substantially concerning the medium access control level (MAC) [10], [9].

ZigBee

ZigBee is a standard containing a suite of protocols using low-power radio based on the IEEE 802.15.4 standard. Mesh network configuration of ZigBee[11] is considered to be a cost-effective solution for the use in different industrial applications. The nodes consume low energy and support different topologies. The standard uses the 868 MHz band in Europe, 915 MHz in the United States, and 2.4 GHz globally. Offset Quadrature Phase Shift Keying (OQPSK) modulation is used at 2.4 GHz, whereas binary phase shift keying (BPSK) modulation is used for the 868 MHz and 915 MHz. Generally, ZigBee targets applications that require low data rate and long battery life, making it a good candidate for wireless sensor and control applications. In order to increase the IEEE 802.15.4 network lifetime, the nodes are usually required to transmitt at a low power level. This makes them more vulnerable to noise, interference, and multipath distortion.

WirelessHart

WirelessHart is a sensor network technology based on the highway addressable remote transducer (HART) protocol. It uses the 802.15.4 standard with O-QPSK modulation at 2.4 GHz achieving a data rate of up to 250 kb/s. Direct sequence spread spectrum [9] (DSSS) is used together with a time-slotted channel hopping (TSCH) technique [9] to mitigate the effect on interference and multipath distortion on packet-by-packet basis. Unlike WirelessHART [11], ZigBee systems uses the same static channel and are thus more susceptible to noise, interference, and multipath effects.This makes WirelessHart more robust than ZigBee for deployment in harsh industrial environments.

In addition to DSSS and TSCH, link-level acknowledgment (ACK) for package re-transmission, graphic routing for path diversity, and transport protocol with end-to-end acknowledgments are used to achieve reliable data transmission.

ISA100

This standard for industrial automation and control applications was developed by the International Society of Automation (ISA). The ISA100 physical layer is based on the IEEE 802.15.4 standard at 2.4 GHz with DSSS and FHSS, providing a data rate of up to 250 kb/s. The main difference between ISA100 and WirelessHART is the main goals of each standard. While WirelessHART is designed to address issues such as reliability, security and interoperability. In general, ISA100 is designed to have broad coverage of industrial automation networks and aims to converge/assimilate existing networks using different communication protocols.

2.2.3 RF Interference

As mentioned previously, industrial applications have higher QoS requirements than typically found in homes and offices. More communication devices are involved and their number is variable. It is necessary to meet specific safety and security requirements, and performance must be deterministic with certain degradation. Coupled with the harsh environment, this means that the spectrum resources vary over time and space. This situation may be exacerbated by device mobility and traffic fluctuations. When different radio signals exist in the same place, at the same time and in a common frequency range, then *Radio Frequnecy Interference (RFI)* occurs[9]. This is particularly a problem when using devices that operate in the ISM and Unlicensed National Information Infrastructure (U-NI) bands, which are both unlicensed and used for different networks including Wireless Personal Area Networks (WPANs), such as WSN, and Wireless Local Area Networks (WLANs). This can be exacerbated by poor frequency planning and a overly crowded frequency spectrum. WirelessHART, ISA 100.11.a, ZigBee, Wi-Fi,

Bluetooth device operates in the 2.4 GHz ISM band, as do other devices. The following table [9] gives a comparison of different industrial wireless platforms that need to coexist.

	IWLAN	**ZigBee**	**WirelessHart**	**ISA 100.11a**	**WISA**
Bandwidth	22 MHz	2 MHz	2 MHz	2 MHz	1 MHz
Channels. Selection	14, static	16, static	15, dynamic	15, dynamic	15, dynamic
Data Rate	11-54 Mps	250 kps	250 kps	250 kps	1 Mps
Frequency Band(s)	2.4 GHz, 5 GHz	2.4 GHz	2.4 GHz	2.4 GHz	2.4 GHz
MAC Layer	IEEE 802.11	IEEE 802.15.4	Proprietary	Proprietary	Proprietary
Radio	IEEE 802.11b/g/a	IEEE 802.15.4	IEEE 802.15.4	IEEE 802.15.4	IEEE 802.15.1

Table 2.1: Wireless Industrial Standards

The effect of the interference in these coexistence standard could also be seen in the channel overlapping as shown in the following figure 2.2.

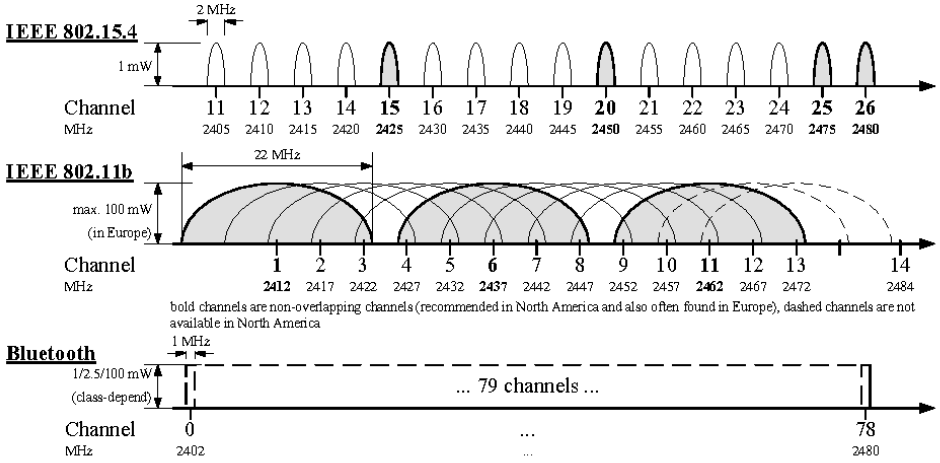

Figure 2.2: Overlapping of IEEE 802.15.4 and IEEE 802.11b and Bluethooth channel allocation

IEEE 802.11

IEEE 802.11 is a family of progressively improved wireless local area network (WLAN) standards [11]. In 1999, the IEEE 802.11b standard was published to operate in the unlicensed 2.4 GHz band with a maximum transmission rate of 11 Mb/s. IEEE 802.11a was released later in the same year to support up to 54 Mb/s in the 5 GHz frequency band. Further enhancements of the standard, IEEE 802.11g, allowed the same rates (54 Mb/s) to be obtained in the 2.4 GHz band, while traffic differentiation mechanisms, were introduced in a subsequent amendment, referred as IEEE 802.11e. Several important improvements were further introduced in IEEE 802.11n to support multiple-input multiple-output (MIMO) and channel bonding capabilities to increase the reliability, coverage, and transmission rate (up to 6000 Mb/s).

The above discussed IEEE 802.11 family of standards have been considered for the use in various industrial applications. However, one of the main challenges of these systems is the issue of interference.

In the 2.4 GHz band, three non-overlapping channels are used to create a micro-cellular architecture, which is not enough to sufficiently isolate micro-cells on the same channel. A significant amount of system capacity is thus lost to co-channel interference. IEEE 802.11n reduces the interference by utilizing more channels in the 5 GHz band, but cannot entirely eliminate the issue. The use of antenna beamforming in IEEE 802.11n dramatically increases the range but also leads to increased interference over distance. Although the IEEE 802.11 family of standards can provide relatively high data rates, the issue of interference will continue to be a key challenge in designing robust highly reliable industrial wireless networks. In addition, the reduction in performance of these systems with regard to the propagation characteristics of harsh industrial environments could be significant.

2.2.4 Multipath Fading

Compared to other indoor and outdoor, the industrial radio channel is usually harsher with respect to radio propagation. In addition, the wireless channel in many industries is non-stationary in the long term, which can cause abrupt changes in the characteristics of the channel over time. Common characteristics of these environments are affecting by large-scale and small-sale fading causing multipath. Multipath is the propagation phenomenon that results in radio signals reaching the receiving antenna from two or more paths. The effects of multipath include constructive and destructive interference, phase shifting of the signal cause errors and impact of the quality of the communications.

Fading

The term ***fading*** refers to rapid fluctuations of the amplitude, phase of a radio signal over a short period o short travel distance.

In principle, the following are the main multipath effects:

1. Rapid changes in signal strength over a small travel distance or time interval.

2. Random frequency modulation due to varying Doppler shifts on different multipath signals.

3. Time dispersion or echoes by multipath propagation delays.

In wireless communications, the presence of reflectors in the environment surrounding a transmitter and receiver create multiple paths that a transmitted signal can traverse. As a result, the receiver sees the superposition of multiple copies of the transmitted signal, each traversing a different path. Each signal copy will experienced differences in attenuation, delay and phase shift while travelling from the source to the receiver. This can result in either constructive or destructive interference, amplifying or attenuating the signal power seen at the receiver.

Thus, in communication it could produce several problems:

- Inter-symbolic Interference
- Large fluctuation of received power

These drawbacks bring errors in receiving symbols.

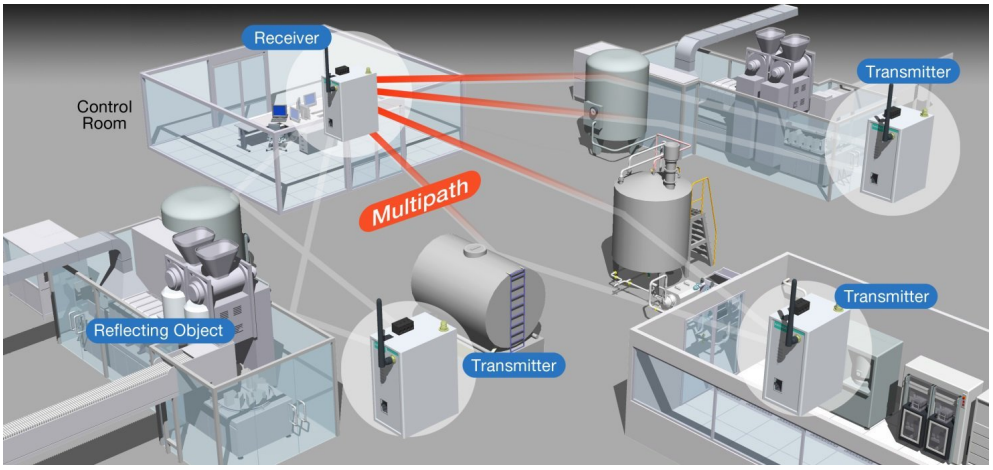

Figure 2.3: Multi-path fading in Wireless Communication

2.2.5 Slow and Fast Fading

Mathematically, fading is usually modeled as a time-varying random change in the amplitude and phase of the transmitted signal [4]. The terms **slow** and **fast** fading refer to the rate at which the magnitude and phase change imposed by the channel on the signal changes. The coherence time is a measure of the minimum time required for the magnitude change of the channel to become decorrelated from its previous value.

- **Slow Fading:** Slow fading arises when the coherence time of the channel is large relative to the delay constraint of the channel. In this regime, the amplitude and phase change imposed by the channel can be considered roughly constant over the period of use. Slow fading can be caused by events such as **shadowing**, where a large obstruction such as a hill or large building obscures the main signal path between the

transmitter and the receiver. The amplitude change caused by shadowing is often modeled using a log-normal distribution with a standard deviation according to the Log Distance **Path Loss Model**.

$$PL = P_T + G_T + G_R - L_T - L_R - P_R \tag{2.1}$$

In 2.1:

- P_T: is de transmitter power in dBm.
- G_T and G_R: are the T_x and R_x antenna gain in dBi.
- L_T and L_R: are the antenna cable losses in dB.
- P_R: is the local mean received power.

- **Fast Fading:** : Fast fading occurs when the coherence time of the channel is small relative to the delay constraint of the channel. In this regime, the amplitude and phase change imposed by the channel varies considerably over the period of use.

Temporal Fading

Temporal fading is defined as the variability of the received power over time at a fixed location in the propagation environment [12]. To determine temporal fading properties of the industrial environment, the measurement cart was placed at fixed locations in specific areas containing a lot of movement. These areas exhibit worst-case temporal fading and will limit the performance of an industrial wireless communication system the most. It was shown the received signal envelope overt time in a fixed location in the industrial environment exhibits Ricean fading properties. So, it is clear that temporal fading behaviour is not determined by stationary physical characteristics of the environments such as LOS.

Several models have been proposed to explain this phenomenon, but the most common model is considering fading from a statistical point of view. The most popular of these models are the Rayleigh and Rician model.

2.2.6 Rayleigh Distribution

This model is used to describe the NLOS propagation. Let there be two multipath signals *S1* and *S2* received at two different time instants due to the presence of obstacles. Now there can be either constructive or destructive interference between the two signals. Let E_n be the electric field and Θ_n be the relative phase of the various multipath signals. So we have:

$$\widetilde{E} = \sum_{n=1}^{N} E_n e^{j}\theta_n \tag{2.2}$$

Now if N $\rightarrow \infty$ (i.e are sufficiently large number of multipaths) and all the E_n are IID distributed, then by Central Limit Theorem we have:

$$\lim_{N\to\infty} \widetilde{E} = \lim_{N\to\infty} \sum_{n=1}^{N} E_n e^{j}\theta_n = Z_r + jZ_i = Re^{j\phi} \tag{2.3}$$

where Z_r and Z_i are the Gaussian Random variables. For the above case:

$$R = \sqrt{Z_r^2 + Z_i^2} \tag{2.4}$$

and

$$\phi = \tan^{-1}\frac{Z_i}{Z_r} \tag{2.5}$$

For all practical purpose, we assume that the relative phase Θ_n is uniformaly distributed.

$$E[e^{j\theta_n}] = \frac{1}{2\pi}\int_0^{2\pi} e^{j\theta_n} d\theta = 0 \tag{2.6}$$

It can be seen that E_n and Θ_n are independent. So,

$$E[\widetilde{E}] = E[\sum E_n e^{j\theta_n}] = 0 \tag{2.7}$$

$$E[|\widetilde{E}|^2] = E[\sum E_n e^{j\theta_n} \sum E_n^* e^{-j\theta_n}] = E[\sum_m \sum_n E_n E_m e^{j(\theta_n - \theta_m)}] = \sum_{n=1}^{N} {E_n}^2 = P_0 \tag{2.8}$$

where P_0 is the total power obtained. To find the Cumulative Distribution Function (CDF) of R:

$$F_R(r) = P_r(R \leq r) = \int_A \int f_{Z_i,Z_r}(z_i, z_r) dz_i dz_r \tag{2.9}$$

where A is determined by the values taken by the dummy variable r. Let Z_i and Z_r be ze mean Gaussian RVs. Hence, the CDF can be written as

$$F_R(r) = \int_A \int \frac{1}{\sqrt{2\pi}\sigma^2} e^{\frac{-(Z_r^2+Z_i^2)}{2\sigma^2}} dZ_i dZ_r \tag{2.10}$$

Let $Z_r = p\cos(\Theta)$ and $Z_i = p\sin(\Theta)$. So we have:

$$F_R(r) = \int_0^{2\pi} \int_0^{2\pi} \frac{1}{\sqrt{2\pi}\sigma^2} e^{\frac{-p^2}{2\sigma^2}} p dp d\theta = 1 - e^{\frac{-r^2}{2\sigma^2}} \tag{2.11}$$

Above equation is valid for all r $\geq$ 0. The PDF, 2.4 is known as Rayleigh distribution, can be written as 2.12 and is shown in following figure 2.4 with different σ values.:

$$f_R(r) = \frac{r}{\sigma^2} e^{\frac{-r^2}{2\sigma^2}} \tag{2.12}$$

This equation too is valid for all r $\geq$ 0.

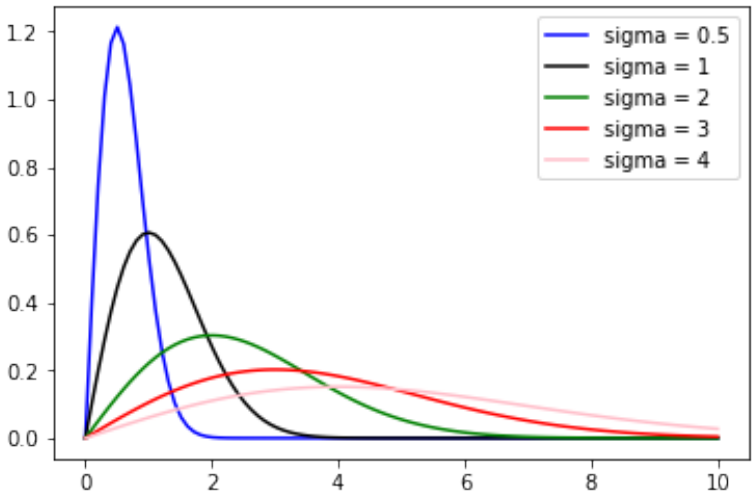

Figure 2.4: Rayleigh distribution

2.2.7 Rice Distribution

Rician Fading originates from the presence of a direct path. It is described by the Rice distribution, which is:

$$f_R(r) = \frac{r}{\sigma^2} e^{\frac{-(r^2+A^2)}{2\sigma^2}} I_0(\frac{A_r}{\sigma^2}) \tag{2.13}$$

for all A $\geq$ 0 and r $\geq$ 0. Here A is the peak amplitude of the dominant signal and $I_0(.)$ is the modified Bessel function of the first kind and zeroth order. We can define the K factor for this distribution as

$$K_{dB} = 10 \log \frac{A^2}{2\sigma^2} \tag{2.14}$$

As A $\to$ 0 then $K_{dB} \to \infty$

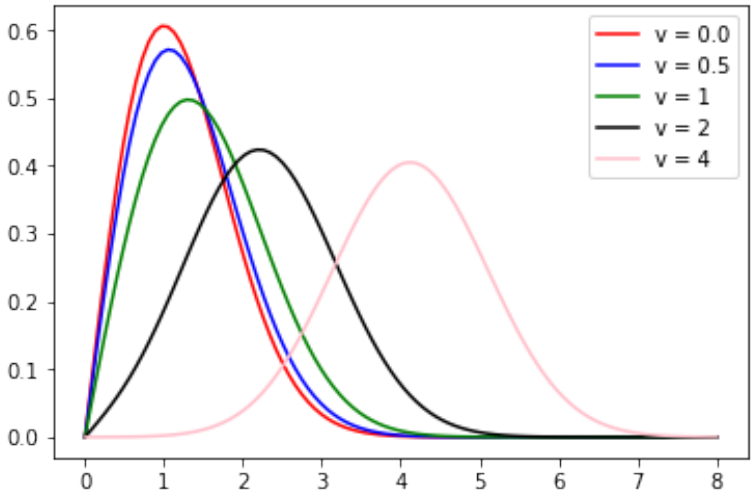

Figure 2.5: Rice distribution

In conclusion, in the absence of a dominating component the envelope of the received signal is shown to be Rayleigh-distributed (NLOS), while in the presence of a dominant, static component, typically a line-of-sight path, the envelope obeys a Rice distribution. In [12] it is shown how temporal fading is Ricean distributed.

2.3 Signal Representation

In order to take advantage of the theoretical-distribution model, and to understand the statistical properties of the radio channel state, we use the standard in-phase/quadrature (I/Q) representation for band-pass signal analysis.

2.3.1 I/Q Data

I/Q data shows the changes in magnitude (or amplitude) and phase of a sine wave. If amplitude and phase changes occur in an orderly, predetermined fashion, since in digital modulations, the information is conveyed by means of appropriate amplitude and phase changes in the transmitted signal [12], I/Q provides a convenient representation for the analysis of these signals [13]. Modulation modifies certain characteristics of a higher frequency carrier signal according to a lower frequency message, or information, signal. I/Q representation is therefore is highly prevalent in Radio Frequency communications systems, and more generally in signal modulation.

Background on Signals

Signal modulation encodes information by changing the parameters of a sine wave. The equation representing a sine wave is as follows:

$$A_c cos(2\pi f_c t + \phi) \tag{2.15}$$

where A_c is the amplitude, f_c the frequency and ϕ is the phase. The equation of sine wave above, shows that it is limited to making changes to the amplitude, frequency and phase of a sine wave to encode information. Frequency is simply the change rate of the phase of a sine wave (frequency is the first derivative of phase), so frequency and phase of the sine wave equation can be collectively referred to as the phase angle. Therefore, we can represent the instantaneous state of a sine wave with a vector in the complex plane using amplitude (magnitude) and phase coordinates in a polar coordinate system.

We consider now the polar representation of the sine wave in Figure 2.6:

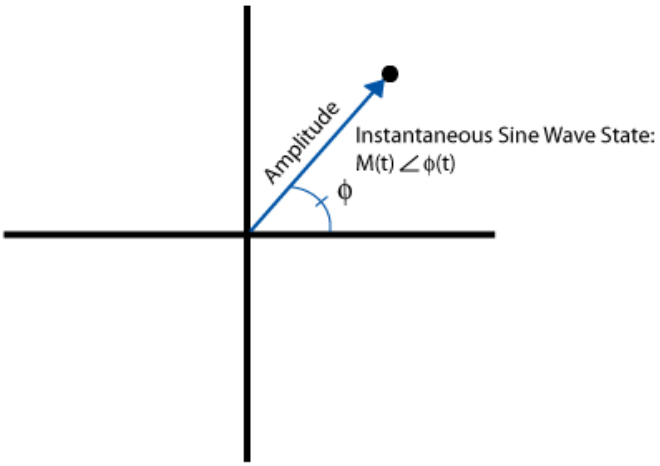

Figure 2.6: Polar representation

It is shown that the distance from the origin to the black point represents the ***amplitude*** (magnitude) of the sine wave and the angle from the horizontal axis to the line represents the ***phase***. Thus, the distance from the origin to the point remains the same as long as the amplitude of the sine wave is not changing (modulating). The phase of the point changes according to the current state of the sine wave. If the amplitude does not change during one revolution, the dot maps out a circle around the origin with radius equal to the amplitude along which the point travels at a rate of one cycle per second. Because phase is a relative measurement, imagine that the phase reference used is a sine wave of frequency equal to the sine wave represented by the amplitude and phase points. If the reference sine wave frequency and the plotted sine wave frequency are the same, the rate of change of the two signals' phase is the same, and the rotation of the sine wave around the origin becomes stationary.

In this case, a single amplitude/phase point can represent a sine wave of frequency equal to the reference frequency. Any phase rotation around the origin indicates a frequency difference between the reference sine wave and the sine wave being plotted.

In fact, I/Q data is merely a translation of amplitude and phase data from a polar coordinate system to a Cartesian (X,Y) coordinate system. Using trigonometry, it is possible to convert the polar coordinate sine wave information into Cartesian I/Q sine wave data 2.16 and 2.17.

$$I(t) = M(t)\cos(\phi(t)) \tag{2.16}$$

$$Q(t) = M(t)\sin(\phi(t)) \tag{2.17}$$

IQ data in Communication System

To better understand why IQ data are used in communication systems, it is necessary understand modulation basis. Signal modulation can be divided into two broad categories: analog modulation and digital modulation. Analog or digital refers the representation (digital or analog) of the source data to be modulated. Both analog modulation and digital modulation involve changing the carrier wave amplitude, frequency or phase (or combination of amplitude and phase simultaneously) according to the message data. Amplitude modulation (AM), frequency modulation (FM) or phase modulation (PM) are all examples of analog modulation. With amplitude modulation, the carrier sine wave amplitude is modulated according to the message signal. The same idea holds true for frequency and phase modulation. For AM, the message signal is the blue sine wave that forms the "envelope" of the higher frequency carrier sine wave. For FM, the message data is the dashed square wave.

As the following figure 2.7 illustrates, this would be the resulting carrier signal changes between two distinct frequency states [12].

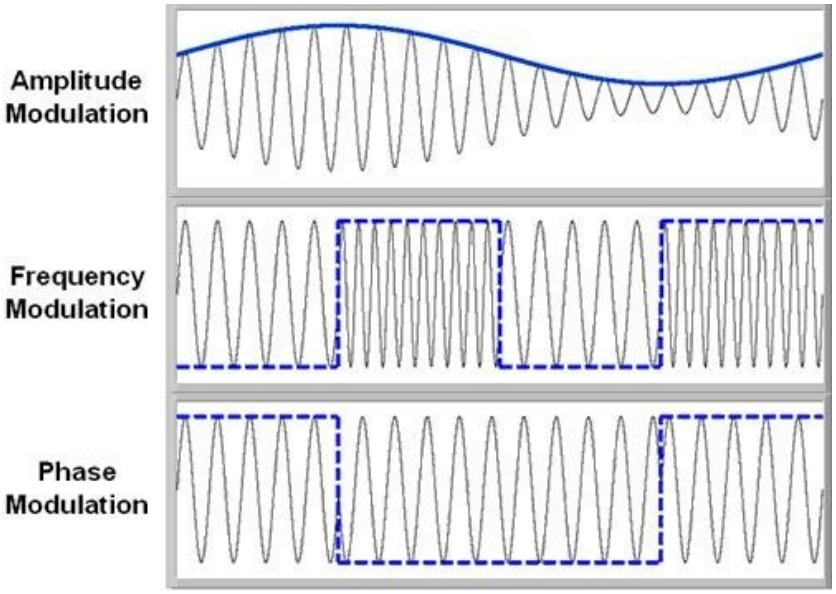

Figure 2.7: Time domain representation of AM, FM, and PM Signals [13]

Each frequency state represents the high and low state of the message signal. In the case of the message signal being a sine wave, there would be a more gradual change in frequency, which would be more difficult to see. For PM, notice the distinct phase change at the edges of the dashed square wave message signal.

As mentioned earlier, if only the carrier sine wave amplitude changes with respect to time (proportional to the message signal), as is the case with AM modulation, the I/Q plane graph changes only with respect to the distance from the origin to the I/Q points.

The Figure 2.8 shows that the I/Q data points vary in amplitude only with the phase fixed at 45 degrees. It is not possible to tell much about the message signal, only that it is amplitude modulated. However, the I/Q data points vary in magnitude with respect to time, so, essentially, it is possible see a representation of the message signal.

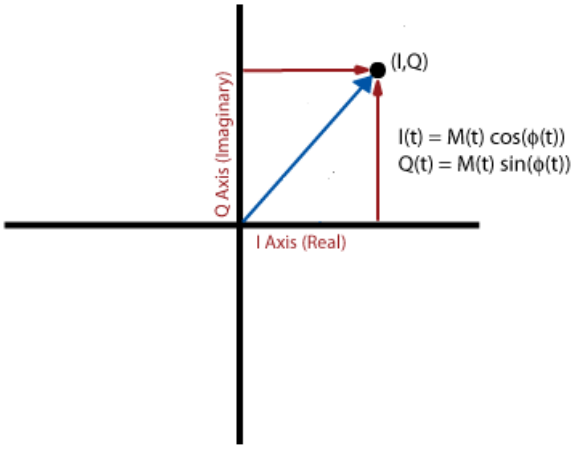

Figure 2.8: I/Q Data in the complex plane

Why IQ data?

Using a specific tool like LabVIEW's 3D graph control, the third axis of time to illustrate the message signal can be seen [13]. In result the I/Q data represents the message signal. Because the I/Q data waveforms are Cartesian translations of the polar amplitude and phase waveforms, it may have trouble determining the nature of the message signal. Because amplitude and phase data seem more intuitive, it is better use polar amplitude and phase data instead of Cartesian I and Q data. However, practical hardware design concerns make I and Q data the better choice. According to the trigonometric identity [13]:

$$\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta) \tag{2.18}$$

$$A\cos(2\pi f_c t + \phi) = A\cos(2\pi f_c t)\cos(\phi) - A\sin(2\pi f_c t)\sin(\phi) \tag{2.19}$$

$$= I\cos(2\pi f_c t) - Q\sin(2\pi f_c t) \tag{2.20}$$

where $I = A\cos(\phi)$ and $Q = A\sin(\phi)$ are respectively the amplitude of the **in-phase carrier** and amplitude of the **quadrature-phase carrier** .

Taking 2.20 and the fact that difference between a sine wave and a cosine wave of the same frequency is 90° phase offset between them into consideration, the amplitude, frequency and phase of a modulating carrier sine wave can be controlled by simply manipulating the amplitudes of separate I and Q input signals. With this method, it is not necessary to directly vary the phase of an RF carrier sine wave, but it is possible to achieve the same effect by manipulating the amplitudes of I an Q components. Of course, the second half of the equation is a sine wave and the first half is a cosine wave,so the hardware design must ensure a 90-degree phase shift between the carrier signals used for the I and Q mixers, but this addition is a simpler design issue than the aforementioned direct phase manipulation.

2.4 Cognitive Radio

We must keep in mind that the radio environment where a typical wireless sensor network operates in is becoming more and more complex. This is due to new access to radio technologies and the ongoing densification of network infrastructure in order to meet exponentially increasing capacity demands. Spectrum sharing and coexistence issues both in co-channel and adjacent channel deployments are thus becoming more and more important for radio resource management, in order to avoid excessing interference between wireless networks.

Cognitive wireless networking principles have become an increasingly investigated approach for solving such complex interference management problems in an automated fashion, in particularly without introducing explicit signaling protocols for each combination of interacting wireless technologies. However, using such cognitive approaches requires accurate estimation of the state of the radio environment, in particular forming an understanding of which kinds of wireless technologies are deployed and are causing interference with the managed network. To improve the quality of wireless communication in IWSNs we introduce signal analysis methods and a classification algorithm to identify radio channel disturbances typical for industrial environments in the physical layer. Signals analysis methods based on IQ data

and statistical analysis of the magnitude received signal properties help to recognize the temporal disturbances affecting the signal propagation in a radio channel.

2.5 Machine Learning

Machine learning techniques provide powerful tools for signal analysis and classification [14]-[15].

There are many different types of machine learning algorithms, and they are typically grouped by either learning style (i.e. supervised learning, unsupervised learning, semi-supervised learning) or by similarity in form or function (i.e. classification, regression, decision tree, clustering, deep learning). After a couple of AI winters and periods of false hope over the past four decades, rapid advances in data storage and computer processing power have dramatically changed the game in recent years. Machine learning was born from pattern recognition and the assumption that computers can learn without being programmed to perform specific tasks.

Machine learning is the science of getting computers to act without being explicitly programmed, but instead letting them learn a few tricks on their own. [17]

The fundamental goal of machine learning algorithms is to generalize beyond the training samples i.e. successfully interpret data that it has never 'seen' before. From a mathematical point of view, machine learning covers problems where we want to learn the "best" mapping f between the input X and output Y by observing a subset $X_{train} \subset$ X. The meaning of "best" depends on the problem to be solved and if the desired output Y is known or not.

The following image shows a general view of machine learning algorithms, in particular based on classification:

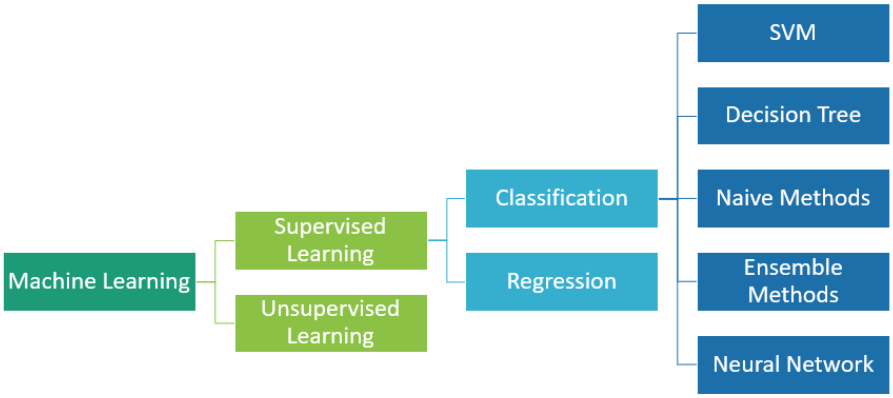

Figure 2.9: General view of machine learning algorithms, with specific interest towards the supervised learning approaches

2.5.1 Unsupervised Learning

Unsupervised machine learning is more closely aligned with what some call true artificial intelligence — the idea that a computer can learn to identify complex processes and patterns without a human to provide guidance along the way. Although unsupervised learning is prohibitively complex for some simpler enterprise use cases, it opens the doors to solving problems that humans normally are not able to tackle. It is used in cases where the input X is known and the output Y is unknown. The aim of unsupervised learning is to obtain more information on the input and to learn more about it. For this reason it is difficult to define a correct answer.

2.5.2 Supervised Learning

Supervised learning is the more commonly used learning algorithms. The term supervised learning derives from the requirement that the algorithm's input (X) and possible outputs (Y) are already known. For example, a clas-

sification algorithm will learn to identify animals after being trained on a dataset of images that are properly labeled with the species of the animal and some identifying characteristics. In that case, the purpose is to minimize the error between the predicted target $\tilde{y}^{(i)} = f(x^{(i)})$ $\{x^{(i)} \in X\}$ and the real/desired output $y^{(i)} \in Y$. So, in Supervised learning all the examples must be labelled, each case by hand, therefore it may not be a feasible approach for every problem. However, the advantage of labeling the data by hand is that we can decide on the desired outcome of system. It is an convenient approach when historical data is very likely to predict future events.

Regression

The goal of the regression problem is to find an approximating function f from input variables X to a continuous real-valued output variable Y. In these problems, the quantity like amount, price, size is defined such as predicted, therefore the model must be evaluated using an error, like the mean squared error or other metrics. Linear regression is the simplest case of regression with the goal to find a linear function that maps inputs and outputs. Focusing on the single-variate case linear regression could be qualified in the following formal way:

$$wx + b = y \tag{2.21}$$

where w and b are respectively **weight** and **bias** and they are the parameter to learn. However, multivariate regression is a more common approach and the model can be derived from the previous one,

$$\sum_{k=0}^{K} w_k x_k + b_k = \mathbf{w}^T \mathbf{x} + b = y \tag{2.22}$$

where $\mathbf{w}$ and $\mathbf{x}$ are vectors and b is the sum of all biases.

Classification

In machine learning and statistics, classification is a supervised learning approach in which the computer program learns from the data input given to

it and then utilizes this learning in order to classify new observations. This data set may either be simply bi-class (like identifying whether the person is male or female or that the mail is spam or non-spam) or it may be multi-class. Some examples of classification problems are speech recognition or handwriting recognition. In general, classification models can predict the class directly or continuous values that are then converted into probabilities. In the last case, the class with the highest probability is the predicted label. The labels are commonly represented by one-hot encoding, a sparse binary vector with a 1 in the i-th position to indicate the actual class; an example has usually only one class. Since the output is discrete, the classification **accuracy** can be computed and used as an evaluation measure for the model. Various classifiers can be found in literature, the most common ones are the following [18]:

- **Naive Bayes Classifier (Generative Learning Model):** It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability. The naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, naive Bayes is known to outperform even highly sophisticated classification methods.

- **SVM:** A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. The operation of the SVM algorithm is based on finding the hyperplane that gives the largest minimum distance to the training examples. Secondly, this distance receives the important name of margin within SVM's theory. Therefore, the optimal separating hyperplane maximizes the margin of the training data.

- **Decision Trees:** The concept of decision trees is that it builds classification or regression models in the form of a tree structure. It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node has two or more branches and a leaf node represents a classification or decision. The top-most decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.

- **Ensemble Methods:** Ensemble methods are meta-algorithms that combine several machine learning techniques into one predictive model in order to decrease variance (bagging), bias (boosting), or improve predictions (stacking).

- **Neural Network:** A Neural Network (NN) consists of units (neurons), arranged in layers, which convert an input vector into some output. Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer. Generally. the networks are defined to be feed-forward: which means that a unit feeds its output to all the units on the next layer, but there is no feedback to the previous layer. Weightings are applied to the signals passing from one unit to another and it is these weightings which are tuned in the training phase to adapt a neural network to the particular problem at hand.

Chapter 3

Related Work

Our focus is on system reliability with specific interest towards classification of channel disturbances. Different models have been used for interference/disturbance classification in wireless communication literature. Work such as [14]-[15]-[16] use different techniques to classify disturbance like Support Vector Machine or Random Forrest or Decision Tree classifier that have been already successfully employed for improve wireless communication. All off these, explain how it could possibly include machine learning and wireless communication in different level of art. For example in [15], by combining signal bandwidth, envelope information and their slender extraction, with the intelligently tailored supervised-learning the Authors enable on-board burst-based interference identification, predominantly in real-time. In [14] the focus is instead on the impact of interference from modulated signals and the influence of realistic wireless channel conditions on classification performance. The paper [3] introduces signal analysis and a classification algorithm to identify radio channel disturbance, based on probability density function (PDF) analysis and spectrum analysis. In particular the Authors have arranged an n-dimensional feature vector exploiting shape parameter coming out of Histogram analysis, obtained from I/Q data [13], and RF objects from the spectrum of received signal. Regarding the classification algorithm, they use a practical and simple approach - *Nearest Neighbour Rule*: samples that are close in feature space are likely to belong the same class. Anyway, the

considered environment it is not for typical industrial environments. In [6] the detection of channel state is based on the detection of the shape of the amplitude histogram of received signal, which reveals the nature of interferences in the radio channel. The Authors classify nominal and disturbed channel models as well as disturbance schemes and compare the measured signal against these. The reference channel model classes are acquired by channel simulations.The real environments tests have also been run promising results for disturbances identifications were achieved. The idea is that the most significant has been noticed and that the most significant source of radio signal distortion can be characterized by the shape of the amplitude histograms of the received signal. For example, multipath propagation phenomenon result in the broadening of the amplitude of the histogram's shape, due to the radio waves destructive and constructive interference in which two waves superimpose to form a resultant wave of greater or lower amplitude [6]. In order to compare the histograms of the received signal and predefined channel models' histograms, we use the chi-squared measure which is typical for expressing similarities/difference of histograms, helping to identify the disturbance.

Our work consists of a mixed approach between [3] and [6]. It means that when considering a typical industrial environments were there will be both large-scale and temporal fading, we introduce a novel histogram approach (of I/Q data) based on PDF analysis and amplitude analysis. Thus, obtaining all features necessary for our classifier. Moreover, we introduce an alternative approach to classify the channel disturbances based on idea [5], that is analyzing the Bit Error Pattern for future improvements of reliability and coexistence of radio systems in these harsh conditions.

Chapter 4

Radio Link Characterization

4.1 Intuition

The idea of our work is based on the concept of introducing some intelligence to each node of the network so that the reliability of the system can be improved. To this end, the intuition could be defined from the following steps:

1. **Monitoring:** Consists of channel sensing, which means that it has the ability to sense, measure, learn and have an awareness of channel features, the working environment and other requirements. For this purpose, it is useful analyze I/Q data.

2. **Problem Detection:** During the monitoring phase it is possible to understand that there are problems in the transmission. Thus, it will be necessary to define some metrics that allows us to understand when problems occurs.

3. **Problem Identification:** When the system detects a problem, the next step is to identify the kind of problem. About that, in our work we define a classification algorithms to understand the signal disturbance.

4. **Improvement:** Following the identification of the problem, it has been identified that one inelligent step is to apply some countermeasures to improve the reliability and the quality of service and security.

4.2 Classification of the Radio-Link State

In our implementation we focus only in the problem identification because, as mentioned before, our goal is to achieve and classify the properties about the channel propagation.

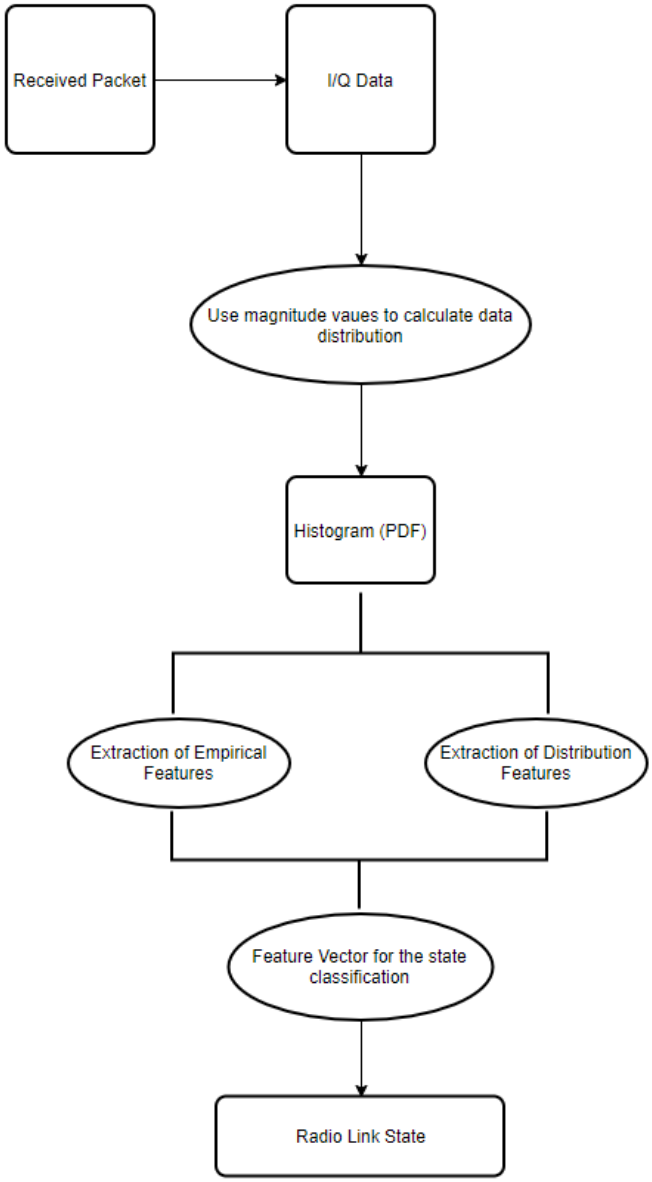

Figure 4.1: General pattern for classification procedure

The recognition and identification of the channel properties is based on the signal magnitude analysis with a classifier. The general pattern for classification procedure includes the following phases: pre-processing of measured data, feature extraction from the measured data and classification.

4.2.1 Data Analysis

The data analysis is based on I/Q data of the received packet. In order to take advantage and information from this data representation we need to understand which could be the possible features and properties that reflect an improvement in the classification accuracy.

Amplitude Histogram

A histogram provides a graphical record of the shape of the data distribution. The x-axis shows all the possible values, while y-axis presents the percentage of input samples that each had as a corresponding value. The continuous-valued counterpart of the histogram is the PDF. The histogram is simply an approximation of a PDF; the count of how many samples each possible value has in the range. The strength of the received signal changes with time, due to fluctuations in the gain of the channel caused by multipath fading or electromagnetic interference [6]. The amplitude values of the received signal form a probability density function of a specific form, or in discrete case a specific shape of histogram. By analyzing the features of the histogram shapes of the received signal and comparing them with characteristics being predefined for different radio channels with different characteristics, we can classify histograms into two main categories histograms related to signal; multipath fading (Ricean and Rayleigh distributed) and histograms related to radio interference.

To express the shape of the histogram in a precise way, it useful define four feature values:

1. **Mean**: It is the mean value of the data.

$$\bar{x} = \frac{\sum x_i}{N} \tag{4.1}$$

   where x_i is the value of data point i and N is the number of data points

2. **Spread**: It is the average squared deviation of the variable's values from the mean

$$spread = \frac{\sum (x_i - \bar{x})^2}{N} \tag{4.2}$$

   where x_i is the value of data point i and N is the number of data points and $\bar{x}$ is the mean value.

3. **Skewness**: The skewness reflects the shape or asymmetry of the distribution: if the value is negative, the data spreads out more to the left of the mean, if it is positive, the data spreads out more to the right, and if it is zero, it is symmetric about the mean

$$skewness = \frac{\sum (x_i - \bar{x})^3}{N} \tag{4.3}$$

4. **Peakedness**: A single-peaked histogram with a lot of distributional weight in the center and in the tails, but not in the shoulders is said to have "fat tails"; or it is "highly peaked."

$$peakedness = \frac{\sum (x_i - \bar{x})^4}{N} \tag{4.4}$$

PDF Analysis of the Received Signal

The PDF analysis is based on the idea that changes in statistical properties of the received signal are strongly correlated with the transitions of the channel states [3].

Thus, matching the PDF of the received signals to the theoretical distributions: Ricean 2.5, Rayleigh 2.4, and deriving the PDF´s shape parameters, allows us to numerically characterize the effects of environments disturbances to the channel state.The PDF shape parameters are estimated from sample data by fitting a probabilistic distribution object to the data. To estimate the distribution parameters from the sample data the Maximum likelihood is used. As previously stated, there are two probability models in literature that are commonly used for the propagation channel:

- Rayleigh PDF:

$$f_\rho(\rho|a,\sigma) = \frac{\rho}{\sigma^2}\exp(-\frac{\rho^2}{2\sigma^2}) \tag{4.5}$$

- Rician PDF:

$$f_\rho(\rho|a,\sigma) = \frac{\rho}{\sigma^2}\exp(-\frac{\rho^2+a^2}{2\sigma^2})I_u(\frac{a\rho}{\sigma^2}) \tag{4.6}$$

4.2.2 Classification Model

In our work, the received digitized *I(n)* and *Q(n)* samples are pre-processed and analyzed. For each received packet, the PDF of the signal magnitude is estimated as a signal magnitude histogram and match to the theoretical PDF 4.2. In parallel, the amplitude of the histogram is computed and all features are created. How the block diagram 4.1 shows. In particular. from a single received packet: *mean*, *spread*, *skewness*, *peakedness* can be computed based on the amplitude of the packet. All of these statistical properties, called also *moments* in literature, represent the so-called **empirical features** because they are obtained from the empirical data, the histogram. On the other hand, the theoretical PDF of the packet is generated, using the empirical/histogram distribution and, from that one, the theoretical distribution is computed by the Rician model.

So, considering 4.6 equation we have that:

- $\rho = \sqrt{I^2(t) + Q^2(t)}$ denotes the magnitude value measured at the receiver.
- σ standard deviation of the real part of the time-harmonic multipath component.
- a corresponds to the signal amplitude due to the LOS path in the absence of all other multipath components.

Although, previously, we distinguished between Rician model and Rayleigh model, the Rician model in our case ensures a better fit to the data[12]. From the Rician PDF, the fitting parameter: ***beta***, ***loc***, ***shape*** 4.6 are estimated. These are used like features to describe a single packet and in addition from them the goodness of the distribution is computed using the **Kologrov-Smirnov test**.

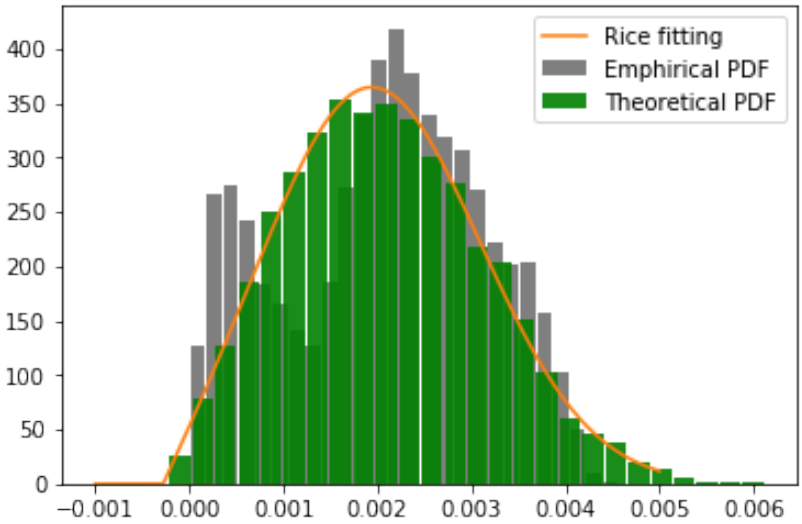

Figure 4.2: Empirical and Theoretical distribution

In statistics, the Kolmogorov–Smirnov test (K–S) is a nonparametric test of the equality of continuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution

(one-sample K–S test), or to compare two samples (two-sample K–S test). The Kolmogorov–Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples [19]. Hence, the value of Kolmogorov-Smirnov test it is also used as a feature. All these properties obtained from the curve fitting could be called like **distribution features**, because it gives us more information about the general data distribution of the packet.

In conclusion, the feature vector is composed by:

- ***Empirical Features:***

Mean	**Spread**	**Skewness**	**Peakdness**

Table 4.1: Empirical Features

- ***Distribution Features:***

Beta	**Location**	**Scale**	**KS-test**

Table 4.2: Distribution Features

When all data are collected with the features that allow us to identify the different conditions in the channel and a dataset has been created, the next step is to use it to classify the packet that are affected by disturbances in supervisioned way. The focus of our work is to classify:

- **LOS**: Line of sight (LoS) is a type of propagation occurring when the transmitter and receiver are in sight, meaning that there is no obstruction of size d such that $d \gg \lambda$, with $\lambda = \frac{1}{f}$ the wavelenght of considered signals.
- **NLOS**: Non-line of sight (NLOS) refers to the path of propagation of a radio frequency (RF) that is obscured (partially or completely) by obstacles, thus making it difficult for the radio signal to pass through.

> Common obstacles between radio transmitters and radio receivers are tall buildings, trees, physical landscape and high-voltage power conductors. While some obstacles absorb and others reflect the radio signal; they all limit the transmission ability of signals.

It could seem as a perfect case where NN would work very well, but generally, neural networks show high variance, high complexity while their set up can be cumbersome. A successful approach to reducing the variance of neural network models is to train multiple models instead of a single one, and to combine the predictions from these models.Furthermore the applicability of NN to low-cost WSN hardware is questionable since of the limited computatoinal resources. This is true, specially, if are thinking of using this classification in a online way, where the model has to be loaded every time for each packet. So, it is better to use **Ensemble Methods** that, according to Ockham's razor *Simplicity leads to greater accuracy*, indeed these assume that it uses a combination of models to increase the accuracy, and in addition, they are more soft and easily manageable.

Chapter 5

Experimental Set-up

In this chapter, we talk about the software and the technologies used, then we describe the data collection and preparation.

5.1 LabVIEW

Laboratory Virtual Instrument Engineering Workbench (LabVIEW) is a system-design platform and development environment for visual programming language from National Instruments [20].
LabVIEW is commonly used for data acquisition, instrument control and industrial automation on a variety operating systems including Windows, various versions of Unix, Linux and macOS. LabVIEW can create programs that run on those platforms and a variety of embedded platforms, including Field Programmable Gate Arrays (FPGAs), Digital Signal Processors (DSPs), and microprocessors.

The LabVIEW program development environment is different from standard C or Java development systems in one important aspect: While other programming systems use text-based languages to create lines of code, LabVIEW uses a graphical programming language, often called **G**, to create programs in a pictorial form called a block diagram. Graphical programming allows you to concentrate on the flow of data within your application, because its simple syntax doesn't obscure what the program is doing.

A LabVIEW program consists of one or more virtual instruments (VIs). Virtual instruments are called as such because of their appearance and operation often imitate actual physical instruments. A VI has three main parts:

- **Front panel:** It is the interactive user interface of a VI, so named because it simulates the front panel of a physical instrument. The front panel can contain knobs, push buttons, graphs, and many other controls (which are user inputs) and indicators (which are program outputs).
- **Block diagram:** It is the VI's source code, constructed in LabVIEW's graphical programming language, G. The block diagram is the actual executable program. The components of a block diagram are lower-level VIs, built-in functions, constants, and program execution control structures. You draw wires to connect the appropriate objects together to define the flow of data between them. Front panel objects have corresponding terminals on the block diagram so data can pass from the user to the program and back to the user.
- **Icon:** In order to use a VI as a subroutine in the block diagram of another VI, it must have an icon. A VI that is used within another VI is called a subVI and is analogous to a subroutine. The icon is a VI's pictorial representation and is used as an object in the block diagram of another VI.

5.1.1 USRP

Universal Software Radio Peripheral (USRP) is a range of Software-Defined Radios (SDR) designed and sold by Ettus Research and its parent company, National Instruments used for RF applications [20]. USRP transceivers can transmit and receive RF signals in several bands, and you can use them for applications in communications education and research. It supports Linux, MacOS, and Windows platforms. Several frameworks including GNU Radio, LabVIEW, MATLAB and Simulink use UHD. The functionality provided by UHD can also be accessed directly with the UHD API, which provides native

support for C++. Any other language that can import C++ functions can also using UHD. This is accomplished in Python through SWIG, for example. Paired with the LabVIEW development environment, USRPs provide an affordable solution that lets you validate wireless algorithms with over-the-air signals.

Figure 5.1: USRP device

5.2 Development Libraries

The development environment chosen is Python, integrated on LabVIEW, for its completeness as to libraries and support for data analysis and the use of tools such as the machine learning. Below are some of the libraries used for development of our solution.

5.2.1 Scikit-Learn

This library, also called sk-learn, provides development tools for data mining and data analysis; makinh available almost all the tools for machine learning, such as classification, regression, clustering and others [21]. This library is open source. Regarding the classification it makes available the implementations of simple use of the most famous algorithms such as Neural Networks, Random Forrest and Gradient Boosting. This library is built and based on 3 other Python libraries: NumPy, SciPy e matplotlib.

5.2.2 Keras

Keras [22] is a library for neural networks and machine learning methods written in Python. It can operate above TensorFlow, Microsoft Cognitive Toolkit, Theano or MXNet. This library in fact does not provide a high high-level interface to use in a way as simple as the aforementioned libraries. It focuses precisely on being user-friendly, modular and extensible. Since 2017 Google has decided to support Keras in the main library of TensorFlow. Keras contains numerous implementations blocks commonly used in neural networks and in common machine learning algorithms.

5.3 Software Design

The software design was based on the aforementioned libraries and programs. It is capable of collecting data using USRP devices integrated in LabVIEW. Moreover using LabView with Python scripts the software can classify real time the packet thus helping us to improve the reliability and the quality of the system.

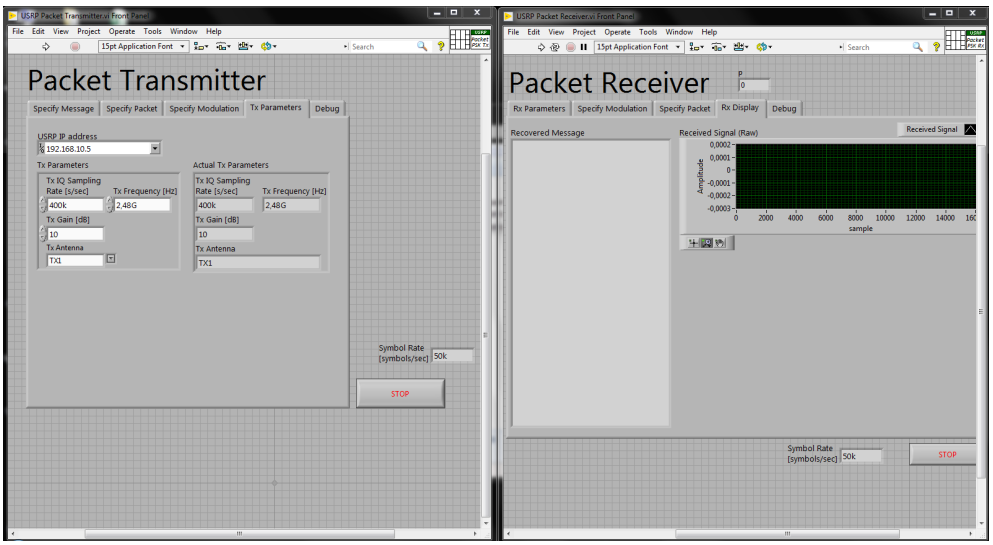

Figure 5.2: Front panel LabView

5.3.1 Data Collection and Preparation

The data collection operation is performed by placing the transmitter and receiver at various distances in different industrial and lab environments-

We collected the I/Q data using LabVIEW for each received packet in a CSV file. The CSV file with the I/Q data is processed using Python functions to create datasets with all features mentioned in Chapter 4. After creating the dataset, exploiting Python's libraries, the classifier has been generated and the best model is stored for the following step.

Implemented Classification Methods

As mentioned before, the choice of the classifier has been done considering the simplicity of the Ensemble methods.

The goal of ensemble methods is to combine the predictions of several base estimators built with a given learning algorithm, in order to improve generalizability/robustness over a single estimator. Two families of ensemble methods are usually distinguished:

- **Bagging**: Building multiple models (typically of the same type) from different subsamples of the training dataset.
- **Boosting**: Building multiple models (typically of the same type) each of which learns to fix the prediction errors of a prior model in the chain.

Random Forest

Random Forests [23] are trained via the bagging method. It consists of randomly sampling subsets of the training data, fitting a model to these smaller data sets, and aggregating the predictions. This method allows several instances to be used repeatedly for the training stage given that we are sampling with replacement. Tree bagging consists of sampling subsets of the training set, fitting a Decision Tree to each, and aggregating their result. The Random Forest method introduces more randomness and diversity by applying the bagging method to the feature space. That is, instead of searching greedily for the best predictors to create branches, it randomly samples elements

of the predictor space, thus adding more diversity and reducing the variance of the trees at the cost of equal or higher bias. This process is also known as "feature bagging" and it is this powerful method what leads to a more robust model.

Gradient Boosting

Gradient Boosting methods is included in the boosting family. Boosting [24] is a method of converting weak learners into strong learners. In boosting, each new tree is a fit on a modified version of the original data set. Gradient Boosting trains many models in a gradual, additive and sequential manner. Performs similarly by using gradients in the loss function. The loss function is a measure indicating how good model's coefficients are at fitting the underlying data. One of the biggest motivations of using gradient boosting is that it allows one to optimize a user specified cost function, instead of a loss function that usually offers less control and does not essentially correspond with real world applications.

5.3.2 Real-time Classification

When the classifier was ready with good result in terms of prediction's accuracy, the LabVIEW code has been executed in real-time way to classify the radio channel state (LOS or NLOS) for each packet calling Python script. The interaction between Python and LabVIEW is fairly problematic because not all type of LabVIEW data are supported in Python.

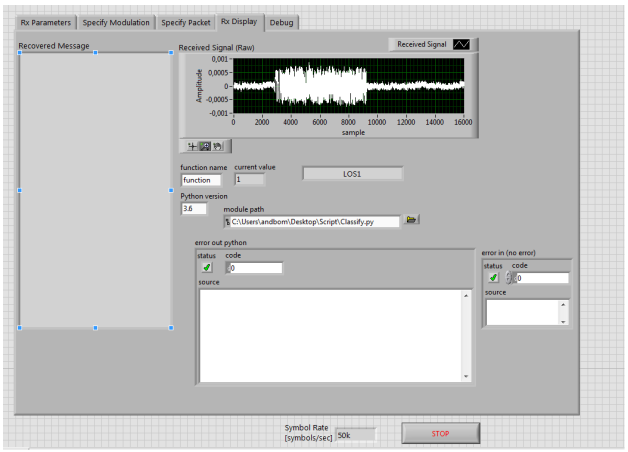

Figure 5.3: LabView front panel Classification LOS packet real-time

The idea is: during the communication of the USRP devices, using the LabView code integrate with Python script, we analyze the transmission. So, exploiting the I/Q data for a single packet and recall our trained classifier, it is possible classify in an online way the channel properties. So, we can define and apply some metrics to improve the system. The countermeasures could be different: for example, let us to consider in the receiver side and we have a tool that allow us to calculate how many packets of the last k packets have been lost. Using this information, after k lost packets, to improve the system applying some changing:

- **LOS packet:** if the packet is labelled like LOS the idea could be the increasing of the transmitting power or change the frequency channel
- **NLOS packet:** otherwise if the packet is NLOS the countermeasures could be changing the channel or changing the structure of the network, that means changing the transmitter (o receiver) device.

Chapter 6

Experiments

In this chapter we report the experimental results that we obtained testing our models on datasets collected from different environments. All the results were obtained by employing 70% of the values in the data set as training set and the 30% as test set. In the effort to maximize the validity of obtained results we have tested the proposed method with data collected in environments with different characteristics: an industrial plant, a semi-industrial workshop and a lab environment.

6.1 Industrial Environment Dataset - Imerys Mineral AB

At the beginning, the first classifier was created based on a real industrial environment where I/Q data was stored in a large binary file with a specific format. This data was obtained in the Imerys Mineral AB facotry in Sundsvall. In particular, the I/Q data in consideration was obtained from channel 26 and placing the transmitter and receiver devices at different distances: 4, 5, 6, 12 and 15 meters. From the binary files, the I/Q data has been processed, using a Python script that allows us to extract the features and create a data set, as mentioned in the previous chapters. The first data set contains just 1637 packets labeled as LOS and 1474 as NLOS so 3111 examples. This data set is divided in two different sets, **training set** and

test set, that contain respectively 2177 and 934 examples.
The following table shown the result of the classifier accuracy.

	Random Forest Classifier	**Gradient Boosting Classifier**
Training Set Accuracy	**99 %**	**87.6 %**
Test Set Accuracy	**81 %**	**79.1 %**

Table 6.1: Training and Test Accuracy from Industrial Dataset

From the table it is very clear that both Random Forest and Gradient Boosting classifiers were not able to achieve good classification accuracy.y. In particular, from the difference of the accuracy values it is clear that the model suffers ***overfitting*** problem.

It is also confirmed by analyzing the *learning curves* [25] shown in the following picture.

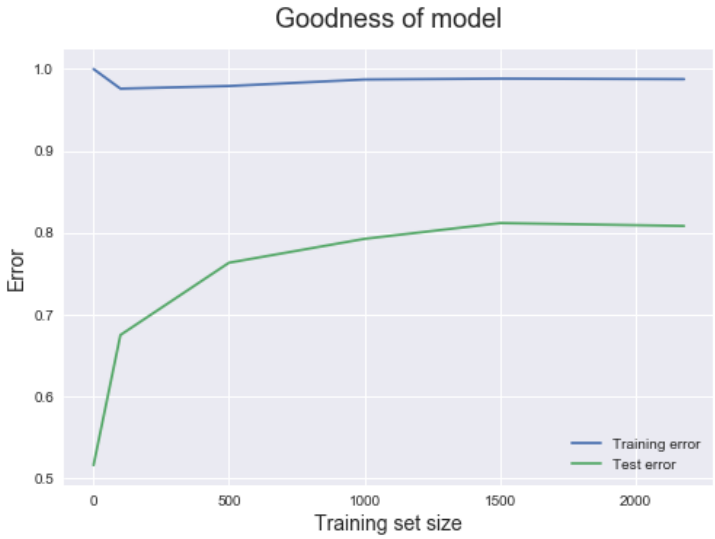

Figure 6.1: Goodness of the model for Industrial Environment

In practice, the learning curves help us understand how our model can be generalized varying the training set size and comparing the train and test error curves. The large gap and the low training error also indicates an overfitting problem.

Overfitting happens when the model performs well on the training set, but poorerly on the test set. One more important observation we can make here is that by adding new training instances it is very likely to lead to better models. But, unfortunately, it is not possible to add new data because, this data were collected one year ago.

6.2 Lab Environment Dataset

Due to insufficient data, the next step was to collect a new datasets in a different environment.

Figure 6.2: Lab Environment

Initially, using LabVIEW, we collected the data and then exploiting Python as before we extracted the necessary features to build different new data set by positioning the transmitter and receiver USRPs at variable distance, i.e., 2, 3, 4, 5, 6 meters.

From this datasets we obtain very good classifier using the previous parameter and model and considering a training and test set composed by 6696 and 2870 packets.

	Random Forest Classifier	**Gradient Boosting Classifier**
Training Set Accuracy	**100 %**	**100 %**
Test Set Accuracy	**99.9 %**	**99.9 %**

Table 6.2: Training and Test Accuracy from Lab Dataset

As the table 6.2 shows, the accuracy of the classifiers is almost 100%, and this is also shown by the learning curve in that case.

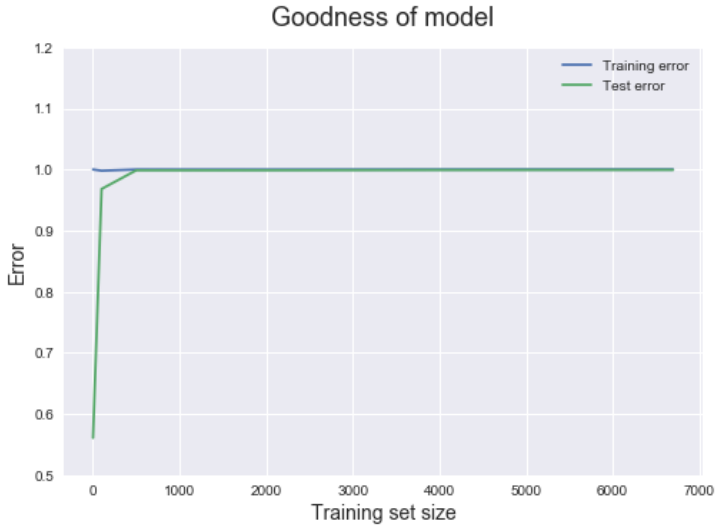

Figure 6.3: Goodness of the model for lab Environment

We can see that the training and test error converge almost in a same line. This effect might be due to the particularly simple environment, while we expect fairly different results from harsher radio-environments.

6.3 Mixed Dataset

An intelligent test to understand better the data collected in IMERYS MINERAL AB was to mix this data set with the Lab Data set. In that case the number of packet examples is very high: 8873 for the training set and 3804 for the test set. The following table shows the scarcity and the unreliability of the data in IMERYS MINERAL AB data set.

	Random Forest Classifier	**Gradient Boosting Classifier**
Training Set Accuracy	**99.6%**	**95.2 %**
Test Set Accuracy	**95 %**	**94 %**

Table 6.3: Training and Test Accuracy from Mixed Dataset

The values of the accuracy are very good in this case and this means that we can add more data using these training models and these features. As it is also shown in the learning curve figure 6.4.

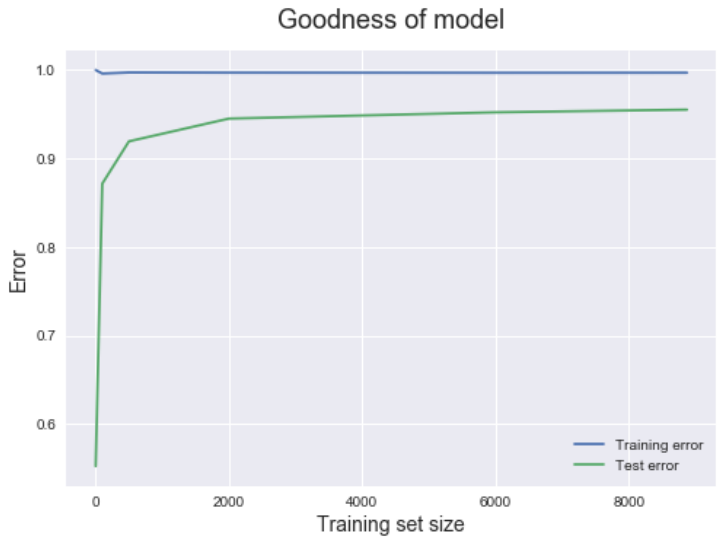

Figure 6.4: Goodness of the model for Mixed Environment

This is the common trend for the learning curve when the model can generalize and classify in a good way. Also in this case we have a gap between the two error curves but this is due from the irreducible error.

6.4 Semi-industrial Environment - Mechanical Workshop

In order to understand if the model could be used in a real industrial environment, we have moved all the system set up in in a work shop inside Mid Sweden University. The work shop, is used for a Design Course where student work using expensive machinery that processes materials, mainly wood, creating high value projects and tools. This environment represent the typical industrial field where people are moving around and working in a machinery with reflective surface as shown in the figure.

Figure 6.5: Workshop Environment

As in the previous experiment, the procedure is the same as in the Lab Environment, collect the data using LabView, exploit Python libraries to extract features and create the data set, upon which the machine learning model based on the radio channel properties are classified.

The data set is created considering different distance: 6, 8, 10 and 12 meters between the transceiver and receiver USRP. It is composed by 6400 examples, where 3200 are labeled as LOS and 3200 as NLOS. As in the previous environments the data set is separated in 70% training and 30%test set. So, respectively, we have 4480 examples for the former and 1920 the latter.
The following table shows the classifier accuracy in this environment.

	Random Forest Classifier	**Gradient Boosting Classifier**
Training Set Accuracy	**99.7%**	**98.1%**
Test Set Accuracy	**97.4%**	**97.3%**

Table 6.4: Training and Test Accuracy from Work shop Dataset

These results show that the classification accuracy is quite good and the model is able to generalize and classify in excellent way.

6.5 Discussion

The results obtained in these different environment and using this methodology based on I/Q data are very interesting and reliable to classify the radio link channel properties.

6.5.1 Features Separation

In the following tables we analyze the behaviour of the classifiers in Lab and Work shop environment separating the two class of features: *Empirical Features* 4.1 and *Distribution features* 4.2.

	Random Forest		**Gradient Boost**	
	Train	**Test**	**Train**	**Test**
All the features	0.999	0.991	0.997	0.991
Empirical features	0.999	0.991	0.997	0.991
Distribution features	0.997	0.962	0.951	0.953

Table 6.5: Accuracy variation and features separation in Lab Environment

	Random Forrest		**Gradient Boost**	
	Train	**Test**	**Train**	**Test**
All the features	0.998	0.974	0.981	0.973
Empirical features	0.979	0.969	0.977	0.971
Distribution features	0.996	0.946	0.949	0.946

Table 6.6: Accuracy variation and features separation in the Work-shop Environment

The table 6.5 and the table 6.6 show interesting aspect: the main contribute is obtained from the Empirical Features, but, in the other hands, despite the distributions they do not have a notable informative contribution, they confirm that the use of the theoretical Rician distribuiton fitting is a proper model choice for the specific problem..

6.5.2 Bit Error Pattern Approach

The common lack of I/Q data availability, as well as the low computational capabilities of common low-cost WSN nodes can limit the applicability of the previous approach. Therefore, we explore in this Paragraph a more lightweight classification method. The basic idea of this approach is then to analyze the bit error pattern of the received packets. To this end, it is necessary compute the distribution of the errors, it means a distribution based on the distance between the errors in the bit output stream of the packet and take advantage from that one to obtain information about the radio link state (LOS or NLOS) as the block diagram shows 6.6. The bit error is computed as:

$$error_i = bit_{tx} \bigoplus bit_{rx} \tag{6.1}$$

That is, the error is calculated applying the XOR operator between the *transmitted bit* and *received bit* for all bits in the data packet. For simplicity, we consider the errors in the payload of the packet only. Since, the operation it is very simple, it will reduce considerably the computational effort. Moreover, it allow us to improve the velocity of the classification.

The histogram of the empirical bit-error distribution is computed for each packet, and an aggregation of histogram is made, considering 1500 packets. So, to compute the bit error pattern useful to understand the properties in the channel the Empirical features of the aggregated histogram are computed.

Based on this approach, a data set that containned just 480 examples was created in the Work shop Environment, and using the same procedure as before a classifier is trained. But, since, the data is sparse, just Random Forest classifier is used.

However, even dividing the data set in training and test set with the correct proportion, the model suffers of overfitting.

	Random Forest Classifier
Training Set Accuracy	**0.964**
Test Set Accuracy	**0.638**

Table 6.7: Training and Test Accuracy from workshop using Bit Error Pattern

To solve the problem of overfitting in our model we need to increase flexibility of our model. But too much of his flexibility can also spoil our model, so flexibility shold such that it is optimal value.
To increase flexibility we can tune some parameters of the Random Forest algorithm, like:

- **Number of estimators**
- **Max depth of the trees**
- **Max node on the leaf**

Thus, applying some practical tuning in these parameters:

	Random Forest Classifier
Training Set Accuracy	**0.607**
Test Set Accuracy	**0.618**

Table 6.8: Training and Test Accuracy from workshop using Bit Error Pattern to overcome overfitting

In the 6.8 table, we show that even tuning the aforementioned parameters of the classifier, the accuracy of the method remains in the range of 60%. This, in turn means that, while the approach is appealing since its simplicity, further efforts are needed in the design of this particular variant of the link-state classification system.

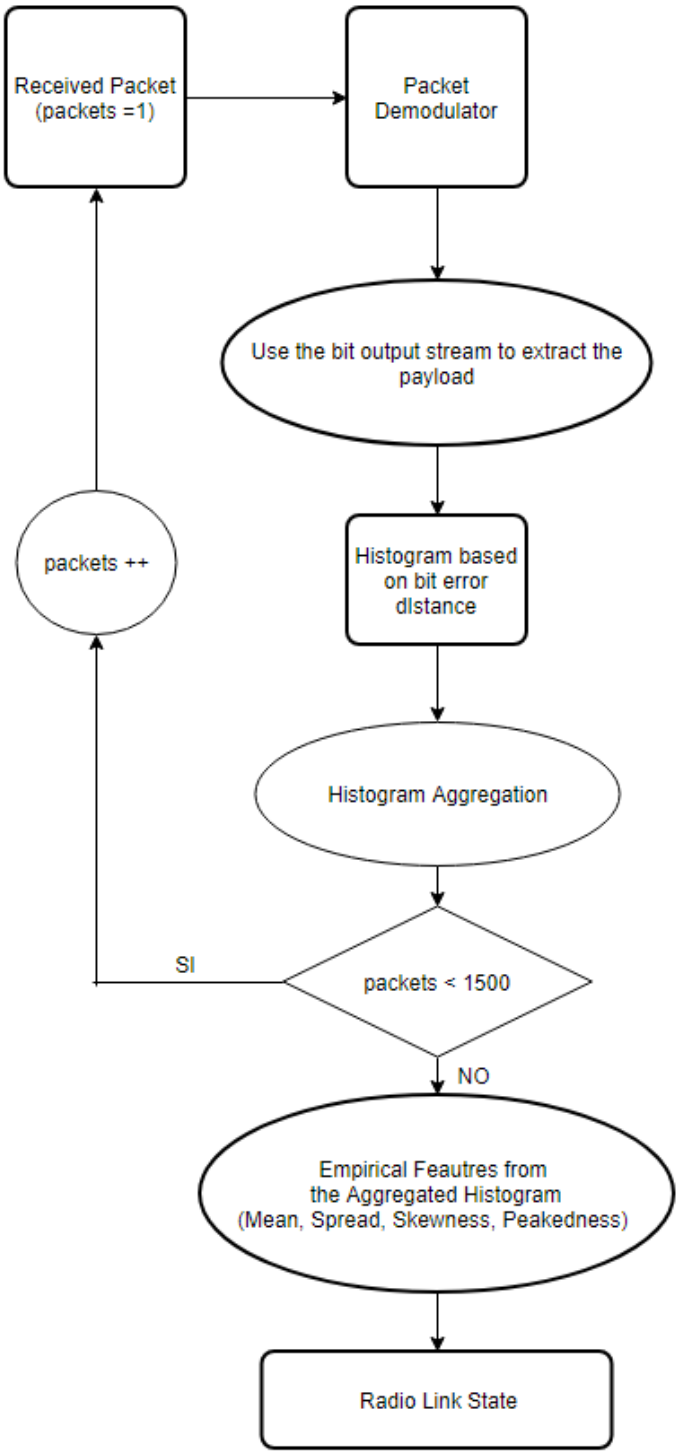

Figure 6.6: Block Diagram Bit Error Pattern

Chapter 7

Conclusion

In this thesis we show how machine learning models can be successfully employed to identify the channel properties in a radio link system.
In particular, we proposed two different approaches. The first method, based on I/Q data, show the potential and the gain of this data. The related classification accuracy obtained with I/Q data source are remarkable, while the method shows a non negligible computational effort. While we proved the implementability of the proposed solution in dedicated hardware (such as software-define radios), the implementation in low-cost WSN nodes appears problematic, since of the limited computational resources an power constraints, as well as the lack of I/Q data in many plattforms. With this respect, we proposed a second method based on bit-error distribution, which is more lightweight and potentially implementable in resource-contrained WSN nodes, while we leave to future works the effort to experimentally validate this strategy.

7.1 Future Work

Unfortunately, for the second method, it was not possible to collect more data because of the thesis timeline. However, the future plan is to continue the second approach based on Bit Error Pattern analysis and to collect more dataset to overcome the overfitting problem. In addition, the concept of

RF interferences will be introduced into the project and it will be shown if our model is able to distinguish the disturbance in the channel (MFA or Interferences) using both approaches.
In particular, considering [5] paper model as a classifier able to classify these properties using the Bit Error Pattern approach.

.

Chapter 8

Acknowledgements

First of all, I would like to thank the Professor Stefano Mattoccia, who believed in me and my potential, giving me confidence and acting as supervisor in this thesis project. Afterwards, I thank Mid-Sweden University because, by providing excellent services, it allowed me to learn a lot. In particular, I thank Professor Gidlud Mikael, Aamir and Simone who followed me in this 6 month journey supporting me and helping me with their availability and patience everyday. In addition, I thank the University of Bologna and my professors who have allowed me to grow and enrich my personal background by providing opportunities and courses that are appropriate to my career. I would like to thank all of my classmates who have been close to me between fun and suffering, Gianna, Buro, Tab, Albe, Andra who have endured me and who still bear me still: how many exams have there been without me asking you for help? Speaking of endurance, I could only thank my roommates Fucio, Edo, Lori and Mardo who saw the anxious and scary side of Andrea, but at the same time they know me more than anyone else. I wanted to test myself in a new and high-level context: I couldn't be happier with the choice, it was an experience that gave me much more than I expected. For this I thank all the erasmus guys that made this project less tiring and more fun and more enjoyable. We create a real family, Spectacular Family: Niklas, Viet, Francesco, Sotos, Alex, Benedetta, Naima, Susan, Svea, Anna, Marie, Janiek (Gianiek), Pauline. In addition, I has been very very lucky to meet

another group of wonderful guys who made me feel at home in the last period of my stay in Sweden, an infinite thank you: Bérénice, Margot, Andrea, Andrea (SSS), Alberto, Kilian , Xabi, Yan, Ilona, Mayra, Julia. Thanks to the friends of Bologna: now I can not even imagine a path without you, without you having made me go wild and get even more into the world of this city, you have made me spend unique moments with simplicity and genuineness. Thanks: Andri, Renna, Go, JD, Jonny, Lorenzone, Old, Nick, Ale, Peco. Thanks to the friends of Monteguiduccio Brosio, Lele, Lupo, Ciando, Serra who have always made my every come back, fun, and unforgettable.

Finally, I would like to thank my family that has always supported me, endured through every examination, for my every decision making me become what I am and giving me important values. I will always be grateful to them for allowing me to do this foreign experience, by making me meet wonderful people and to know and interact with new cultures and with a new world. Your support has been more than essential, you have taken charge of my many moments of anxiety, fear and above all stress: Thank you. I could not be happier about how I am now knowing that you will always be there for me. I love you Romina, Lino and Alice.

Innanzitutto, rivolgo i miei ringraziamenti al professore Stefano Mattoccia che ha creduto in me e nelle mie potenzialità dandomi fiducia e facendomi da relatore in questo progetto di tesi. Successivamente ringrazio l'Università Mid-Sweden in quanto, fornendo ottimi servizi, mi ha permesso di imparare tanto. In particolar modo ringrazio il professor Gidlud Mikael, Aamir e Simone che mi hanno seguito in questo percorso di 6 mesi supportandomi e aiutandomi nei momenti di difficoltà con la loro disponibilità e pazienza odierna. Inoltre, ringrazio l'Università di Bologna e i miei professori che mi hanno permesso di crescere e di arricchire il mio bagaglio personale fornendo opportunità e corsi adeguati alla mia carriera. Vorrei poi ringraziare a tutti i miei compagni di corso che mi sono stati vicini tra divertimenti e sofferenze, grazie Gianna, Buro, Tab, Albe, Andra che mi hanno sopportato e che mi sopportano tutt'ora d'altronde: quanti esami ci sono stati senza che vi chiedessi aiuto? Parlando di sopportazione non potrei che ringraziare i

miei coinquilini Fucio, Edo, Lori e Mardo che hanno visto il lato ansioso e pauroso di Andrea, ma che allo stessso tempo mi hanno conosciuto più di chiunque altro. Volevo mettermi alla prova in un contesto nuovo e di alto livello: non potrei essere più contento della scelta, è stata un'esperienza che mi ha dato molto di più di quello che mi aspettassi. Per questo ringrazio a tutti i ragazzi erasmus che hanno reso questo progetto meno faticoso e più divertente e ancora più piacevole creando una specie di famiglia Spectacular Family: Niklas, Viet, Francesco, Sotos, Alex, Benedetta, Naima, Susan, Svea, Anna, Marie, Janiek (Gianiek), Pauiline. In aggiunta, ho avuto la fortuna di conoscere un'altro gruppo di ragazzi meravigliosi che mi hanno fatto sentire a casa l'ultimo periodo della mia permanenza svedese, un grazie infinito: Bérénice, Anna, Andrea, Andrea (SSS), Alberto, Kilian, Xabi, Yan, Ilona, Margot, Mayra, Julia. Un grazie agli amici di Bologna: ormai non saprei nemmeno immaginare un percorso senza di voi, senza voi che mi avete fatto svagare e entrare ancora di più nel mondo di questa città e che mi avete fatto passare dei momenti unici con semplicità e genuinità. Grazie: Andri, Renna, Go, JD, Jonny, Lorenzone, Vecchio, Nick, Ale, Peco. Un grazie agli amici di Monteguiduccio Brosio, Lele, Lupo, Ciando, Serra che hanno sempre reso ogni mio rientro movimentato, divertente e indimenticabile.

Infine, vorrei ringraziare la mia famiglia che mi ha sempre sostenuto, sopportato per ogni esame per ogni mia decisione facendomi diventare quello che sono e dandomi importanti valori. Gli sarò sempre grato per avermi permesso di fare questa esperienza estera facendomi incontrare persone magnifiche e di conoscere e interagire con nuove culture e con un nuovo mondo. Il vostro supporto è stato più che essenziale, vi siete fatti carico dei miei tanti momenti di ansie, paure e soprattutto stress: Grazie. Non potrei essere più felice di come sono ora sapendo che voi per me ci sarete sempre. Vi voglio bene Romina, Lino e Alice.

Bibliography

[1] Chiara Buratti, Andrea Conti, Davide Dardari,Roberto Verdone, *An Overview on Wireless Sensor Networks Technology and Evolution*, August 2009

[2] Sahar Ben Yaala, Fabrice The'oleire, Ridha Bouallegue, *Cooperative resynchronization to improve the reliability of colocated IEEE 802.15.4 - TSCH networks in dense deployments*, 2017

[3] Marina Eskola, Tapio Heikkilä, *Classification Radio Channel disturbances*, 2015, VTT Technical Research Centre Of Finland Ltd

[4] Daniele Puccinelli and Martin Henggi, Network Communication and Information Processsing Laboratory University of Notre Dame , IN, USA, 2006 *Multipath Fading in Wireless Sensor Networks: Measurements and Interpretation*

[5] Filip Barac, Mikale Gidlund, Tingting Zhang *Scrutinizing Bit- and Symbol-Errors of IEEE 802.15.4 Communication in Industrial Environments*, 2014

[6] Marina Eskola, Tapio Heikkilä, Tero Peippola, 2013 *Identification of radio disturbances of Wireless Sensor Networks*, Technical Research Centre of Finland

[7] International Electrotechnical Commission, *Internet of Things: Wireless Sensor Networks*

[8] Diego V.Queiroz , Marcelo S.Alencar, Ruan D.Gomes, Iguatemi E.Fonseca, Cesar Benavente-Peces, 2017 *Survey and systematic mapping of industrial Wireless Sensor Networks*

[9] Tapiwa M. Chiwewe, Colman F.Mbuya, Gerhard P.Hancke, Senior, 2015 *Using Cognitive Radio for Interference Resistant Industrial Wireless Sensor Network. An Overview*

[10] `https://www.electronics-notes.com/articles/connectivity/ieee-802-15-4-wireless/basics-tutorial-primer.php` *IEEE 802.15.4*

[11] Michael Cheffena, *Industrial Wireless Communications over the Millimeter Wave Spectrum: Opportunities and Challenges*, IEEE Communications Magazine, September 2016

[12] Enmeric Tanghe, Wout Joseph,Leen Verloock, Luc Martens, Henk Capoen, Kobe Van Herwegen, and Wim Vantomme, *The industrial Indoor Channel: Large Scale and Temporal Fading at 900, 2400 and 5200 MHz*, 2008

[13] http://www.ni.com/tutorial/4805/en *What is I/Q Data?*, National Instruments

[14] Arnau Mata Llenas, Janne Riihijärvi, Marina Petrova, 2017 *Performance Evaluation of Machine Learning based Signal Classification using Statisitical and Multiscale Entropy Features*, RWTH Aachne University, Institute for Networked Systems

[15] Simone Grimaldi, Aamir Mahmood, Mikael Gidlund, 2018 *Real-time Interference Identification via Supervised Learning: A coexistence Framework for Massive IoT Networks*, 2018

[16] Simone Grimaldi, Aamir Mahmood and Mikael Gidlund, Journal of Sensor and Actuator Networks , Department of Information Systems and Technology, Mid Sweden University, 851 70 Sundsvall, Sweden, 2017 *An*

SVM-Based Method for Classification of External Interference in Industrial Wireless Sensor and Actuator Networks

[17] Dr. Danko Nikolic, CSC and Max-Planck Institute *Definition of Machine Learning*

[18] *Types of classification algorithms in Machine Learning*, Mandeep Sidana, Technology Consultant, Sifium Technologies

[19] `http://www.physics.csbsju.edu/stats/KS-test.html` *Kolmogorov-Smirnov test*

[20] `http://www.ni.com/sv-se.html` *National Instrument*

[21] https://scikit-learn.org/stable/ *Scikit-Learn*

[22] `https://keras.io/` *Keras*

[23] `https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html` *Random Forest Classifier*

[24] `https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html` *Gradient Boosting Classifier*

[25] `https://www.dataquest.io/blog/learning-curves-machine-learning` *Learning Curves for Machine Learning*