Alma Mater Studiorum - Università di Bologna

ENGINEERING AND ARCHITECTURE SCHOOL
Second cycle degree in Telecommunications Engineering
Electromagnetic Propagation for Wireless Systems M

RAY-TRACING ASSESSMENT OF THE ROBUSTNESS OF PHYSICAL LAYER SECURITY KEY GENERATION PROTOCOL

MASTER THESIS OF:

SIMONE DEL PRETE

SUPERVISOR:

Prof. Eng. FRANCO FUSCHINI

CO-SUPERVISOR:

Prof. Eng. MARINA BARBIROLI

Eng. MARCO ZOLI

Academic Year 2020/2021

Session II

Contents

Introduction **1**

I Fundamentals of Physical Layer Security 3

1 Introduction to information security **5**
1.1 Wireless security 6
1.1.1 Authenticity 6
1.1.2 Availability 7
1.1.3 Confidentiality 7
1.1.4 Integrity 7
1.2 Cryptography 8
1.2.1 Symmetric encryption 9
1.2.1.1 One Time Pad. 10
1.2.2 Asymmetric encryption 11
1.3 Thesis work 12
1.4 Summary 12

2 Physical Layer Security **15**
2.1 Basics of information security 15
2.2 Key-less Security 17
2.2.1 Artificial noise aided security 18
2.3 Physical Layer based-Key generation 19
2.3.1 Information theory of key generation 21
2.4 Physical Layer-Key generation protocol 22
2.4.1 Channel Probing 23
2.4.2 Quantization 25
2.4.2.1 Absolute Value Based-Quantization 25
2.4.2.2 Differential-Based Quantization 26
2.4.3 Information Reconciliation 27
2.4.4 Privacy Amplification 28
2.4.5 Evaluation metrics of the PL based-Key generation 28
2.4.5.1 Cross-correlation 28

2.4.5.2 Autocorrelation function (ACF) . . . 29
2.4.5.3 Key Disagreement Rate (KDR) . . . 29
2.4.5.4 Secrecy Key Rate (SKR) . . . 29
2.4.5.5 Key Generation Rate . . . 29
2.4.5.6 Randomness . . . 29
2.5 Filter-bank model . . . 31
2.6 Summary . . . 32

II Ray Tracing assessment 35

3 Ray tracing as a tool 37
3.1 Ray Tracing . . . 37
3.1.1 Narrowband output . . . 40
3.2 Ray Tracing Attack . . . 43
3.2.1 Vector Network Analyzer . . . 44
3.2.2 Antennas . . . 45
3.2.3 Software tools for post processing . . . 46
3.3 Ray tracing as a simulation tool for the PL based-Key generation . . . 46
3.4 Summary . . . 47

4 Ray tracing attack 49
4.1 Description of the environments . . . 49
4.1.1 Empty room . . . 49
4.1.2 Room with obstacles . . . 54
4.2 Procedure for the Ray Tracing attack assessment . . . 56
4.2.1 Measurements and simulations parameters . . . 58
4.2.1.1 Channel parameters . . . 58
4.2.1.2 Position of the measurements . . . 58
4.2.1.3 Ray Tracing parameter . . . 58
4.2.1.4 Physical Layer based-Key generation . . . 60
4.2.2 Evaluation metrics . . . 60
4.2.2.1 Entropy . . . 60
4.2.2.2 Percentage of random keys . . . 62
4.2.2.3 Pearson coefficient . . . 62
4.2.2.4 Root mean square error (RMSE) . . . 62
4.2.2.5 Mutual information . . . 63
4.2.2.6 Percentage of guessed bits . . . 64
4.2.3 Metrics for the time domain evaluation . . . 64
4.3 Results for the empty room . . . 65
4.3.1 Time domain . . . 78
4.4 Results for room with obstacles . . . 80
4.4.1 Time domain . . . 93
4.5 Summary . . . 94

5 Ray Tracing as a design tool for PL-key generation **95**
5.1 Description of the simulation environment 95
5.2 Procedure and metrics for the simulation assessment 96
5.2.1 Channels parameters 97
5.2.1.1 Antennas 97
5.2.1.2 Ray Tracing 98
5.2.2 Evaluation Metrics 98
5.3 Results of the simulation 99
5.4 Evaluation in the empty room 100
5.5 Evaluation in the complex room 104
5.6 Spatial decorrelation assessment 107
5.6.1 Spatial assessment in the empty room 108
5.6.2 Spatial assessment in the complex room 122
5.7 Summary 135

Conclusions **137**

Bibliography **139**

Introduction

Nowadays, information security is a very important topic. In particular, wireless networks are experiencing an ongoing widespread diffusion, also thanks the increasing number of *Internet Of Things* devices, which generate and transmit a lot of data: protecting wireless communications is of fundamental importance, possibly through an easy but secure method. Physical Layer Security is an umbrella of techniques that leverages the characteristic of the wireless channel to generate security for the transmission. In particular, the Physical Layer based-Key generation aims at allowing two users to generate a random symmetric keys in an autonomous way, hence without the aid of a trusted third entity. Physical Layer based-Key generation relies on observations of the wireless channel, from which harvesting entropy: however, an attacker might possesses a channel simulator, for example a Ray Tracing simulator, to replicate the channel between the legitimate users, in order to guess the secret key and break the security of the communication. This thesis work is focused on the possibility to carry out a so called *Ray Tracing attack*: the method utilized for the assessment consist of a set of channel measurements, in different channel conditions, that are then compared with the simulated channel from the ray tracing, to compute the *mutual information* between the measurements and simulations. Moreover, the measured and simulated channels are used to generate random keys to see how many bits the Ray Tracing attacker is able to guess. Furthermore, it is also presented the possibility of using the Ray Tracing as a tool to evaluate the impact of channel parameters (e.g. the bandwidth or the directivity of the antenna) on the Physical Layer based-Key generation, as well as to evaluate to what extent the keys generated from the channel are actually random or not. The measurements have been carried out at the Barkhausen Institut gGmbH[1] in Dresden (GE), in the framework of the existing cooperation agreement between BI and the Dept. of Electrical, Electronics and Information Engineering "G. Marconi" (DEI) at the University of Bologna.

The work is subdivided in two main parts. The first, includes an introduction to cryptography and the description of Physical Layer Security, encompassing both the theory behind it and the techniques proposed to safely protect the exchange

[1] `https://www.barkhauseninstitut.org/en`

of data. In particular, the Physical Layer based-Key generation is explained and the particular method considered in this work called *Filter-bank model*. Then, the second part provides the description of the methodology for the assessment and the results obtained. In particular, the third chapter explains the reasons why the *Ray Tracing attack* can be a threat, at least in principle, for the Physical Layer based-Key generation and the tools utilized for the assessment. In the fourth chapter there are the results relative to the Ray Tracing Attack and in the fifth there is the presentation of the possibility of using the Ray Tracing as a design tool for the Physical Layer based-Key generation.

Part I

Fundamentals of Physical Layer Security

1

Introduction to information security

This chapter outlines the main issues about security of data communication, with a focus on wireless communication. First, there will be an explanation of the security threat to cope with. Then, a brief introduction about cryptography, with a special attention on wireless communication.

During the past decades, the number of wireless devices experienced a huge increase. In particular, *Internet Of Things* (IOT) devices are more and more present in our life: smart home application, smart cities or medical devices [1] are being installed thanks to the progresses in wireless communications. As the number of internet user increases, the number of cyber attacks increases as well. [2] reported that 56% of the internet consumers have experienced a cyber crime, and 46% of them lost money from the attack: although the increasing number, security is still not seen as big issue by people and companies, even though recently there has been paid more and more attention to the problem.

The actual worries have often concerned the security of data stored in *data centres*, trying to prevent their theft, without taking too much into account the security of a wireless transmission. In addition, security issues of IOT devices have been particularly overlooked, seen as something "nice to have" rather than "must to have" [3]. In fact, some works highlighted the ease with which wireless devices can be attacked [4]: in [5] authors showed the possibility to hack implantable medical devices and in [6] researchers proved that automotive remote key-less systems can be cracked by low cost wireless modules. Therefore, more research effort must be invested in order to find new security mechanism for wireless network. Moreover, security means additional energy consumptions and computational capacity, which are constrained in IOT devices. Therefore, there is the need to pay more and more attention to the security of wireless communications, considering the future increase of wireless IOT devices and it is important to start working on security *ex-ante*, not after the failure of the system, in order to prevent in advance possible security problems.

1.1 Wireless security

In general, a secure transmission should meet these four *security traits*:

- **Authenticity**.
- **Availability**.
- **Confidentiality**.
- **Integrity**.

The meaning of the traits will be given in the following and a brief summary can be found in Table 1.1. As a matter of fact, wireless networks are extremely vulnerable owing to the broadcast nature of the wireless medium: this requires additional efforts and specific solution for the protection of the wireless communication. A malicious user might be able to carry out two different types of attack:

- **Passive attack**: an eavesdrop of the communication, which aims at stealing the information exchanged over the channel.
- **Active attacks**: an action with the goal of disrupting the communication (jamming) or to steal the identity of an authorized user.

1.1.1 Authenticity

Authenticity aims at ensuring that the user who is transmitting or requiring an information is actually who he claims to be. A device is equipped with a wireless network interface and has a unique *Medium Access Control (MAC)* address that can be used for authentication. In addition, user authentication might be obtained at different layer of the protocol stack, exploiting modern cryptographic techniques. A malicious user can perform an attack called *MAC spoofing* [7], i.e. changing its assigned address in order to mask its identity. Furthermore, he can listen to the wireless channel, observe the traffic and steal a legitimate address. This specific action is called *identity theft attack*. At the network layer, similar attacks can be carried out using the IP address. Therefore, they are called *IP spoofing* and *IP hijacking*, which aims at taking over another legitimate IP address [7].

In the field of wireless communication, there is a promising technique called *radio frequency fingerprint* that allows to authenticate users based on observable unique feature of the transmitted signal. Mored details can be found in [8] and [9].

1.1.2 Availability

Availability means that an information stored in a remote point must always be available for the access. From the network point of view, the availability of a server can be disrupted through a *Denial of Service (DoS)* or *Distributed Denial of service (DDoS* attack. In a *DoS* attack a malicious user sends a lot of requests to the server until its queues are full and the server becomes unavailable or start acting in a strange way, while the *DDoS* follows the same principle but many users coordinate themselves to attack the server simultaneously.

As for a wireless communication, availability implies that the authorized users are capable of accessing a wireless network anytime and anywhere upon request [7]. The easiest way to undercut a communication is to perform a *Jamming Attack*. The general idea of a Jamming attack is to continuously sends a signal over the user bandwidth to increase the level of interference and eventually make the communication impossible. This can be done both by sending only a carrier frequency or sending a dummy signal. An easy way to prevent from a Jamming attack is to use a *spread spectrum modulation*, which is known to be more robust against interferences [10]. Furthermore, it is possible to use a *frequency hop* technique to change the transmission frequency to react against a Jamming attack [11], as soon as the users realize that the communication is demoted over the frequency range in use.

1.1.3 Confidentiality

Confidentiality aims at limiting the data access only to the intended users. This is a characteristic difficult to be achieved at physical level due to the inner broadcast nature of the channel: in fact, once the signal is transmitted over the channel it can be heard, at least in principle, by anyone who is in coverage and can be decode. Instead, a transmission over a cable or an optical fiber is more difficult to be intercepted at physical level. Therefore, it is of fundamental importance to protect the confidentiality of the wireless transmission.

Confidentiality is typically achieved at the application layer by *encrypting* the message. The branch of computer science that studies the techniques to encrypt a message is called *cryptography* and in section 1.2 there will be a brief introduction to cryptography. *Physical Layer Security* (PLS) emerges as a solution to protect the confidentiality by achieving *information theoretic security* using the unpredictable fading characteristics of the wireless channel. More details will be provided in chapter 2.

1.1.4 Integrity

Information exchanged over the wireless channel must be received without alteration and falsification. Integrity aims at protecting the message from possible alteration. Typically, to guarantee the integrity of a message it is transmitted together

Authenticity	Confirm the identity of a user, can be obtained at physical layer through *RF fingerprint*.
Availability	Information stored in a remote point must always be available for the access.
Confidentiality	Aims at protecting the privacy of a message, the main goal of Phsyical Layer Security.
Integrity	Avoid modifications of the transmitted data. Code recovery process [7] might be employed to correct possible alterations.

Table 1.1: *Summary of the security treats*

with a *fingerprint* of the message itself. The fingerprint is obtained by means of a *cryptographic secure hash function*. Data integrity can be violated by so called *insider attacks* [12]

In general, it is difficult in a wireless scenario to detect attacks on the integrity. A solution to detect compromised nodes is to utilize the automatic code update and recovery process, which guarantees that the nodes are periodically patched and a compromised node may be detected, if the patch fails. The compromised nodes can be repaired and revoked through the so-called code recovery process [7].

1.2 Cryptography

Security threats are tackled by means of **cryptography**, the branch of computer science which studies the techniques for a secure communication between two users. In particular, there are two main techniques mostly used for cryptography named *symmetric encryption* and *asymmetric encryption*. Cryptography makes use of known standard algorithms and protocols, based on a secret key (symmetric) or a pair of keys (asymmetric): once the message is encrypted, only the users in possession of the right key will be able to decrypt the message and retrieve the original message. Crypthographic techniques often exploit complex mathematical computation and are referred to be *computational secure*: a user with enough computational capacity, even if it has not the right key, might be able to revert the algorithm employed, or guessing the key by means of a lot of trials (brute force attack). Of course, this situation is not really possible since algorithms are designed such that modern computers will take an infeasible time to break the security (something like hundreds of year). In fact, in order to invert the algorithms the attacker has to solve complex mathematical problems like *discrete logarithm* and *prime number factorization*, which modern5 computers now are not able to solve in a feasible time. However, this condition may not be true anymore in the future with the introduction of *quantum computers*, which are expected to have a huge computational capacity. Researches in

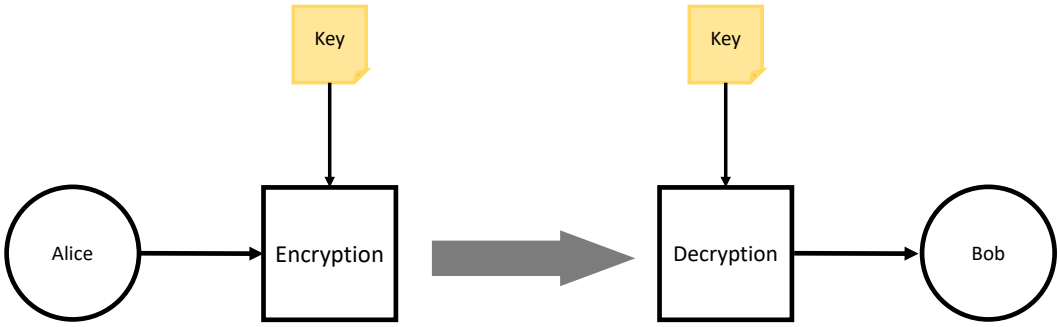


Figure 1.1: *Symmetric encryption scheme*

the field of cryptography are pushing toward *quantum resistant* algorithm and also *information theoretic security*. Just a brief introduction to cryptography is provided here, and for additional information refer to [13] [14]

1.2.1 Symmetric encryption

Symmetric encryption make use of a single secret shared key named **symmetric key**. As shown in Fig. 1.1, Alice encrypts the data with the key and send them to Bob, who knows the encryption key and can easily decrypt the data. The data over the channel cannot be decrypted by a possible eavesdropper who doesn't know the key. The encryption mechanism is based on the principles of permutation and substitution and the algorithm used are quite efficient from the computation point of view. A simple but inefficient method to break the security is to use a *brute force attack*, so try all the possible combinations for the key until the right one is found. To make this attack infeasible, the National Institute of Standards and Technology (NIST) [15] has set a minimum key length to 128 bit: in this way, a brute force attack will take hundreds of year to have success, considering the computational resources nowadays possibly available. In addition, the current standard *AES* (Advanced Encryption Standard) is also considered quantum resistant, so an algorithm that cannot be break by a quantum computer [16], provided that the key is 256 bit long. There are some properties that the symmetric key must respect:

- The key must be *random*, so generated through a true random generator.
- The key should be *changed frequently*.
- The key should be *long enough* for the target level of security.

The main drawback of symmetric encryption lies in the key distribution phase and in the number of keys involved in the system. In fact, in case N users are present in the system and want to securely communicate, each one should store $\frac{N(N-1)}{2}$ keys, which could be a huge number in case N is large. In addition, there should be a reliable system to distribute the keys: the most common system is the *Diffie-Hellman*

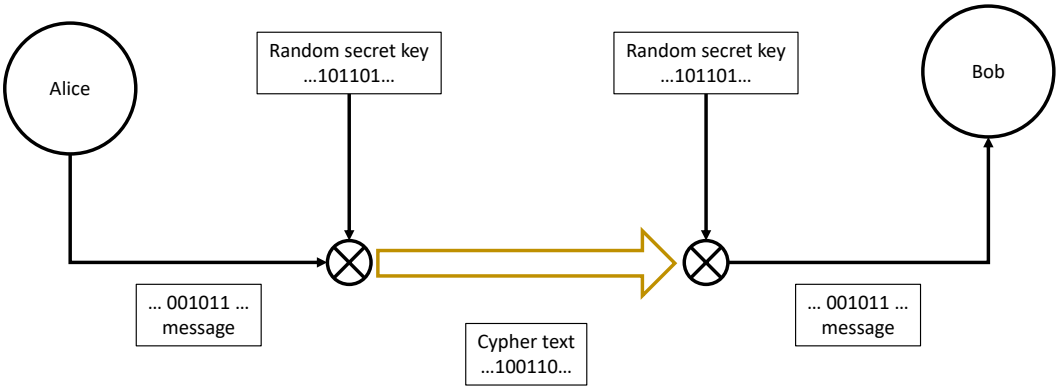


Figure 1.2: *Scheme of One Time Pad mechanism*

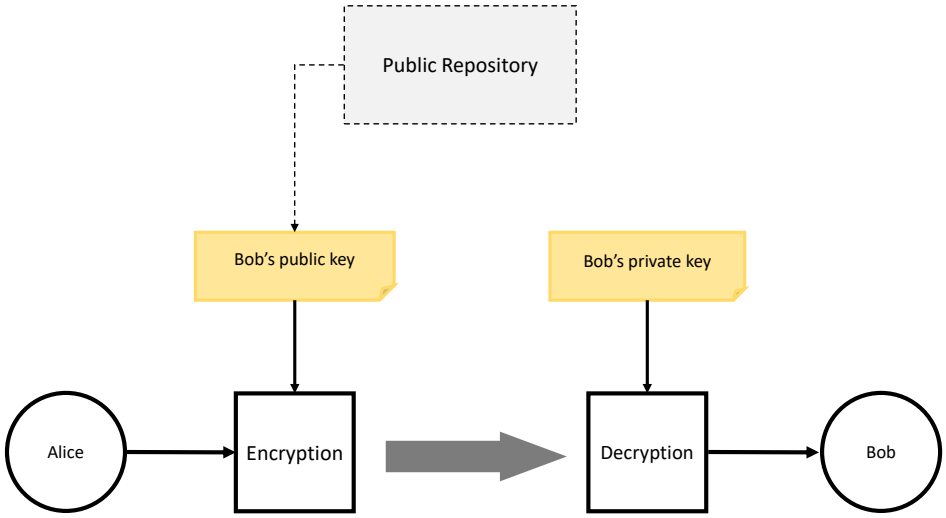
protocol, which allows two users to agree on a symmetric keys by exchanging some information on a public and insecure channel. Another possibility is to use a trusted *third authority* that is in charge of distributing the keys to the users, but they need to have a pre-shared key with the third authority. The key distribution is a delicate procedure, since the key must be kept secret. In IOT networks, sometimes a common method is to use pre-shared keys loaded inside the device through USB before the deployment of the network. This is neither a secure nor efficient method: symmetric keys should be changed frequently, and considering the huge amount of devices that could be present in the network [7] it might become an hard task. Eventually, symmetric encryption guarantees only confidentiality and can hardly provide also integrity and authentication.

1.2.1.1 One Time Pad.

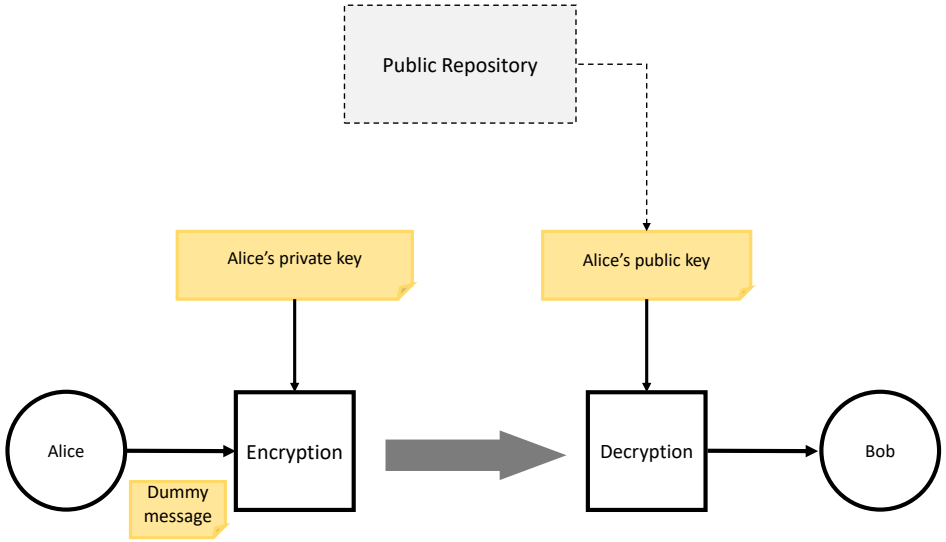
In 1919 G. S. Vernam published his work about One Time Pad (OTP) [17]. It is a simple encryption mechanism that allows to achieve **perfect secrecy**, and works as shown in Fig. 1.2: first, Alice generate a message, represented as a sequence of bits and she encrypts it through a bit wise `xor` between the key and the message itself. When Bob receives the text he will perform the same operation using the same key as Alice, and recover the original plain text. In order for OTP to be unbreakable the key must follow these rules:

- The key must be at least long as the message to be encrypted.
- The key must be random.
- A key must only be used once.
- The key must be kept secret.

OTP impose a challenge in the key distribution. In fact, it is difficult every time to distribute random and long enough keys to the users: this has forced, so far, a



(a) Confidentiality scheme



(b) Authentication scheme

Figure 1.3: *Schemes for asymmetric cryptography where Alice is transmitting and Bob is receiving.*

strong limitation on the practical usage of OTP. A common trade off is to use *AES* and a session key, so that can be used only for a session of the communication, which may range from one to few exchanged packets.

1.2.2 Asymmetric encryption

Asymmetric cryptography, also called Public-key cryptography, is a cryptographic system that employs two different keys per user, one called *public key* and the other called *private key*. The public key is stored in a public repository while the private key must be kept random. In order to generate the two keys *RSA* algorithm is employed, then one key is used to encrypt the message and the other to decrypt. Asymmetric encryption is able to guarantee confidentiality and authentication. In

addition, the encryption methods are based on *one-way functions*, functions that are very hard to be inverted. If Alice wants to transmit to Bob as shown in Fig. 1.3, then *confidentiality* and *authentication* can be achieved as follow:

- Confidentiality (Fig. 1.3a): Alice encrypts the message with Bob's public key, which can be obtained from the public repository. Then, only Bob will be able to decrypt the message since only he knows his private key.

- Authentication (Fig. 1.3b): Alice wants to authenticate herself to Bob. She produces a known message, then she encrypts it with her private key. Then, Bob receives the encrypted message and decrypts it with Alice's public key. If the message can be decrypted correctly then Bob knows that the sender was actually Alice, since only Alice could have encrypted it with her private key.

Now, in a network of N users the number of keys to be shared are just N and they are public, so can be exchanged over an insecure channel. However, the repository should be a trusted third party and this increases the complexity of the network. In addition, asymmetric encryption is not as efficient as symmetric encryption and is not the best solution for wireless low power networks. Plus, RSA uses the discrete logarithm as one-way function and can be broken by a quantum computer that uses the Shor's algorithm [18]: this introduces an in important future vulnerability, since asymmetric encryption is vulnerable to attacks from quantum computers.

1.3 Thesis work

This thesis work is focused on the Physical Layer Security key generation method, a novel way to guarantee *perfect secrecy* at a low expense, which makes key generation suitable for low power devices. The thesis work includes both an experimental part and a simulation work. All the works have been carried out in collaboration with the Barkhausen Institut gGmbH[1] (BI) in Dresden (GE), in the framework of the existing cooperation agreement between BI and the Dept. of Electrical, Electronics and Information Engineering "G. Marconi" (DEI) at the University of Bologna. In particular, the experimental part was developed in Dresden, where I spent three months and a half working in the laboratory of the BI.

1.4 Summary

This chapter has briefly surveyed the main requirements to build a secure wireless transmission. All the techniques now used are usually based on computation: for this reasons, the "traditional" cryptography is said to be *computational secure*.

[1] https://www.barkhauseninstitut.org/en

Moreover, as time goes by, the computational power required in order to use for the encryption is increasing more and more: this could be a problem for low power devices which are designed to have low computational capacity to reduce energy consumption. However, as IOT will become more and more pervasive in our lives, there will be the need to make the communication secure without the need of much computational power. In this sense, *Physical Layer Security* can be a promising solution, as will be explained in chapter 2.

2

Physical Layer Security

This chapter is an introduction to Physical Layer Security (PLS). First, the main elements of information theory at the base of PLS will be outlined. Then, PLS is defined and discussed. Finally, the Filter Bank approach will be presented, as it represents the method for secret key generation referred to in the experimental and simulation part of the work.

2.1 Basics of information security

The *classical cryptography* approach for information security has been introduced in chapter 1. However, computer power is increasing at a very fast peace and attacks like brute force attacks, that were deemed to be infeasible, are now possible and this situation will be worse with the arrival of quantum computers. In addition, classical cryptography requires a trusted third part to distribute the keys and complex and expensive protocol from the computational point of view. Furthermore, in classical cryptography there are no precise metric to establish a *security level*: usually the security level is given in term of time that the adversary takes to break the security of the chiper-system.

Another possibility is given by *Physical Layer Security* (PLS), which is able to achieve *information-theoretic security*, or *perfect secrecy*, by exploiting the unpredictable features of the fading channel [3]. The concept of **perfect secrecy**, already referred to in Chapter 1, was established in 1949 by Shannon [19]. Suppose to have a message M to be transmitted and an encryption mechanism: it possible to encode the message into a codeword C such that the knowledge of C does not bring any information about the initial message. This can be formulated as:

$$\mathbb{H}(M|C) = \mathbb{H}(M) \tag{2.1}$$

where $\mathbb{H}(\cdot)$ indicates the entropy. PLS can be realized through two main techniques:

Security Route	Technique	Feature	Algorithm
Cryptography	Asymmetric Encryption (Public Key Cryptography)	Use same public key but different private keys; based on mathematical problems	Encryption: RSA Key distribution: Diffie Hellman key exchange Digital signature: ElGamal cryptosystem
	Symmetric Encryption	use the identical key at both users.	DES, RC4, AES
Physical Layer Security	Keyless Security Transmission	Confidential wireless transmission by employing the channel advantage	Beamforming, artificial noise
	Key Generation	Automatic key generation by leveraging unpredictable wireless fading	A four-stage protocol

Figure 2.1: *Summary of information security methods, from [3]*

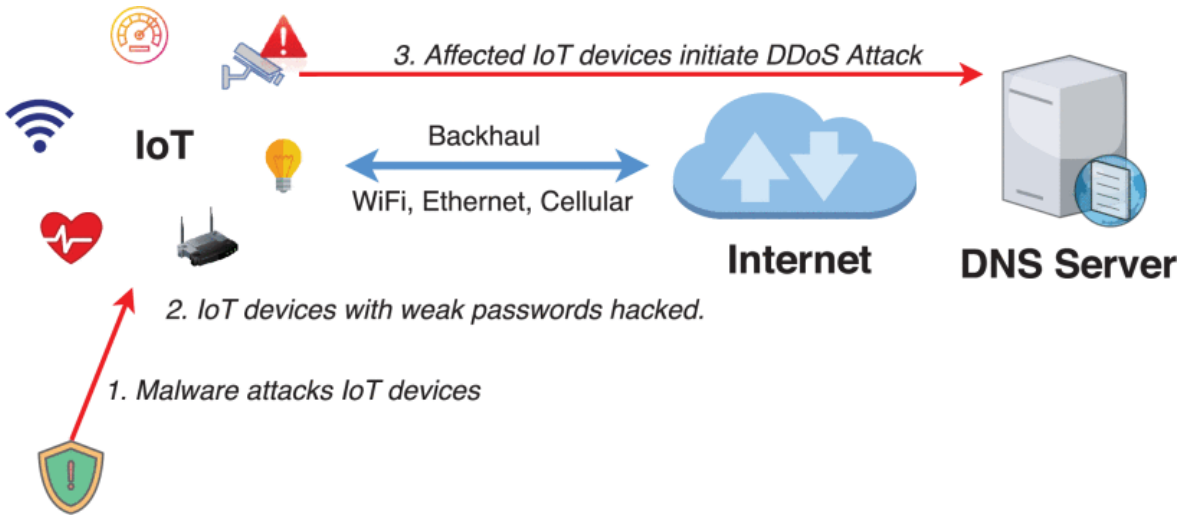


Figure 2.2: *IoT cyber-attack. Compromised nodes in the IoT network can initiate a DDoS attack ot the internet. Picture from [3]*

- **Key-less Security**, which utilizes the properties of the wireless channel to transmit in a secure way without encrypting the message.
- **Physical Layer based-Key generation**, which allows two users to generate a symmetric encryption key that will be employed in a classical cryptographic algorithm.

PLS, in particular the Physical Layer based-Key generation, is a security mechanism specific for wireless channel and that works together with common cryptographic methods. A summary of the techniques for information security is reported in Fig. 2.1. In addition, PLS is a lightweight but strong solution that can be suitable to secure IoT communication and to understand the importance of protecting the edge node communications. Consider the model represented in Fig. 2.2: in case the IOT node is not protected, or uses a weak password, an attacker might inject a *malware* in the network that can start a *DDoS* attack with the goal of attacking internet. Therefore, it is of fundamental importance to protect the edge section of the network and PLS is a well suited candidate for this aim. In this thesis work, PLS is explored for a radio frequency wireless channel, but there is the possibility of using PLS also on a quantum channel through *quantum key distribution* [20] [21].

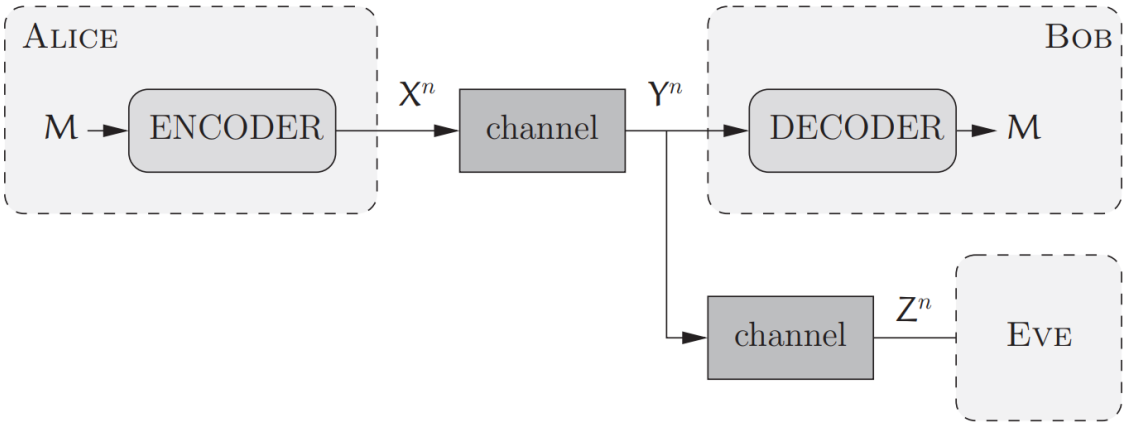


Figure 2.3: *General case of the Wyner's wire-tap channel, with two users and one eavesdropper. Picture from [23]*

However, PLS is based on average information measures: the security is given in term of probability, but there is not the possibility to achieve confidentiality with probability one. In addition, the information-theoretic model is based on assumptions on the wireless channel that might not be true in reality (e.g. Rayleigh fading). Moreover, the physical layer is just one of the layer of the protocol stack: each layer has its own security mechanism with a specific goal in mind. In this sense, PLS security emerges as an additional layer of security at a low expense. Also, it might be integrated with other security mechanism to realize *cross layer security*, a layered model in which the layers exchange information about the security and contribute to develop an efficient security mechanism specific for the wireless channel [7].

2.2 Key-less Security

Key-less security is capable of achieving perfect secrecy without encrypting the message. A.D. Wyner was the pioneer of the key-less security when he published his work about the *wire-tap channel* [22] and introduced a new condition for perfect secrecy.

Consider the model depicted in Fig. 2.3. Alice wants to reliably send data toward Bob while keeping confidentiality of the data. At the same tame, an eavesdropper Eve wants to intercept the message sent by Alice. Alice encodes her message M into a codeword X^n which is then sent to Bob through a noisy wireless channel. Bob receives the codeword Y^n, a noisy observation of X^n. Since the wireless channel is inherently broadcast, the message arrives to Eve who observes Z^n, a noisy observation of Y^n. Wyner introduced the concept of **equivocation rate** $\left(\frac{1}{n}\right)\mathbb{H}(M|Z^n)$ which must be arbitrarily closed to the entropy of the message $\left(\frac{1}{n}\right)\mathbb{H}(M)$ [23]. Similar to

(2.1), the condition for perfect secrecy can be formalized in this way:

$$\left(\frac{1}{n}\right)\mathbb{H}(M|Z^n) \xrightarrow[n\gg 1]{} \left(\frac{1}{n}\right)\mathbb{H}(M) \tag{2.2}$$

The model in (2.2) suggest the existence of a set of codes that *asymptotically* achieve perfect secrecy, and including the Shannon's theory about communications, they can also achieve arbitrarily small error probability: such codes are known as *wiretap codes*.

Starting from the model in (2.2), in [24] the concept of **secrecy capacity** C_s has been introduced. With respect to Fig. 2.3, suppose that the Alice-Bob channel has a capacity C_{AB} and the Alice-Eve channel has a capacity C_{AE}, the secrecy capacity C_s is defined as:

$$C_s = C_{AB} - C_{AE} \tag{2.3}$$

C_s indicates the amount of information that can be securely exchanged on the channel without any possibility for the eavesdropper to decode the message. Respecting this bound it is possible to achieve perfect secrecy.

2.2.1 Artificial noise aided security

A practical implementation of key-less security can be the *Artificial noise aided security*: it is a technique that uses both *beamforming* and interfering signals to increase the secrecy capacity. Considering the previous model, suppose that Alice is equipped with a directive antenna and implements a beamforming algorithm, as shown in Fig. 2.4: the main lobe will be directed toward the direction of Bob, while Eve receive the communication from the side lobes. In this way, most of the power will be directed toward Bob and Eve will experience a low *Signal to Noise ratio* (SNR). Moreover, it is possible for Alice to send in the side lobes an interfering signal, or directly noise, in order to further reduce the SNR. In the end, the Alice-Eve capacity (C_{AE}) will be small compared to the Alice-Bob capacity (C_{AB}), which leads to a high secrecy capacity. In [25] there is a solution for jointly optimizing power allocation and noise in the beams. In the end, this method follows the model for perfect secrecy, since it will be impossible for Eve to decode the signal.

However, in case of a dynamic channel it would be difficult to follow the variation of the capacity and adapt the communication rate to respect the bound of C_s. Furthermore, determining the eavesdropper's channel capacity is not straightforward. These are the main practical limitation of the key less security.

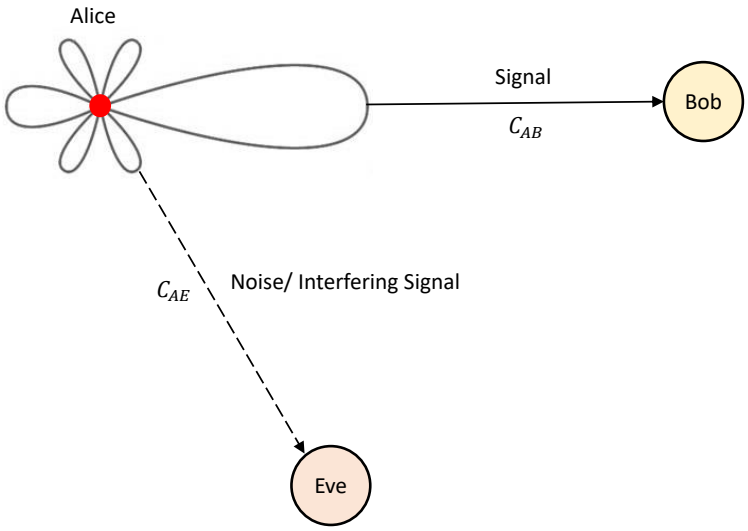


Figure 2.4: *Artificial noise aided scheme*

2.3 Physical Layer based-Key generation

Physical Layer based-Key generation aims at providing a protocol to autonomously generate a symmetric encryption key between two users and it exploits the randomness inside the wireless channel to extract random bit. Small scale fading is a good candidate as a source of randomness, since it varies in a small spatial scale and changes fast in time due to mobility of the terminals or of the objects inside the environment. In order to extract the randomness, the two users initiate a public discussion over the channel and by means of pilot signals they observe the random fluctuations of the fading.

The multipath channel (Fig. 2.5) can be modelled as a superposition of several components and the *channel impulse response* (CIR), $h^{ab}(\tau, t)$, from transmitter a (Alice) to receiver b (Bob) can be written as:

$$h^{ab}(\tau, t) = \sum_{l=1}^{L^{ab}(t)} \alpha_l^{ab}(t) e^{-j\phi_l^{ab}(t)} \delta\left(\tau - \tau_l^{ab}(t)\right) \tag{2.4}$$

where $\alpha_l^{ab}(t)$ is the amplitude attenuation of the l-th path, $phi_l^{ab}(t)$ the phase shift, $L^{ab}(t)$ the total number of paths, $\delta(\cdot)$ is the Dirac function, t represents the time while τ the propagation delay. Then, if a signal $s(t)$ is transmitted via the multipath channel, the received signal $y(t)$ is given by the convolution:

$$y(t) = \int_0^{\tau_{max}} h^{ab}(\tau, t) s(t - \tau)\, d\tau \; + \; n^b(t) \tag{2.5}$$

where $n^b(t)$ is the noise at the receiver and τ_{max} is the maximum delay of the echoes.

The same can be written in frequency considering the Channel Transfer Function

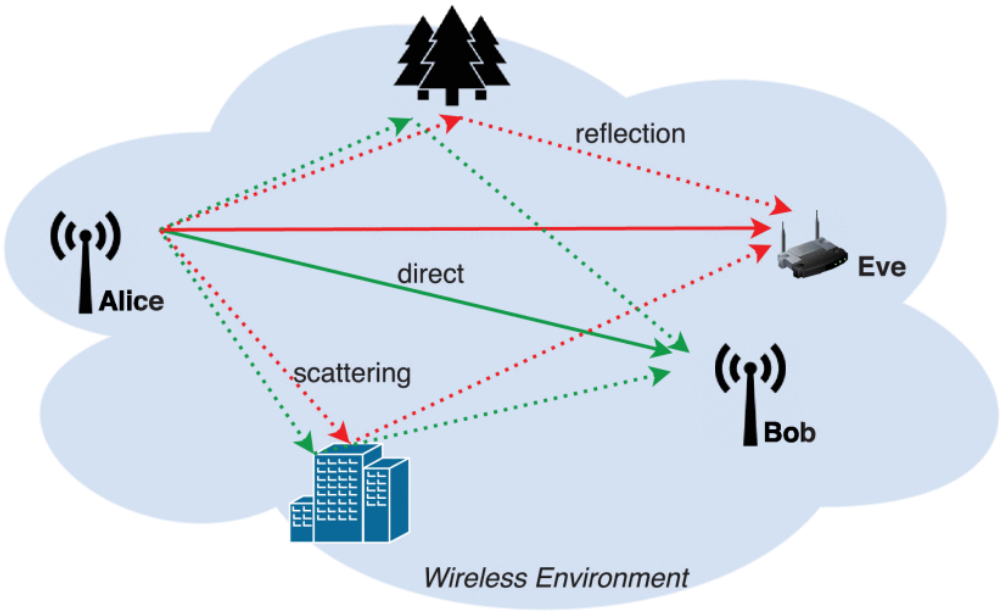


Figure 2.5: *Multipath effect inside the channel. The multipath configuration changes in the different point and varies in time in case the objects in the channel move.*

(CTF) $H^{ab}(f,t)$, which is the Fourier transform of the CIR, $H^{ab}(f,t) = \mathscr{F}\left[h^{ab}(\tau,t)\right]$:

$$Y(f,t) = H^{ab}(f,t)S(f,t) + w^{b}(f,t) \tag{2.6}$$

with $S(f,t)$ and $Y(f,t)$ the spectrum of the transmitted and received signals.

The spatial distribution of reflectors and scatterers in the channel determine its randomness by changing the multipath configuration. Physical Layer based-Key generation wants to observe the features of the channel and use them as a source of randomness.

Key generation leverages the following (expected) properties of the wireless channel:

- **Channel reciprocity**: the channel gains and phases shifts are the same for the transmitter and the receiver. In this way, every pair of legitimate users can extract the same information from the channel. Channel reciprocity holds in case the system is using Time Division Duplexing, as long as the legitimate users sample the propagation channel within the same *fading coherence time*. Instead, with Frequency Division Duplexing the channel might not be reciprocal anymore and key generation becomes thorny: some hints can be found in [3].

- **Spatial decorrelation**: in general, the spatial auto-correlation of fading decreases with distance and becomes negligible after a proper coherence distance. According to Jake's model, in a rich multipath scenario the fading samples are decorrelated after a distance of 0.4λ.

- **Fading randomness**: due to multipath and mobility, channel properties (e.g. Received Signal Strength (RSS), CTF) undergo random-like fluctuations in the spatial / frequency / temporal domain, to an extent that is usually related to the degree of "multipath richness". The greater the multipath effects, the more the channel appears as random.

2.3.1 Information theory of key generation

To justify the usage of the PL based-Key generation, it is now important to spend few words about the information theory behind it. Physical Layer based-Key generation is proved to be information-theoretically secure in [26] and [27]. Consider the model illustrated in Fig. 2.6 (which will be the reference model in the following): in order to extract the random key, Alice and Bob have to exchange some information s over the public channel, which can be overheard by Eve. Alice, Bob and Eve acquire the channel observations $X^A = [x^A(1), x^A(2), \cdots, x^A(n)]$, $X^B = [x^B(1), x^B(2), \cdots, x^B(n)]$, $X^E = [x^E(1), x^E(2), \cdots, x^E(n)]$. For any ε and sufficiently large n there exists a key generation protocol $K_{ir}^A = g_A(X^A)$ and $K_{ir}^B = g_B(X^B, s)$ which satisfies ([3]):

$$P(K_{ir}^A \neq K_{ir}^B) < \varepsilon \tag{2.7}$$

$$\frac{1}{n}\mathbb{I}(K_{ir}^A; s, X^E) < \varepsilon \tag{2.8}$$

$$\frac{1}{n}\mathbb{H}(K_{ir}^A) > R - \varepsilon \tag{2.9}$$

$$\frac{1}{n}\log_2|\mathcal{K}| < \frac{1}{n}\mathbb{H}(K_{ir}^A) + \varepsilon \tag{2.10}$$

where $\mathbb{H}(.)$ is the entropy, $\mathbb{I}(.)$ denotes the mutual information, and $\mathcal{K}$ is the key's alphabet. R is the achievable key rate: the maximum rate at which Alice and Bob can agree on a secret key while keeping the rate at which Eve obtains information arbitrarily small [27]. To give an explanation of the expression above:

- (2.7) represents the channel reciprocity: Alice and Bob can get the same key with an high probability.

- (2.8) represents the spatial decorrelation: Eve cannot infer the key based on her observation of the public discussion.

- (2.9) is referred to the capability of generating the key while avoiding Eve to steal useful information: for sufficiently large n it is possible to achieve the maximum key rate.

- (2.10) represents the temporal variability, which ensures having a uniformly

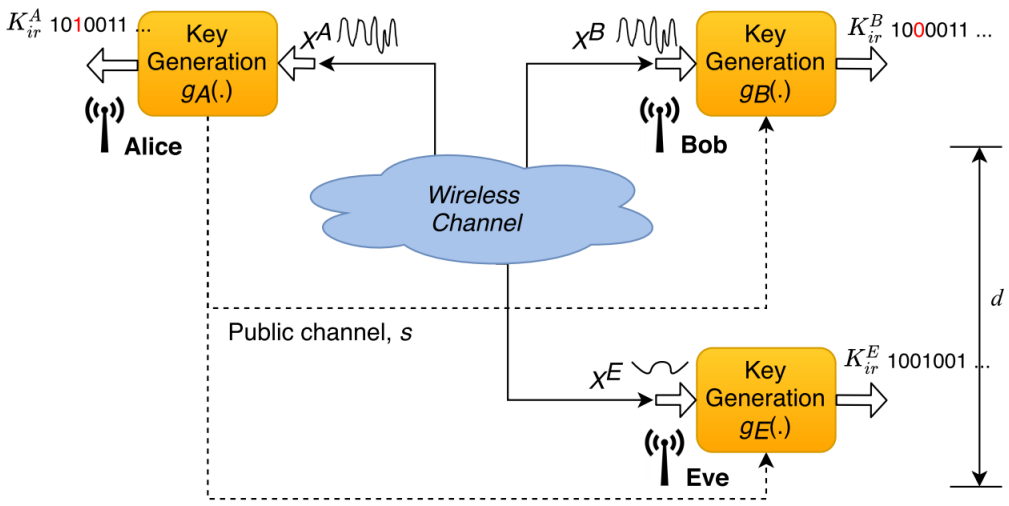


Figure 2.6: *Physical Layer based-Key generation model. Alice and Bob want to generate a common key from the wireless channel, Eve wants to observe the same channel and try to extract the same key. However, thanks to the spatial decorrelation properties, Eve will unlikely extract the same key, since the observations are different from the ones of Alice and Bob. Picture from [3]*

distributed key.

2.4 Physical Layer-Key generation protocol

In general a Physical Layer based-Key generation protocol relies on 4 main stages, as shown in Fig. 2.7:

- Channel probing: in this phase Alice and Bob exchange some information to measure the channel and observe the random features.
- Quantization: both users quantize the features observed to extract a random sequence of bits.
- Information reconciliation: Alice sends to Bob a public message in order for Bob to correct possible mismatch in the string of bits.
- Privacy amplification: both users perform some operation on the bit sequence to improve the randomness and produce a usable key.

In the end, Alice and Bob are able to generate the same key and use it for encryption. In the following, a detailed explanation of the main phases is provided.

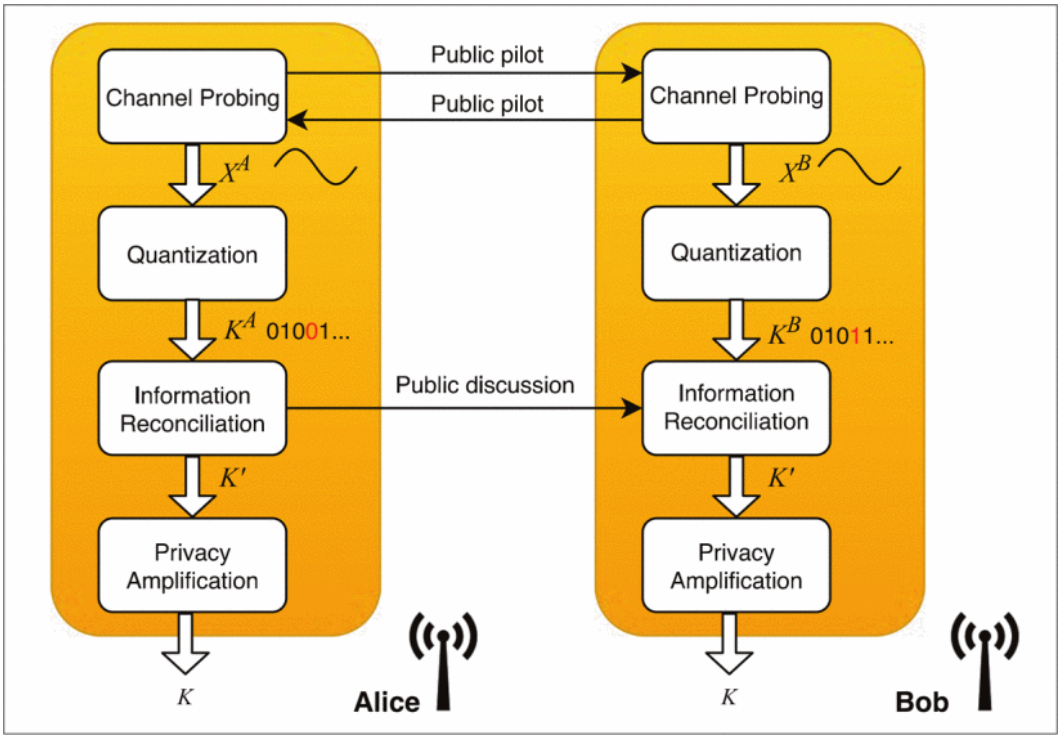


Figure 2.7: *Scheme of the key generation protocol. Picture from [8]*

2.4.1 Channel Probing

Alice and Bob alternately measure the common channel through the exchange of dummy packets. The goal of this part is to observe the features of the channel in order to harvest entropy from it: features can be observed in time and/or frequency and this phase mainly determine the key generation time. In fact, a system using the fading fluctuation in time will have to probe the channel many time and with a certain delay greater than the coherence time of the fading. Features commonly employed are:

- Radio Signal Strength Indicator (**RSSI**): this quantity is already available in many IOT system like LoRA or IEEE 802.15.4. The RSSI is a measure of the received power of the packet and it is computed for each packet. This feature changes in time thanks to the fading fluctuation due to mobility in the channel. A system using the RSSI as feature will take some time to generate the key and the users will have to exchange a lot of packets to have enough bits for the key.

- Channel State Information (**CSI**): this is a fine-grained quantity which can generate more information than the RSSI. CSI can be either the CIR $h^{ab}(\tau, t)$ or the CTF $H^{ab}(f, t)$. CSI are complex quantities and contain both the amplitude and the phase of the channel. In addition, while the RSSI is a narrowband quantity, CSI can take into account the dispersive properties of the channel, which means that can take advantage from the time/frequency selectivity of the channel. For example, OFDM systems already estimate the CSI in order to

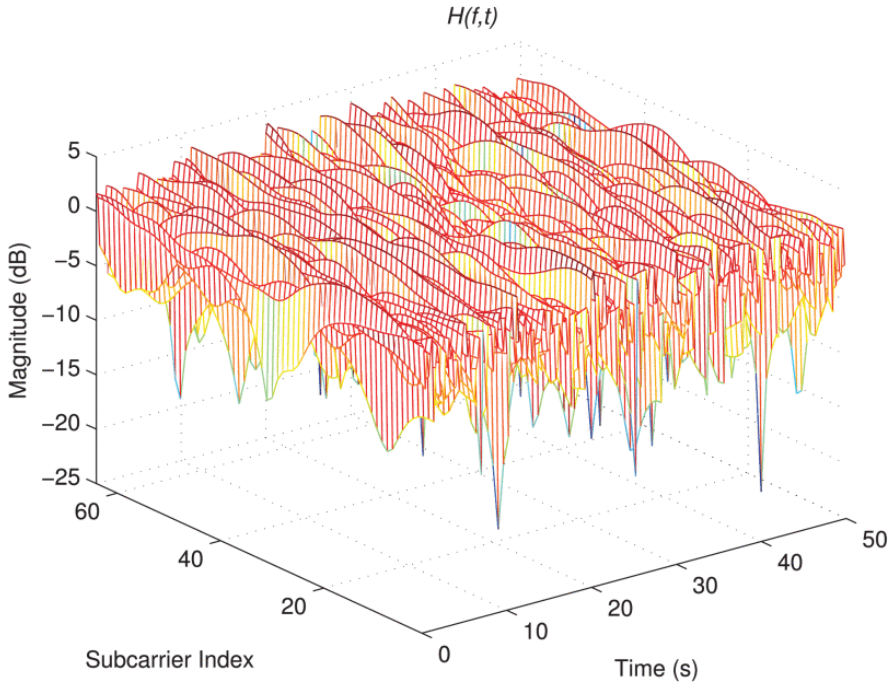


Figure 2.8: *CTF of OFDM signals. CTF varies in frequency and time, allowing to exploit both the characteristics. Picture from [3]*

equalize the transmission, although usually these information are not readily available and thus usable. With reference to the equation (2.6), simple estimation can be formulated as:

$$\widehat{H}^{ab}(f,t) = \frac{Y(f,t)}{S(f,t)} = H^{ab}(f,t) + \widehat{w}^{b}(f,t) \tag{2.11}$$

In case the original signal $S(f,t)$ is a known pilot signal. For the sake of this project, the CTF will be mainly referred to. In case the CTF is used, it is possible to exploit both time and frequency variability. As shown in Fig. 2.8, IEEE 802.11 OFDM signals can take advantage from both frequency and time in order to generate more bits for the key.

Although most common methods for Physical Layer based-Key generation are usually based on RSSI, on the CTF (see [28] in particular Table 2) or the CIR (see [29]), any other propagation marker can be used to harvest entropy from the channel, provided that it is symmetric between the two users and it is random: for example, there have been some proposal that take advantage from the Doppler Effect as a source of randomness [30]. Moreover, channel Probing leverages the channel reciprocity: the channel is almost reciprocal in case the communication system employs TDD. In case FDD is utilized, the channel might not be the same on the two frequencies. Therefore, specific effort must be put into action in case of FDD. A comprehensive description of the problem of FDD can be found in section V of [3].

Despite channel reciprocity, the estimate of the features might not be the same for

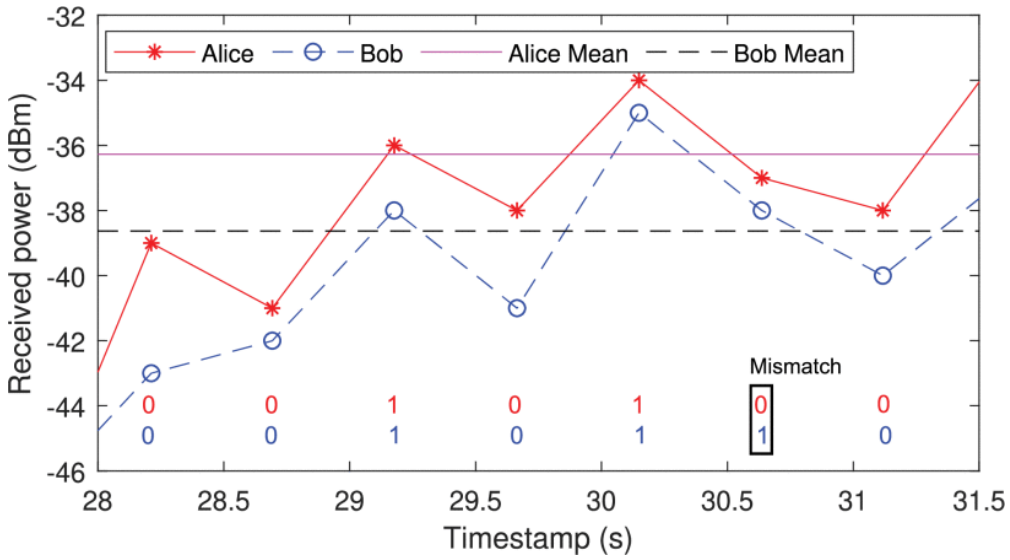


Figure 2.9: *Mean and Standard Deviation-Based Quantization. Note that in this case, Alice and Bob compute two different threshold, but what matter are the relative values of the power, not the absolute values.*

both the users: after the probing an additional phase called *signal pre-processing* is envisaged. Moreover, autocorrelation might be observed between channel samples, mainly due to coherence time and coherence bandwidth of the channel. Therefore, the subsequent signal pre-pocessing aims at correcting possible mismatch and auto-correlation between samples: the procedure is carried out locally inside the devices without exchange of information.

2.4.2 Quantization

The features extracted from the channel are analog quantities: in order to create a string of bits, one need to quantize the analog values into digital values, similarly to what is done with an analog-to-digital converter. With reference to Fig. 2.7, after the channel probing phase Alice and Bob come up with the channel observation X^A and X^B, the quantization phase produces the sequence K^A and K^B. There are two simple quantization algorithm that can be employed: Absolute Value Based-Quantization and Differential-Based Quantization.

2.4.2.1 Absolute Value Based-Quantization

It uses some threshold computed based on the statistics of the observation. A simple mechanism can be the *Mean and Standard Deviation-Based Quantization*: the users compute the mean value μ and the standard deviations σ of the channel samples. Then, they compute two thresholds:

$$\eta_+ = \mu + \alpha \times \sigma \tag{2.12}$$

$$\eta_- = \mu - \alpha \times \sigma \tag{2.13}$$

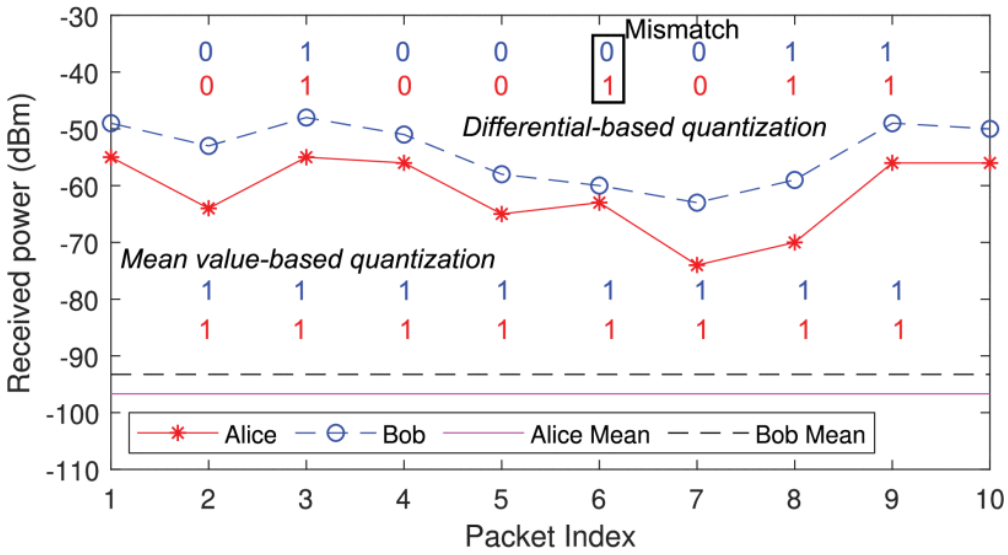


Figure 2.10: *Differential-Based Quantization compared to Mean and Standard Deviation-Based Quantization. Here the discrimination between* 0 *and* 1 *stands in the difference between two consecutive values.*

where α is a parameter to be tuned: what falls between η_+ and η_- is discarded, what is above η_+ is quantized in a 1 and what is below η_- is a 0. In case $\alpha = 0$, the two thresholds coincide, which is the situation depicted in Fig. 2.9. A more advanced quantization scheme is the *Cumulative Distribution Function Based Quantization*: it envisage also the possibility of using multiple threshold to have multi-bit quantization, the thresholds are computed based of the Cumulative Distribution Function of the observations. In order to map the bit to the quantization levels, Grey Code is used to reduce the mismatch between the quantization of Bob and Alice.

2.4.2.2 Differential-Based Quantization

Here no threshold is employed. Instead, the choice between 0 and 1 stands in the difference between two consecutive channel samples. Let's consider the case in which Key generation utilizes the RSSI: Alice and Bob exchange many packets in time and measure the RSSI from each one. Then, they look at two consecutive power values: if the current sample is greater than the previous then a 1 is produced, instead a 0 is produced. Furthermore, usually an additional margin is considered in order to protect from possible fluctuations of the noise. In terms of algorithm it can be written as:

if $x(i) > x(i-1) + \varepsilon$ **then**
  $K(i) = 1$
else
  if $x(i) < x(i-1) - \varepsilon$ **then**
    $K(i) = 0$
  end if
else

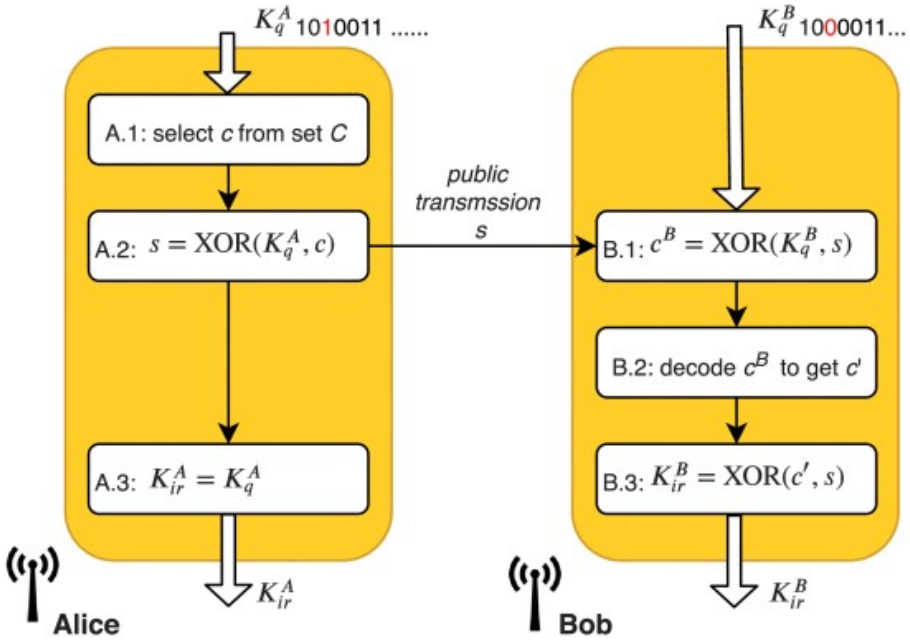


Figure 2.11: *Scheme of Information Reconciliation. Picture from [3]*

$x(i)$ is discarded
end if

This method is suitable for channel with low variability: in fact, even a small variation of the channel results in a different bit, which may not be true in the Absolute Value Based-Quantization. At the end of the quantization, Alice and Bob produces a the keys K_q^A and K_q^B.

2.4.3 Information Reconciliation

Despite channel reciprocity, the extracted features may may be slightly different, which results in possible mismatch between the two generated keys: errors are caused by thermal noise, not perfect channel reciprocity, possible interferences and hardware imperfections. Information Reconciliation phase allows the users generating the key to correct the errors in the key via a public discussion. A comprehensive summary of the possible schemes can be found in [31].

Reconciliation is a specific case of Forward Error Correction (FEC) and it can be implemented by using BCH codes, Reed-Solomon, turbo codes, polar codes, LDPC code. The choice of the code impacts on the reconciliation effectiveness: in fact, it is equal to the error correction capability of the code. As an example, a BCH(n, k, t) has a n-bit codeword, a k-bit message and can correct up to t errors: a BHC(15,3,3) can correct 20 percent mismatch. In case the errors in the key are larger than the reconciliation capability of the code, the keys cannot be corrected. Moreover, it is worth mentioning that the Reconciliation methods, but also the Privacy Amplific-

ation, are borrowed from the Quantum Key Distribution [32].First of all, Alice and Bob must agree on a code to be used and a codebook C. Then, as shown in Fig. 2.11, the reconciliation works as follow:

1. Alice selects a random codeword c and sets $K_{ir}^{A} = K_{q}^{A}$.
2. Alice computes $s = \texttt{XOR}(K_{q}^{A}, c)$ and sends it to Bob.
3. Bob receives s and computes $c^{B} = \texttt{XOR}(K_{q}^{B}, s)$
4. Then Bob decodes the codeword c^{B} into a codeword c', which is the original codeword c chosen by Alice
5. In the end, Bob can generate the reconciliated key $K_{ir}^{B} = \texttt{XOR}(K^{B}, c')$

2.4.4 Privacy Amplification

During the key generation process there is an exchange of public information: during the channel probing the pilot signals, but also the codeword exchanged during the Information Reconciliation phase. For this reason, the privacy amplification is a mandatory phase to protect against possible information leakage. In fact, from the public discussion Eve may be able to infer the key, or at least identify some of the bits of the key. Even gaining the knowledge of few bits represents a negative occurrence, since it reduces the search space in case Eve performs a *brute force attack*.

Privacy amplification employs the so called *universal hash families*, such as the *leftover hash lemma* [33], the *cryptographic hash function* [34] and the *Merkle-Damgard hash* [35]. Therefore, it distils a shorter key from the one generated after the Information Reconciliation, in order to reduce Eve's attack capability and to spread the entropy along the key. Hence, Alice and Bob might want to generate a longer key from the previous phase, which is then shorted but still a usable and secure key (e.g. with 256 bits).

2.4.5 Evaluation metrics of the PL based-Key generation

It is useful now to spend few words about the evaluation metric in the key generation area.

2.4.5.1 Cross-correlation

First, it is possible to compute the cross-correlation (also called the *Pearson coefficient*) to evaluate the similarity between the measurements of two users a and b (e.g. Alice-Bob or Bob-Eve)

$$\rho^{ab} = \frac{\mathbb{E}\{X^{a}X^{b}\} - \mathbb{E}\{X^{a}\}\mathbb{E}\{X^{V}\}}{\sigma^{a}\sigma^{b}}$$

2.4.5.2 Autocorrelation function (ACF)

The ACF is used to quantify the correlation among the channel samples. If the channel is represented by the random process $X(t)$, then the ACF is written as:

$$r(t, \delta t) = \frac{\mathbb{E}\{(X(t) - \mu)(X(t + \delta t) - \mu)\}}{\sigma_u^2}$$

where μ is the mean value of $X(t)$.

2.4.5.3 Key Disagreement Rate (KDR)

An important metric is the KDR: it quantifies the mismatch between the keys generated by two users after the quantization phase K_q^a and K_q^b. It is expressed as:

$$KDR^{ab} = \frac{\sum_{i=1}^{n_k} |K_q^a(i) - K_q^b(i)|}{n_k}$$

This quantity should be less than the correction capacity of the code used in the information reconciliation phase.

2.4.5.4 Secrecy Key Rate (SKR)

SKR is the upper bound on the number of bits per channel observation that Alice and Bob can extract from the channel, without the possibility for Eve to obtain any useful information. Of course, this is the theoretical limit and it should be interpreted in the same way as the channel capacity defined by Shannon. Maurer provided an upper and lower bound fot the SKR [27]:

$$R(X^A, X^B \parallel X^E) \geq \max[\mathbb{I}(X^A; X^B) - \mathbb{I}(X^A; X^E), \mathbb{I}(X^A; X^B) - \mathbb{I}(X^B; X^E)] \quad (2.14)$$

$$R(X^A, X^B \parallel X^E) \leq \min[\mathbb{I}(X^A; X^B), \mathbb{I}(X^A; X^B \mid X^E)] \quad (2.15)$$

2.4.5.5 Key Generation Rate

This is the actual number of bits that can be generated in a unit of time using a specific key generation method: it is usually measured in $\mathrm{bit/s}$. A well designed protocol can achieve a KGR close to the SKR.

2.4.5.6 Randomness

The key generated is employed in a symmetric encryption scheme: one of the requirements of the key is to be truly random, in order to prevent against cryptoanlitic attacks. NIST [15] provides a suite of fifteen tests used to evaluate the randomness of a *random number generator* (RNG). Each test returns a P-value which is compared to a statistical significance level α: in case the P-value $> \alpha$ the test is passed. Moreover, some tests requires long sequences of bits to evaluate the RNG. A summary of the test is shown in Fig. 2.12

Test	Purpose	Recommended key size n_k
Frequency (monobit) test	Proportion of zeros and ones for the entire sequence	100
Frequency test within a block	To determine whether the frequency of ones in an M-bit block is approximately $M/2$	100
Runs test	Total number of runs in the sequence, where a run is an uninterrupted sequence of identical bits	100
Longest run of ones in a block test	Longest run of ones within M-bit blocks	n_k=128, M=8
Binary matrix rank test	Check linear dependence among fixed length substrings of the original sequence	38,912
Discrete fourier transform (Spectral) test	To detect periodic features(i.e., repetitive patterns that are near each other)	1000
Non-overlapping template matching test	To detect generators that produce too many occurrences of a given non-periodic (aperiodic) pattern	Not specified
Overlapping template matching test	To detect generators that produce too many occurrences of a given non-periodic (aperiodic) pattern	10^6
Maurer's universal statistical test	To detect whether or not the sequence can be significantly compressed without loss of information	387,840
Linear complexity test	To determine whether or not the sequence is complex enough to be considered random	10^6
Serial test	The frequency of all possible overlapping m-bit patterns across the entire sequence	Choose m and n such that $m < \lfloor(\log_2 n_k - 2)\rfloor$
Approximate entropy test	To compare the frequency of overlapping blocks of two consecutive/adjacent lengths(m and m+1) against the expected result for a random sequence	Choose m and n such that $m < \lfloor(\log_2 n_k - 5)\rfloor$
Cumulative Sums test	The maximal excursion (from zero) of the random walk defined by the cumulative sum of adjusted (-1, +1) digits in the sequence	100
Random excursions test	The number of cycles having exactly K visits in a cumulative sum random walk	10^6
Random excursions variant test	To detect deviations from the expected number of visits to various states in the random walk	10^6

Figure 2.12: *Summary of the NIST random test. Picture from [3]*

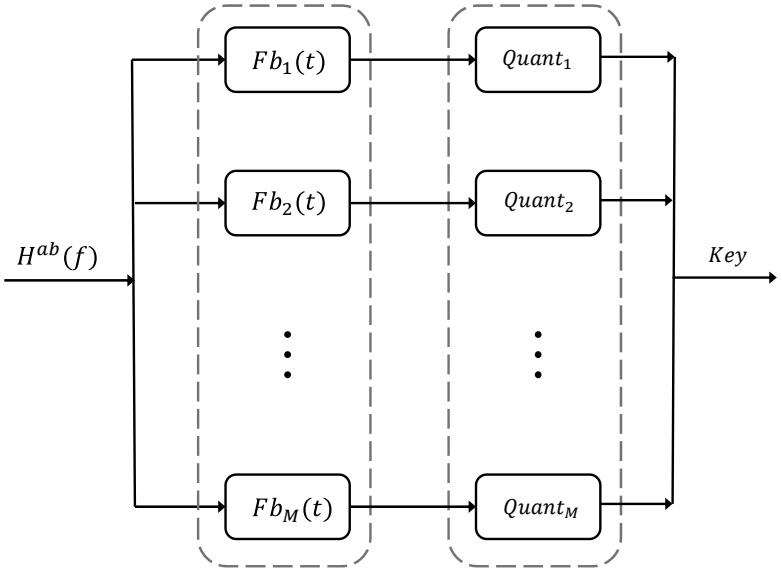


Figure 2.13: *Representation of the Filter-bank model, with the set of M parallel filters and the subsequent quantizers.*

2.5 Filter-bank model

Filter-bank (FB) model is one of the techniques for channel probing and it is the one considered in the following. Suppose that during the channel probing Alice and Bob exchange N packets (frame) with a time sample of t_p and compute the spectrum of the received signal. Frames contain *pilot symbols*, so it is possible also to evaluate the CTF. FB is a set of M parallel filters that process the received frame with a sampling frequency f_p [36]. A representation of the FB is contained in Fig. 2.13. Passing the frame through the filters allows to have $M \times N$ samples to be quantized: in this way, the KGR increases since both frequency and time are exploited, but also one can introduce more degree of freedom in the generation phase, which allows to tune the key generation process to the specific channel. Moreover, both M and N can be chosen dynamically, in case the received frames are correlated in time and frequency: t_p should be kept greater than the coherence time of the fading and f_p greater than the coherence bandwidth. However, it is important to bear in mind that due to the *data processing inequality* FB cannot increase the entropy of the observation, but rather it allows to better exploit the randomness in channel by observing both the time and frequency.

An abstract representation of the FB model is depicted in 2.14: the matrix $\overline{\overline{C}}$ repres-

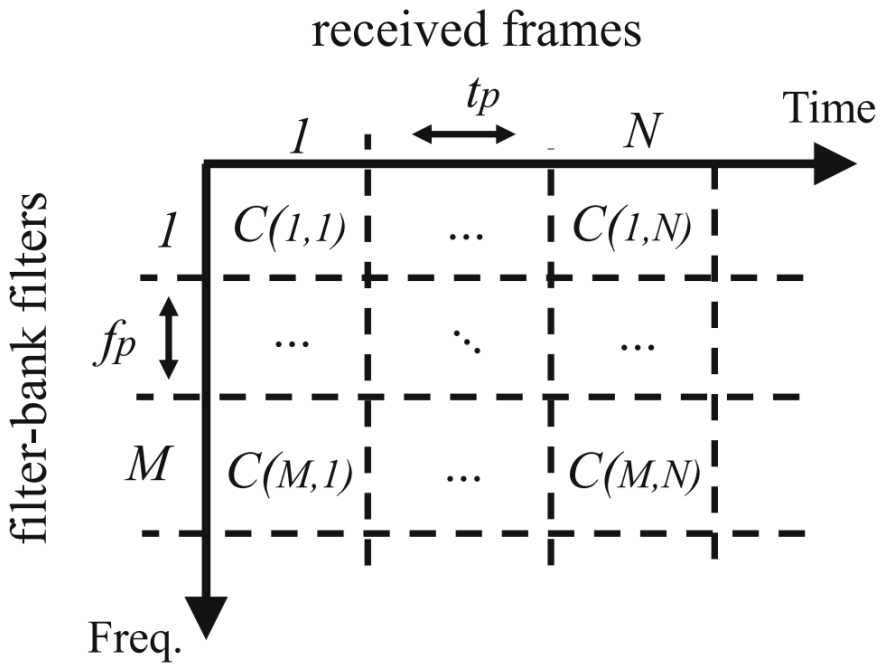


Figure 2.14: *Representation of the Filter-bank time-frequency plane. Picture from [36]*

ents the FB output in time frequency and it is obtained according to:

$$C_{(m,n)} = \frac{1}{\gamma} \int_{(n-1)t_p}^{nt_p} y(t) * Fb_m(t)\, dt \quad \forall n = 1, \ldots, N, m = 1, \ldots, M \tag{2.16}$$

where $C_{(m,n)}$ is the FB matrix, m is the frequency index (the filter), n the frame index, $y(t)$ is the received signal, $Fb_m(t)$ is the impulse response of the m-th filter, $*$ is the convolution operation and $\frac{1}{\gamma}$ is an arbitrarily normalization factor. In practise, the FB takes the CTF and passes it through the filters, for each filters, a coefficient is generated, as shown in Fig. 2.15.

The FB matrix is then quantized to obtain the key:

$$Key = Quant\left(\begin{bmatrix} C_{(1,1)} & \cdots & C_{(1,N)} \\ \vdots & \ddots & \vdots \\ C_{(M,1)} & \cdots & C_{(M,N)} \end{bmatrix}\right) \tag{2.17}$$

FB works well in wideband systems: some IOT systems are too narrowband (e.g. LoRa), but other communication standards have enough bandwidth such as WiFi, UWB, 5G [36].

2.6 Summary

This chapter described the Physical Layer Security, providing also elements of the information theory. In particular, the focus was on the Physical Layer based-Key

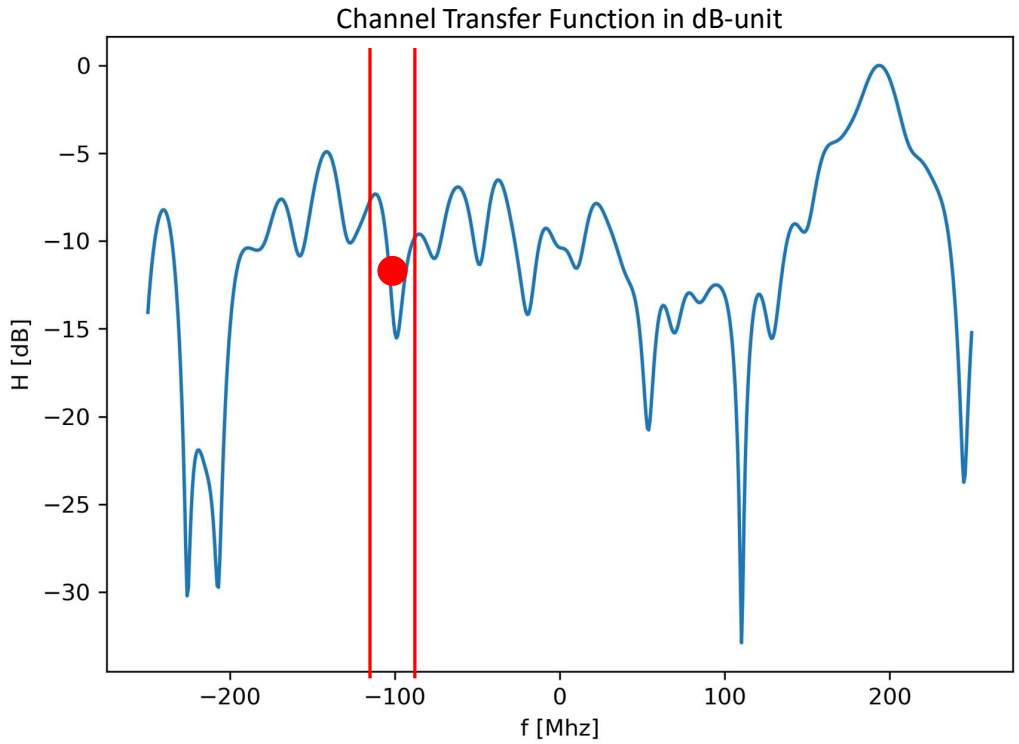


Figure 2.15: *Example of a filtering. The CTF is passed through one of the filters which outputs a single coefficient. In this example, the filter is assumed to be an ideal pass-band filter which outputs the mean value of the CTF.*

generation method, thanks to which two wireless devices can agree on the same symmetric key without the aid of a third unit. Moreover, there is an explanation of the key generation protocol, the metrics utilized to evaluate the performance of a protocol. Finally, the *Filter-bank model* has been introduced, which is the method considered for the key generation in this work.

This ends the first part related to the description of the theories behind this thesis work: starting from the next chapter, there will be an explanation of the activity conducted, the methodology followed and the results obtained.

Part II

Ray Tracing assessment

3

Ray tracing as a tool

This chapter wants to depict the goals of the thesis work and the tool utilized for the assessment. First, there is an introduction to the Ray Tracing, a software tool used to simulate electromagnetic propagation. Then, there is the description of the *Ray Tracing Attack*, a possible vulnerability for thePhysical Layer based-Key generation and the tools utilized to assess the possibility of success: the results are present in Chapter 4. Eventually, there is an hint about using the Ray Tracing as a design tool to evaluate the impact of different channel conditions on the Key generation method, which will be analysed in Chapter 5.

3.1 Ray Tracing

Ray Tracing (RT) is a numerical simulation of electromagnetic propagation according to Geometrical Theory of Propagation (GTP). A RT simulator computes some of the paths linking the transmitter and the receiver in a given modelled environment, in order to compute the channel in a deterministic way: the geometry and the field of the rays must satisfy GTP rules. In particular, the rays may experience reflections, diffractions, scattering, transmission through objects, usually referred as *events*: the simulator allows to set a maximum number of events sometimes referred as **prediction order** N_{ev}. Ray models can be fully 3-dimensional or can resort to simplified 2-dimensional modelling: the RT utilized in this work is a fully 3-dimensional model.

In order to be accurate, the Ray model requires in input a detailed and complete description of the environment, called **environment database**, and the description of the radiation properties of the transmitting and receiving antennas. A ray model follows the procedure sketched in Fig. 3.1:

- RT takes in input the environment database and the antenna description. The

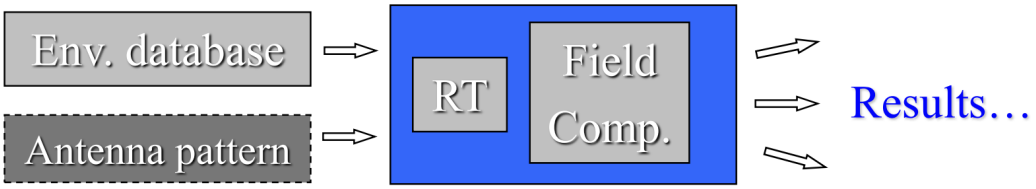


Figure 3.1: *Workflow of a Ray Tracing simulator. It takes in input the environment database and the radiation pattern of the antennas, first trace the rays and then compute the field along each ray. In the end it is able to compute the received filed and the power.*

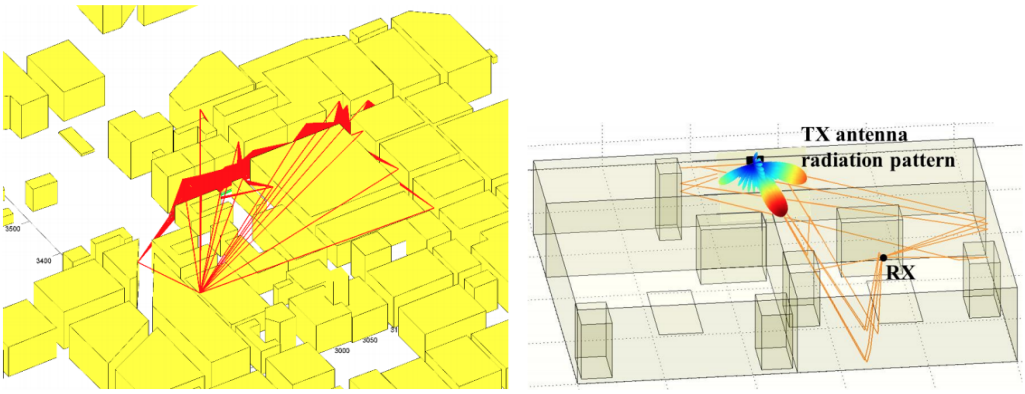
database contains the geometrical description of the environment and the electromagnetic characteristic of the materials. The antenna description consists of the radiation pattern of the antenna. In addition, RT takes in input the frequency over which the simulation must be carried, the transmitted power and the position of the transmitter and receiver. If necessaries, it is possible to define multiple transmitter or receivers.

- Then, the RT identifies the paths from the transmitter to the receiver, taking into account the possible events that a ray may experience. One can set both a limit on the total number of events and a limit on the number of events of a specific kind: for example, the number of diffraction that a ray can experience is usually kept low, since diffraction introduces a substantial attenuation. It is also possible to neglect a phenomena, e.g. just considering reflection and diffraction without allowing a ray to experience scattering.

- Once the rays trajectories have been determined, the field along the rays is computed, taking into account the interaction with the objects and the propagation losses. Furthermore, RT can compute the phase of the field up to some extent. However, the phase computed may not be trustworthy, since RT requires a model to compute the phase shift due to the interactions, that might not be accurate or just be a scalar model.

- The simulation outputs the ray pattern, i.e. a list of the tracked rays with some parameters as the received power of the component, the field's components, the propagation delay, the phase shift, the angle of the departure and arrival. Therefore, since the radiation pattern of the antennas are known, it is possible to compute the total received power and field as a coherent sum between the components.

The RT utilized in this thesis work has been developed at the University of Bologna and it is a 3D model that support also the simulation of the scattering [37]. The objects in the environment database are described as plain polygons with four or three edges, determined by the coordinates of the vertices. Moreover, it is possible



(a) *Example of urban environment.* ***(b)*** *Example of indoor environment.*

Figure 3.2: *Example of environment database.*

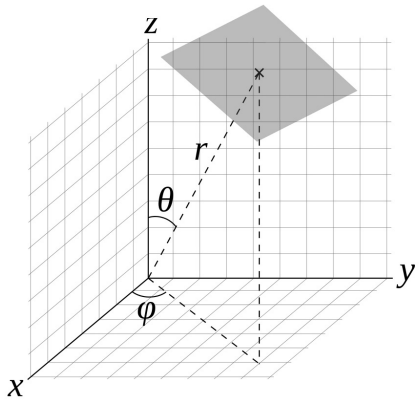


Figure 3.3: *Spherical coordinate system*

to represent outdoor or indoor environments: this representation is suitable for a wall in a room, while it imposes some limitation for other objects. In fact, obstacles inside the room are represented as parallelepipeds, but they are empty inside, so it is possible to insert only hollow objects by inserting the face of the parallelepipeds. Moreover, it is not possible to represent curved surfaces or complex walls: as shown in Fig. 3.4, a complex wall is represented as a superposition of polygon, neglecting the garnish of the building. Once the coordinates of the walls have been provided, the electromagnetic characteristics of the material must be inserted: the material is characterized by its *relative dielectric permeability* (ε_r) and *electric conductivity* (σ). In addition, the *thickness* can be specified: although the description of the surfece is in two dimensions only, this parameter is utilized to compute the propagation losses inside the material, so to somehow account for the real thickness of the item.

As for the antenna representation, it can be given in two ways:

- 3-dimensional: with reference to the spherical coordinate system in Fig. 3.3, for each partial direction (φ, θ), the antenna radiated complex field must be

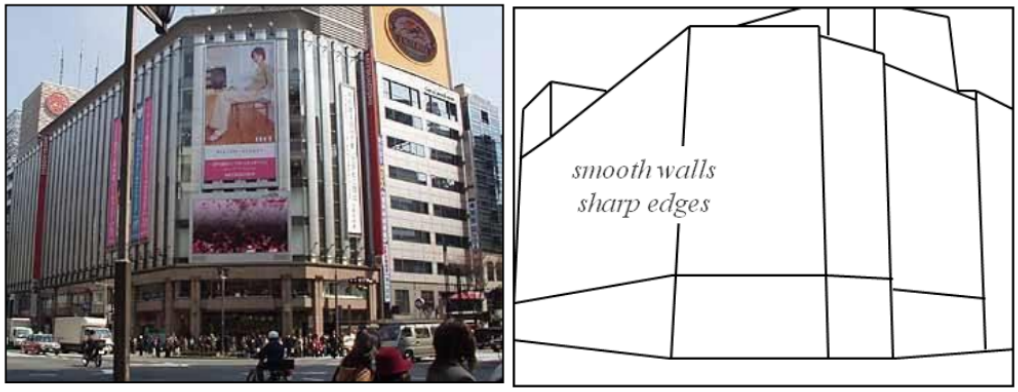


Figure 3.4: *Example of complex wall that must be approximated as superposition of plain surfaces.*

given. The values of φ an θ are discretized with a certain angular step.

- 2-dimensional: the antenna is represented by the gain function over the E-plane and H-plane. The Ray Tracing then utilizes the *Correia Method* [38] to extrapolate the radiation pattern in all the directions (φ, θ). This representation is simpler than the 3-D, yet being also less precise.

A a proper choice of the material's characteristics, a good representation of both the environment and antenna characteristics contribute to a better simulation. Eventually, the RT is able to simulate only an environment without moving objects.

3.1.1 Narrowband output

A Ray model is able to find the paths from the transmitter to the receiver respecting some constraints on the interactions undergone by the rays. As modelled in equation (2.4), the CIR can be computed based on the rays information: the RT provides in output the list of of the tracked rays, together with the main propagation parameters of the field propagating along each rays (like intensity, phase, propagation delay, direction of departure/arrival, number and kind of experienced interactions, etc.). All the outputs of the simulation are collected in some text based files. Therefore, in first place, the RT is able to simulate the CIR of the environment. However, the RT is able to simulate only a static environment, hence in the CIR one has to neglect the temporal variability of the channel. Moreover, even though the Ray Tracing sounds the channel over a single frequency, its approach can also investigate the wideband properties of the propagation channel, provided that the analysis is limited to the frequency range where the materials' electromagnetic parameters are basically constant and the radiation properties of the antennas are also unchanged. For the sake of completeness, the CIR without the temporal dependency is reported

here:

$$h^{ab}(\tau) = \sum_{l=1}^{L^{ab}} \alpha_l^{ab} e^{-j\phi_l^{ab}} \delta\left(\tau - \tau_l^{ab}\right) \tag{3.1}$$

Now, the attenuation α_l^{ab}, the phase shift ϕ_l^{ab} and the delay τ_l^{ab} depend only on the considered path.

In order to compute the CIR, a post processing tool (like Matlab or Python) is employed, which works in the digital domain: for this reason, it is mandatory to work with a discrete version of the CIR. The procedure to consider the discrete version of the CIR employed in this work is the following:

1. Define a sampling period T_s.

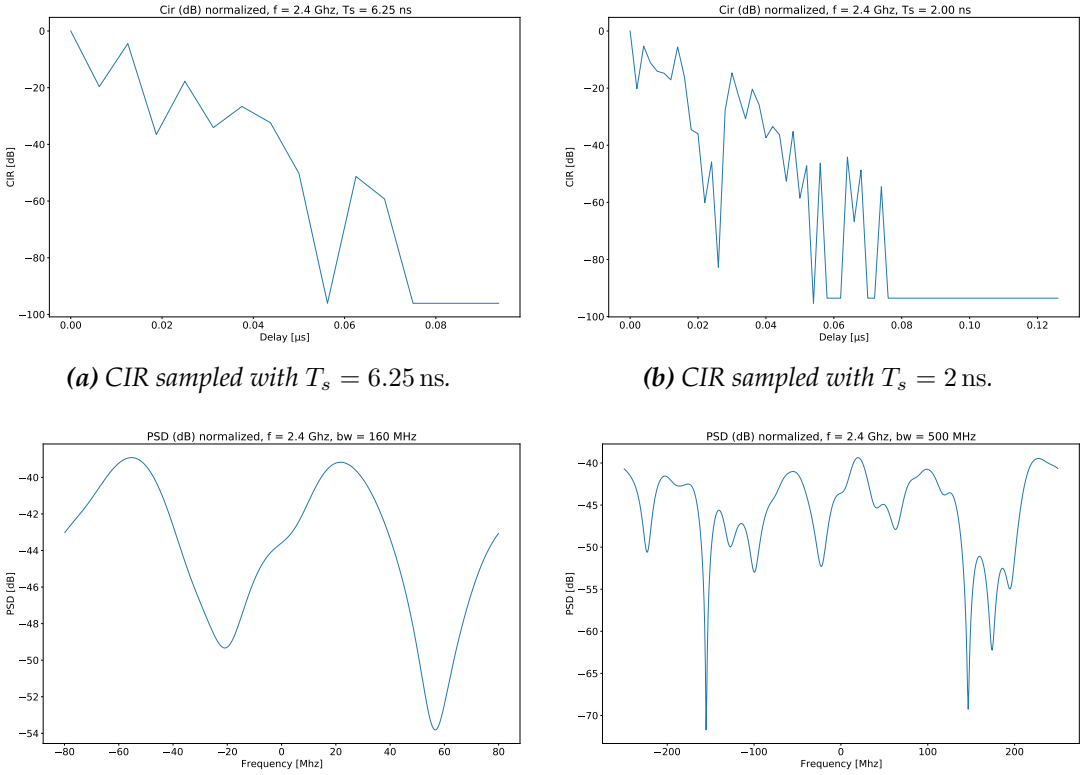
2. Initialize to 0 an array long enough to store the samples.

3. For each ray, take the relative delay with respect to the first path ($\Delta\tau$) and convert the delay in a sample number ($n_s = \lfloor \Delta\tau / T_s \rfloor$).

4. Once the time sample has been found, add the ray contribution to the correspondent array position. The rays are considered as complex numbers computed according to (3.1). In case two rays have the same time sample, they are coherently summed in the array, so simply the two complex number are added.

In the end, the discrete CIR is built as a complex-valued array. To give an interpretation, it is like the receiver is able to resolve in time only some components, hence does not have the time resolution to capture all the single rays, but rather a coherent sum of the contributions arriving in the same time slot. To have a better comprehension of the effect of T_s, consider Fig. 3.5a and Fig. 3.5b: both pictures show the same CIR extracted from the same RT simulation, but the first is sampled with $T_s = 6.25\,\mathrm{ns}$ while the second with $T_s = 2\,\mathrm{ns}$. Clearly, the CIR in Fig. 3.5b shows more peaks since with a lower sampling period it is possible to resolve more components.

Despite the narrowband simulation, there is still a possibility to evaluate the wideband channel. In fact, the CTF can be estimated by applying the Fourier Transform to the CIR: this method does not provide the exact CTF, since the CIR is computed through a narrowband measurement. In addition, the CTF, but also the CIR in case of low time resolution, is highly influenced by the phase of the rays, which determines the position and the depth of the null over the CTF. However, the phase shift of the components is determined by the total propagation distance and by the interaction with the materials, which depends also on their permittivity and conductivity. If the electromagnetic characteristics of the material are roughly known, then the computed phase shift will be inaccurate. Furthermore, the Power Spectral Density



(a) *CIR sampled with* $T_s = 6.25\,\mathrm{ns}$.

(b) *CIR sampled with* $T_s = 2\,\mathrm{ns}$.

(c) *Correspondent Power Spectral Density with* $B = 160\,\mathrm{MHz}$.

(d) *Correspondent CTF with* $B = 500\,\mathrm{MHz}$.

Figure 3.5: *CIR and correspondent Power Spectral Density from the same RT simulation but with different sampling period. Both the CIR and Power Spectral Density are shown in* dB*-unit and normalized to* $0\,\mathrm{dB}$

(PSD) of the channel can be computed by squaring the CTF. In order for this method to be reliable, two assumptions must be done:

- The phase shift remains constant over the considered bandwidth. In this way, a single frequency is representative of the entire bandwidth. This is a strong assumption, since the phase shift depends also on the frequency.

- The electromagnetic characteristics of the materials are constant over the considered bandwidth. This is not a stringent assumptions, since dispersion effects are negligible for communications up to few GHz.

Under these two assumptions, the CTF can be obtained through a Fourier Transform, which can be implemented via a Fast Fourier Transform (FFT). Actually, the output of the FFT is the *pass-band equivalent* CTF. Since the CIR has been sampled with a sampling period T_s, the resulting CTF will have a bandwidth (pass-band) $B = \frac{1}{T_s}$. The results of the sampling period of the CIR on the CTF can be seen in Fig. 3.5c and Fig. 3.5d: the first is obtained from the CIR sampled with a lower sampling period, resulting in a lower bandwidth; the second instead is obtained

from the CIR with more resolved paths and has an higher bandwidth.

3.2 Ray Tracing Attack

As explained in section 2.3, Physical Layer based-Key generation relies on some channel features, in order to harvest entropy from the wireless channel: a lot of key generation methods exploits the CIR or the CTF. However, simulation tools are able to infer the channel and create a possible vulnerability for the key generation method. In fact, a malicious user that posses a simulation tool might be able to simulate the channel between Alice and Bob and guess the key generated, or at least extract some information that will be used in a cryptoanalytic attack. In particular, it might be possible to carry out a so called *Ray Tracing Attack*, in which the adversary exploit a RT tool to simulate the channel and infer the key. In order for the Ray Tracing Attack to be successful, the adversary should:

- Have a precise knowledge of the geometry of the environment and of the electromagnetic characteristic of the materials.
- Know the position of of Alice and Bob in the environment, the radiation pattern of their antennas and the key generation protocol used.
- Possess adequate computational power in order to execute the RT simulation in a short time.
- Be able to eavesdrop the correction messages to correct the extracted key.

Since the RT is able, at least currently, to simulate only a static environment, a Ray Tracing attack is suitable for the static case, without excluding a future improvement in the Ray-based models.

This thesis work focuses on the possibility to carry out a Ray Tracing Attack against the PL based-Key generation , which in this case uses the *filter-bank* model in a wideband scenario. Moreover, it also consider the possibility of using the RT as a tool to evaluate the sensibility of the PLS to the channel conditions, as the multipath, type of antennas, bandwidth, etc. The filter-bank model makes use of the CTF as channel feature, so the RT will be used to compute the CTF of the channel, with all the limitation expressed before. The step followed to evaluate the effectiveness of the Ray Tracing attack are:

1. Choose an environment to be used;
2. Identify some positions for Alice and Bob;
3. Create the environment database for the RT tool with the geometry of the loc-

Figure 3.6: *Picture of the R&S®FSH-8, the VNA used for the measurements.*

ation and the material characteristics;

4. Measure the channels between different positions of Alice and Bob;

5. Simulate the same channels with the RT;

6. Repeat the same measurements and simulations for different central frequencies and channel bandwidth;

7. Choose some metrics to evaluate the similarity of the channel, possibly generate the key from the channel and compute the KDR between the key extracted from the actual channel and the key generated from the simulation.

The measurement campaign and evaluation of the Ray Tracing Attack have been performed at the Barkhausen Institut gGmbH[1](BI) in Dresden (GE). A similar evaluation of the ray tracing attack can be found in [29], where the authors explored the possibility of the ray tracing attack against UWB, which utilizes the CIR as a feature to extract the key.

3.2.1 Vector Network Analyzer

In order to measure the channel between Alice and Bob, a **Vector Network Analyzer** (VNA) has been used. The VNA available at the BI was the R&S®FSH-8[2] (Fig. 3.6). A VNA has two ports and allow to measure the *scattering matrix* S_t of the device under test: in particular, it performs a sweep from a starting frequency to a final frequency, probing the channel with a certain frequency step. For the aim of the

[1] https://www.barkhauseninstitut.org/en

[2] https://www.rohde-schwarz.com/products/test-and-measurement/handheld/rs-fsh-handheld-spectrum-analyzer_63493-8180.html, https://www.rohde-schwarz.com/manual/r-s-fsh4-8-13-20-operating-manual-manuals-gb1_78701-29159.html

project, the VNA was utilized to measure the S_{21} in the power domain, or the S_{12} since the wireless channel can be assumed to be reciprocal: the S_{21} is equal to the *Power Spectral Density* of the channel. Regardless of the sweep width, the VNA utilized was able to probe the channel only in 631 points: the wider the sweep, the larger the frequency step. In addition, the power of the output port was set to $0\,\mathrm{dBm}$.

In the considered scenario, two antennas were attached to the port in order to measure the wireless channel. Moreover, from the datasheet of the instrument [39] it is possible to read the average noise level, normalized to $1\,\mathrm{Hz}$ and specified for different operational bands: for the aim of the project, when considering the noise level of the instrument the value $-140\,\mathrm{dBm}$ has been utilized. Eventually, the VNA can be remotely controlled from the PC by means of a proprietary software briefly described in section 3.2.3.

3.2.2 Antennas

The antennas utilized for the measurements were the Mobile Mark MGRM-WHF[3] (Fig. 3.7) omnidirectional wideband dipole, working from $1.7\,\mathrm{GHz}$ to $6\,\mathrm{GHz}$. Since the radiation pattern of this antenna is not available online, it was necessary to characterize the radiation pattern of the antenna to be used in the RT simulations, especially the vertical plane (or E-plane) of the dipole. To do this, two antennas were employed one placed vertically, the other placed horizontally: the first was fixed and fed by a function generator, the other was rotating in order to measure the vertical plane of the vertical dipole. In addition, two absorbing plane were placed behind the antennas to reduce the effect of echoes, as shown in Fig. 3.8. In the end, the measured characteristics are available in Fig. 3.9. The E-plane has been measured for three frequencies: $2.4\,\mathrm{GHz}$, $3.7\,\mathrm{GHz}$, $5.5\,\mathrm{GHz}$, that are the central frequencies considered for the Ray Tracing Attack. The antennas are omnidirectional, so the horizontal plane (H-plane) has not been measured. Having the E-plane and the H-plane, the antennas in the RT are inserted by means of the 2D representation.

Figure 3.7: *Picture of the antenna used for measuring the channel.*

[3] https://www.mobilemarkantennas.com/pub/media/solwin/productattachment/attachment/file/m/g/mgrm-whf_spec_sheet.pdf

Figure 3.8: *Antenna on top of a tripod with the absorber wall behind. However, in this picture the antenna was placed vertically, while for the measurements it was placed horizontaly.*

3.2.3 Software tools for post processing

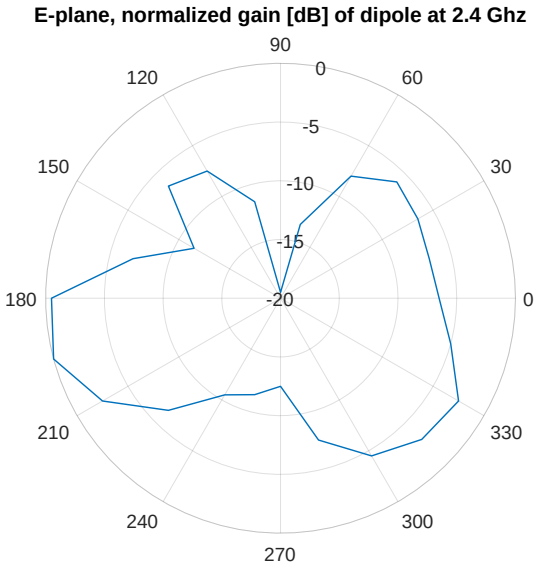
First, the VNA can be controlled remotely via a software produced by Rhode&Schwarz called **R&S®Instrument View** [4]. The software allows to remotely see the display of the instrument (figure Fig. 3.10), change the measurement parameters, start the frequency sweep of the VNA and store the results of the measurements as a `.csv` file: in this way, the measurement can be performed without being physically in the environment and consequently interfere with the measurements.

Second, **Anaconda**[5] based on **python™** 3.8.8 [6] has been used as a post processing tool, to create a routine for launching the Ray Tracing simulation and for the comparison between Ray Tracing channels and measured channels.

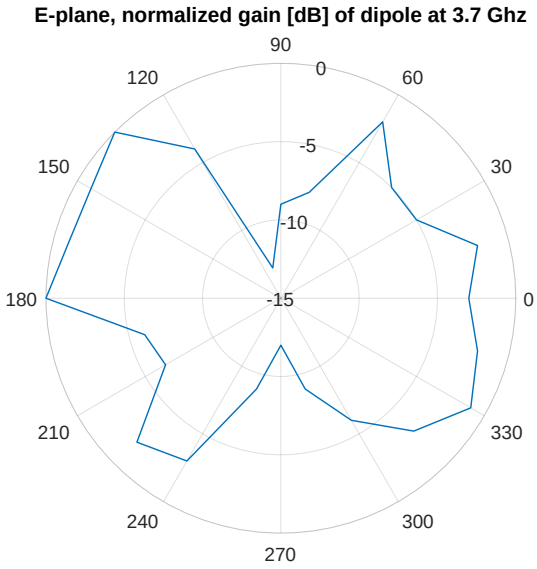
3.3 Ray tracing as a simulation tool for the PL based-Key generation

Key generation in susceptible to different channel conditions, that may vary based on the specific environment in which the system is deployed. However, a measurement campaign to assess the impact of the channel might be long and expensive. Therefore, the RT emerges as a possible tool to generate realistic channels [37] according to the characteristics and geometry of the environment. As for this work, the impact of the channel bandwidth, the central frequency and the type of antenna have been considered to evaluate their impact on the key generation.

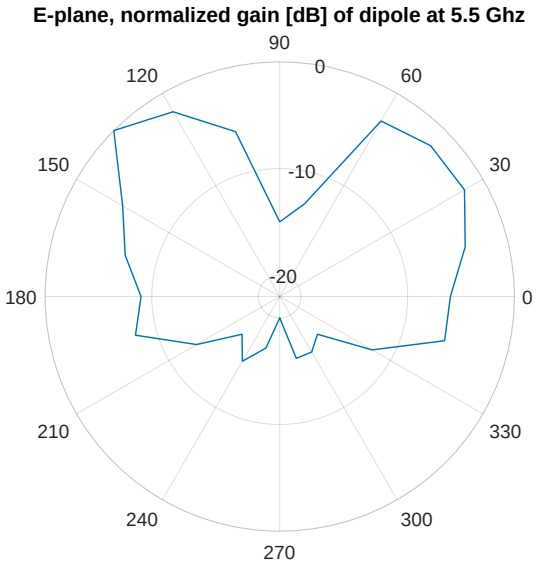
[4] `https://www.rohde-schwarz.com/software/instrumentview/`
[5] `https://www.anaconda.com/`
[6] `https://www.python.org/`



(a) E-plane of the dipole measured at 2.4 GHz.



(b) E-plane of the dipole measured at 3.7 GHz.



(c) E-plane of the dipole measured at 5.5 GHz.

Figure 3.9: *Normalized gain of E-plane of the Mobile Mark MGRM-WHF in* dB *for the different frequencies.*

3.4 Summary

This chapter described the Ray Tracing attack and the tools utilized for the assessment: the VNA, the antennas and briefly the software used. Then, it gave some hints about the Ray Tracing as a design tool. In the next two chapters, there will be the presentation of the results of the assessment.

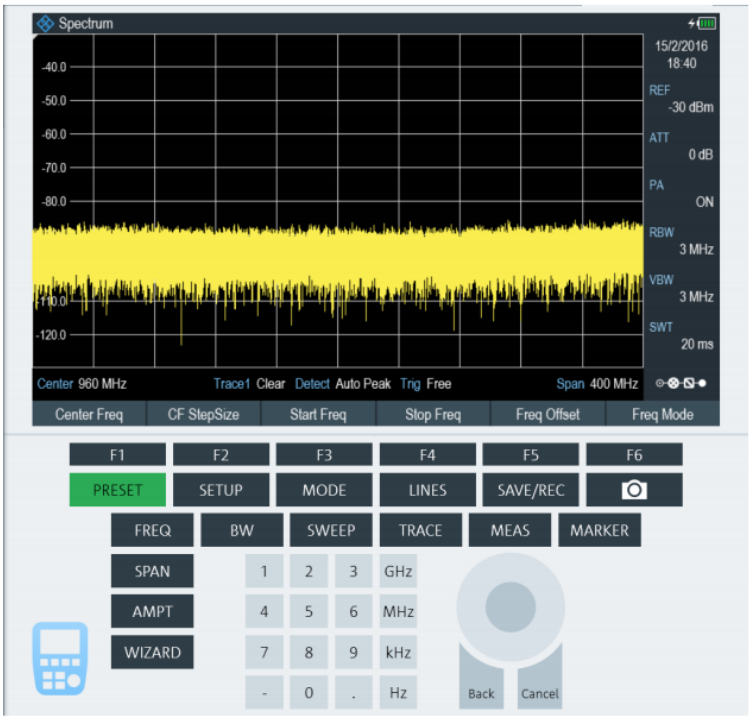


Figure 3.10: *Picture of the remote display feature of R&S®Instrument View*

4

Ray tracing attack

The goal of this chapter is to describe the activity conducted for the assessment of the Ray Tracing attack. First, there is the description of the two environments. Then the procedure utilized for the assessment, with all the channel parameters under consideration and the metrics utilized for the evaluation. Measurements and simulations have been carried out in two environments, three different central frequencies and two different channel bandwidths. Although the key generation method considered makes use of the channels in the frequency domain, there is also a quick comparison of the channel in the time domain. In addition, some indication are given on the impact of channel conditions on the key generation procedure, which is the topic of the next chapter.

4.1 Description of the environments

An empty room has been considered at first for the sake of simplicity, i.e. with propagation always occurring in Line of Sight (LOS). Then, some obstacles have been placed in order to prevent direct visibility between the antennas. The environment must be surveyed in order to create the database for the ray tracing, with the dimensions of the room, the positions of the objects and of the measurement points. The dimensions of the room have been taken by means of a laser meter and a measuring tape.

4.1.1 Empty room

The room tested is the *Mondrian Room* at the Barkhausen Institut gGmbH, which was emptied by all the furnitures and objects inside, in order to have a simple environment and with a strong Line of Sight component. A picture of the room can be seen in Fig. 4.1: it is a square of 5×5 meters and with an height of 3.54 meters with a wooden floor and ceiling, the walls are assumed to be made of concrete; there

Figure 4.1: *Overview of the empty room. In the picture, there are also the antennas with their wooden supports*

are two wooden doors in the room (Fig. 4.2) that have been modelled as a single layer of wooden on the wall. Two glass windows are present on the other two walls and inserted in a cavity: they have two glass layer and a wooden structure, but for the ray tracing they are modelled as a single glass layer without the wooden structure, since it is not expected to affect too much the electromagnetic propagation. Moreover, below each window there is a radiator that is modelled as a Perfect Electric Conductor (PEC) metal panel, without the side surfaces and without anything behind: this choice is justified by the fact that the number of interaction of the rays has been set quite low, so it is very unlikely that in the simulation a ray experiences a multiple bounce between the back of the metal panel and the wall behind. On the wall, at 0.62 meters from the ground, there is a cable duct made of metal (details are shown in figure Fig. 4.3). The cable duct is modelled as a parallelepiped, with only the upper lower and front surfaces, assumed to be PEC; the cable inside has not been represented in the ray tracing since PEC acts as a perfect reflector. On the floor, there are the measurement positions for Alice and Bob indicated by the sign made by tape. As for the electromagnetic parameter of the material, the values of the *relative dielectric permeability* (ε_r) and the *electric conductivity* (σ) are shown in Table 4.1. The values are specified for three frequencies: $2.4\,\mathrm{GHz}$, $3.7\,\mathrm{GHz}$ and $5.5\,\mathrm{GHz}$, which are the three central frequencies utilized for the evaluation.

In the end, the digital representation of the room for the ray tracing is shown in Fig. 4.4. Moreover, Fig. 4.5 shows a graphical description of an example of a ray tracing simulation in the room: the rays experience reflection on the walls and the windows, diffraction on the edge of the cavity with the windows. The number of the rays in the figure has been limited to make the plot more clear since the simulation considered generated over 7000 rays: in particular, in the figure there are the rays with a maximum delay of $0.017\,\mu\mathrm{s}$.

Figure 4.2: *Detail of the wooden door, with the column and the handle.*

Frequency	2.4 [GHz]		3.7[GHz]		5.5[GHz]	
Parameter	ε_r[F/m]	σ[S/m]	ε_r[F/m]	σ[S/m]	ε_r[F/m]	σ[S/m]
Concrete	4.5	0.007	6.0	0.0017	5.5	0.087
Wood	3.0	0.0	1.8	0.012	2.5	0.009
Glass	6.0	0.016	6.0	0.0016	5.5	0.016

Table 4.1: *Table of the electromagnetic parameters of the material, values obtained from a survey of the literature.*

(a) *Cable duct.*

(b) *Detail of the cables inside the duct.*

Figure 4.3: *Cable duct inside the room.*

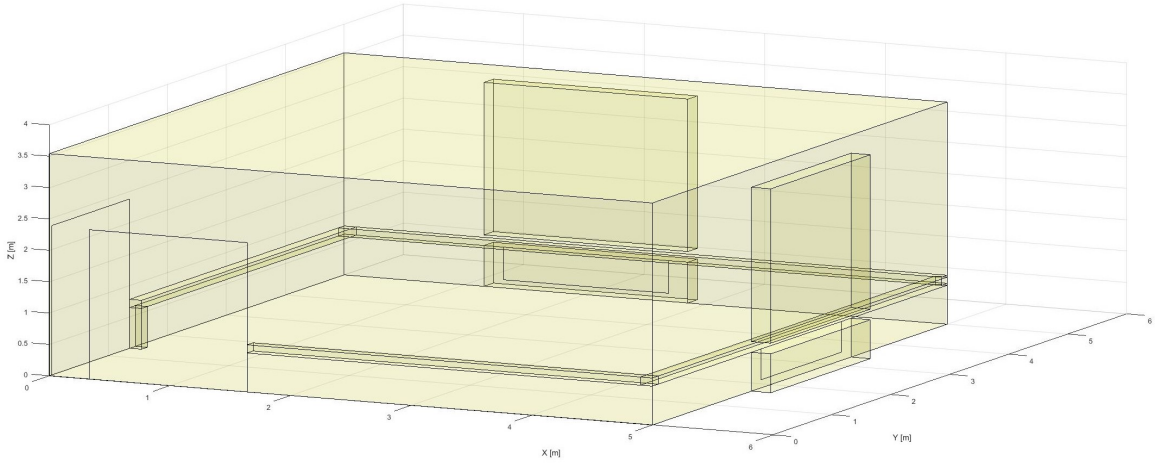


Figure 4.4: *Ray tracing environment of the Mondrian Room.*

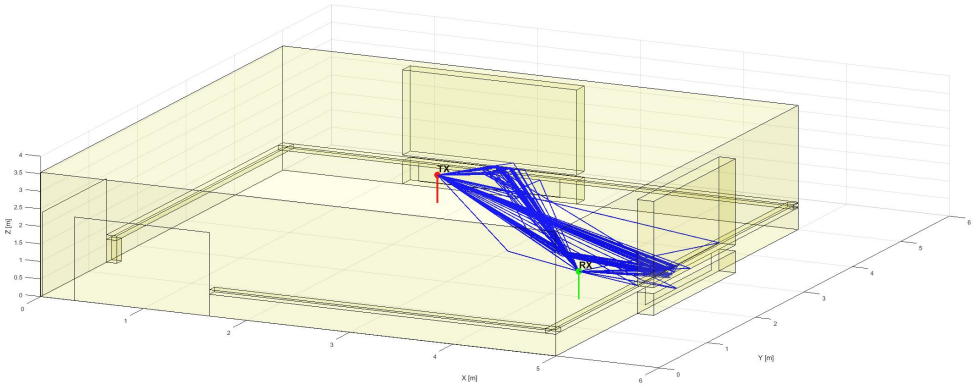


Figure 4.5: *Example of a ray tracing simulation inside the room.*

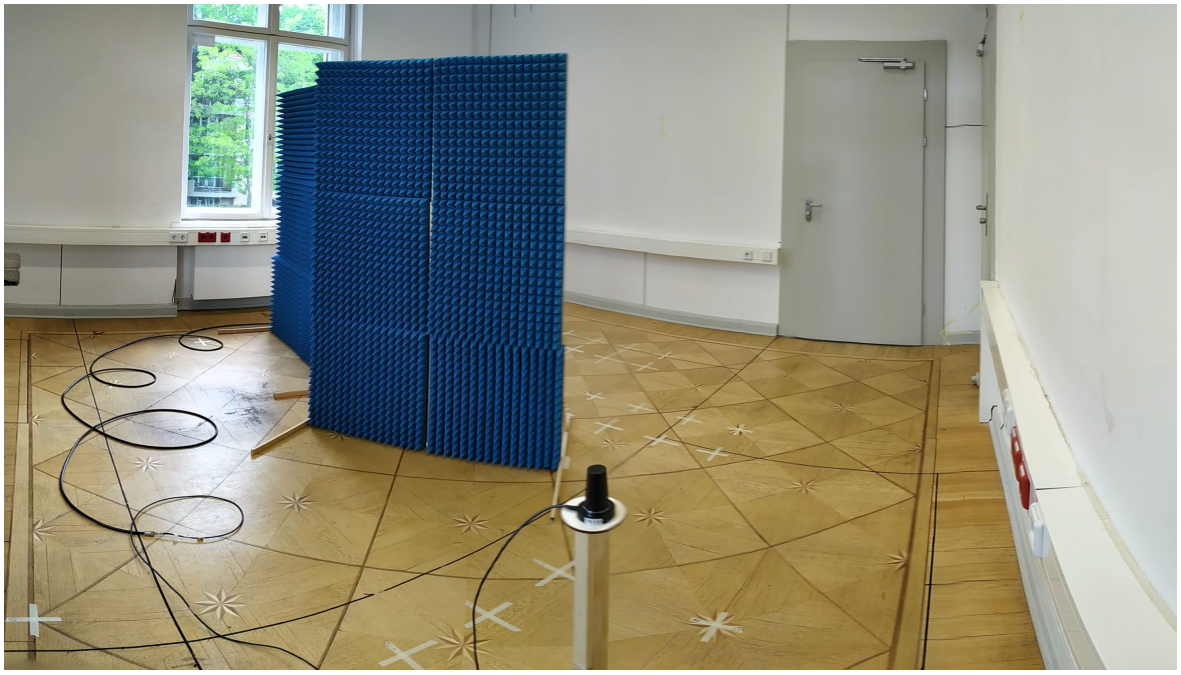


Figure 4.6: *Mondrian Room with the two absorber panel in the middle.*

4.1.2 Room with obstacles

The second environment considered is the same room as before, but with some obstacles in the middle in order to cut the Line Of Sight component: in this way, the communication would relies on the multipath components. The obstacles chosen were two absorbing panels placed in the middle of the room, held by a wooden support (Fig. 4.6 and Fig. 4.7).

In the environment database the absorbing panel are modelled as perfect absorber on both surfaces and the wooden support has not been included: they are just represented as rectangles without thickness, as shown in Fig. 4.8. A ray based representation of propagation in the room with obstacles is sketched in Fig. 4.9: as the LOS path is blocked, the rays experience reflection on the walls and diffraction on the edges present in the room. The positions used for the measurements are the same as before.

Figure 4.7: *Back surface of the absorber.*

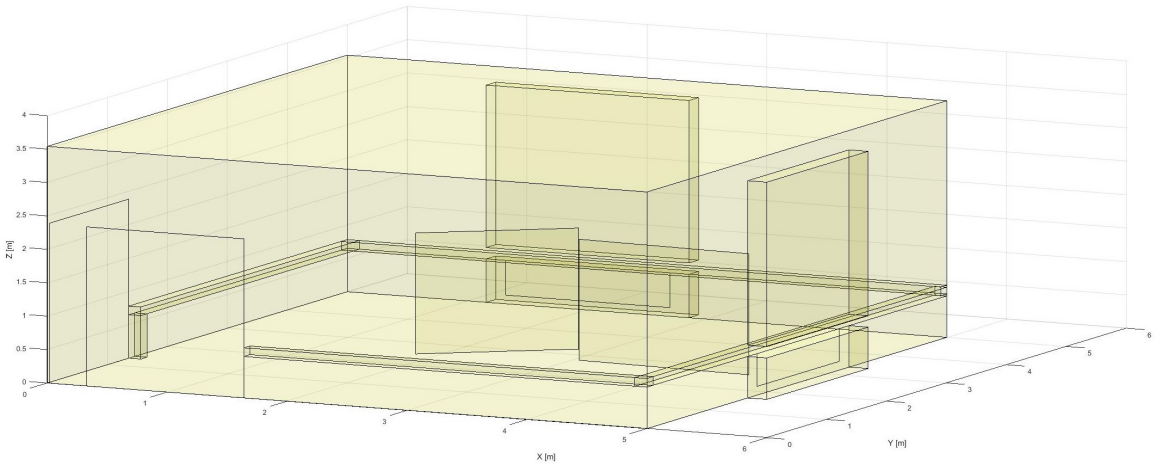


Figure 4.8: *Ray tracing environment of the Mondrian Room with the absorber.*

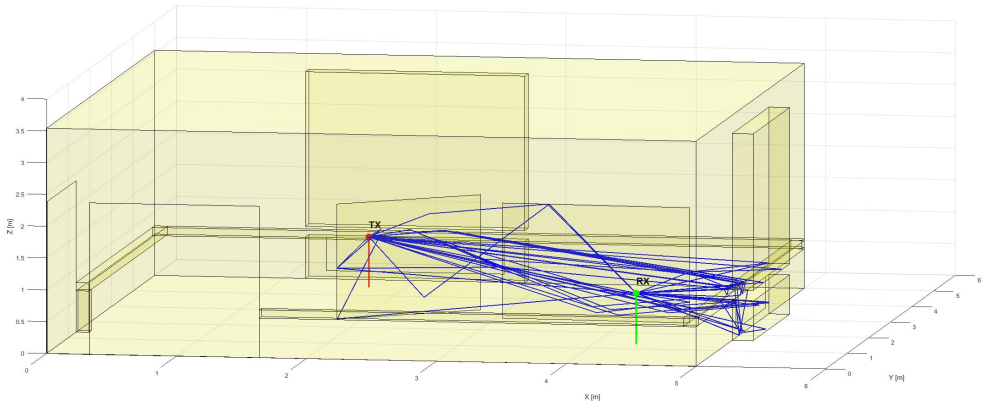


Figure 4.9: *Example of a ray tracing simulation inside the room with the absorber. The maximum delay of the ray drawn is* $0.02\,\mu s$.

4.2 Procedure for the Ray Tracing attack assessment

The procedure chosen to evaluate the effect of the Ray Tracing attack is schematized in Fig. 4.10. First, all the simulations and the measurements have been performed: the output of the ray tracing is a text based file with all the information about the rays, while the measurements are saved as `.csv` files. In addition, both the output power of the VNA and the simulated trasmitted power were set to $0\,\mathrm{dBm}$. Then, the files are read by a routine written in `python`. First, the CIR from the simulation is extracted from the rays, then the CTF is computed and transformed into the *Power Spectral Density*, thus converted in dB-unit and normalized to $0\,\mathrm{dB}$. The CTF is computed through the FFT as described in the section 3.1.1, with two supplements:

- First, the rays with a power lower than the noise level of the VNA (equal to $-140\,\mathrm{dBm}$) have been cut from the analysis.

- Second, the FFT is zero padded in order for the CTF to have the same samples as the VNA sweep, 631 samples.

Moreover, the CIR are stored also in dB-unit and normalized to $0\,\mathrm{dB}$. Thereafter, the `.csv` files of the VNA are read and the PSD, already in dB-unit is extracted and normalized, then the CIR is computed with an inverse Fourier Transform. These steps are executed for all the positions, for all the central frequencies and bandwidth, but the `python` routine has to be separately run for each environment (empty and furnished room, in this case). Eventually, all the data are saved inside the `python` workspace ready to be used for the processing phase.

After all the channel realizations are collected in the workspace, it is possible to proceed to the processing and the evaluation in the different channel conditions, like the central frequency and the bandwidth. For each combination of the channel

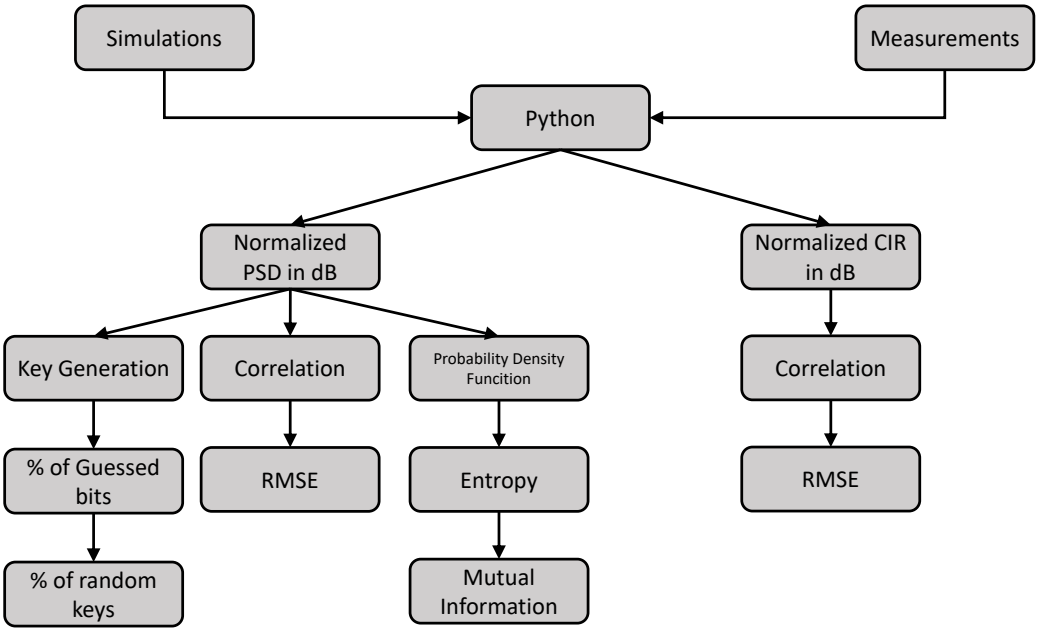


Figure 4.10: *Scheme of the procedure to evaluate the effects of the Ray Tracing attack*

parameter the following quantities have been computed:

- Average correlation (see section 4.2.2.3) and Root Mean Square Error (see section 4.2.2.4) between measured and simulated channel.
- Entropy of both the channels (see section 4.2.2.1).
- Mutual Information between channels (see section 4.2.2.5).

Moreover, for each position other metrics have been used and then averaged for each link condition: correlation (see section 4.2.2.3), root mean square error (see section 4.2.2.4. Then, a key generation method using the Filter-bank model is applied to evaluate the difference between the key generated from the measured channel and the key from the simulated channel. Furthermore, some randomness tests are included to evaluate the usability of the keys generated.

The comparison and the key generation leverage the CTF of the channel. In addition, an analysis in the time domain has also been carried out, similarly to what proposed in [29]. Therefore, an average correlation and the Root Mean Square Error is computed for the different channel condition to evaluate the similarity between the channel in time. In order to evaluate the CIR similarity, an additional step has been employed: once the CIR is normalized, everything falls below $-50\,\mathrm{dB}$ is raised to $-50\,\mathrm{dB}$. This step has been included in order for the simulated CIR to have the same dynamics.

4.2.1 Measurements and simulations parameters

In the following, the measurements and simulations conditions will be described. Then, a better view on the evaluation metric will be provided.

4.2.1.1 Channel parameters

There are two parameters to be changed for evaluating the effect of the Ray Tracing Attack:

- **Central frequency**
- **Channel Bandwidth**, or equivalently the sampling time of the CIR (see section 3.1.1)

Measurements has been carried out on three **central frequencies**: $2.4\,\mathrm{GHz}$, $3.7\,\mathrm{GHz}$ and $5.5\,\mathrm{GHz}$. $2.4\,\mathrm{GHz}$ and $5.5\,\mathrm{GHz}$ are the two central frequencies utilized by IEEE 802.11 standards (WiFi), so they are of particular interest due to the wide presence of devices for personal communications. Moreover, they are unlicensed spectrum, so they are suitable for testing new devices and technologies. Moreover, there are a lot of devices using this band for communication. However, in these bands there could be some interferences, in particular in the first band which is the most overcrowded. Therefore, the $3.7\,\mathrm{GHz}$ band has been utilized, since the BI have a license to use that band for some tests.

As for the **channel bandwidth**, two values have been chosen: $160\,\mathrm{MHz}$ ($T_s = 6.25\,\mathrm{ns}$), which is the maximum allowed bandwidth for the WiFi standard, and $500\,\mathrm{MHz}$ ($T_s = 2\,\mathrm{ns}$). A system using $500\,\mathrm{MHz}$ is already considered an *Ultra Wideband* (UWB).

4.2.1.2 Position of the measurements

A total of 24 positions have been identified to measure the channel, subdivided in three cluster as shown in Fig. 4.11: blue, red and green. Each cluster is composed by 8 receiver spaced $30\,\mathrm{cm}$ one from the other.

4.2.1.3 Ray Tracing parameter

The Ray Tracing tool and the type of parameter for the simulation have been introduced in section 3.1. In particular, it is possible to modify the maximum number of interaction that a ray can experience and some characteristics of the interaction. The following values have been set:

- `N_MAX_INTERRACTIONS`: maximum number of interactions for a ray, considering reflection, diffraction and scattering altogether. Value set to 7.

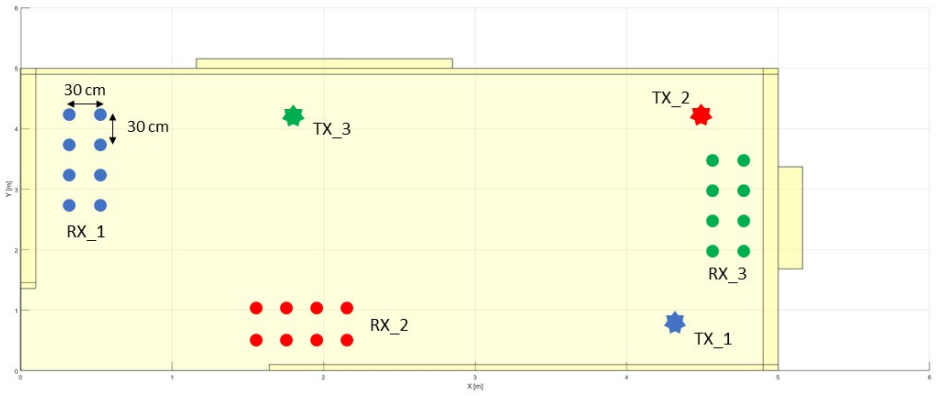


Figure 4.11: *Scheme of the approximate positions of the measurements. The positions are depicted into the empty room, but they are the same also for the room with the absorber. The dots are the receivers and the stars are the transmitters.*

- `N_MAX_REFL`: maximum number of reflections allowed. Value set to 5.
- `N_MAX_DIFF`: maximum number of diffractions allowed. Value set to 3.
- `N_MAX_REFL_AND_DIFF`: maximum number of reflections plus diffractions allowed. The singular limitations still hold. Value set to 5.
- `N_MAX_TRASM`: maximum number of transmissions through the objects. This value does not impact too much on the computational time. Therefore, it has been set to 50.
- `EDGE_LENGTH_MIN`: minimum length in wavelength that an edge must have in order to be considered for the interactions. Value set to 1.
- `WALL_AREA_MIN`: minimum area in squared wavelength that a surface must have in order to be considered for the interactions. Value set to 1.

The number of interactions has been kept low in order to reduce the simulation time. Although having an higher number of interactions may seems to increase the precision of the simulation, actually the rays experiencing an high number of interaction arrives at the receiver with such a low power that they are masked by thermal noise. Furthermore, the scattering has not been utilized in the simulation so far, thus limiting the interactions only to reflections and diffractions.

The ray tracing simulate the CIR, which is then transformed with the Fourier Transform, implemented through the Fast Fourier Transform (FFT) present in the `scipy` package of `python` [40] to obtain the CTF and thus the PSD.

4.2.1.4 Physical Layer based-Key generation

The evaluation of the possible success of a RT attack passes through the comparison of the keys generate from the channel and from the simulations, in order to see if an attacker can actually steal the key or how many bits the attacker is able to guess. To generate the keys, the filter-bank model (see section 2.5) has been utilized with a multi level quantization. To keep things simple, only a static allocation of the filters and of the quantization level has been envisaged. Therefore, from a single channel a fixed length key has been generated. The number of **filters** has been set to 64 and the number of **quantization levels** has been set to 16, leading to keys of $256\,\mathrm{bit}$. The quantization levels are encoded with the **Grey Coding**. The key generation process has been applied both to the PSD in dB-unit and in linear scale, to evaluate also the difference between the two features.

4.2.2 Evaluation metrics

The following metrics are used to evaluate the possibility of a Ray Tracing attack, a summary can be found in 4.2.

4.2.2.1 Entropy

Entropy is a quantity not related to the attack itself, but rather quantify the quality of the channel for the key generation. In this case, the entropy must be intended as the maximum number of secret bits that can be extracted from the channel. In addition, for the sake of this project the **differential entropy** has been employed. If $f(x)$ is the probability density function (PDF) of the random variable X, then the differential entropy is defined as:

$$\boldsymbol{h} = -\int_{-\infty}^{+\infty} f(x)\,\log_2(f(x))\,dx \tag{4.1}$$

In this case, the entropy is measured in bits. The entropy has been computed for both the measured and simulated channel for a pair of channel bandwidth and central frequency and it is computed on the PSD of the channel, since it is the feature used to extract the key. The computation procedure starts with the extraction of the PDF. Considering a specific pair of central frequency and bandwidth, each PSD is an array of 631 values: all the PSD array corresponding to the different positions (for the same frequency, bandwidth and environment) are appended to form a single big-array. A visual representation to better understand this step is shown in Fig. 4.12. Then, the from the array an histogram is built: data are subdivided into 60 bins spanning from $-40\,\mathrm{dB}$ to $0\,\mathrm{dB}$. The histogram is thus normalized to obtain a

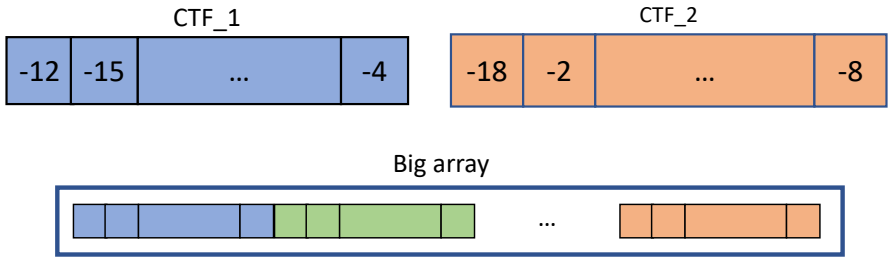
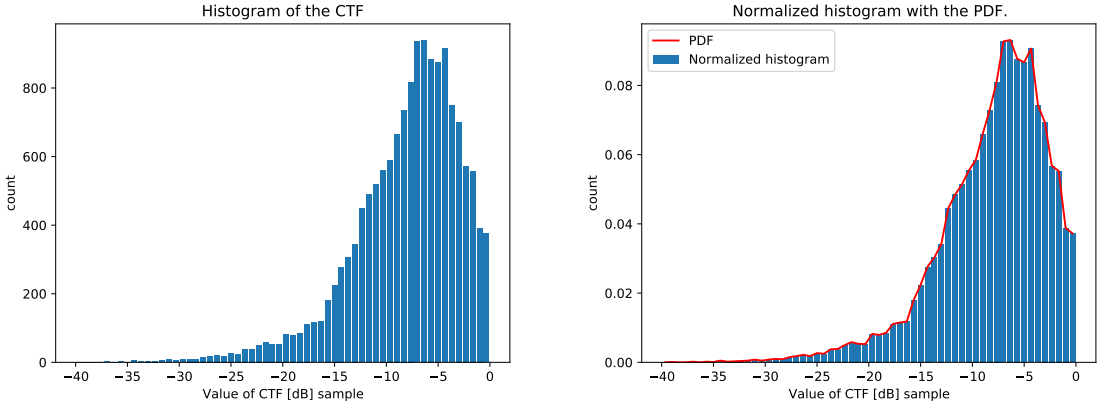


Figure 4.12: *Visual representation of the array built with all the PSD*



(a) *Example of an histogram from the CTF.* ***(b)*** *The normalized histogram that leads to a PDF.*

Figure 4.13: *Example of an histogram and PDF.*

PDF. To perform this computation the function `numpy.histogram` has been used [41]. An example is reported in Fig. 4.13. Once the PDF is computed, it is sufficient to integrate it and obtain the entropy.

This is a rough way to compute the entropy, since the CTFs samples are correlated, at least within the coherence bandwidth. This method does not take into account the correlation among the samples, which provide an imprecise estimate of the entropy. Moreover, this value should be intended as the theoretical maximum number of bits that can be extracted from a filter, provided that the output of the filters are not correlated. Despite the limitations of this method, it was the best that has been found. Another possibility could have been to use a Non Parametric Entropy Estimator Toolbox for `python` [42], but it would have required a lot more realizations of the channels in order to be reliable (in the order of thousands of measurements),

which was prohibitive due to time limitation.

4.2.2.2 Percentage of random keys

In section 2.4.5.6 there is an explanation of the tests that can be used to evaluate the randomness of the key. It is important to evaluate the randomness of a key, since as a requirement for the cryptographic algorithm the key must be random. For the project purposes, 6 tests have been chosen from the NIST suite (the tests returns a `true` or `false` value): frequency monobit test, frequency test within a block, runs test, longest run in a block, serial test, cumulative sums test. All the test are applied on each key and in case the key passes *all* the 6 test, then it can be considered as random. Thereafter, for each channel condition the percentage of keys that are considered random is computed: this allows also to evaluate the impact of the channel parameter on the strength of the key generation method. In addition, the tests have been performed on the keys extracted from the channel in linear scale and in the logarithmic scale.

4.2.2.3 Pearson coefficient

In each point of measure, the Pearson correlation coefficient can be computed between the CTF of the measured channel and of the simulated channel. The Pearson coefficient is defined as:

$$r = \frac{\mathbb{E}[(H^{ab}_{meas}(f) - \mu_{meas})(H^{ab}_{RT}(f) - \mu_{RT})]}{\sigma_{meas}\sigma_{RT}} \tag{4.2}$$

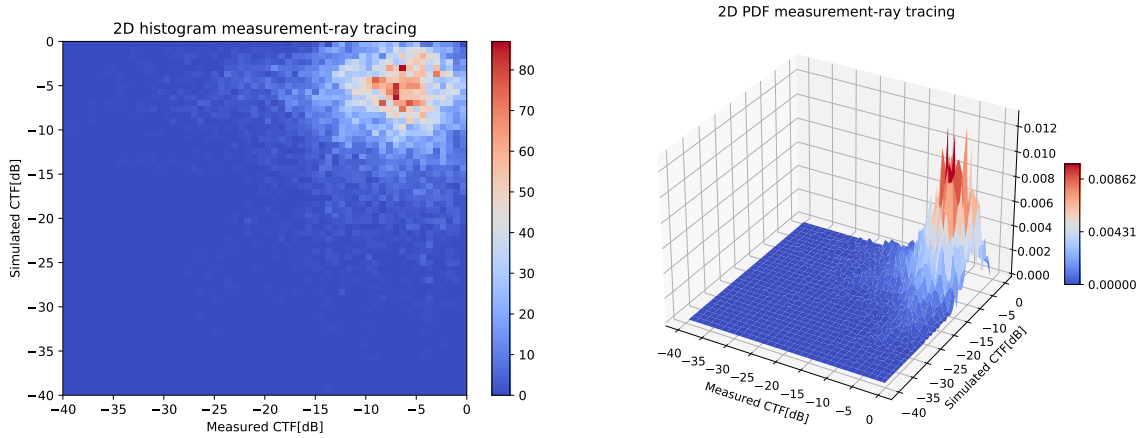
Since it is not so interesting to read a lot of correlation values, the evaluation is based on the average of the correlation values over all the points for a combination of central frequency and bandwidth. In addition, it is worth reminding that the correlation is not always a good parameter to evaluate the similarity of two distributions. In fact, the correlation quantifies the similarity of the trends of the distributions.

4.2.2.4 Root mean square error (RMSE)

Given the two digitized CTF of the measurements and of the ray tracing, both with the same number of samples N_s the RMSE is computed as:

$$\text{RMSE} = \sqrt{\frac{1}{N_s}\sum_{n=0}^{N_s-1}(H^{ab}_{meas}[n] - H^{ab}_{RT}[n])^2} \tag{4.3}$$

RMSE quantifies the differences between two quantities, by making an average of the root of the square of the difference point by point: RMSE is a better metric than the Pearson coefficient to evaluate the difference between quantities like CTFs. RMSE has been chosen to quantify how much the simulated and the measured channel are similar. Even in this case, the RMSE for a channel condition are averaged to obtain a single value for the evaluation.



(a) *Example of an 2D histogram from the CTF.* ***(b)*** *The joint PDF extracted from the histogram*

Figure 4.14: *Example of an 2D histogram and the joint PDF.*

4.2.2.5 Mutual information

The *mutual information* is a measure of the statistical dependence between two random phenomena. Given X and Y, the mutual information $\mathbb{I}(X;Y)$ quantifies the amount of information that is possible to obtain about X by knowing Y and vice versa. In the context of this work, the mutual information quantify the theoretical number of bits that an attacker can infere with a Ray Tracing attack. To compute the mutual information, the following properties is exploited:

$$\mathbb{I}(X;Y) = \boldsymbol{h}(X) + \boldsymbol{h}(Y) - \boldsymbol{h}(X,Y) \tag{4.4}$$

$\boldsymbol{h}(X,Y)$ is the joint entropy, that is defined as $\boldsymbol{h}(X,Y) = -\mathbb{E}_{f(x,y)}[\log_2 f(x,y)]$, where $f(x,y)$ is the joint PDF. X and Y represent the measured channel and the simulated channel. To compute the joint entropy the same procedure as the 1-dimension entropy has been applied, but now with a 2-dimensional histogram, as shown in Fig. 4.14. For this task, still a function of the `numpy` package has been employed, the function `numpy.histogram2d` [43]. The mutual information is an important parameter that is able to measure the effectiveness of the attack, since it works by considering the statistics of the channel and how much the simulated and measured channel are similar. Moreover, consider the equation 2.14 in section 2.4.5.4: if X^A is the channel measured and X^E is the simulated channel, which is the channel that Eve sees through the simulation, the mutual information is the term that lower the Secrecy Key Rate. Therefore, the mutual information can be considered a good metric to evaluate the Ray Tracing attack. A small value of the mutual information is thus desirable.

4.2.2.6 Percentage of guessed bits

Once the measurements and the simulations are performed, it is possible to extract the keys from the channel and compare them in order to see if the key from the simulation (K_{RT}) is similar or not to the key from the actual channel (K_{meas}). To quantify the similarity between the two keys it is sufficient to compute the percentage of equal bits between the two key, which is the opposite of computing the Key Disagreement Rate (see section 2.4.5.3):

$$\%\text{guessed bits } = \frac{1}{N_b} \sum_{k=0}^{N_b-1} K_{RT} \odot K_{meas} \tag{4.5}$$

where $\odot$ is the `XNOR` operation (returns 1 if the two inputs are equal). The optimal value is 50%, since it indicates that the attacker is generating a random sequence, which is equal to say that the mutual information between the two keys is equal to 0. Contrary to a first view, a percentage equal to 0% is not as good as one can think: it means that all bits are wrong, but since the alphabet is binary it is sufficient to invert all the bits to obtain the true key. Moreover, since the errors can be corrected by the attacker by eavesdropping the information reconciliation message from Alice, one can set an upper and lower bound to the guessed bits, by considering the reconciliation capability: using 25% as maximum reconciliation capability (see section 2.4.3), the percentage of guessed bits should be less than 75%. However, since having 0% or 100% of guessed bits is the same, the percentage must be also greater than 25%, since the attacker can easily invert all the bits of the key. In the end, $25\% \leq$ %guessed bits $\leq 75\%$. The percentage has been computed with regard to the key from the channel in dB and from the channel in linear units.

4.2.3 Metrics for the time domain evaluation

In addition to the frequency domain evaluation, a quick assessment has been performed also in the time domain. The CIRs have been considered in dB-unit, normalized to $0\,\mathrm{dB}$. The quantities utilized are:

- Correlation: which is the metric utilized to evaluate the effects of the attack in [29].
- Root Mean Square Error.

For the evaluation the CIRs are processed considering an additional maximum dynamic: starting from the normalized The time domain evaluation was not the core of the project, it has been inserted just for completeness and to replicate a similar work in the literature [29]. Therefore, it has not been deepened as the frequency domain, but this might be the starting point for future work in particular for the UWB key generation.

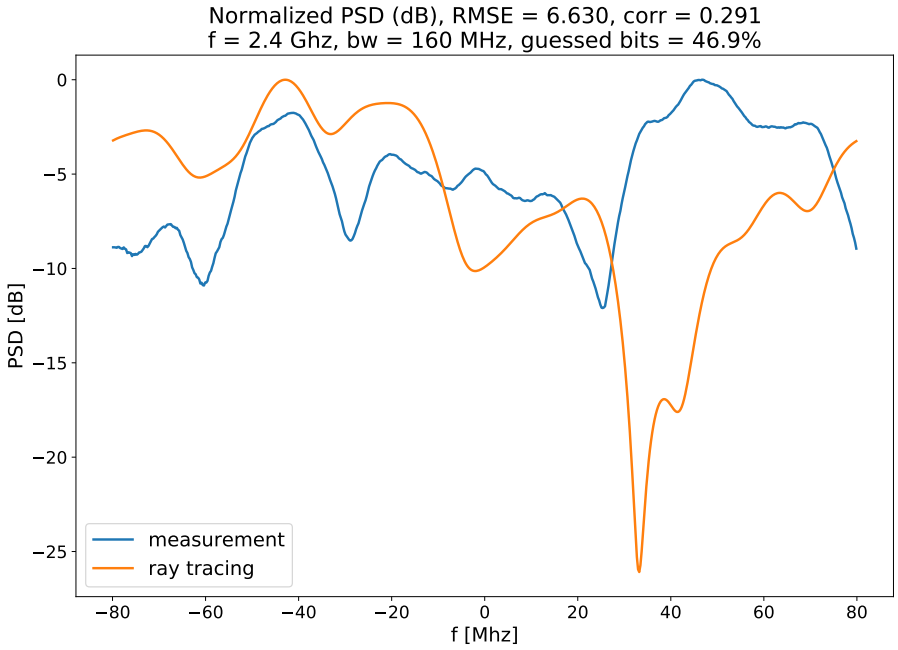
Channel Parameter	Central frequencies [GHz]	2.4, 3.7, 5.5
	Channel bandwidth [MHz]	160, 500
Ray Tracing	Max. number of interactions	7
	Max. reflections	5
	Max. diffractions	3
	Max. diffractions + reflections	5
	Max. transmissions	50
	Min. length of edge (in wavelength)	1
	Min. wall area (in squared wavelength)	1
	FFT coefficients in output	631
	Noise Dynamic	50 dB
Statistic computations	Number of histogram's bins	60
	Range of histogram	-40 to 0 dB
Key generation (Filter-bank)	Number of filters	64
	Number of levels	16
	Key bits	256

Table 4.2: *Summary of the parameter for the assessment*

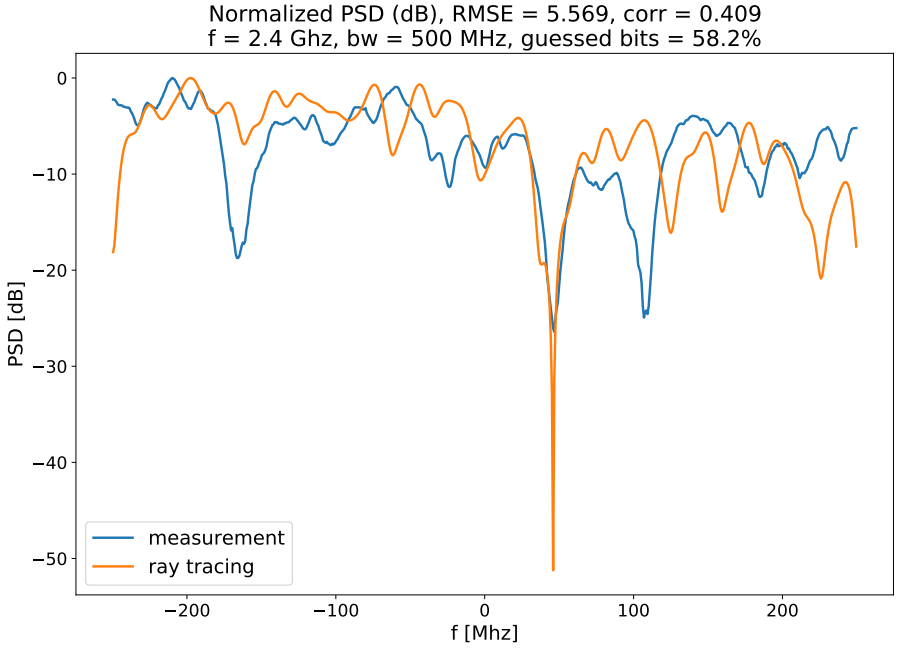
4.3 Results for the empty room

As for the propagation characteristics, the *empty room* discloses a strong LOS component. Moreover, since the antennas are placed at $0.80\,\mathrm{m}$ from ground, they are are aligned to the metal cable ducts, thus presenting a lot of multipath due the interaction with the metal duct.

Let's start by discussing some Power Spectral Densities. All the channels that are reported, both here and in the next section, are extracted from the same position but at different frequencies and bandwidth (Fig. 4.15, Fig. 4.16 and Fig. 4.17) and there is the comparison between the measured and simulated channel. With regard to Fig. 4.15, at a first sight the channels seems to be quite similar, in particular at the higher bandwidth, but looking at the correlation and the RMSE they indicate that the channel are not very similar. Moreover, looking at the guessed bits, the keys generated are quite different: guessing the 46.9% and 58.2% of the bits indicate that the keys are far from being equal, so the ray tracing is not really capable of guessing the key.

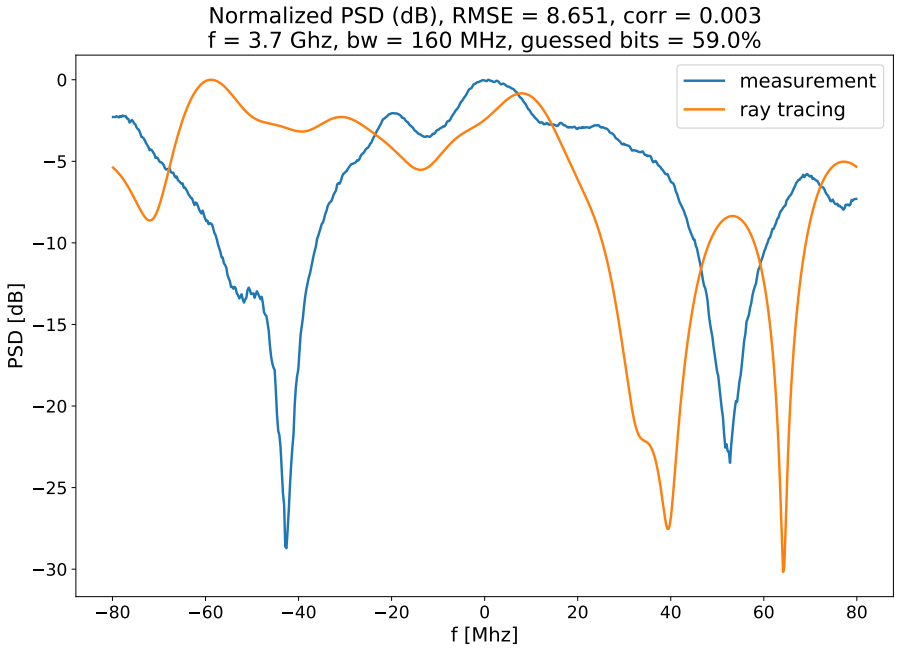


(a) *Power Spectral Density (dB) at* 2.4 GHz *and a bandwidth of* 160 MHz

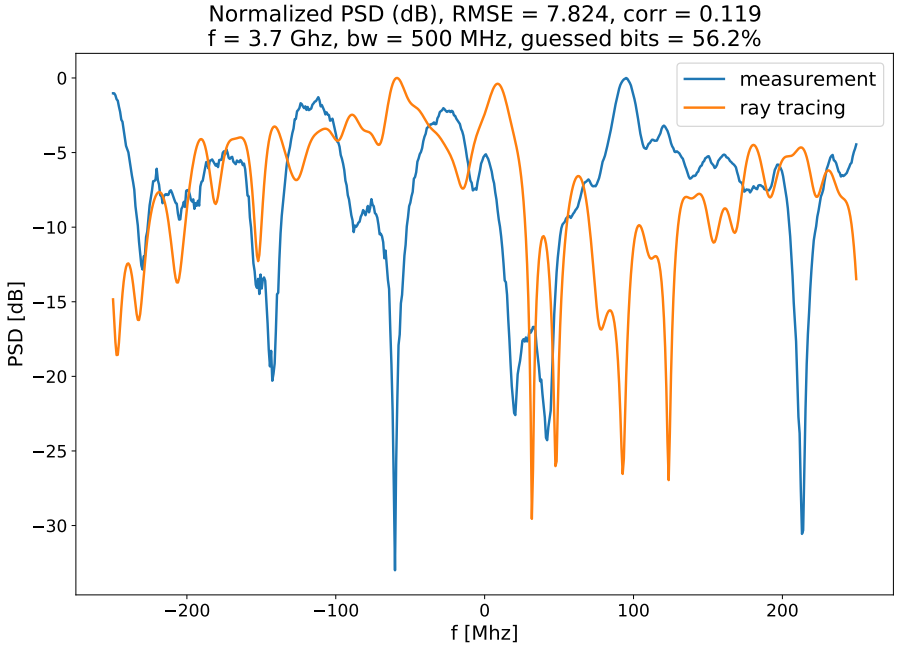


(b) *Power Spectral Density (dB)at* 2.4 GHz *and a bandwidth of* 500 MHz

Figure 4.15: *Examples of Power Spectral Density (dB) in the* ***empty room****, comparison between the measured and the simulated channel, with also the value of the correlation coefficient and the Root Mean Square Error.*

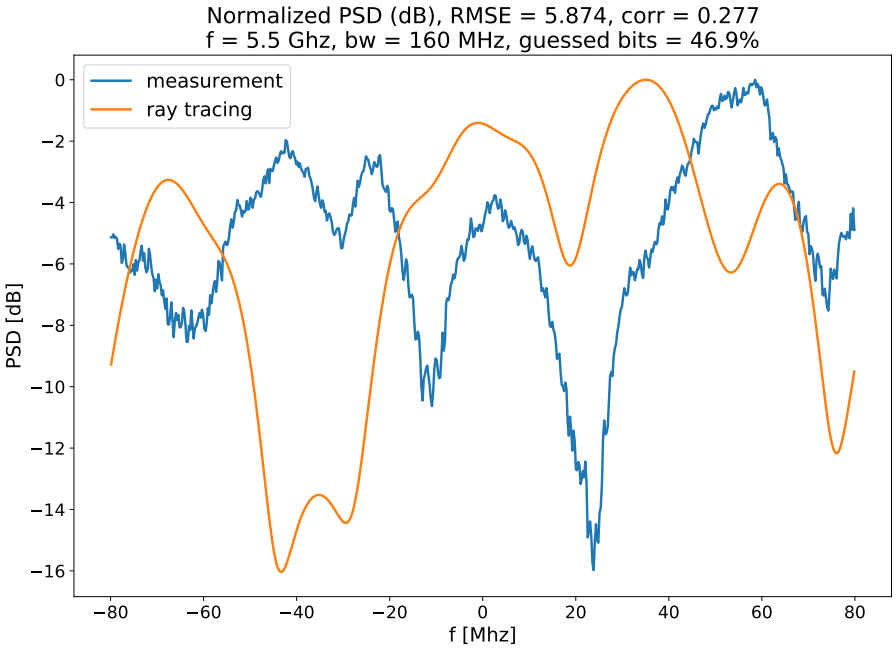


(a) *Power Spectral Density (dB) at* 3.7 GHz *and a bandwidth of* 160 MHz

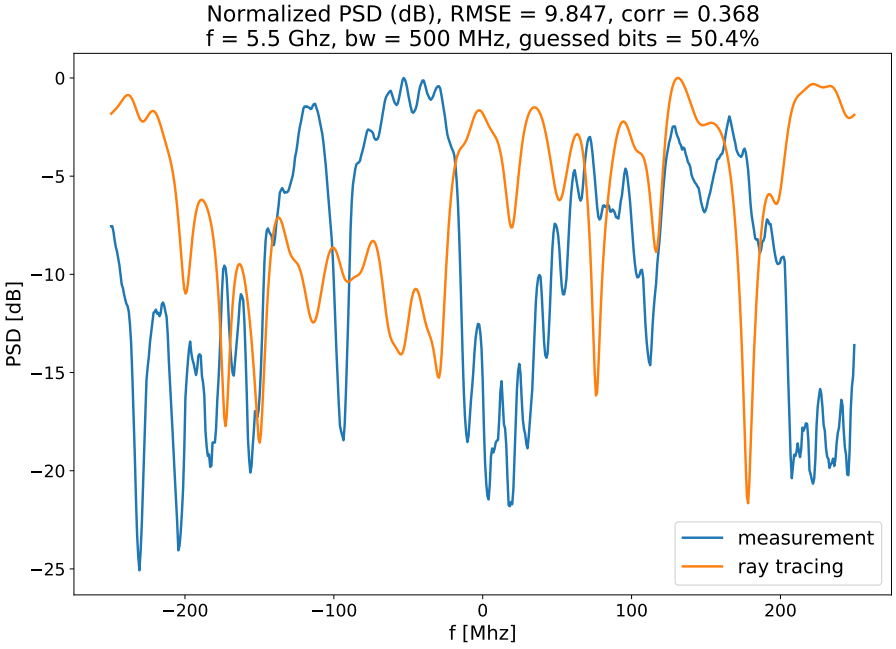


(b) *Power Spectral Density (dB) at* 3.7 GHz *and a bandwidth of* 500 MHz

Figure 4.16: *Examples of Power Spectral Density (dB) in the* ***empty room****, comparison between the measured and the simulated channel, with also the value of the correlation coefficient and the Root Mean Square Error.*

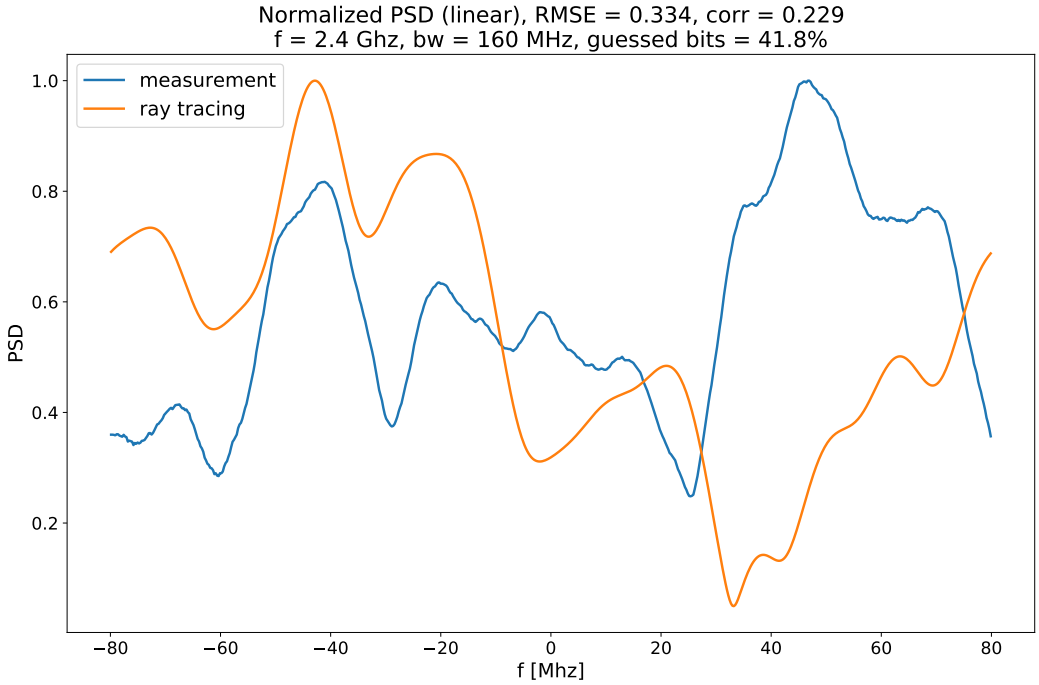


(a) *Power Spectral Density (dB) at* 5.5 GHz *and a bandwidth of* 160 MHz

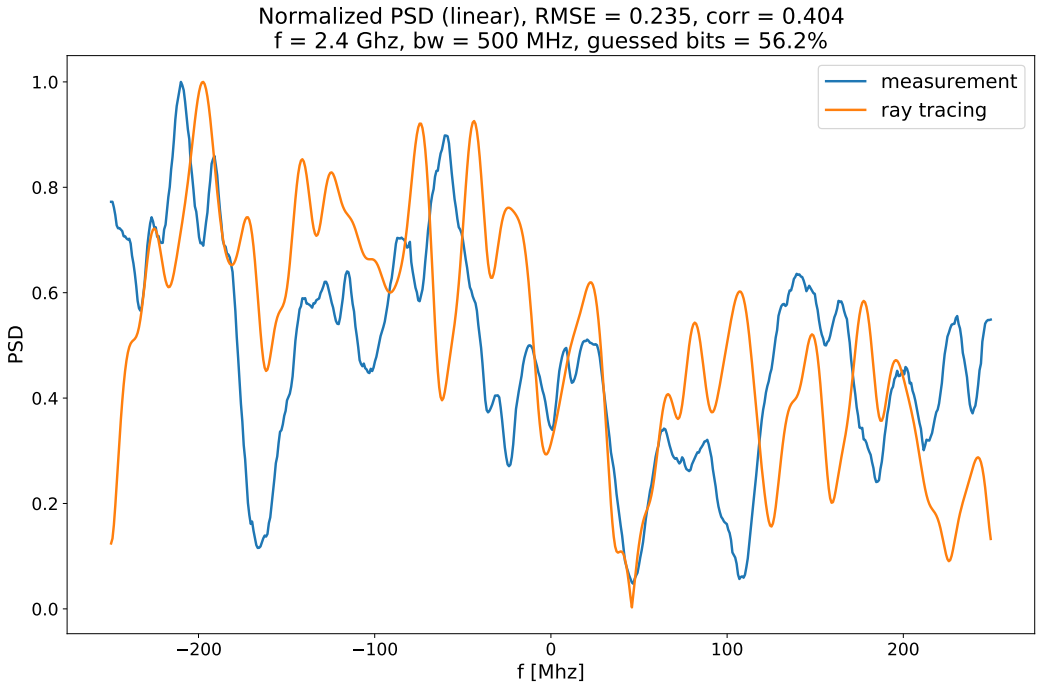


(b) *Power Spectral Density (dB) at* 5.5 GHz *and a bandwidth of* 500 MHz

Figure 4.17: *Examples of Power Spectral Density (dB) in the* ***empty room****, comparison between the measured and the simulated channel, with also the value of the correlation coefficient and the Root Mean Square Error.*

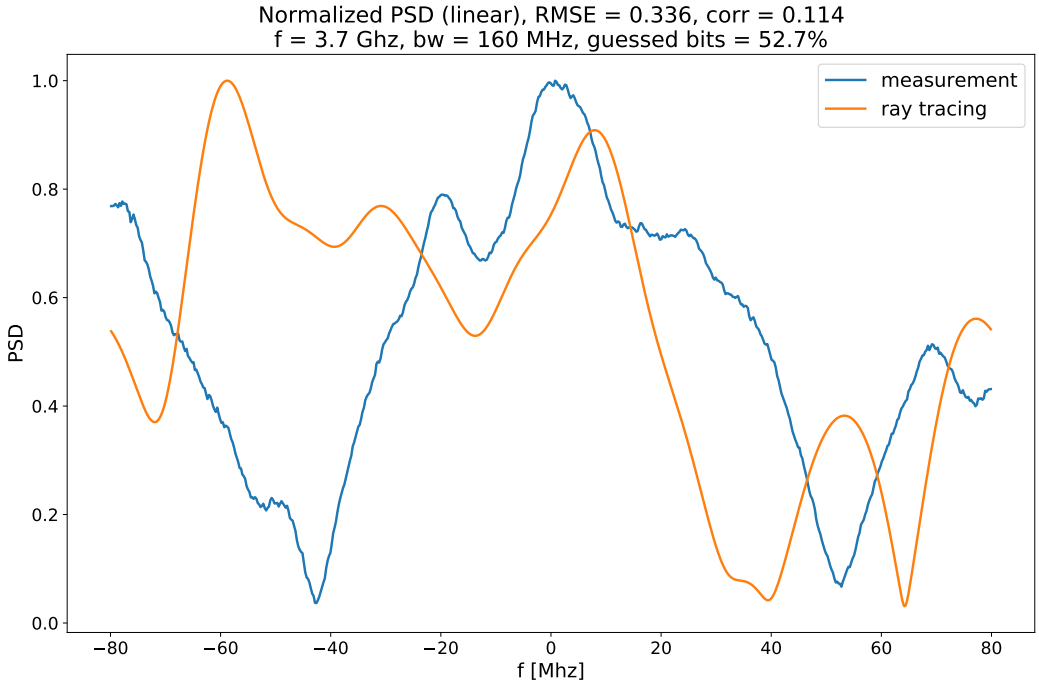


(a) *Power Spectral Density (linear) at* $2.4\,\mathrm{GHz}$ *and a bandwidth of* $160\,\mathrm{MHz}$

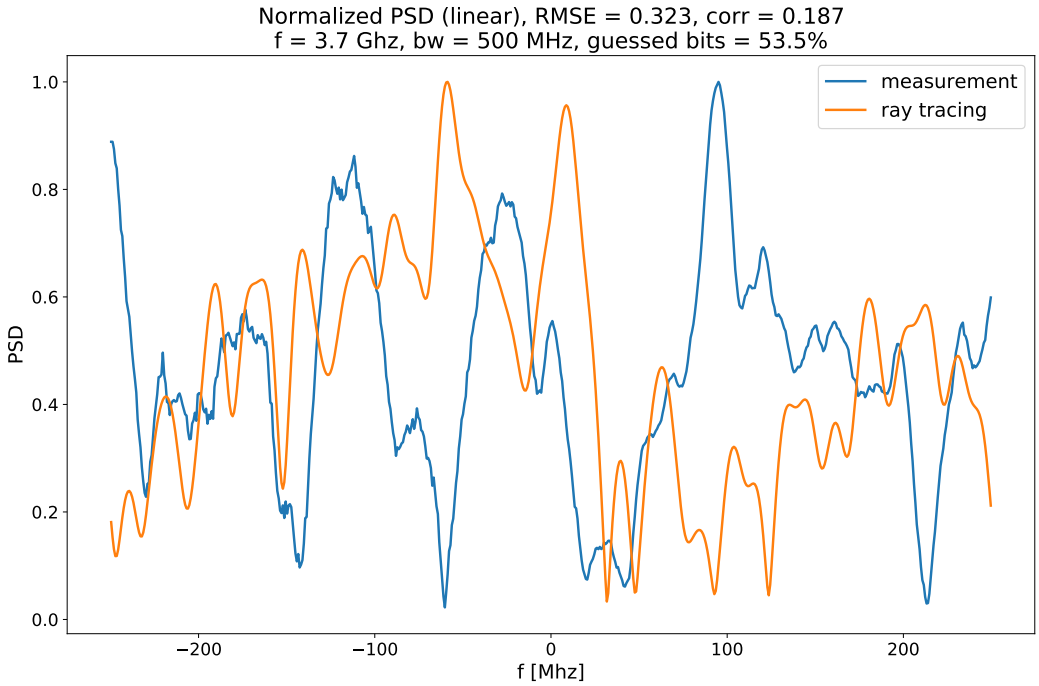


(b) *Power Spectral Density (linear) at* $2.4\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$

Figure 4.18: *Examples of Power Spectral Density (linear) in the* ***empty room****, comparison between the measured and the simulated channel, with also the value of the correlation coefficient and the Root Mean Square Error.*

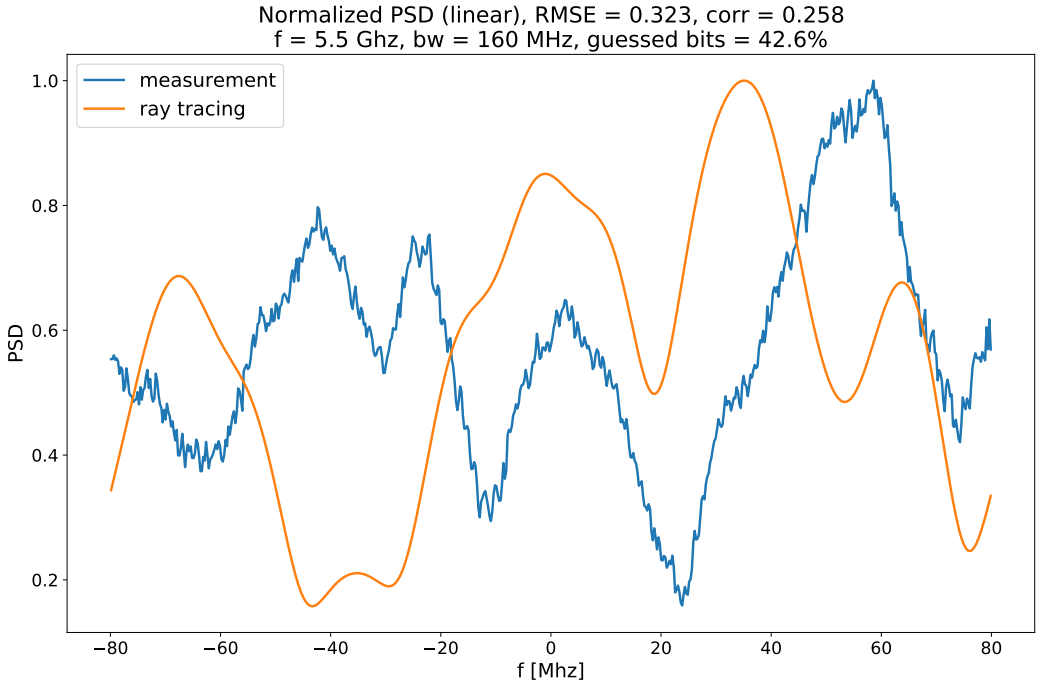


(a) *Power Spectral Density (linear) at* 3.7 GHz *and a bandwidth of* 160 MHz

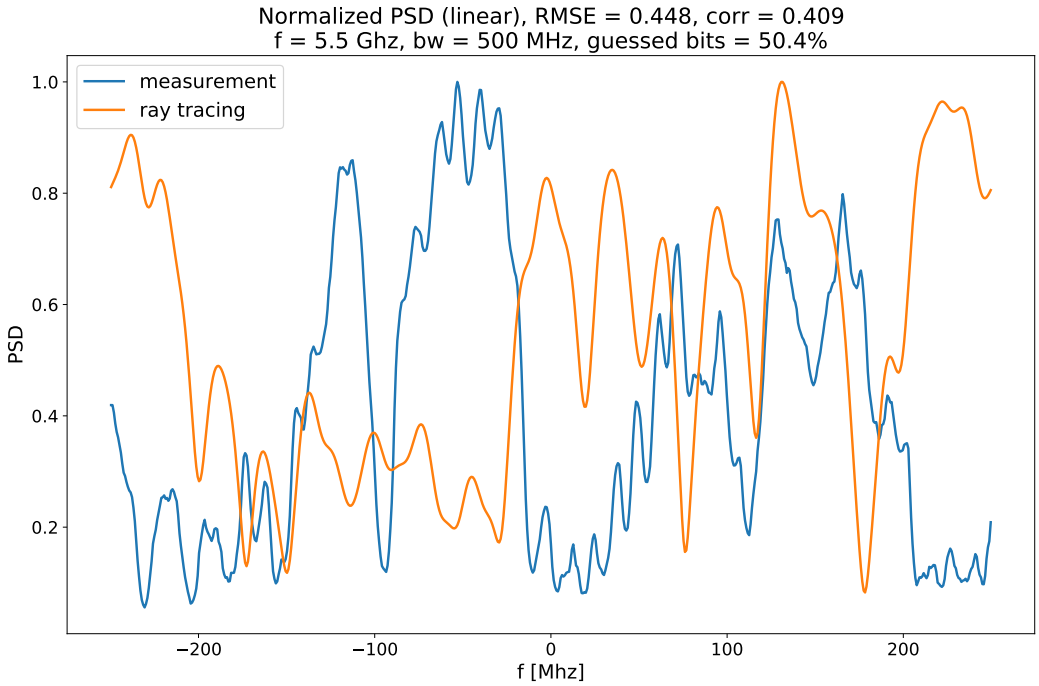


(b) *Power Spectral Density at* 3.7 GHz *and a bandwidth of* 500 MHz

Figure 4.19: *Examples of Power Spectral Density (linear) in the* ***empty room****, comparison between the measured and the simulated channel, with also the value of the correlation coefficient and the Root Mean Square Error.*

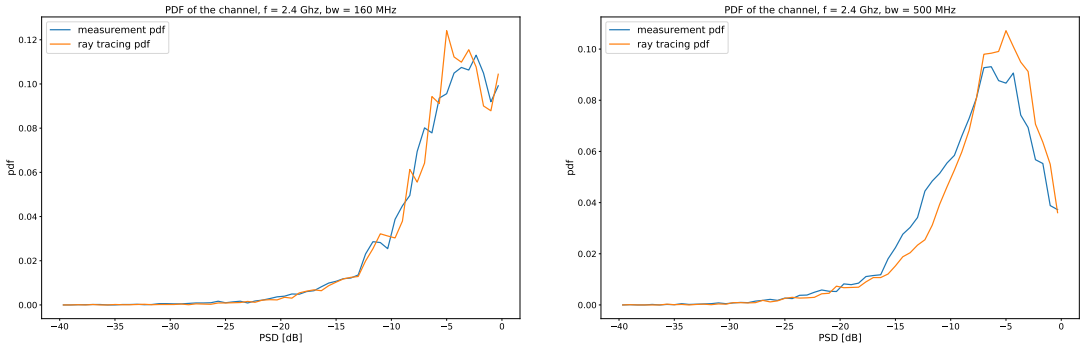


(a) *Power Spectral Density (linear) at* $5.5\,\mathrm{GHz}$ *and a bandwidth of* $160\,\mathrm{MHz}$



(b) *Power Spectral Density (linear) at* $5.5\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$

Figure 4.20: *Examples of Power Spectral Density (linear) in the* ***empty room****, comparison between the measured and the simulated channel, with also the value of the correlation coefficient and the Root Mean Square Error.*



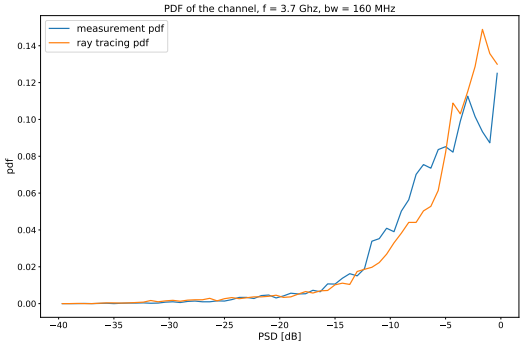
(a) *Probability Density Function of the PSD (dB) at* $2.4\,\mathrm{GHz}$ *and a bandwidth of* $160\,\mathrm{MHz}$

(b) *Probability Density Function of the PSD (dB) at* $2.4\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$

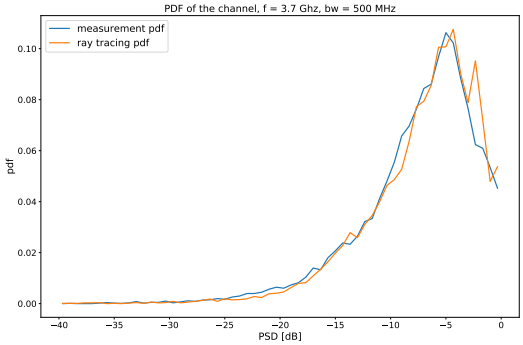
Figure 4.21: *Probability Density Function of the Power Spectral Density in the* ***empty room****, comparison between the measured and the simulated channels.*

As a general consideration, except for Fig. 4.17, the ray tracing seems to be able to follow the trend of the CTF, at least at lower frequency. In fact, looking at the Probability Density Functions of the channels in Fig. 4.21 and Fig. 4.22, the Ray Tracing is able to follow the statistic of the actual channel, which indicates that it is able to simulate a channel with the right propagation characteristic. Moreover, even at $5.5\,\mathrm{GHz}$ in Fig. 4.23 this condition holds, at least in the higher bandwidth. However, due to the imprecisions in the ray's phase computation the position of the fades of the CTF is often wrong. Moreover, an additional imprecision is given by the antenna description: the phase of the emitted field in the different directions is not known, which of course introduce additional errors. For example, in Fig. 4.17b the two CTF are almost opposite, the RMSE is high ($9.847\,\mathrm{dB}$) and indeed the guessed bits are the 50.4%. These considerations hold also for the channels in linear units. Moreover, it is important to notice that if the measurements are repeated, the channel will be always slight different, due to small changes in the environment that cannot be predicted by the ray tracing: the ray tracing instead is a deterministic channel model and will output always the same channel if the environment or the simulation parameters do not change. In the end, even a small change of the environment lead to a different channel, therefore to a different key and following these small variations with the ray tracing is almost impossible.

After seeing some channel comparison with general consideration, it is time to take a look at the assessment results. The values of all the metrics previously introduced are reported in Table 4.3, subdivided by central frequency and channel bandwidth. The entropy of the channel is slightly larger in the higher bandwidth than in the lower bandwidth while it remains more or less constant for the different channel frequency. Moreover, the entropy of simulations follow the same the trend of the measurements, indicating that the ray tracing can be able to at least guess the dynamic of the channel and provide a value of entropy similar to the real one.

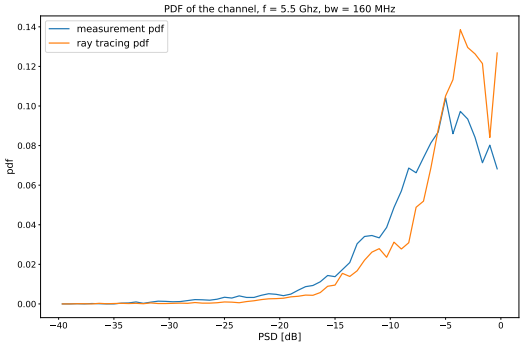


(a) *Probability Density Function of the PSD (dB) at* 3.7 GHz *and a bandwidth of* 160 MHz

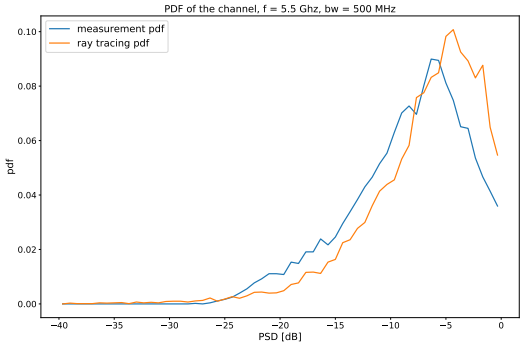


(b) *Probability Density Function of the PSD (dB) at* 3.7 GHz *and a bandwidth of* 500 MHz

Figure 4.22: *Probability Density Function of the Power Spectral Density in the* ***empty room****, comparison between the measured and the simulated channels.*



(a) *Probability Density Function of the PSD (dB) at* 5.5 GHz *and a bandwidth of* 160 MHz



(b) *Probability Density Function of the PSD (dB) at* 5.5 GHz *and a bandwidth of* 500 MHz

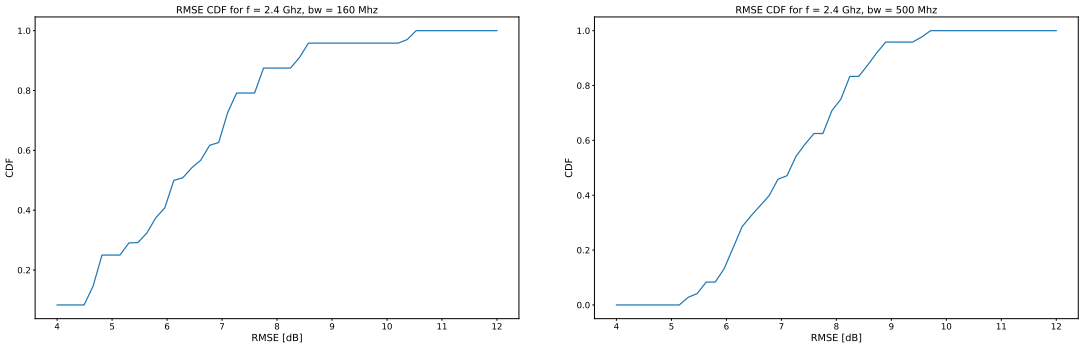
Figure 4.23: *Probability Density Function of the Power Spectral Density in the* ***empty room****, comparison between the measured and the simulated channels.*

Related to the entropy of the channel there is the percentage of random key, only from the measured channel and both for linear and logarithmic units: the keys from the linear channels are more probably random, since for simplicity a uniform quantization was employed, while for logarithmic channel a non-uniform quantization is more advisable. Looking at the percentage for linear unit, with higher bandwidth the percentage is higher than in the lower bandwidth: the key generation method remains the same, with the same number of filters and with $160\,\mathrm{MHz}$ it is very probable that the correlation between the sub-bands is high, which is not the case with the higher bandwidth. However, this trend is not valid for the channel in logarithmic units: the logarithmic scale tends to compress the large variation of the channel. In fact, by looking at the channel (e.g. Fig. 4.15), the PSD in dB-unit are compressed around the smaller values, with some deep fade: therefore, the output of the filters will be quantized with the same bit sequence, which brings to a non random sequence. Eventually, the percentages obtained for the linear channels represent a promising value for the usability of the key generation and with the right attentions (e.g. non-uniform quantization or dynamic allocation of the filters) it might be possible to obtain better performances.

As for the ray tracing attack, the metrics are reported in their average value with the standard deviation. First, the RMSE is quite large, which indicates that the simulated channel is far from being similar to the measured channel. Second, following the same trend the correlation coefficient is quite low: to have a clearer vision of the values under consideration, from Fig. 4.24 to Fig. 4.29 there are the Cumulative Distribution Functions of the Root Mean Square Error and the Pearson Coefficient. Third, the mutual information, which is a very important parameter for the assessment, is quite low and always below $1\,\mathrm{bit}$. As a general trend, RMSE correlation and mutual information is slightly larger in the lower bandwidth than in the higher: this can be due to the flatter channel with respect to higher bandwidth and the lower probability to have a lot of deep fade in the channel, which are difficult to be guessed by the ray tracing. In line with the low mutual information, the percentage of guessed bits remains close to 50%, also with a restrained standard deviation which ensure that the possibility of guessing the key using the ray tracing attack is very low. Furthermore, it is worth reminding that the Ray Tracing simulates a static channel, so in case a system utilizes also the fading in the time domain to generate the key it will be impossible for the Ray Tracing to follow the mobility of all the objects in the environment. In conclusion, the Ray Tracing is not able to predict the fast fading pattern in the environment, even though it is able to guess the statistical distribution of the channel amplitude in frequency.

Frequency [GHz]	2.4		3.7		5.5	
Bandwidth [MHz]	160	500	160	500	160	500
Entropy meas	3.806	4.174	3.872	4.112	4.068	4.241
Entropy RT	3.738	4.038	3.760	4.036	3.655	4.077
% of random key LIN	20.833	58.333	25.000	62.500	25.000	54.167
% of random key DB	29.167	4.167	8.333	4.167	12.500	8.333
RMSE	6.289	7.193	6.728	7.347	6.887	7.586
RMSE std	1.563	1.110	2.058	1.025	1.708	1.165
correlation	0.284	0.186	0.254	0.169	0.264	0.173
std correlation	0.151	0.116	0.167	0.099	0.210	0.124
Mutual Info	0.426	0.224	0.520	0.265	0.441	0.262
% of guessed bits LIN	49.626	50.993	51.400	51.204	51.286	50.749
std % guessed LIN	5.067	3.952	5.781	2.810	5.206	2.263
% of guessed bits DB	54.476	58.936	57.243	58.333	55.713	54.215
std % guessed DB	4.955	4.127	4.574	4.305	6.839	3.555

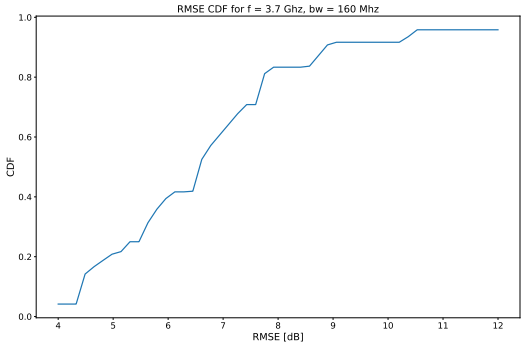
Table 4.3: *Table with the results of the assessment in the* ***empty room*** *in the frequency domain. Results are subdivided by central frequency and channel bandwidth.*



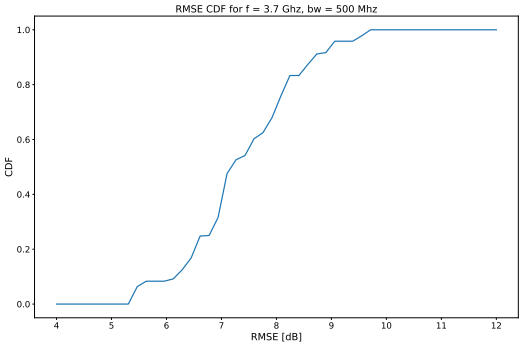
(a) *Cumulative Distribution Function of the Root Mean Square Error at* 2.4 GHz *and a bandwidth of* 160 MHz

(b) *Cumulative Distribution Function of the Root Mean Square Error at* 2.4 GHz *and a bandwidth of* 500 MHz

Figure 4.24: *Cumulative Distribution Function of the Root Mean Square Error in the* ***empty room****.*

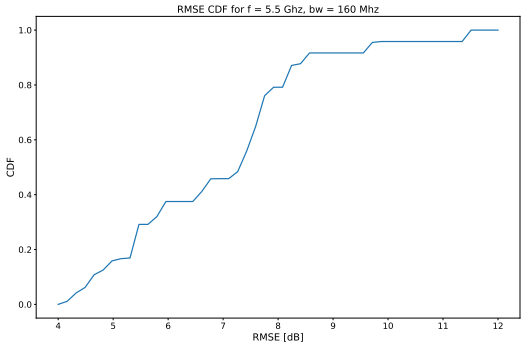


(a) *Cumulative Distribution Function of the Root Mean Square Error at* 3.7 GHz *and a bandwidth of* 160 MHz

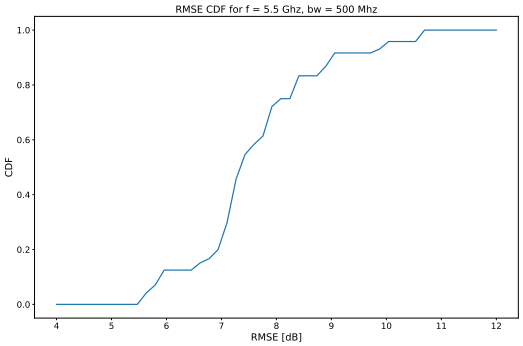


(b) *Cumulative Distribution Function of the Root Mean Square Error at* 3.7 GHz *and a bandwidth of* 500 MHz

Figure 4.25: *Cumulative Distribution Function of the Root Mean Square Error in the* ***empty room****.*

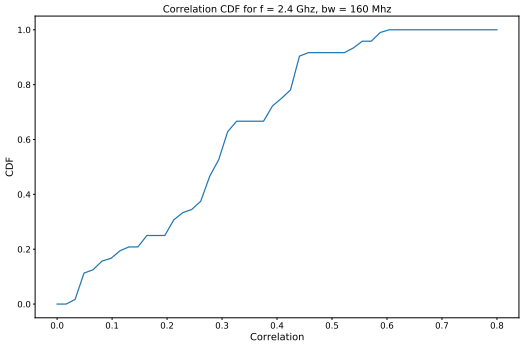


(a) *Cumulative Distribution Function of the Root Mean Square Error at* 5.5 GHz *and a bandwidth of* 160 MHz

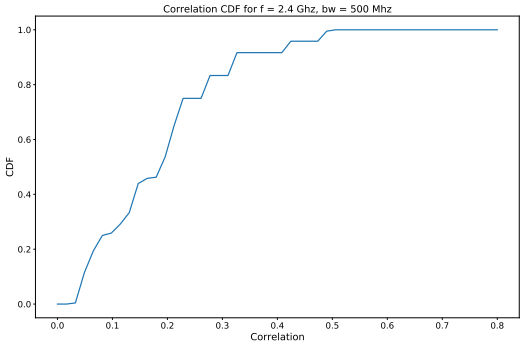


(b) *Cumulative Distribution Function of the Root Mean Square Error at* 5.5 GHz *and a bandwidth of* 500 MHz

Figure 4.26: *Cumulative Distribution Function of the Root Mean Square Error in the* ***empty room****.*

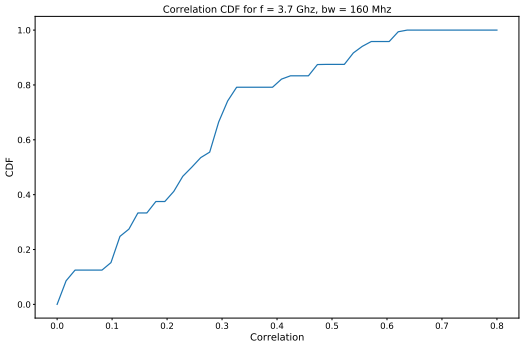


(a) *Cumulative Distribution Function of the Pearson Coefficient at* 2.4 GHz *and a bandwidth of* 160 MHz

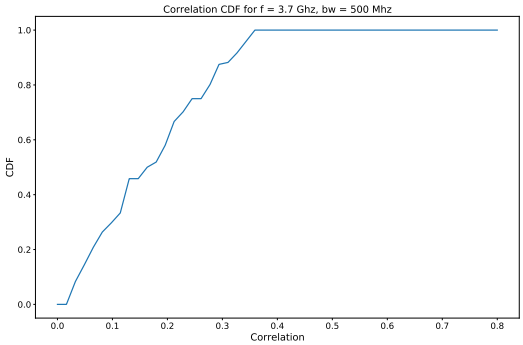


(b) *Cumulative Distribution Function of the Pearson Coefficient at* 2.4 GHz *and a bandwidth of* 500 MHz

Figure 4.27: *Cumulative Distribution Function of the Pearson Coefficient in the* ***empty room****.*

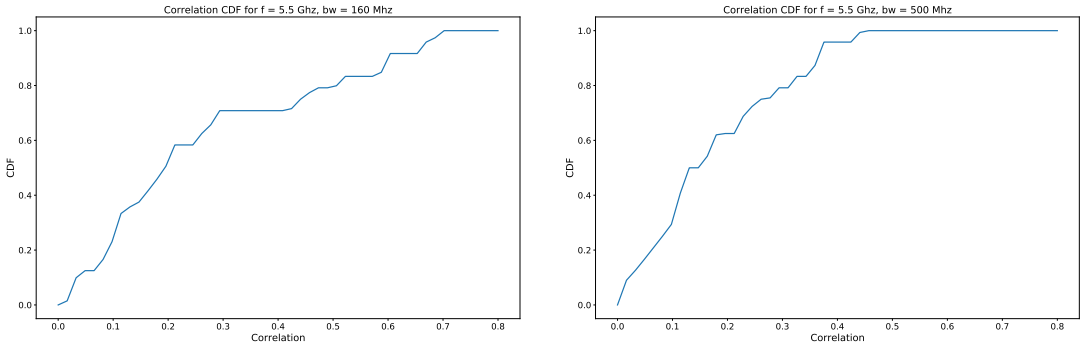


(a) *Cumulative Distribution Function of the Pearson Coefficient at* 3.7 GHz *and a bandwidth of* 160 MHz



(b) *Cumulative Distribution Function of the Pearson Coefficient at* 3.7 GHz *and a bandwidth of* 500 MHz

Figure 4.28: *Cumulative Distribution Function of the Pearson Coefficient in the* ***empty room****.*



(a) *Cumulative Distribution Function of the Pearson Coefficient at* 5.5 GHz *and a bandwidth of* 160 MHz

(b) *Cumulative Distribution Function of the Pearson Coefficient at* 5.5 GHz *and a bandwidth of* 500 MHz

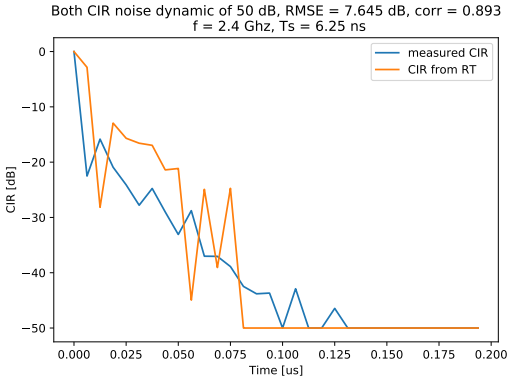
Figure 4.29: *Cumulative Distribution Function of the Pearson Coefficient in the* ***empty room****.*

Frequency [GHz]	2.4		3.7		5.5	
Bandwidth [MHz]	160	500	160	500	160	500
RMSE time	9.034	9.465	8.278	9.626	8.193	10.675
RMSE std time	1.496	0.965	1.954	0.825	1.467	1.164
correlation TIME	0.854	0.782	0.872	0.788	0.882	0.691
std correlation TIME	0.041	0.057	0.059	0.051	0.042	0.062

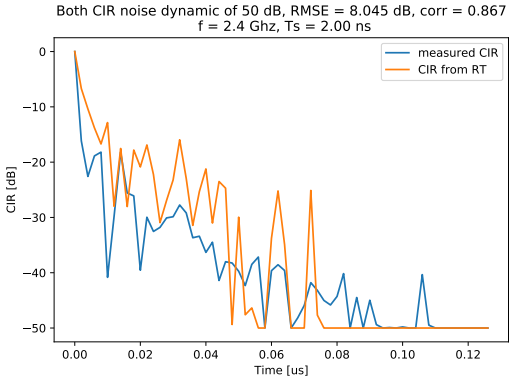
Table 4.4: *Time domain evaluation in the* ***empty room****.*

4.3.1 Time domain

Now, the time evaluation for the empty room will be explained. First, the CIR, correspondent to the previous PSD, are shown in Fig. 4.30 Fig. 4.31 and Fig. 4.32: at a first sight it is possible to notice that the ray tracing is not really able to guess the CIR, sometimes the peaks and the fades correspond or the direction is guessed, but it is not possible to state that the simulated CIR are close to the measured. Moreover, the late components in the measured CIR are not found by the ray tracing. In fact, by looking at Table 4.4, the RMSE is high, even greater than the RMSE for the channel in frequency. Despite the differences, the correlation remains always quite high: this is caused by what the correlation quantify, since it is able to measure the "linear similarity" between two distributions, therefore how similar are the trends between them. In this case, since CIR always follows a decreasing trend, the correlation is larger compared to the frequency domain.

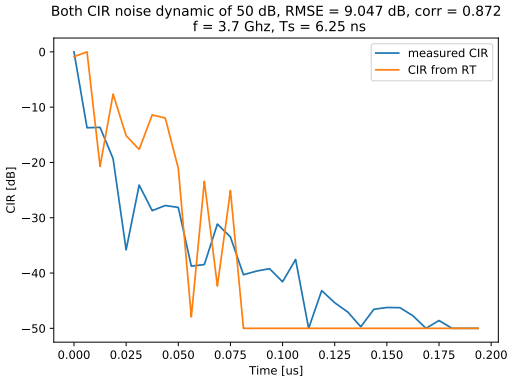


(a) *Channel Impulse Response at* $2.4\,\mathrm{GHz}$ *and a bandwidth of* $160\,\mathrm{MHz}$

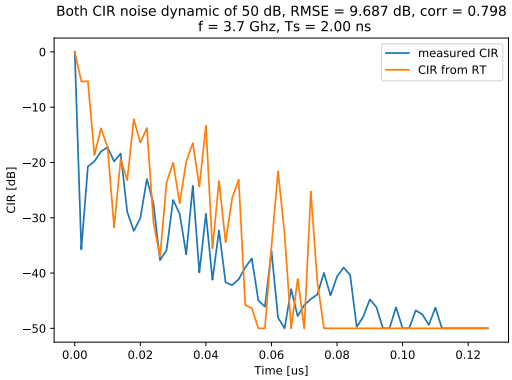


(b) *Channel Impulse Response at* $2.4\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$

Figure 4.30: *Examples of Channel Impulse Response (CIR) in the* ***empty room****, compared considering a noise dynamic of* $50\,\mathrm{dB}$*. In the figures there are also the value of the RMSE and the correlation between the two channels.*

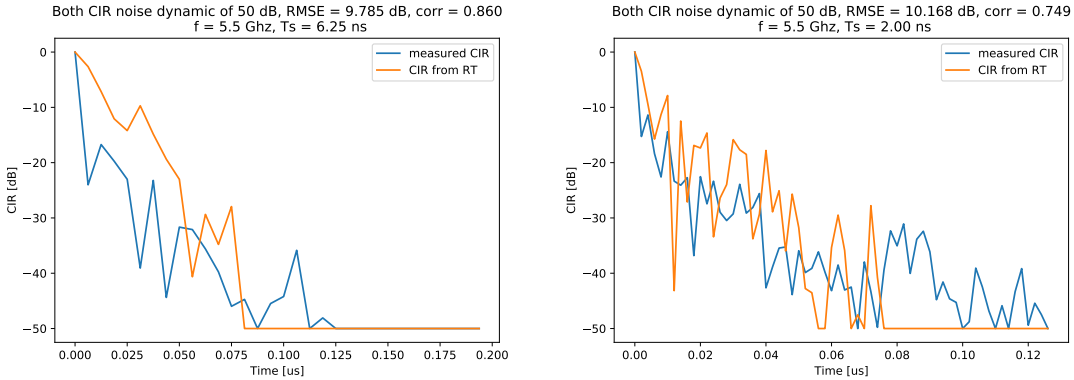


(a) *Channel Impulse Response at* $3.7\,\mathrm{GHz}$ *and a bandwidth of* $160\,\mathrm{MHz}$



(b) *Channel Impulse Response at* $3.7\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$

Figure 4.31: *Examples of Channel Impulse Response (CIR) in the* ***empty room****, compared considering a noise dynamic of* $50\,\mathrm{dB}$*. In the figures there are also the value of the RMSE and the correlation between the two channels.*



(a) *Channel Impulse Response at* $5.5\,\mathrm{GHz}$ *and a bandwidth of* $160\,\mathrm{MHz}$

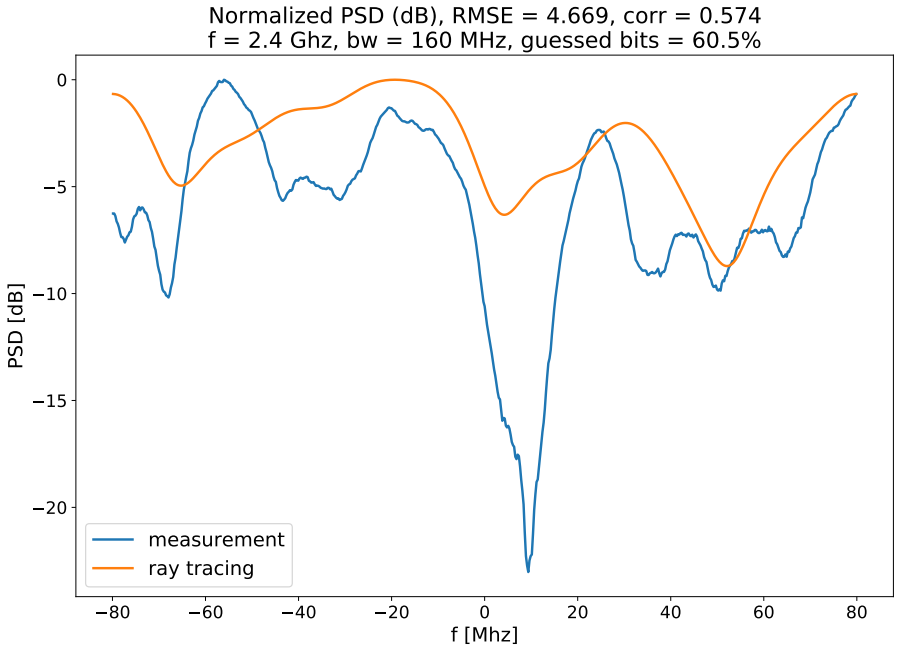
(b) *Channel Impulse Response at* $5.5\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$

Figure 4.32: *Examples of Channel Impulse Response (CIR) in the* ***empty room****, compared considering a noise dynamic of* $50\,\mathrm{dB}$*. In the figures there are also the value of the RMSE and the correlation between the two channels.*

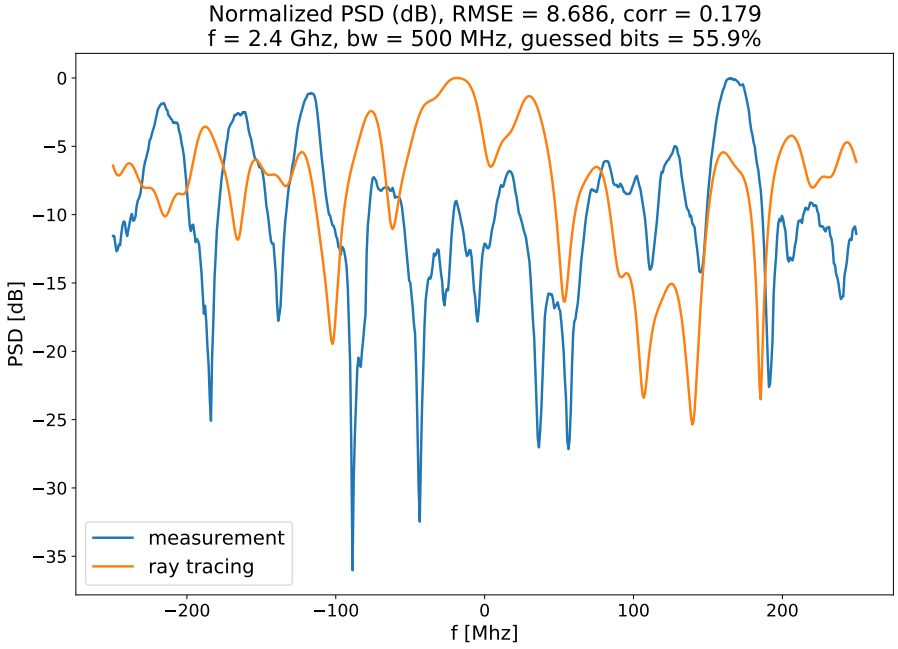
4.4 Results for room with obstacles

Now, the propagation characteristics are different with respect to the empty room, since there are the two absorbers in the middle that block the LOS component, plus they may add diffractions on the edges, whose phase shift is difficult to be predicted.

Looking at the Power Spectral Densities in Fig. 4.33, Fig. 4.34, Fig. 4.35 and the correspondent in linear units in Fig. 4.36, Fig. 4.37 and Fig. 4.38. What first comes into view is the fact that the channel oscillates more with respect to the previous environment, even though the depth of the fade is more or less the same. Indeed, this is caused by the absence of the LOS component, the propagation happens mainly due to multipath components, hence the variability in frequency increases. Also in this case, the ray tracing seems to be able to follow the general trend of the PSD: for example, at $2.4\,\mathrm{GHz}$ and $160\,\mathrm{MHz}$ (Fig. 4.33a, where the correlation reaches a value of 0.574) the simulated channel follows the shape of the actual channel but the fades are in different positions, or even at $2.4\,\mathrm{GHz}$ with $500\,\mathrm{MHz}$ (Fig. 4.33b where some fades and peaks are in the correct position. However, RMSE and correlation still indicates that, in general, the simulated channels are quite different from the actual channels, which is confirmed also by the percentage of random keys that remains around 50%. Moreover, as it is possible to see in Fig. 4.39, Fig. 4.40 and Fig. 4.41, the statistics of the channel are quite different, contrary to the previous environment. In fact, the absorber might not have been modelled correctly, their positions may be slightly wrong: since it is a more complex environment than before, the errors and approximations of the environment database can have an impact on the precision of the simulations.

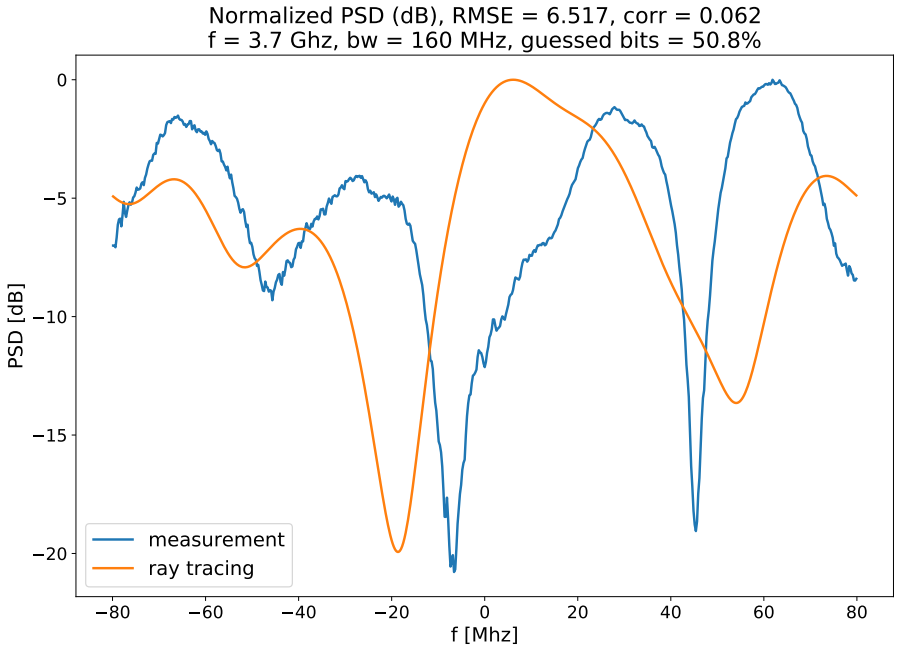


(a) *Power Spectral Density (dB) at* $2.4\,\mathrm{GHz}$ *and a bandwidth of* $160\,\mathrm{MHz}$

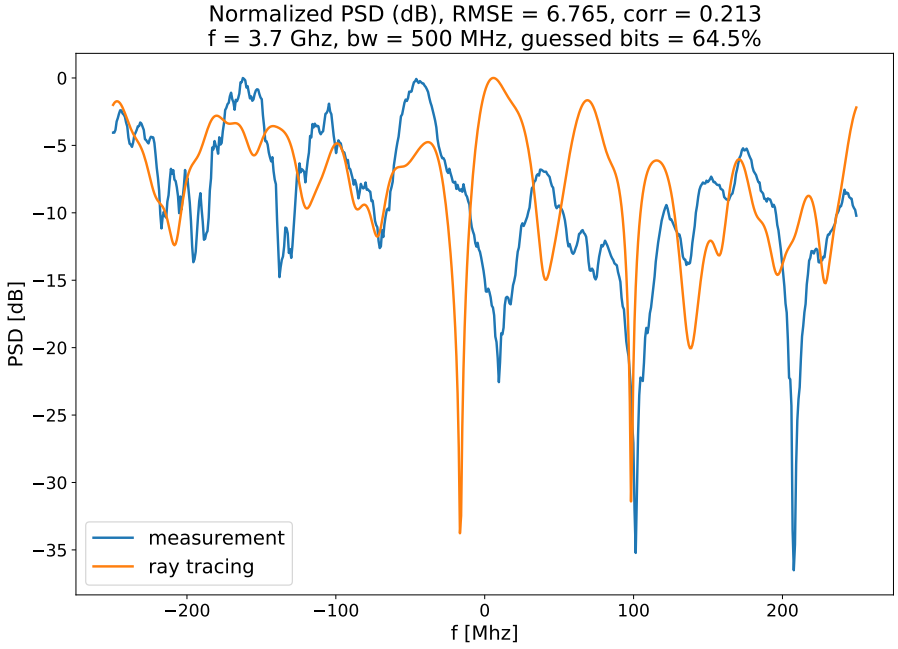


(b) *Power Spectral Density (dB)at* $2.4\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$

Figure 4.33: *Examples of Power Spectral Density (dB) in the* ***room with obstacles****, comparison between the measured and the simulated channel, with also the value of the correlation coefficient and the Root Mean Square Error.*

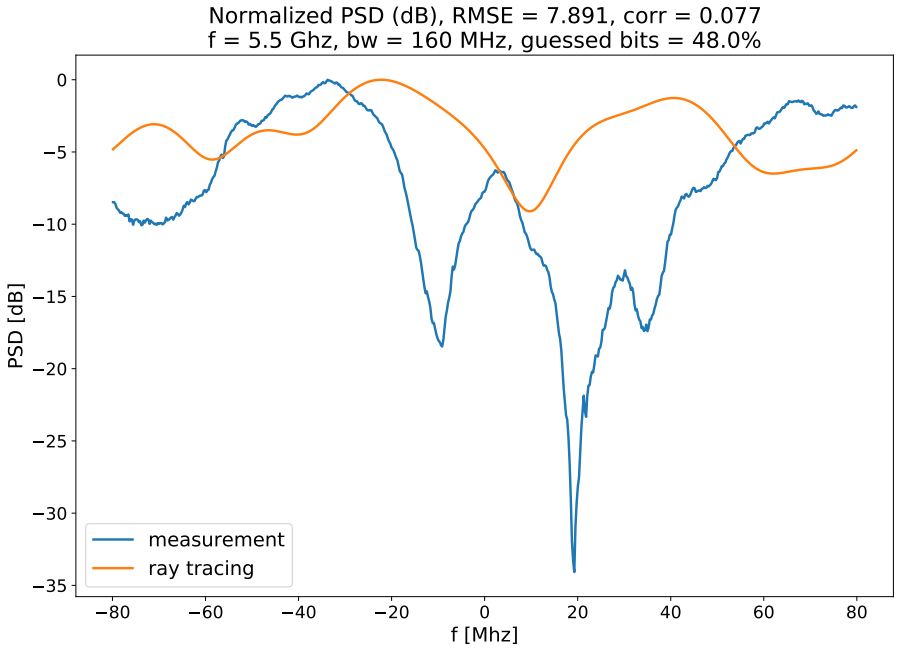


(a) *Power Spectral Density (dB) at* 3.7 GHz *and a bandwidth of* 160 MHz

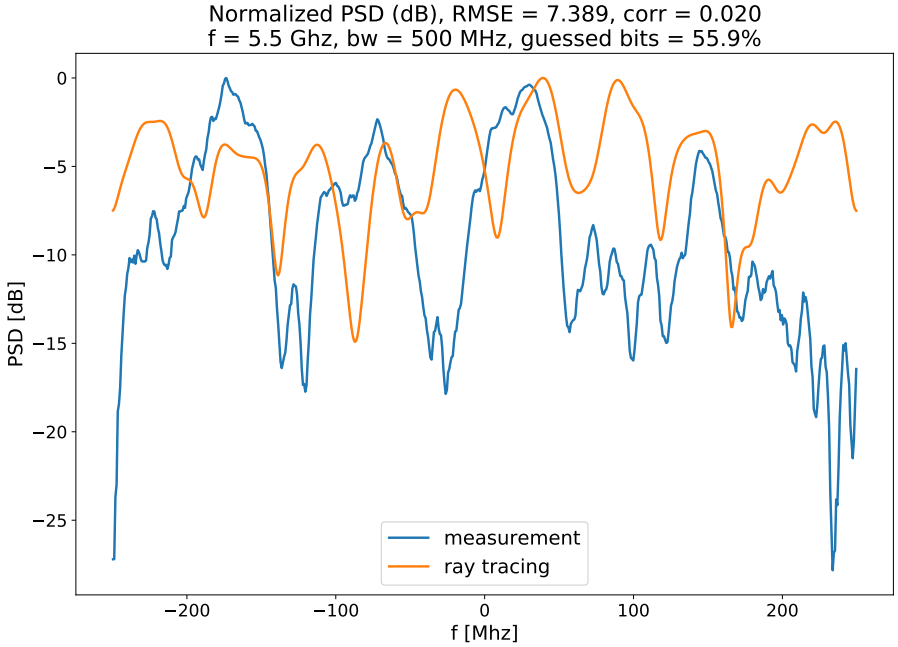


(b) *Power Spectral Density (dB) at* 3.7 GHz *and a bandwidth of* 500 MHz

Figure 4.34: *Examples of Power Spectral Density (dB) in the* ***room with obstacles****, comparison between the measured and the simulated channel, with also the value of the correlation coefficient and the Root Mean Square Error.*

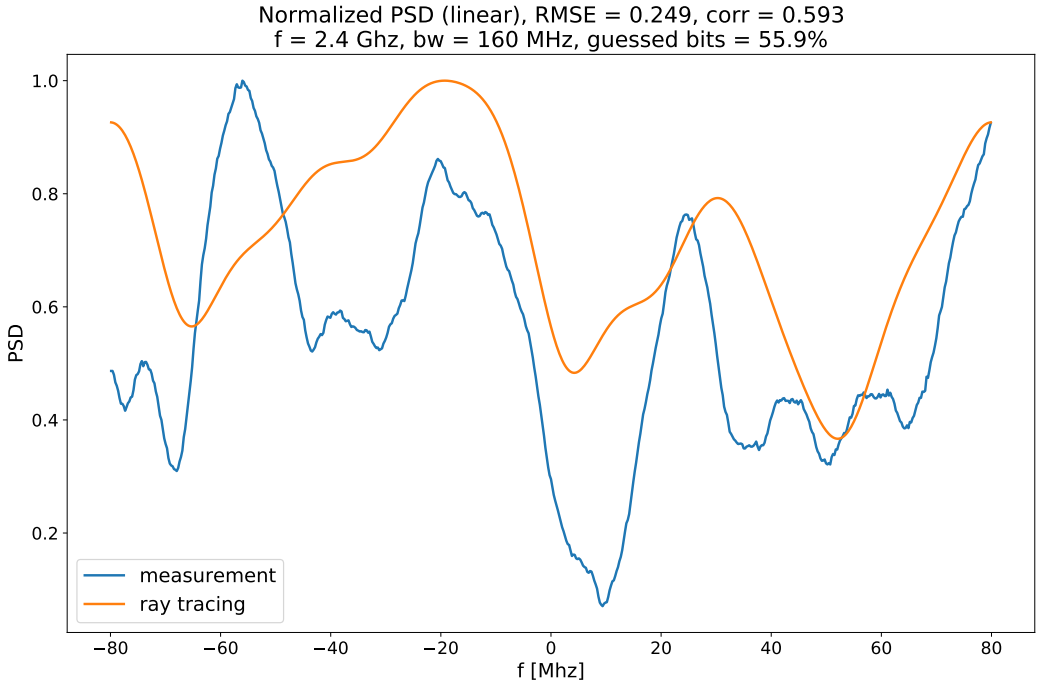


(a) *Power Spectral Density (dB) at* $5.5\,\mathrm{GHz}$ *and a bandwidth of* $160\,\mathrm{MHz}$

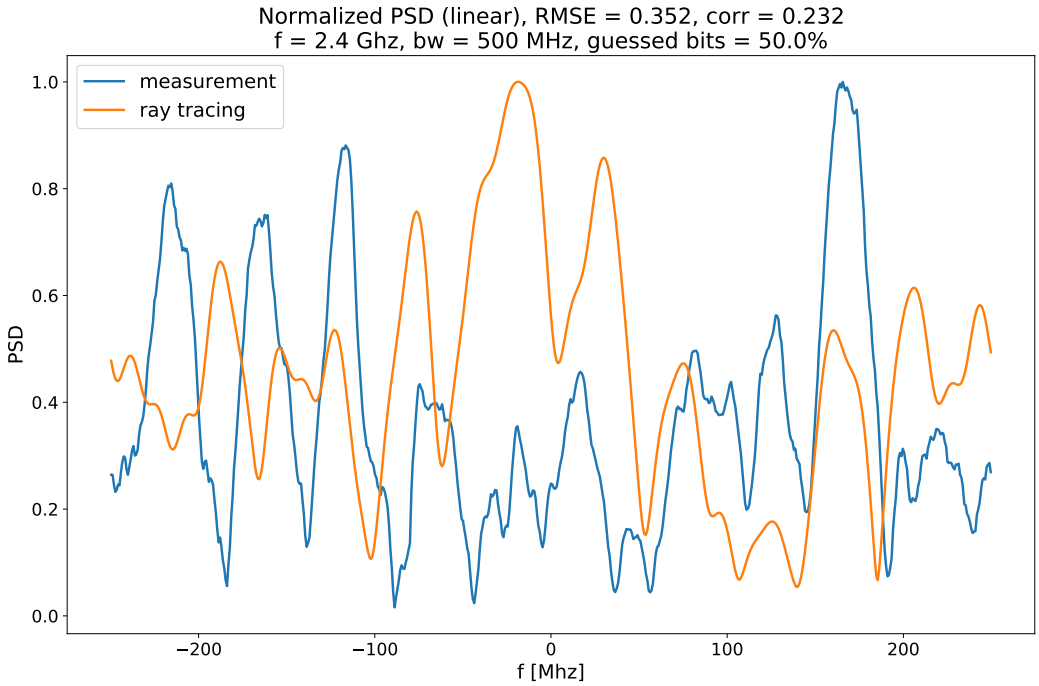


(b) *Power Spectral Density (dB) at* $5.5\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$

Figure 4.35: *Examples of Power Spectral Density (dB) in the* ***room with obstacles****, comparison between the measured and the simulated channel, with also the value of the correlation coefficient and the Root Mean Square Error.*

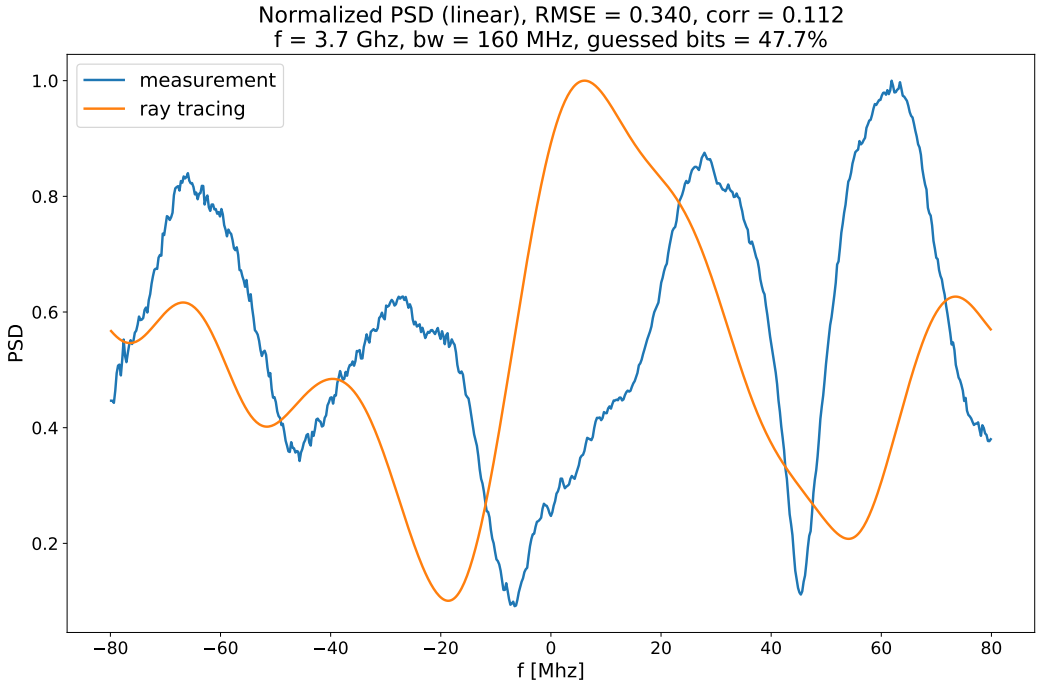


(a) *Power Spectral Density (linear) at* 2.4 GHz *and a bandwidth of* 160 MHz

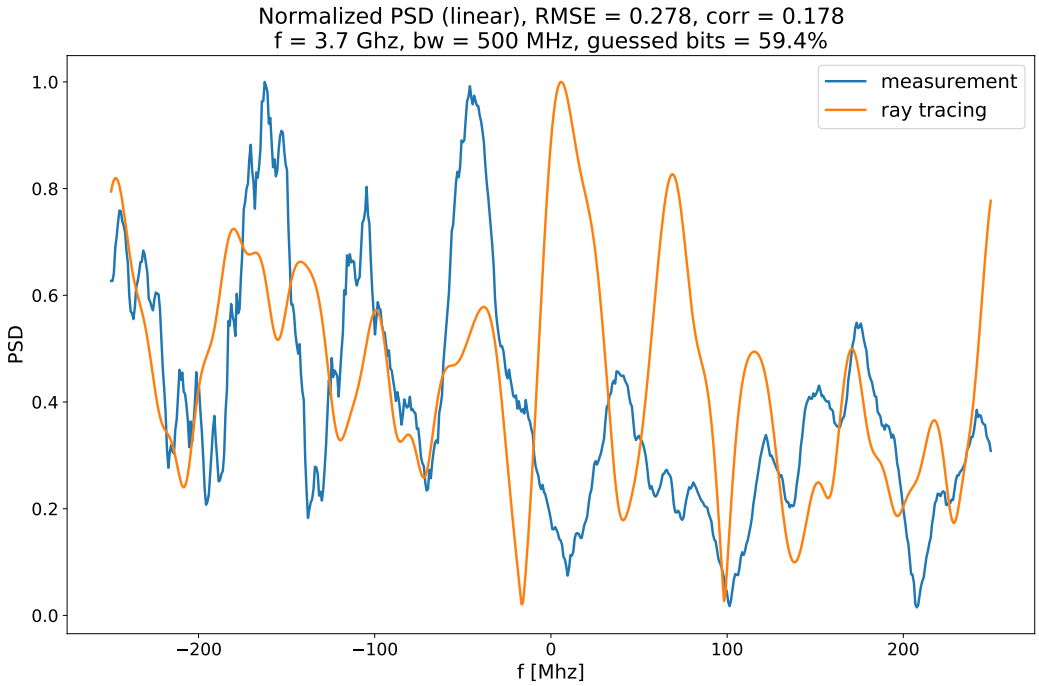


(b) *Power Spectral Density (linear) at* 2.4 GHz *and a bandwidth of* 500 MHz

Figure 4.36: *Examples of Power Spectral Density (linear) in the* ***room with obstacles****, comparison between the measured and the simulated channel, with also the value of the correlation coefficient and the Root Mean Square Error.*

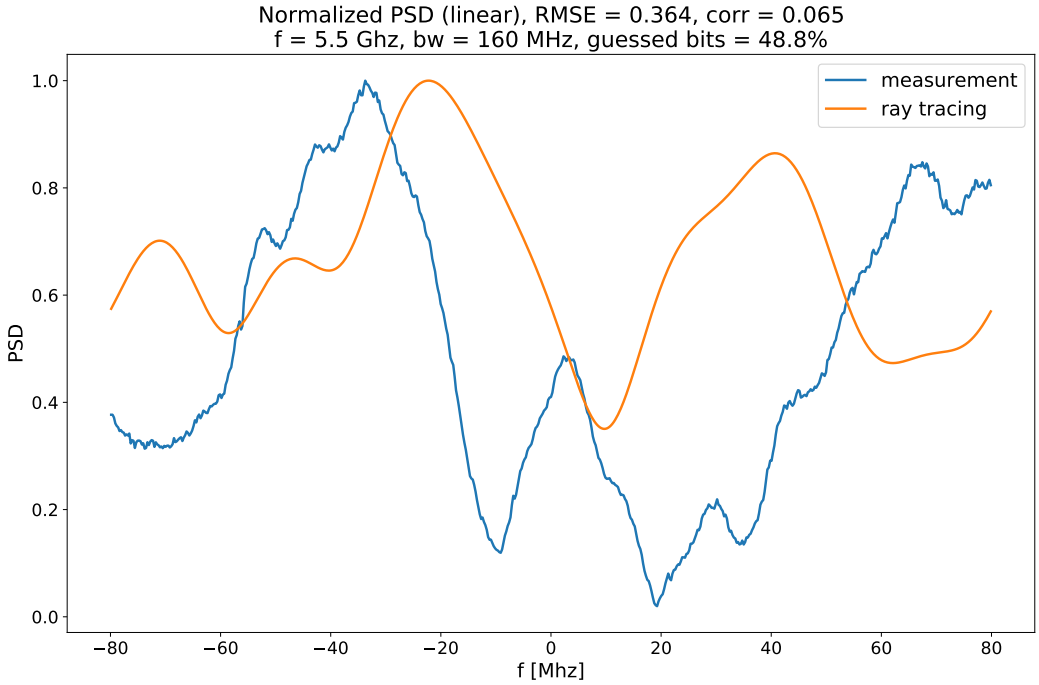


(a) *Power Spectral Density (linear) at* 3.7 GHz *and a bandwidth of* 160 MHz

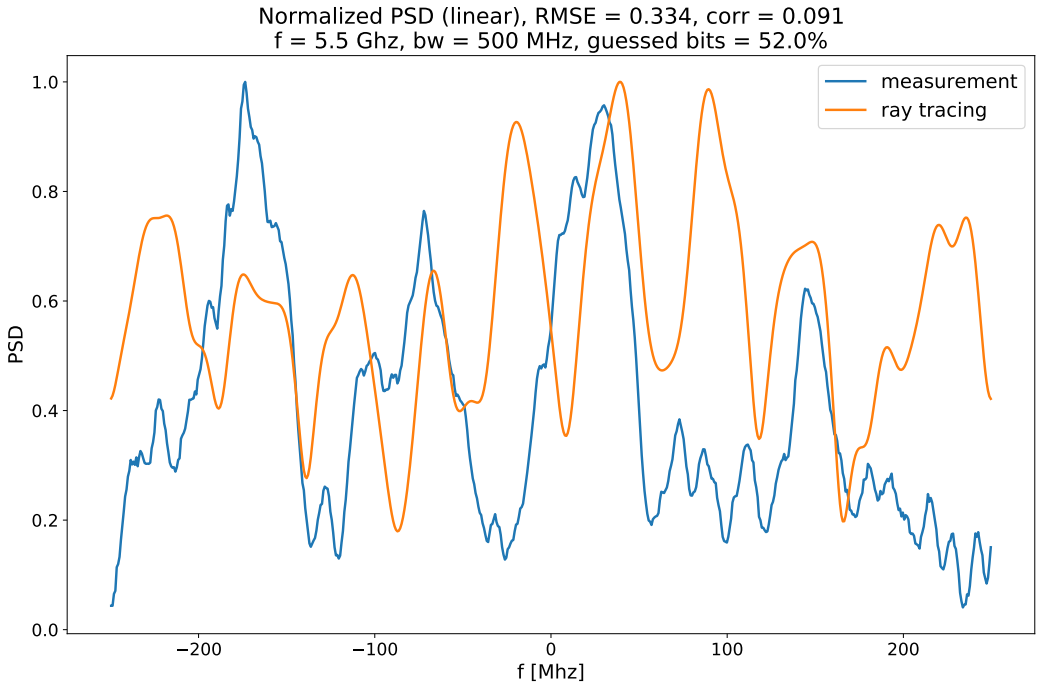


(b) *Power Spectral Density at* 3.7 GHz *and a bandwidth of* 500 MHz

Figure 4.37: *Examples of Power Spectral Density (linear) in the* ***room with obstacles****, comparison between the measured and the simulated channel, with also the value of the correlation coefficient and the Root Mean Square Error.*

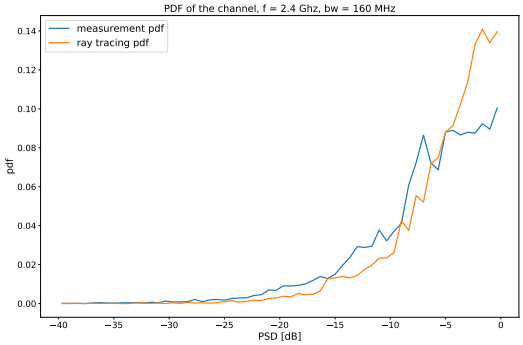


(a) *Power Spectral Density (linear) at* 5.5 GHz *and a bandwidth of* 160 MHz

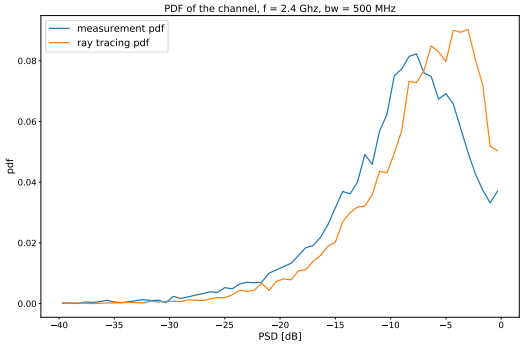


(b) *Power Spectral Density (linear) at* 5.5 GHz *and a bandwidth of* 500 MHz

Figure 4.38: *Examples of Power Spectral Density (linear) in the* ***room with obstacles****, comparison between the measured and the simulated channel, with also the value of the correlation coefficient and the Root Mean Square Error.*

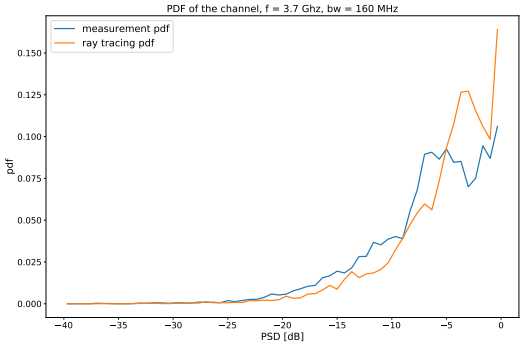


(a) *Probability Density Function of the PSD (dB) at* 2.4 GHz *and a bandwidth of* 160 MHz

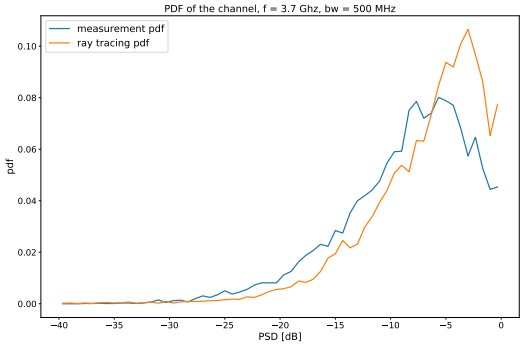


(b) *Probability Density Function of the PSD (dB) at* 2.4 GHz *and a bandwidth of* 500 MHz

Figure 4.39: *Probability Density Function of the Power Spectral Density in the* ***room with obstacles****, comparison between the measured and the simulated channels.*

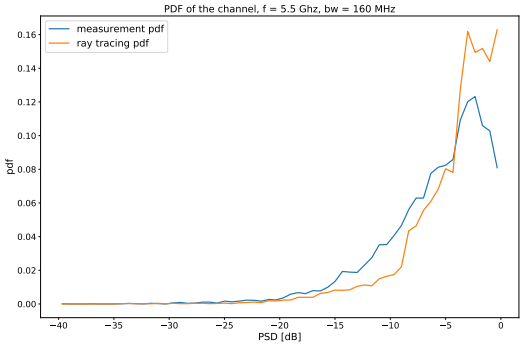


(a) *Probability Density Function of the PSD (dB) at* 3.7 GHz *and a bandwidth of* 160 MHz

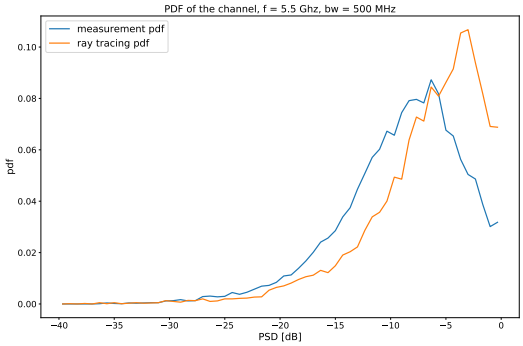


(b) *Probability Density Function of the PSD (dB) at* 3.7 GHz *and a bandwidth of* 500 MHz

Figure 4.40: *Probability Density Function of the Power Spectral Density in the* ***room with obstacles****, comparison between the measured and the simulated channels.*



(a) *Probability Density Function of the PSD (dB) at* 5.5 GHz *and a bandwidth of* 160 MHz



(b) *Probability Density Function of the PSD (dB) at* 5.5 GHz *and a bandwidth of* 500 MHz

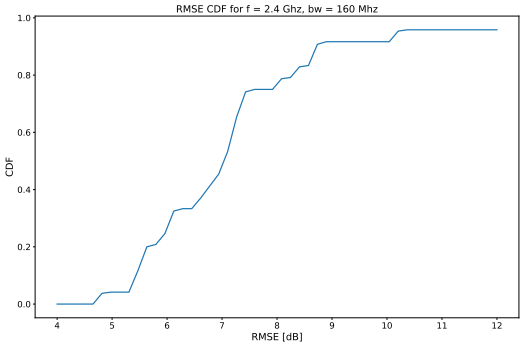
Figure 4.41: *Probability Density Function of the Power Spectral Density in the* ***room with obstacles****, comparison between the measured and the simulated channels.*

Frequency [GHz]	2.4		3.7		5.5	
Bandwidth[MHz]	160	500	160	500	160	500
Entropy meas	4.056	4.303	3.999	4.320	3.863	4.305
Entropy RT	3.584	4.040	3.656	4.037	3.407	4.042
% of random key LIN	37.500	75.000	29.167	70.833	45.833	62.500
% of random key DB	16.667	0.000	0.000	12.500	16.667	0.000
RMSE	6.765	7.635	6.844	7.790	6.235	7.608
RMSE std	1.641	1.210	1.322	1.102	1.892	1.383
std correlation	0.180	0.111	0.204	0.082	0.164	0.130
Mutual Info	0.485	0.275	0.493	0.326	0.513	0.288
% of guessed bits LIN	49.398	51.188	51.774	50.765	50.537	52.002
std % guessed LIN	6.141	4.293	5.773	4.511	5.260	3.958
% of guessed bits DB	52.930	57.715	57.145	58.138	55.046	57.682
std % guessed DB	4.845	4.411	5.314	4.127	6.000	3.790

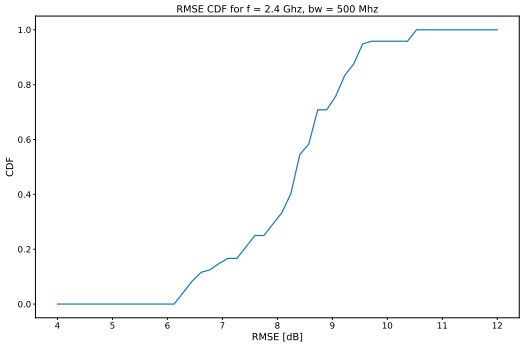
Table 4.5: *Table with the results of the assessment in the* ***room with obstacles*** *in the frequency domain. Results are subdivided by central frequency and channel bandwidth.*

The values of the metrics are reported in Table 4.5. First, the values of the entropy are slightly larger than before, but such a small difference does not imply that this environment is better than the empty room, due to the simple computational method for the entropy. In addition, the entropy is larger for the higher bandwidth. Second, the percentage of random keys follows the same trend as before, with more random keys in the higher bandwidth and extracted from the channel in linear units. Furthermore, as for the percentage in the higher bandwidth from the linear channels, it is highlighted a decreasing trend with the central frequency: since the propagation happens thanks to multipath components, as long as the frequency increases the phase difference between the phase of the rays tends to diminish, resulting in a flatter channel than at the lower central frequencies.

As for the ray tracing attack, the values of the metrics are similar to the previous case. The RMSE shows an average around 6 or 7 dB but with the possibility to reach higher and lower values, as depicted in Fig. 4.42, Fig. 4.43 and Fig. 4.44. Analogous considerations can be made for the Pearson Coefficients, whose Cumulative Distribution Functions are shown in Fig. 4.45, Fig. 4.46 and Fig. 4.47. In line with the values of the RMSE and the correlation, the mutual information remains low, with a slightly larger value in the 160 MHz for the reasons explained before. Along with the low mutual information, the percentage of guessed bits remains around 50% or little more, indicating that it is impossible for the ray tracing to guess the exact key.

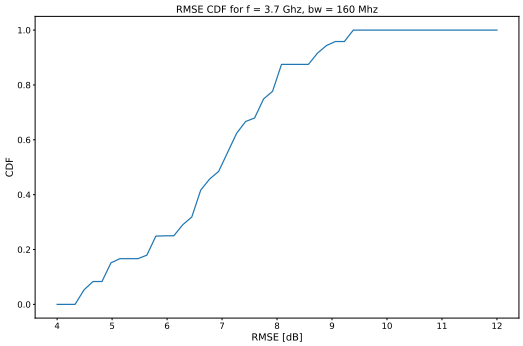


(a) *Cumulative Distribution Function of the Root Mean Square Error at* 2.4 GHz *and a bandwidth of* 160 MHz

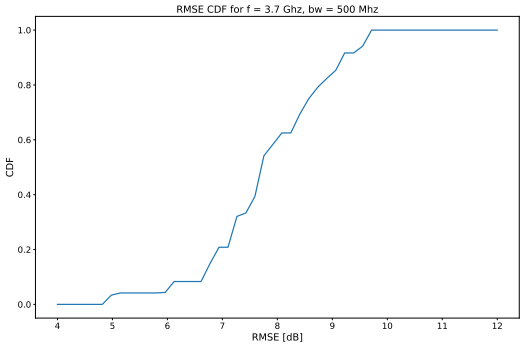


(b) *Cumulative Distribution Function of the Root Mean Square Error at* 2.4 GHz *and a bandwidth of* 500 MHz

Figure 4.42: *Cumulative Distribution Function of the Root Mean Square Error in the* ***room with obstacles****.*

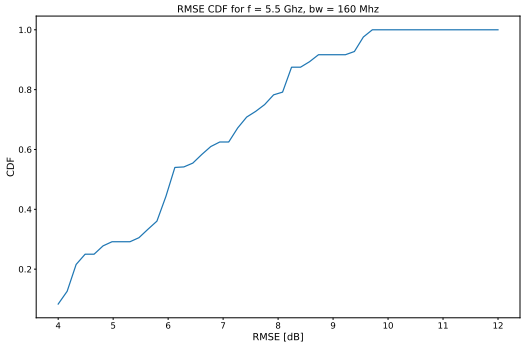


(a) *Cumulative Distribution Function of the Root Mean Square Error at* 3.7 GHz *and a bandwidth of* 160 MHz

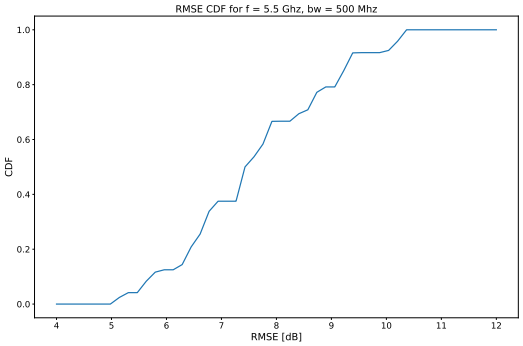


(b) *Cumulative Distribution Function of the Root Mean Square Error at* 3.7 GHz *and a bandwidth of* 500 MHz

Figure 4.43: *Cumulative Distribution Function of the Root Mean Square Error in the* ***room with obstacles****.*

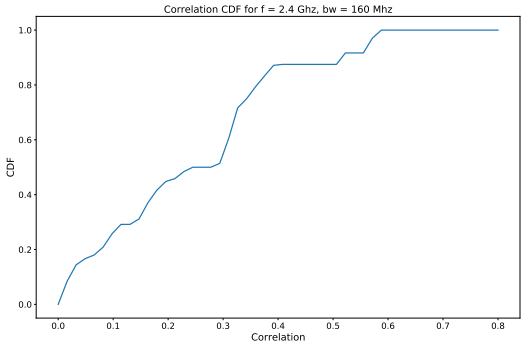


(a) *Cumulative Distribution Function of the Root Mean Square Error at* 5.5 GHz *and a bandwidth of* 160 MHz

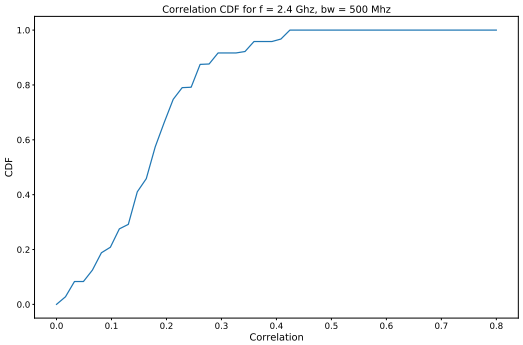


(b) *Cumulative Distribution Function of the Root Mean Square Error at* 5.5 GHz *and a bandwidth of* 500 MHz

Figure 4.44: *Cumulative Distribution Function of the Root Mean Square Error in the* ***room with obstacles****.*

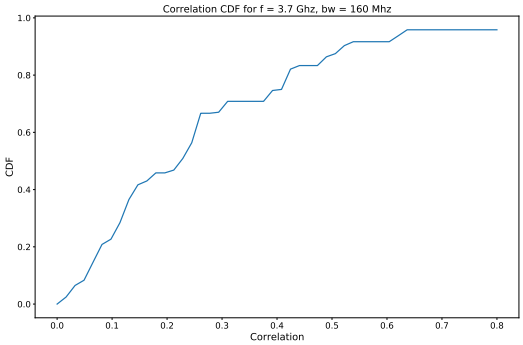


(a) *Cumulative Distribution Function of the Pearson Coefficient at* 2.4 GHz *and a bandwidth of* 160 MHz

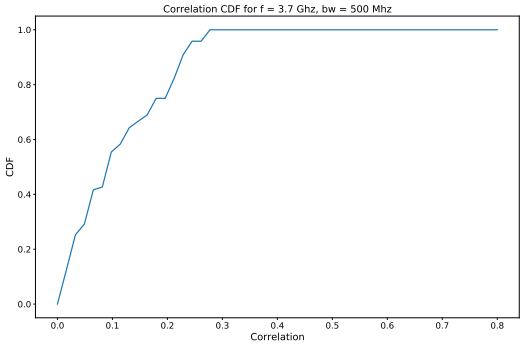


(b) *Cumulative Distribution Function of the Pearson Coefficient at* 2.4 GHz *and a bandwidth of* 500 MHz

Figure 4.45: *Cumulative Distribution Function of the Pearson Coefficient in the* ***room with obstacles****.*

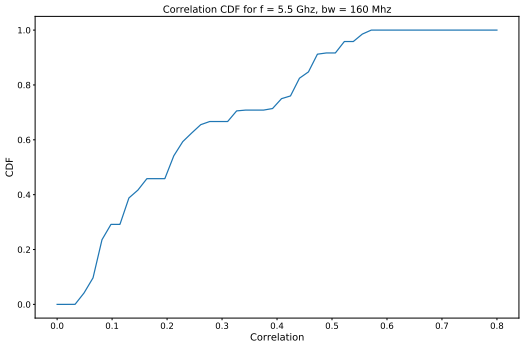


(a) *Cumulative Distribution Function of the Pearson Coefficient at* 3.7 GHz *and a bandwidth of* 160 MHz

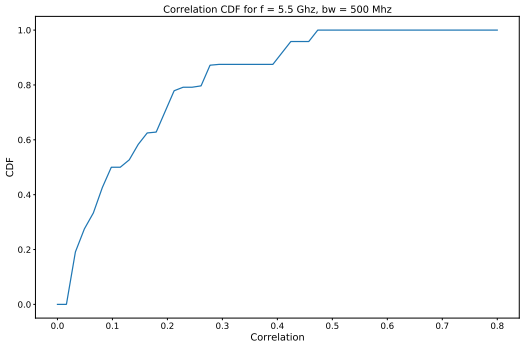


(b) *Cumulative Distribution Function of the Pearson Coefficient at* 3.7 GHz *and a bandwidth of* 500 MHz

Figure 4.46: *Cumulative Distribution Function of the Pearson Coefficient in the* ***room with obstacles****.*

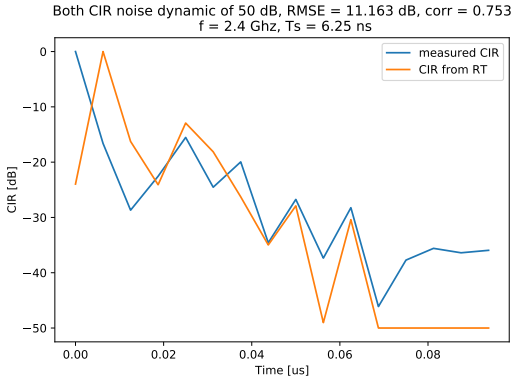


(a) *Cumulative Distribution Function of the Pearson Coefficient at* 5.5 GHz *and a bandwidth of* 160 MHz

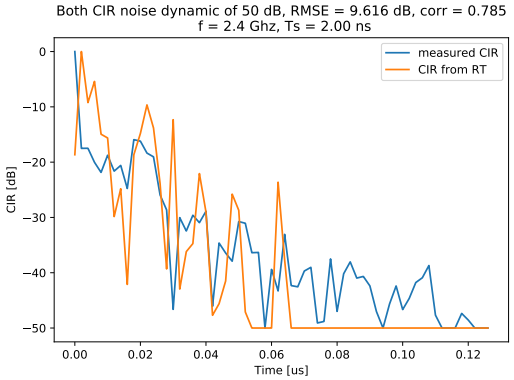


(b) *Cumulative Distribution Function of the Pearson Coefficient at* 5.5 GHz *and a bandwidth of* 500 MHz

Figure 4.47: *Cumulative Distribution Function of the Pearson Coefficient in the* ***room with obstacles****.*

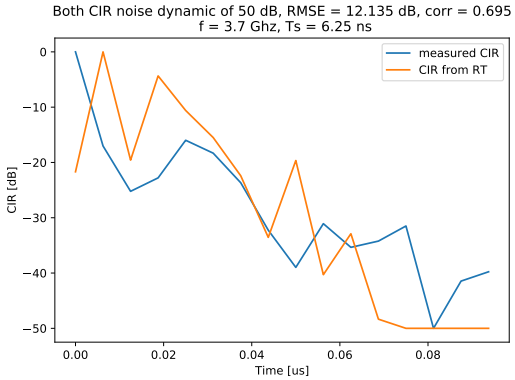


(a) *Channel Impulse Response at* 2.4 GHz *and a bandwidth of* 160 MHz

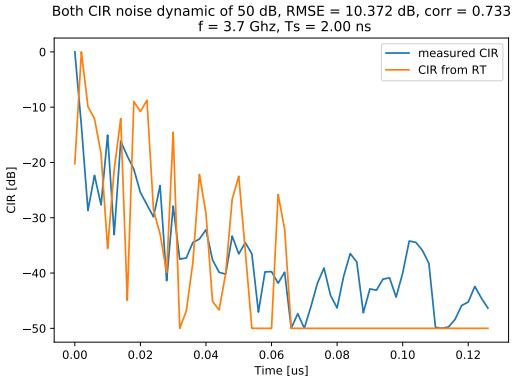


(b) *Channel Impulse Response at* 2.4 GHz *and a bandwidth of* 500 MHz

Figure 4.48: *Examples of Channel Impulse Response (CIR) in the* ***room with obstacles****, compared considering a noise dynamic of* 50 dB*. In the figures there are also the value of the RMSE and the correlation between the two channels.*

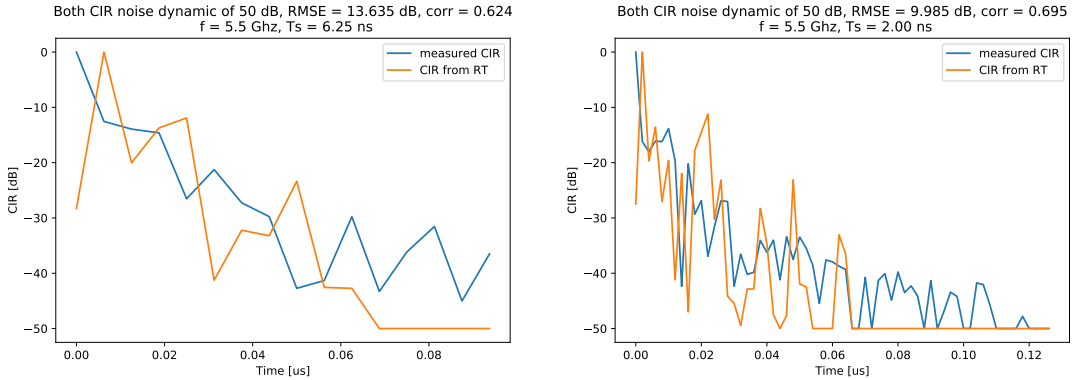


(a) *Channel Impulse Response at* 3.7 GHz *and a bandwidth of* 160 MHz



(b) *Channel Impulse Response at* 3.7 GHz *and a bandwidth of* 500 MHz

Figure 4.49: *Examples of Channel Impulse Response (CIR) in the* ***room with obstacles****, compared considering a noise dynamic of* 50 dB*. In the figures there are also the value of the RMSE and the correlation between the two channels.*



(a) *Channel Impulse Response at* 5.5 GHz *and a bandwidth of* 160 MHz

(b) *Channel Impulse Response at* 5.5 GHz *and a bandwidth of* 500 MHz

Figure 4.50: *Examples of Channel Impulse Response (CIR) in the* ***room with obstacles****, compared considering a noise dynamic of* 50 dB*. In the figures there are also the value of the RMSE and the correlation between the two channels.*

Frequency [GHz]	2.4		3.7		5.5	
Bandwidth [MHz]	160	500	160	500	160	500
RMSE time	11.067	9.439	12.753	11.001	13.609	10.895
RMSE std time	2.006	0.864	1.986	1.035	2.336	0.992
correlation TIME	0.816	0.789	0.737	0.700	0.668	0.692
std correlation TIME	0.084	0.048	0.085	0.038	0.167	0.052

Table 4.6: *Time domain evaluation in the* ***room with obstacles****.*

4.4.1 Time domain

As before, the evaluation has been done also in the time domain and some of the Channel Impulse Responses are reported in Fig. 4.48, Fig. 4.49, Fig. 4.50. It is interesting to start by looking at the channels in Fig. 4.48a: the simulated CIR present a peak not in the first ray resolved, which is caused by the absence of a direct component, therefore the multipath distribution creates such trend. Except for the first peaks, the Ray Tracing is able to guess the trend and also the peaks of the actual CIR. However, the simulation does not find the last component of the channel. The last part might be the reason why the RMSE is very high (11.163 dB) even though the channels are very similar. Moreover, this similarity does not exposes the key generation: in fact, only UWB systems are able to utilizes the CIR as a feature to extract the key from (thanks to the time resolution power of the UWB systems), and looking at the CIR in the higher bandwidth, the channels are far from being similar. This observation is confirmed by the values of the RMSE reported in Table 4.6.

4.5 Summary

The Ray Tracing attack seems not to be an important threat against Physical Layer based-Key generation, since the Ray Tracing is not able to guess the exact structure of the multipath. In addition, the environment considered were two simple environments, and it is expected that in a more complex environment, also with moving objects, the probability of success will be very low. Not only the guessed bits are kept around 50%, even the low value of the mutual information indicates that it is very hard for the attacker to infer the key from the simulated channels. Furthermore, from the time analysis it is possible to see that even thought the general trend is guessed, the RMSE indicates that that they are far from being equal, also due to the lack of the latest components. To conclude, the Ray Tracing is still able to guess the statistics of the channel and the trend of the entropy, which opens the possibility to use it as a design tool, possibility explored in the next chapter.

5

Ray Tracing as a design tool for PL-key generation

In the previous chapter the possibility of a Ray Tracing attack has been explored and it has been shown that the Ray Tracing seems not able to predict the exact pattern of the *fast fading*. Therefore, since the Key Generation relies on the observation of the *fast fading* the Ray Tracing would not be able to break the security of the protocol. However, the Ray Tracing has been proved to predict the statistics of fast fading quite reliably (in the previous chapter and in [44]). Therefore, it can be used as a tool to generate realistic channels and evaluate the impact of some features of the link (e.g frequency, bandwidth, directivity of the antennas, environment) on the Physical Layer based-Key generation protocol. In this chapter, this possibility is explored through a series of simulations performed in two different environments: first, the *empty Mondrian Room* utilized also for the measurements; second, an environment (see Fig. 5.2) with some furnitures inside. Simulations have been performed with different bandwidth, frequency and antennas, in order to evaluate the impact of using also directive antennas. The goals of the simulations are:

- Evaluate the sensibility of the PL-based Key Generation to different link parameters.
- Evaluate the spatial correlation between Bob and Eves, by generating the channels in some positions around Bob and comparing the channel between the series of Eves.

5.1 Description of the simulation environment

Simulations are performed in two different environments, the empty Mondrian room utilized for the measurements, whose description can be found in 4.1.1. The

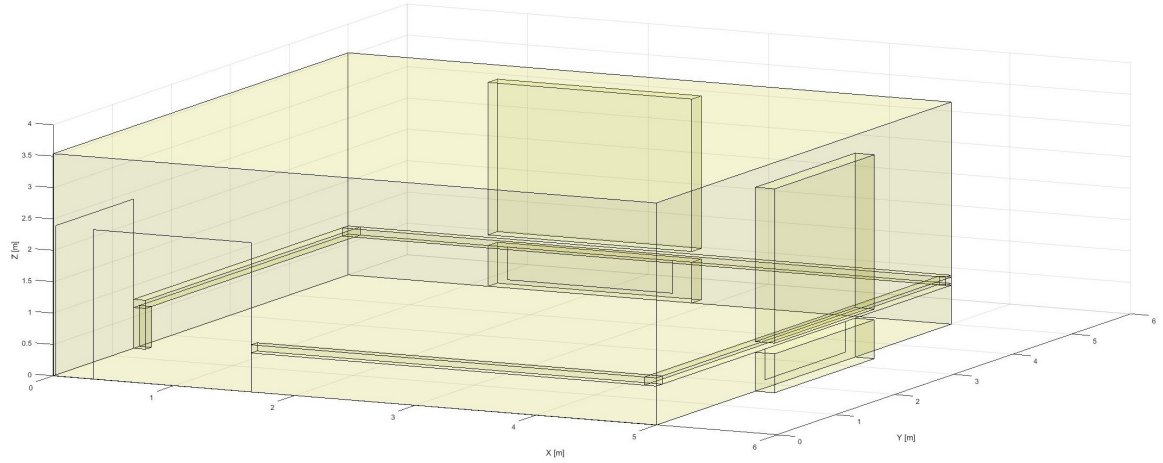


Figure 5.1: *Ray tracing environment of the Mondrian Room.*

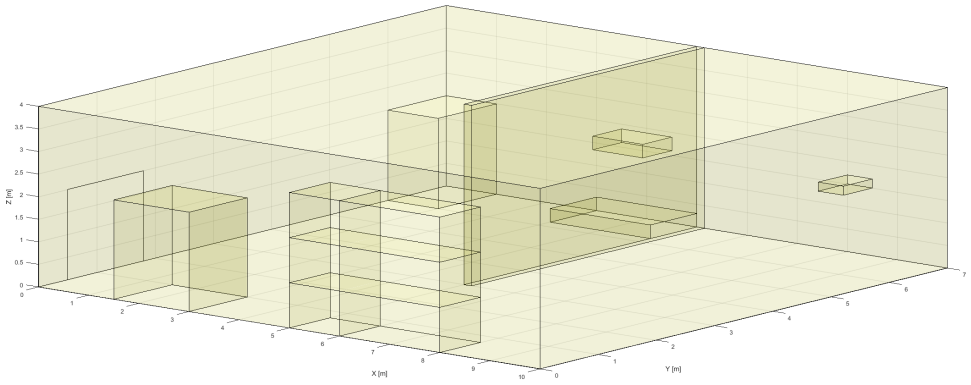


Figure 5.2: *Ray tracing environment of the complex room.*

Ray Tracing environment is reported in Fig. 5.1 for simplicity. Then, a second room has been considered as shown in Fig. 5.2: some furnitures are present in the room like a bookcase, two wardrobes, two floating shelf, a door and the structure of a bed. Moreover, there is a piece of wall in the middle, in order to limit the line of sight component: the wall is supposed to be empty inside and the three surfaces have a thickness of $5\,\mathrm{cm}$. The walls are supposed to be made of concrete and the furnitures of wood, the electromagnetic characteristic of the materials are reported in Table 5.1 and are the same for both the environments.

5.2 Procedure and metrics for the simulation assessment

The simulations follow the steps described in Section 4.2: a set of points in the environment have been chosen and the channel simulated. For each link condition (frequency, bandwidth, antenna), 72 points have been selected. Then, the Power Spectral Density (PSD) is extracted from each channel, in order to:

Electromagnetic characteristics

Frequency	2.4 [GHz]		3.7[GHz]		5.5[GHz]	
Parameter	ε_r[F/m]	σ[S/m]	ε_r[F/m]	σ[S/m]	ε_r[F/m]	σ[S/m]
Concrete	4.5	0.007	6.0	0.0017	5.5	0.087
Wood	3.0	0.0	1.8	0.012	2.5	0.009
Glass	6.0	0.016	6.0	0.0016	5.5	0.016

Table 5.1: *Table of the electromagnetic parameters of the material.*

- Compute the entropy of the channel.
- Extract the key from the channel in linear units and in logarithmic units. Keys are extracted with the filter-bank model, using 64 filters and 16 levels of quantization.
- Apply the randomness tests on the key and compute the percentage of random keys.

These kind of simulations are used to assess the impact of the link feature impact, but other simulations have been made to assess the spatial decorrelation between the Alice-Bob and Alice-Eve channels. The metrics utilized, as will be explained in 5.6, are the Key Disagreement Rate and the Root Mean Square Error.

5.2.1 Channels parameters

Simulations have been performed on three frequencies: $2.4\,\mathrm{GHz}$, $3.7\,\mathrm{GHz}$ and $5.5\,\mathrm{GHz}$, in combination with two bandwidths $160\,\mathrm{MHz}$ and $500\,\mathrm{MHz}$.

5.2.1.1 Antennas

In the previous chapter, only omnidirectional antennas have been considered, but in the Ray Tracing it is possible to define also directive antennas. The first antenna considered is the dipole used also for the measurement, whose radiation pattern in the three central frequencies can be found in Fig. 3.9. The other antennas are two *square planar array*, as schematized in Fig. 5.3. It is important to evaluate the effects of directive antennas, since in future wireless communications will more and more make use of high frequency and beamforming [45]. The first array has 5 elements per side and a directivity of $12.68\,\mathrm{dB}$, the second has 20 elements per side and a directivity of $24.95\,\mathrm{dB}$: the elements are supposed to be isotropic antennas, uniformly spaced at $\lambda/2$ and uniformly fed. Therefore, both are directive arrays but the second one has an higher directivity, thus a narrower main lobe. In addition, both are supposed to be placed above an ideal ground plane, so they radiate only in one half space. In the end, the radiation patterns are reported in Fig. 5.4.

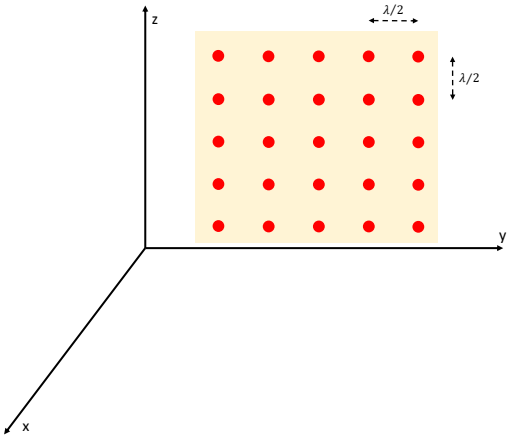
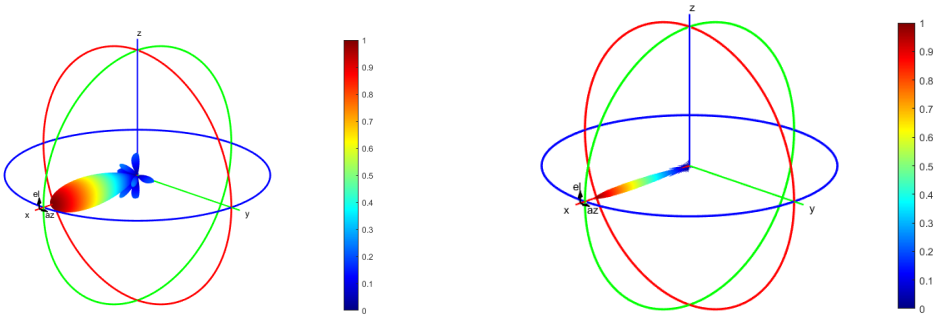


Figure 5.3: *Scheme of the planar array with the ground plane behind. The array is placed on the $z - y$ plane and the main lobe is directed toward the positive x direction*



(a) *Normalized radiation pattern of the 5×5 planar array*

(b) *Normalized radiation pattern of the 20×20 planar array*

Figure 5.4: *Normalized radiation pattern of the planar arrays.*

During the simulation, Bob and Eve have the dipole antenna, while Alice changes her antenna. In case of directive antennas, it is properly rotated to point always toward Bob, hence Eve receive the signal from a side lobe of the antenna

5.2.1.2 Ray Tracing

The ray tracing characteristics are the same of the Ray Tracing attack described in section 4.2, but with a lower number of events, as summed up in Table 5.2.

5.2.2 Evaluation Metrics

The goal of the simulations is to evaluate the impact of different channel conditions on the Key Generation. Therefore, only few metrics are utilized:

- The **entropy** of the channel. When using the directive antennas, the channel

Summary of simulation parameter

Channel Parameter	Central frequencies [GHz]	2.4, 3.7, 5.5
	Channel bandwidth [MHz]	160, 500
	Antenna	real dipole, 5×5 planar array, 20×20 planar array
Ray Tracing	Max. number of interactions	5
	Max. reflections	5
	Max. diffractions	2
	Max. diffractions + reflections	5
	Max. transmissions	50
	Min. length of edge (in wavelength)	1
	Min. wall area (in squared wavelength)	1
	FFT coefficients in output	631
Statistic computations	Number of histogram's bins	60
	Range of histogram	`min`(Array) to 0 dB
Key generation	Number of filters	64
	Number of levels	16
	Key bits	256

Table 5.2: *Summary of the parameter for the simulations*

dynamic is surely reduced with respect to when omnidirectional antennas are employed, since the multipath is reduced. In order to agree to the lower dynamic of the channel, the computation of the entropy is slightly changed with respect to what is written in section 4.2.2.1: the PDF of the channel is computed through the histogram, but now the range of the histogram is changed according to the minimum value of the collection of channels (see section 4.2.2.1). Therefore, the binning is no more fixed but changes according to the dynamic of the channel.

- The **percentage of random keys**, computed both over the channels in dB-unit and in linear unit.

5.3 Results of the simulation

The results are subdivided for the two environmentS.

5.4 Evaluation in the empty room

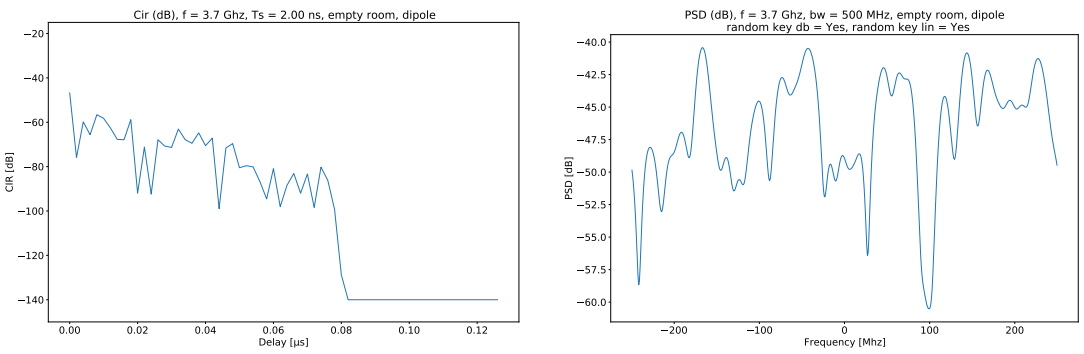
Let's begin with the empty room, even though some considerations on this environment have been already given. In Table 5.3 there is the summary of the evaluation. First, focus on the trend of the entropy: as the directivity increases, the multipath components are attenuated and the channel fluctuates less in frequency. Therefore, the entropy of the channel diminishes as the directivity increases. This effect can be seen in Fig. 5.5, Fig. 5.6 and Fig. 5.7, where the channel with the three antennas are reported. Moreover, in Fig. 5.9 there is the bar plot of the values of entropy, which highlight the decreasing trend of the entropy with the directivity of the antenna.

How could one imagine, since the key generation remains the same regardless of the antenna, the keys should be less random as the entropy diminishes. However, looking at the table, it is possible to see that this trend is not followed by the keys: to have a clearer vision, in Fig. 5.8 there is a bar plot of the random keys. First, it is possible to see that the keys from the channel with larger bandwidth are more probable random: even though the entropies are similar (see Fig. 5.9), the filters are narrower than the higher bandwidth case. Hence, their output are correlated, since the filters are smaller than the coherence bandwidth of the channel. Then, the percentage does not follow a specific trend with respect to the directivity. The reason can be found in the quantization of the channels: every filter's output is quantized with a 16 level quantizations, which is suitable for channel with high dynamic range since it exploits all the entropy inside (e.g. the channel in Fig. 5.5 has a dynamic of more or less $20\,\mathrm{dB}$.) Of course, as the dynamic decreases and the number of quantization levels remains the same, the bits extracted will contain some redundancy: in fact, the situation with the 5×5 array has the least amount of random keys. Nevertheless, when moving to the case with the highest directivity, the percentage comes back to be high: the channel in this case (see Fig. 5.7) has a little dynamic range and the quantization recognizes even a very little change in the channel, since the quantization slots are narrower with respect to the previous cases. However, these little fluctuations can be recognized and utilized in this ideal case, but in a real scenario the small fluctuations will be hidden by the noise, since their amplitude might be comparable with the noise power. Therefore, even though a computer can recognize the little fluctuations, they will be masked by noise making impossible the agreement between Alice and Bob. In addition, the simulations have an high numerical precision and resolution: instead, a real system usually has a smaller resolution and the small fluctuations that the simulator is able to find might not be recognized by a real systems. In this work, the problems related to the symmetry of the channel have been neglect, thus without caring of the disagreement between Alice and Bob. Eventually, the keys generated from the linear channel are more probable random, except for the case in which the 20×20 planar array is employed: the small fluctuations of the channel remains the same in both the linear and logarithmic channels.

Evaluation in the **empty room**

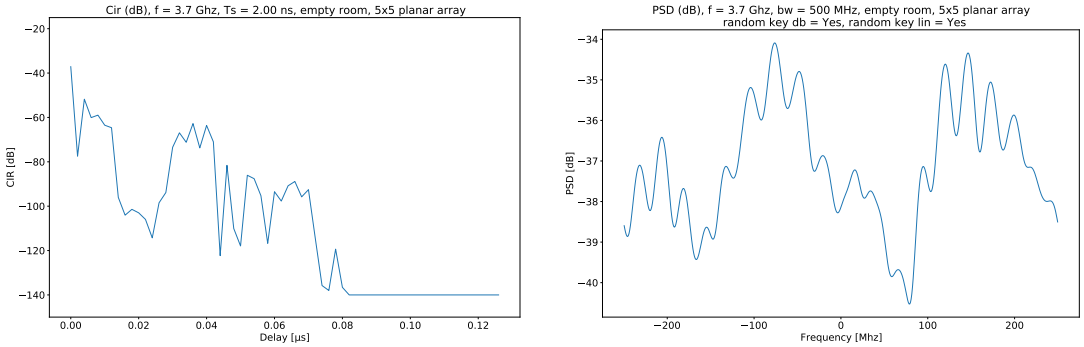
Antenna	Frequency [GHz]	Bandwidth [MHz]	Entropy	% of random keys lin	% of random keys dB
Dipole	2.4	160	3.089	23.611	12.500
		500	3.553	43.056	13.889
	3.7	160	3.407	41.667	23.611
		500	3.735	44.444	12.500
	5.5	160	3.158	27.778	18.056
		500	3.593	44.444	9.722
5×5 planar array	2.4	160	1.766	16.667	13.889
		500	2.395	48.611	34.722
	3.7	160	1.924	33.333	29.167
		500	2.571	41.667	31.944
	5.5	160	1.862	26.389	27.778
		500	2.447	38.889	33.333
20×20 planar array	2.4	160	0.407	12.500	15.278
		500	1.110	45.833	51.389
	3.7	160	0.571	41.667	43.056
		500	1.308	41.667	37.500
	5.5	160	0.526	34.722	30.556
		500	1.104	54.167	47.222

Table 5.3: *Assessment results for the **empty room**, subdivided by the antenna type, central frequency and bandwidth. The table contains the values of entropy and the percentage of random keys. Previous tables always contained the evaluation metrics on the rows and the channel parameters on the columns, now they are inverted in order for the table to be readable.*



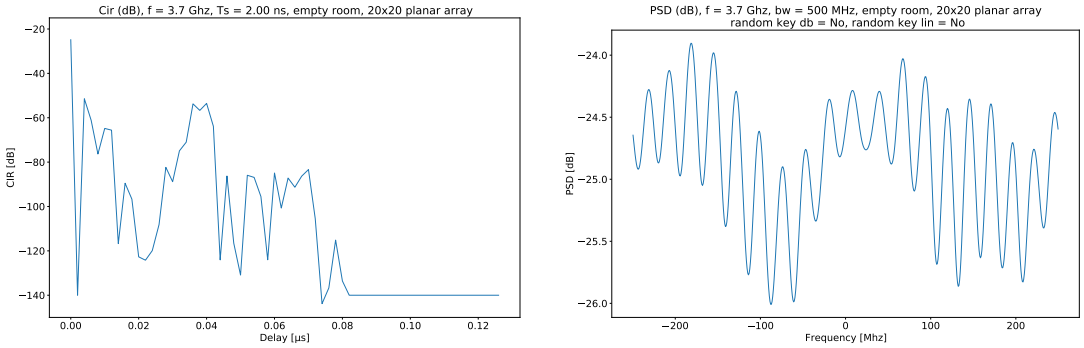
(a) *Channel Impulse Response in dB.* ***(b)*** *Power Spectral Density of the channel in dB*

Figure 5.5: *Channel with the dipole, at* $3.7\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$*, in the **empty room**. In the plot there is also the indication if the key is random or not.*



(a) *Channel Impulse Response in dB.* ***(b)*** *Power Spectral Density of the channel in dB*

Figure 5.6: *Channel with the* 5×5 *planar array, at* $3.7\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$*, in the* ***empty room****. In the plot there is also the indication if the key is random or not.*



(a) *Channel Impulse Response in dB.* ***(b)*** *Power Spectral Density of the channel in dB*

Figure 5.7: *Channel with the* 20×20 *planar array, at* $3.7\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$*, in the* ***empty room****. In the plot there is also the indication if the key is random or not.*

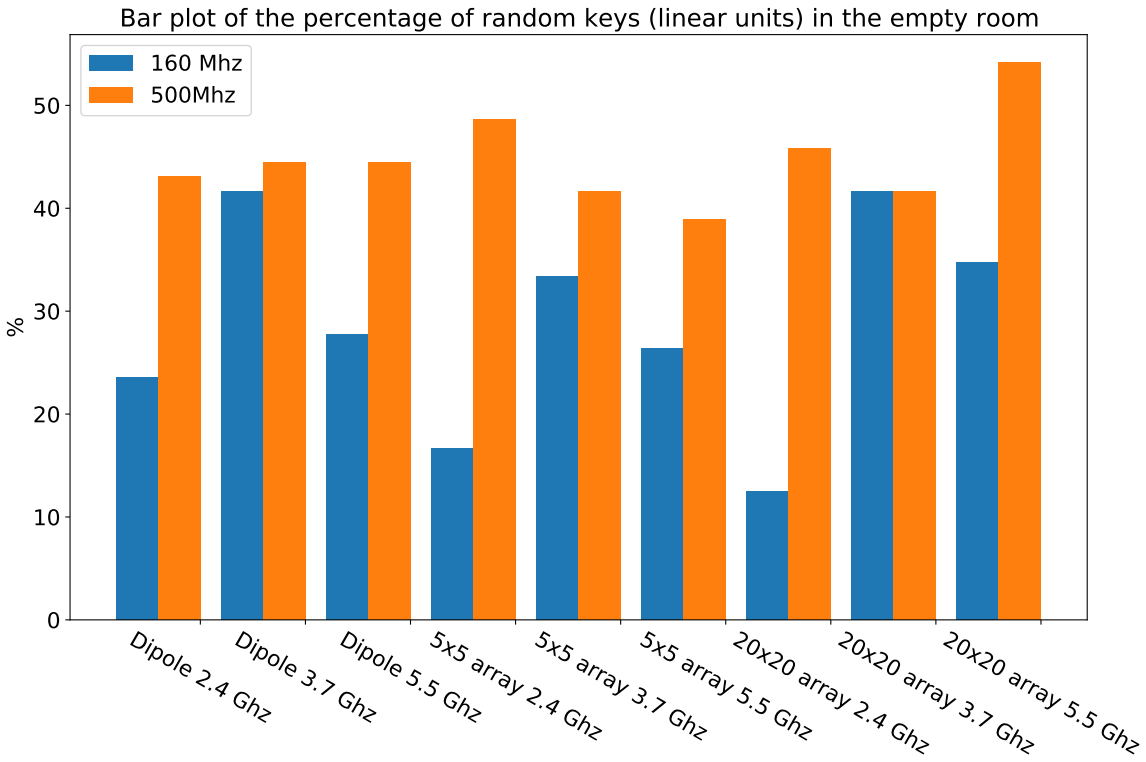


Figure 5.8: *Bar plot of the percentage of random keys in the* ***empty room****, extracted from the channel in linear unit*

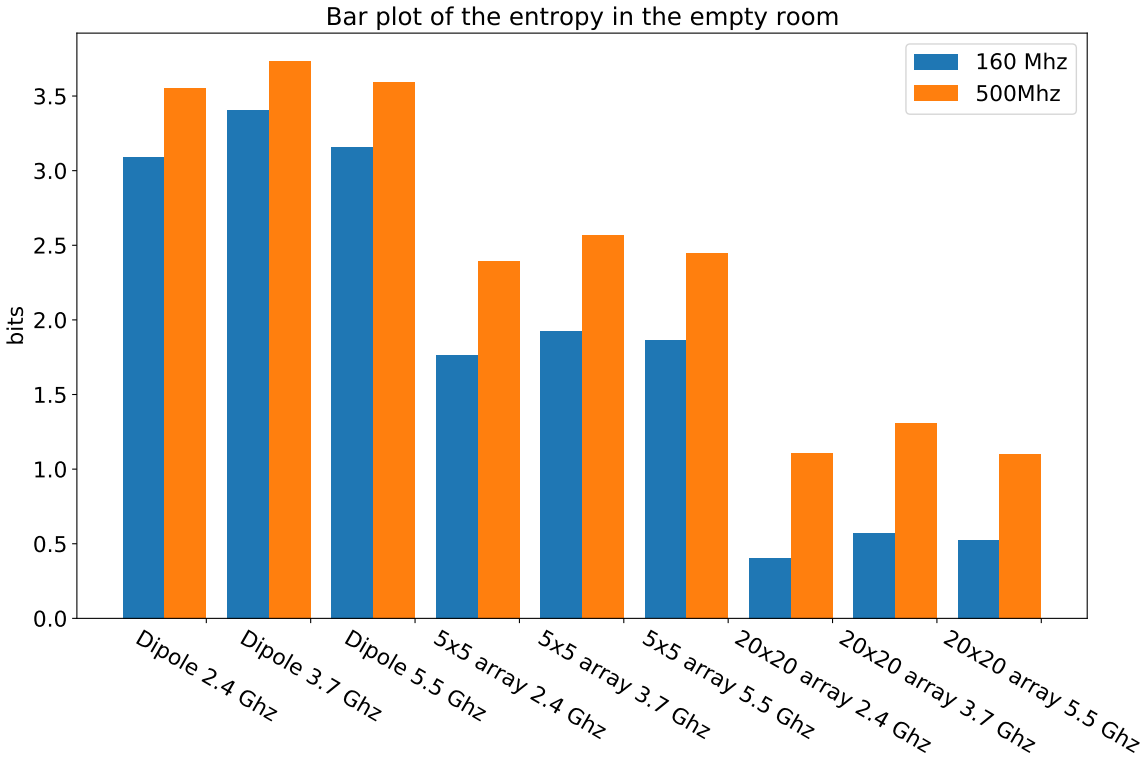


Figure 5.9: *Bar plot of the entropy of* ***empty room***

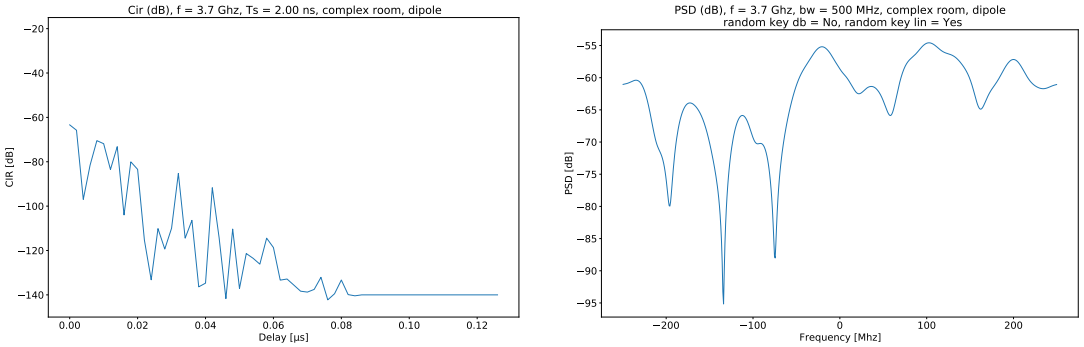
Evaluation in the **complex room**

Antenna	Frequency [GHz]	Bandwidth[MHz]	Entropy	% of random keys lin	% of random keys dB
Dipole	2.4	160	3.010	29.167	25.000
		500	3.403	50.000	23.611
	3.7	160	3.110	16.667	23.611
		500	3.419	65.278	34.722
	5.5	160	3.199	33.333	30.556
		500	3.561	63.889	19.444
$5\times$ planar array	2.4	160	2.310	27.778	23.611
		500	2.561	65.278	41.667
	3.7	160	2.398	30.556	31.944
		500	2.688	63.889	47.222
	5.5	160	2.338	25.000	27.778
		500	2.605	51.389	33.333
20×20 planar array	2.4	160	1.209	25.000	25.000
		500	1.458	72.222	52.778
	3.7	160	1.166	27.778	31.944
		500	1.394	56.944	56.944
	5.5	160	1.342	23.611	36.111
		500	1.530	59.722	52.778

Table 5.4: *Assessment results for the* ***complex room****, subdivided by the antenna type, central frequency and bandwidth. The table contains the values of entropy and the percentage of random keys. Previous tables always contained the evaluation metrics on the rows and the channel parameters on the columns, now they are inverted in order for the table to be readable.*

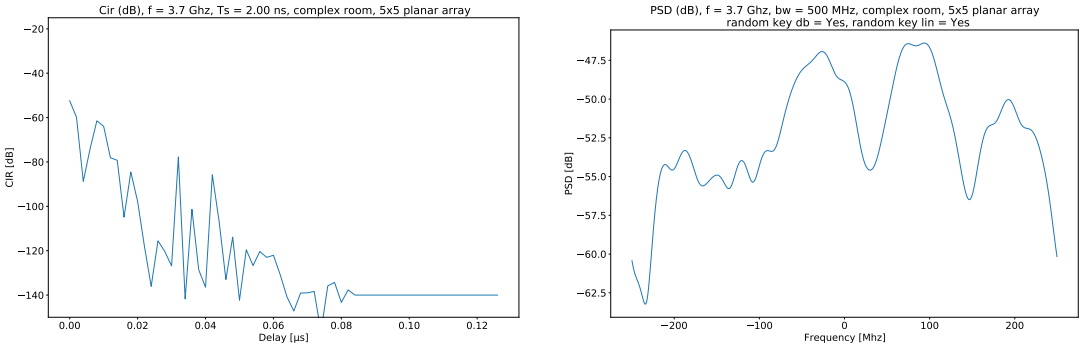
5.5 Evaluation in the complex room

The same evaluation has been performed also for the *complex room* and in Table 5.4 there is the summary of the evaluation. The room has some furniture inside and a wall that limits, but does not block, the LOS condition. The trend of the entropy is similar to the previous environment: the higher bandwidth exhibit a slightly larger entropy, but it diminishes as the directivity of the antenna increases (see Fig. 5.14). This values are slightly larger than the previous environment. Furthermore, the entropy similarity between the two bandwidths can be caused by a limitation of the method used to compute it: the procedure described in section 4.2.2.1 provides just a rough evaluation of the entropy, which seems to be able only to provide a general indication on the trend of the actual value of entropy. Moreover, looking at the random keys (see Fig. 5.13), the percentages reach higher peaks, especially with a central frequency of $2.4\,\mathrm{GHz}$: this can be caused by the environment which limits the power of the LOS component, therefore creating a lot of fluctuations in the channel, which instead are limited in case of higher frequency due to the higher attenuation. In case of the directive antennas, the same statements as before still hold regarding the high percentage of random keys.



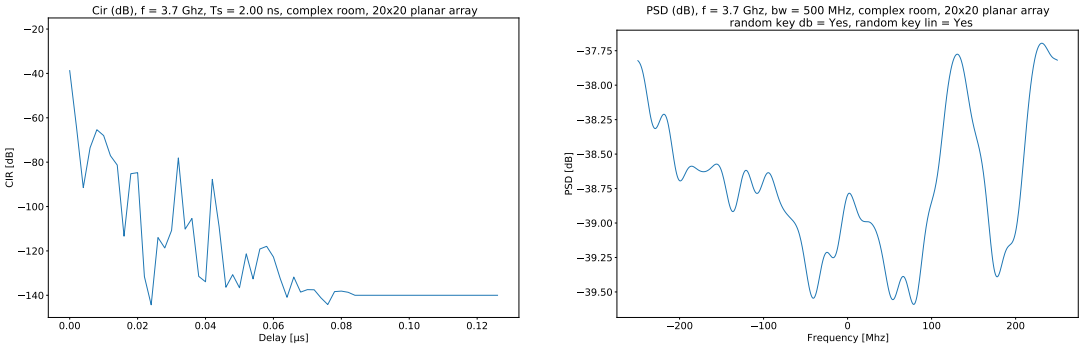
(a) *Channel Impulse Response in dB.* ***(b)*** *Power Spectral Density of the channel in dB*

Figure 5.10: *Channel with the dipole, at* $3.7\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$*, in the* ***complex room****. In the plot there is also the indication if the key is random or not.*



(a) *Channel Impulse Response in dB.* ***(b)*** *Power Spectral Density of the channel in dB*

Figure 5.11: *Channel with the* 5×5 *planar array, at* $3.7\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$*, in the* ***complex room****. In the plot there is also the indication if the key is random or not.*



(a) *Channel Impulse Response in dB.* ***(b)*** *Power Spectral Density of the channel in dB*

Figure 5.12: *Channel with the* 20×20 *planar array, at* $3.7\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$*, in the* ***complex room****. In the plot there is also the indication if the key is random or not.*

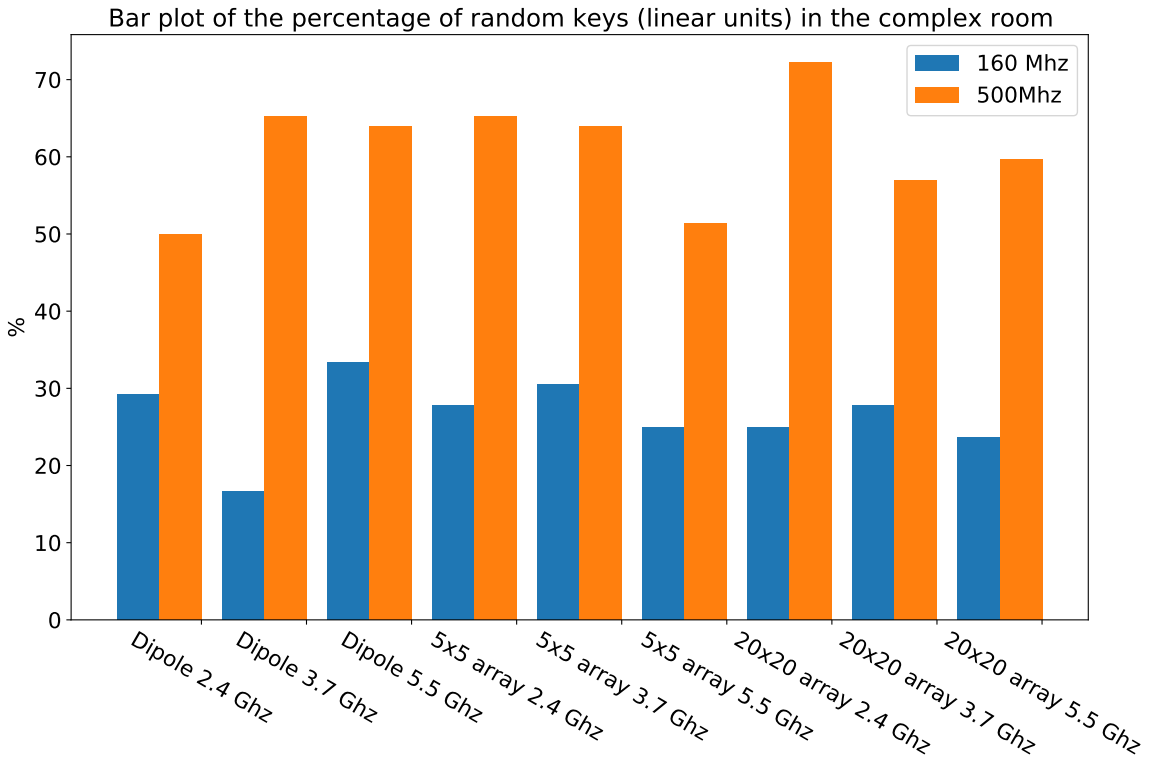


Figure 5.13: *Bar plot of the percentage of random keys in the* ***complex room****, extracted from the channel in linear unit*

Figure 5.14: *Bar plot of the entropy of* ***complex room***

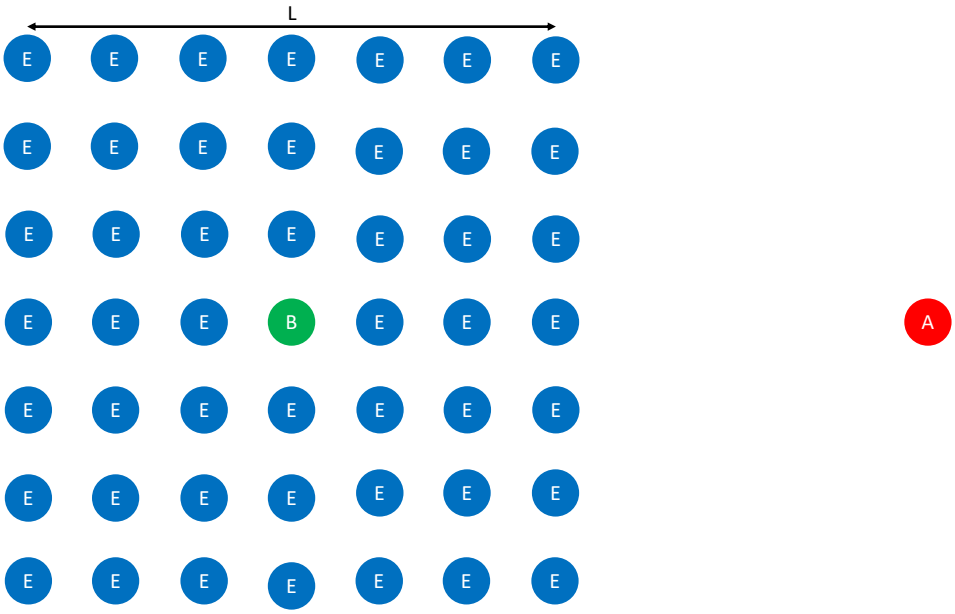


Figure 5.15: *Scheme of the grid of Eves around Bob.*

5.6 Spatial decorrelation assessment

The goal is to evaluate the similarity between the Alice-Bob channel and the Alice-Eve channel: Eve's positions are spread over a square grid of side equal to L, with N Eves per side (see Fig. 5.15). The similarity is assessed through:

- **Root Mean Square Error** between the Alice-Bob channel and the Alice-Eve channels.
- **Key Disagreement Rate** between Bob and Eve: for each Alice-Eve channel the key is extracted according to the filter-bank model and the KDR with respect to the Alice-Bob key.

Moreover, N is preferable to be an odd number, so that one Eve falls in the exact position of Bob, in order to confirm the correctness of the comparison. In fact, that Eve position will have a RMSE equal to 0 and a KDR equal to 0.

As for the simulations, the ray tracing allows to define multiple receivers in the environment, which allows to reduce the computational time with respect to launch-

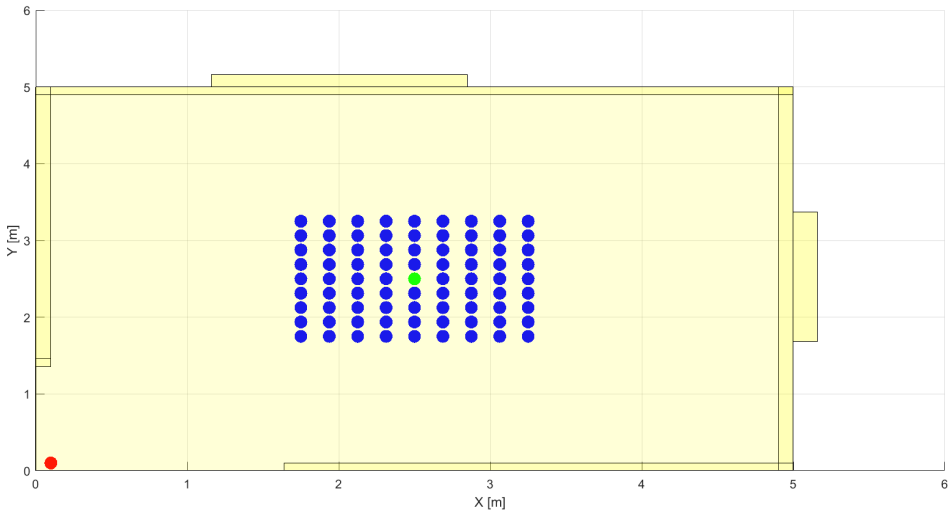


Figure 5.16: *Scheme of the grid of receivers (Eves) in the* ***empty room****.*

ing all the single simulation of the Alice-Eve channel. Simulations are repeated for the three antennas, the three frequencies and two bandwidths. Bob and Eve always utilize the dipole, while Alice changes her antenna: in case of the planar array, it is rotated to point toward Bob, therefore Eve will see the channel through one of the side lobe of the antenna. As for the spatial distribution of Eves, first $N = 9$ and $L = 1.5\,\mathrm{m}$ are considered: in this case, the spacing is $16\,\mathrm{cm}$ along each axis, thus it is larger than $\lambda/2$. The, a second grid with $N = 5$ and $L = 4\,\mathrm{mm}$, hence the spacing is $1\,\mathrm{mm}$. The results are presented in two forms:

- Spatial map: a graphical map with the values of the KDR and RMSE in the Eves' positions with respect to Bob (in the middle of the grid). For this graph, only results with the dipole and the 20×20 array are presented, and only at $3.7\,\mathrm{GHz}$ are shown.

- Cumulative Distribution Functions of the KDR and RMSE, used to compare their distribution with respect to the frequency and bandwidth. Moreover, there are the CDF at $3.7\,\mathrm{GHz}$ and a bandwidth of $500\,\mathrm{MHz}$ with respect to the different antennas.

5.6.1 Spatial assessment in the empty room

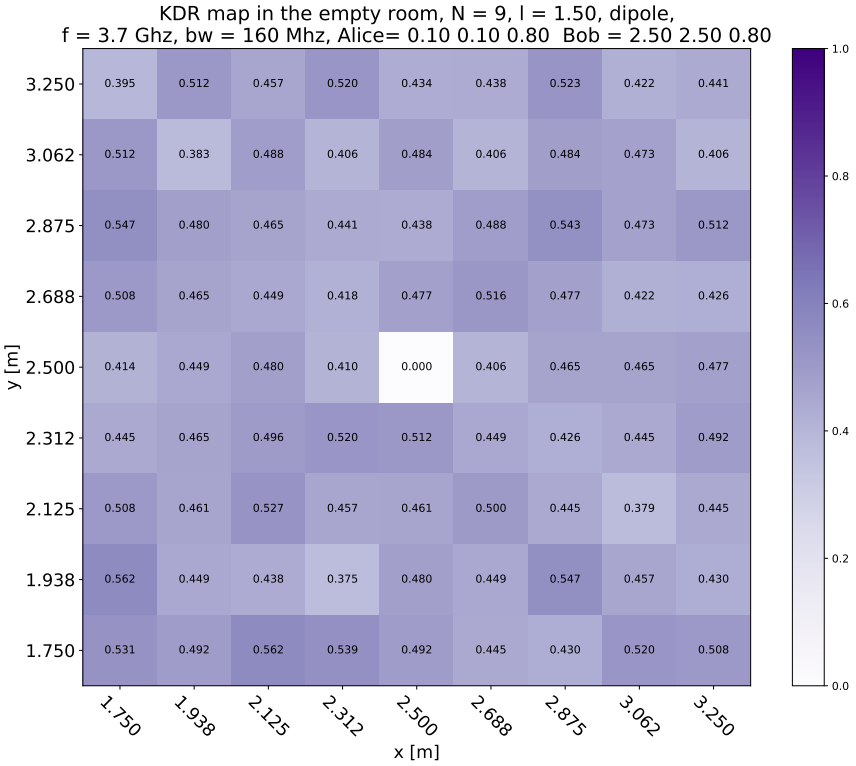
Users are spread in the room as shown in Fig. 5.16: Alice (red dot) is placed in the corner of the room, while Bob (green dot) is in the middle with Eves (blue dots) surrounding him. Eves are $16\,\mathrm{cm}$ distant along x and y. In Fig. 5.17 and Fig. 5.18 it is possible to see the grid of the KDR and the RMSE with the dipole and with both bandwidth: the central point shows the maximum value, since it is in the same position of Bob and the Ray Tracing is a deterministic tool. In spite of the bandwidth

utilized, looking at the KDR, the keys generated are very different, meaning that the channels seen from Eve around Bob are uncorrelated with respect to the Alice-Bob channel, showing that a possible eavesdropper around Bob can hardly get similar channel. In addition, in case an attacker wants to launch a Ray Tracing Attack, a imprecise knowledge of the position of Alice and/or Bob brings an additional difficulty to carry out the attack. This result is also confirmed by the values of the RMSE: although close to Bob the value is quite low (2 or 3 dB), the channels are still enough uncorrelated to generate different keys.

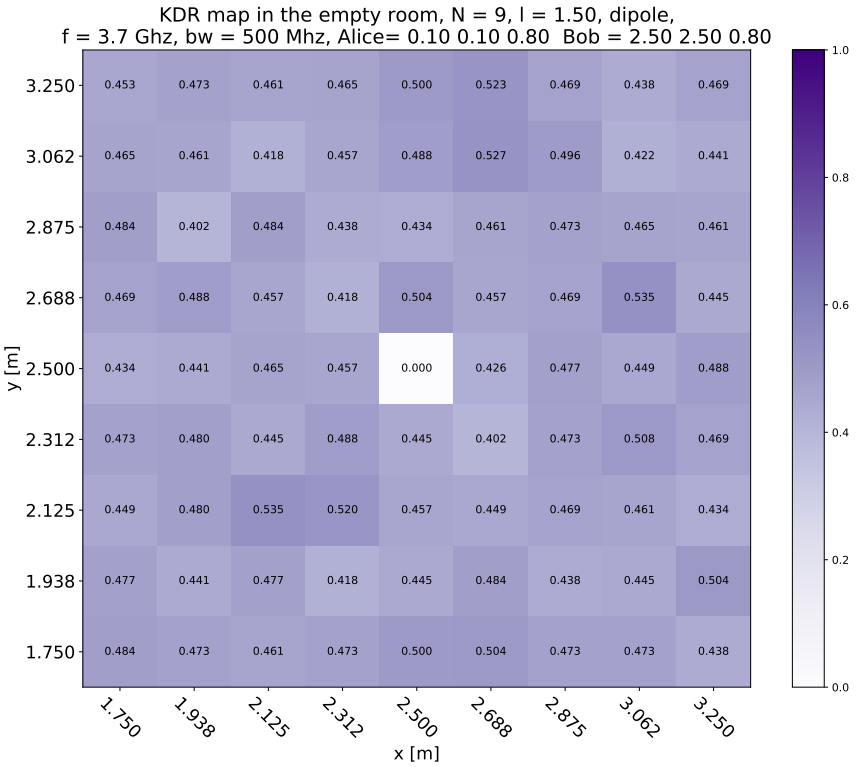
As the channels with the dipole are uncorrelated, the same can not be said about the channels with the planar array. Looking at Fig. 5.19 and Fig. 5.20, in particular at the RMSE maps, it is possible to see that the channels around Bob are quite similar. This is caused by the lack of dense multipath and the presence of a strong component: both characteristics tend to generate similar channels, with a low dynamic, which result in a low RMSE (remember that it is computed from the normalized channels). Regardless of the RMSE, the KDR maintains the same spatial trend and it is always greater than 0.3. Therefore, in this environment, even though the Alice-Eve channels around Bob are similar with respect to the Alice-Bob (due to similar propagation characteristics), the keys generated are different. However, it is important to remind the disadvantages of using high directivity antennas, as explained in section 5.4 and 5.5.

Additional considerations can be made by looking at the CDF of the KDR and RMSE with different link characteristic (Fig. 5.38, Fig. 5.39, Fig. 5.40): looking at the KDR distribution, the KDR always oscillates between 0.4 and 0.6, regardless of the link feature (frequency and bandwidth) and the trend in maintained for the different antennas: therefore, neither the link nor the antenna features have an impact on the KDR distribution. Instead, the trend of the RMSE changes based on the antenna employed: with the dipole (Fig. 5.25) the values are high (in particular at $2.4\,\mathrm{GHz}$ for both the bandwidths), they are grater for the higher bandwidth. As for the 5×5 planar array, the distributions are more concentrated with respect to the previous case, but on average the values are smaller: the dynamic of the channel is smaller, thus it is probable to have similar channels, even though the KDR remains always around 50%. In case of the 20×20 array is considered (Fig. 5.27), the CDFs show a wide distribution of the RMSE: the directivity of the antenna in high and it is always pointed toward Bob, hence Eve sees the channel through a side lobe and it is reasonable to think that the propagation happens mainly thanks to the multipath and the RMSE is high. Moreover, the point in which the RMSE is low (Fig. 5.20) are along the diagonal of the square of Eves, that might be point in which the propagation conditions are similar. Furthermore, by looking at the KDR with respect to the antenna (Fig. 5.28), the dipole and the 5×5 planar array show similar behaviour, since their propagation conditions might be similar. Instead, the 20×20 shows a general higher values of KDR closer to 0.5.

In case the distance between Eve and Bob is very small, the channels are of course similar. In case of the dipole (Fig. 5.21 and Fig. 5.22) the KDR shows small values, in particular along the diagonal, whereas in other positions the KDR is higher, but does not reach values close to 0.5. This trend is confirmed also by the RMSE. Instead, with the 20×20 array (Fig. 5.23 and Fig. 5.24) the channels, even though Eves are very close to Bob, are similar only along the diagonal, while in the the other positions the KDR reaches values close to 0.5: this is due to the narrow beam of the planar array.

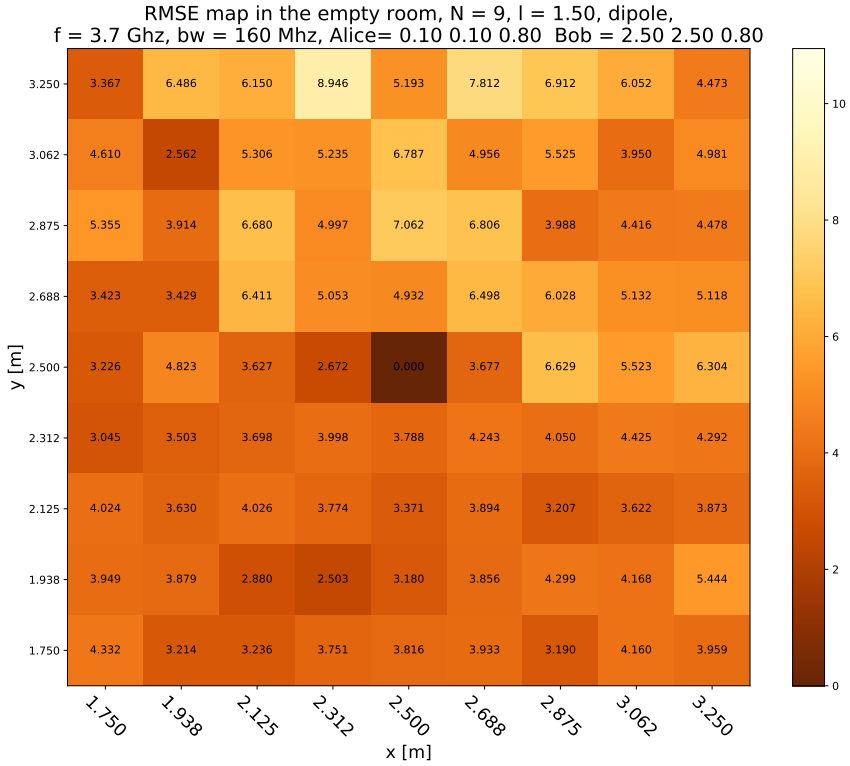


(a) *Key Disagreement Rate map with a bandwidth of* 160 MHz

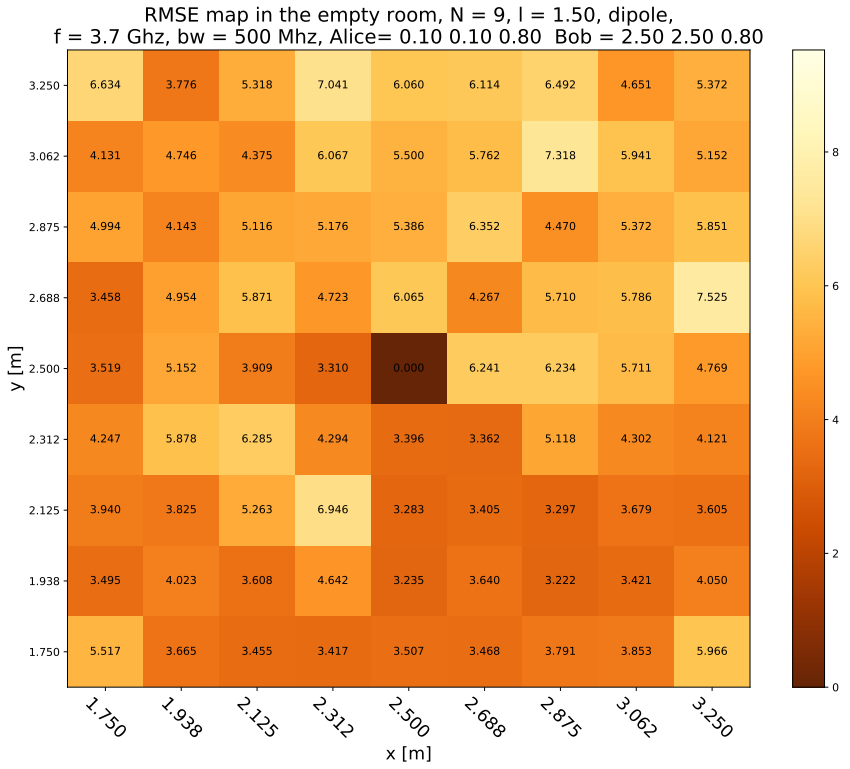


(b) *Key Disagreement Rate map with a bandwidth of* 500 MHz

Figure 5.17: *Spatial map of the KDR of the* ***empty room*** *with the* ***dipole****. Abscissa and ordinate are the positions of Eves.*

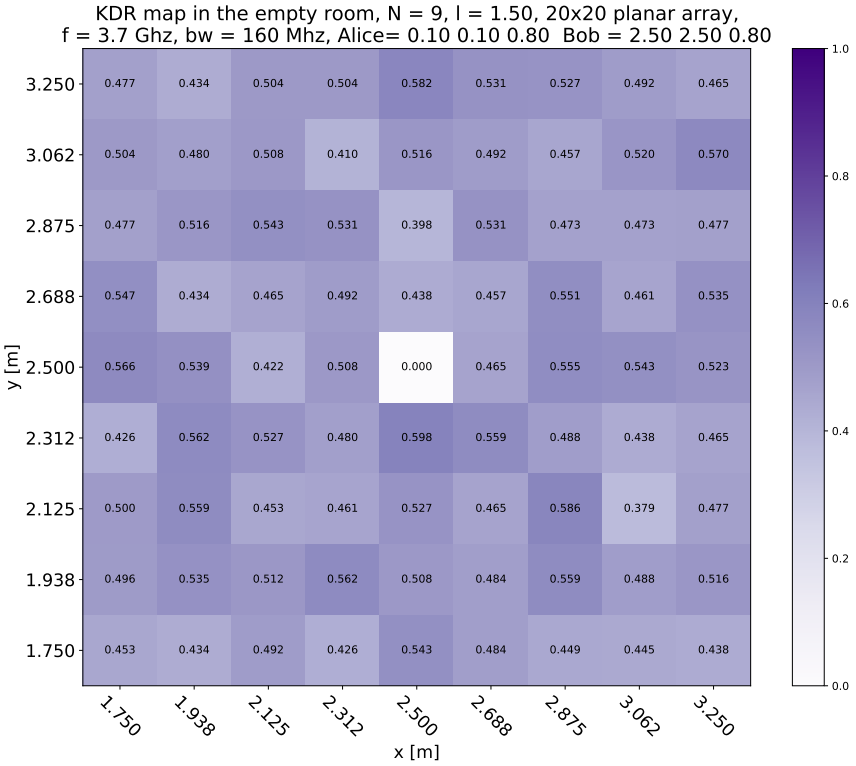


(a) *Root Mean Square Error map with a bandwidth of* 160 MHz

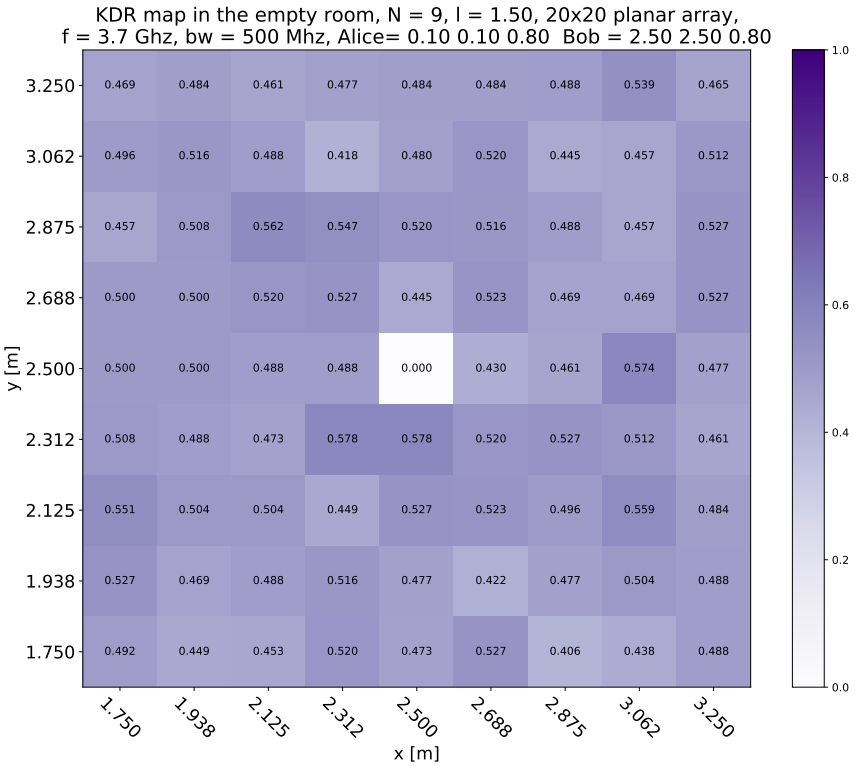


(b) *Root Mean Square Error map with a bandwidth of* 500 MHz

Figure 5.18: *Spatial map of the RMSE of the* ***empty room*** *with the* ***dipole****. Abscissa and ordinate are the positions of Eves.*

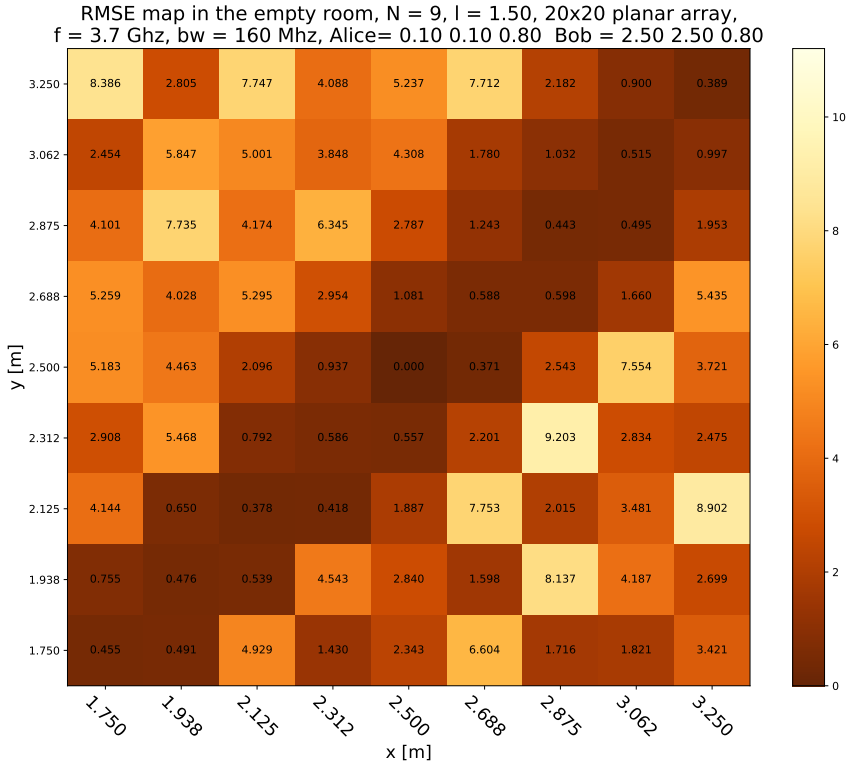


(a) *Key Disagreement Rate map with a bandwidth of* $160\,\mathrm{MHz}$

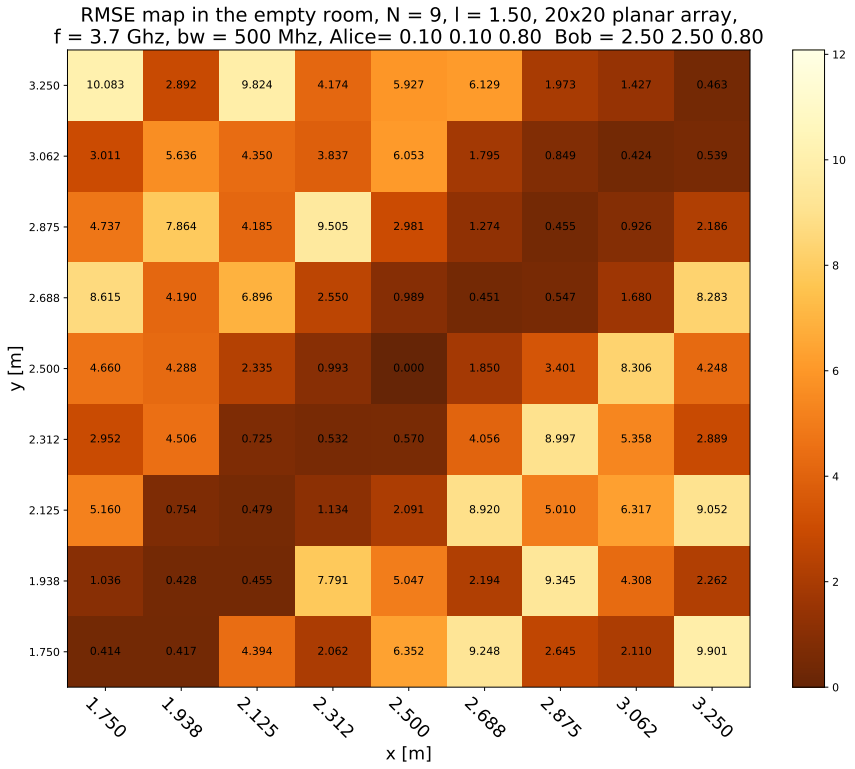


(b) *Key Disagreement Rate map with a bandwidth of* $500\,\mathrm{MHz}$

Figure 5.19: *Spatial map of the KDR of the* ***empty room*** *with the* 20×20 ***array***. *Abscissa and ordinate are the positions of Eves.*

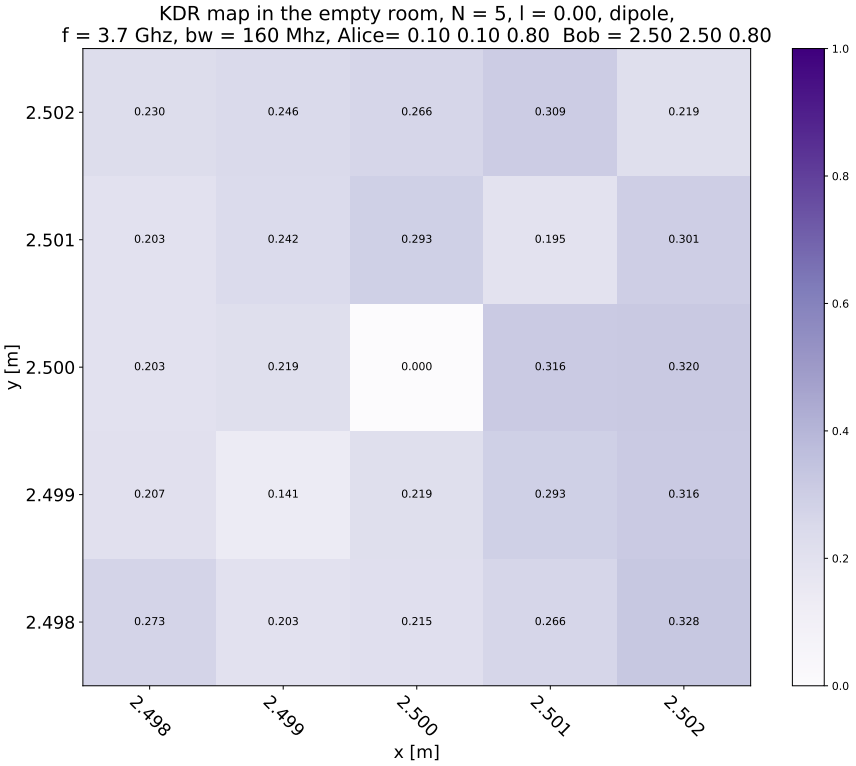


(a) *Root Mean Square Error map with a bandwidth of* $160\,\mathrm{MHz}$

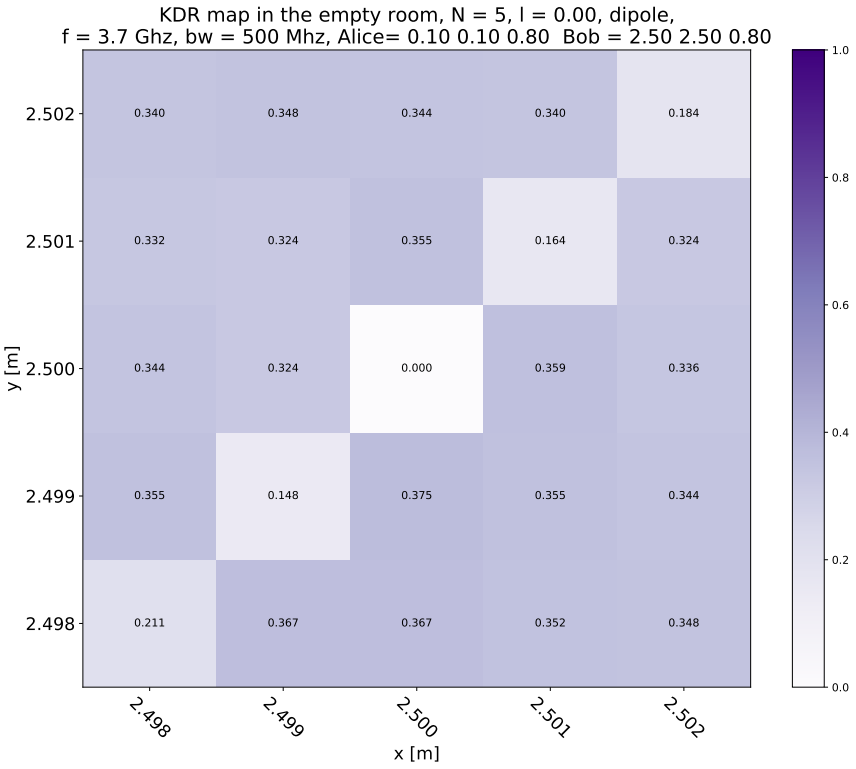


(b) *Root Mean Square Error map with a bandwidth of* $500\,\mathrm{MHz}$

Figure 5.20: *Spatial map of the RMSE of the* ***empty room*** *with the* 20×20 ***array****. Abscissa and ordinate are the positions of Eves.*

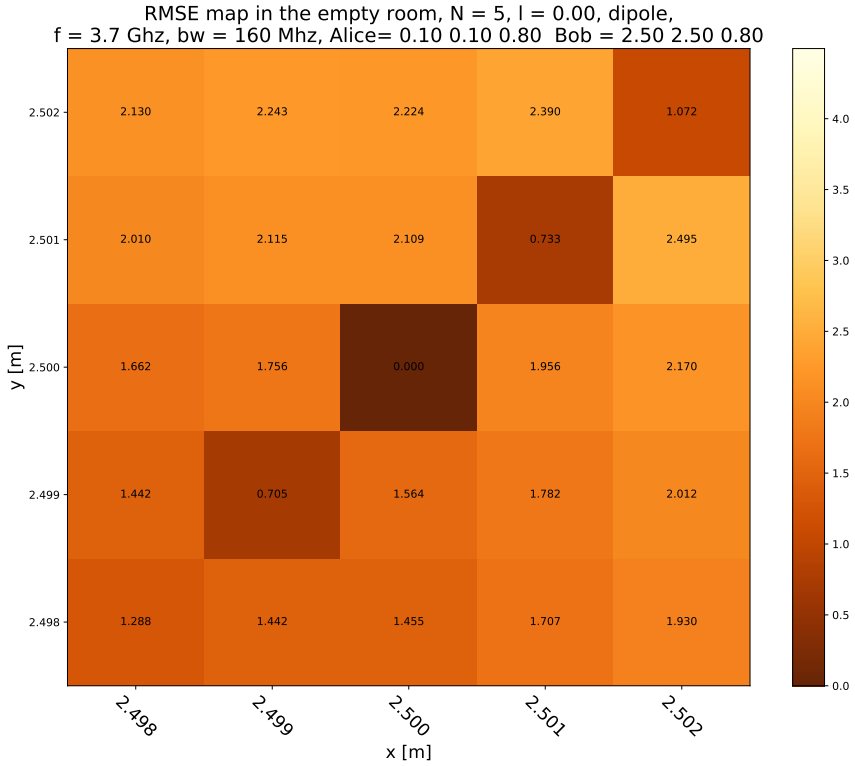


(a) *Key Disagreement Rate map with a bandwidth of* 160 MHz

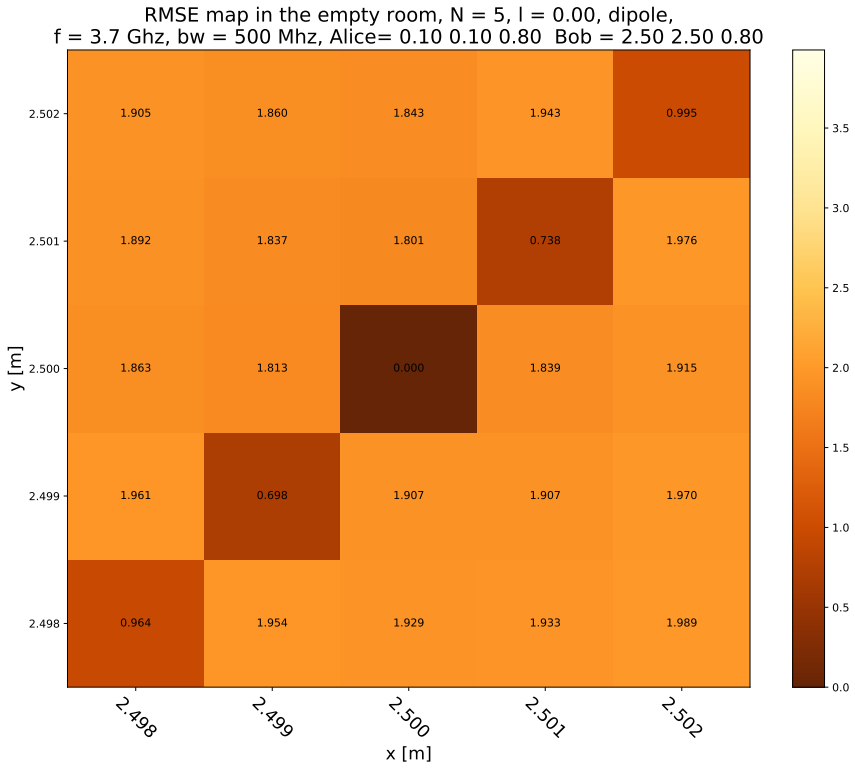


(b) *Key Disagreement Rate map with a bandwidth of* 500 MHz

Figure 5.21: *Spatial map of the KDR of the* ***empty room*** *with the* ***dipole****, distance of* 1 mm. *Abscissa and ordinate are the positions of Eves.*

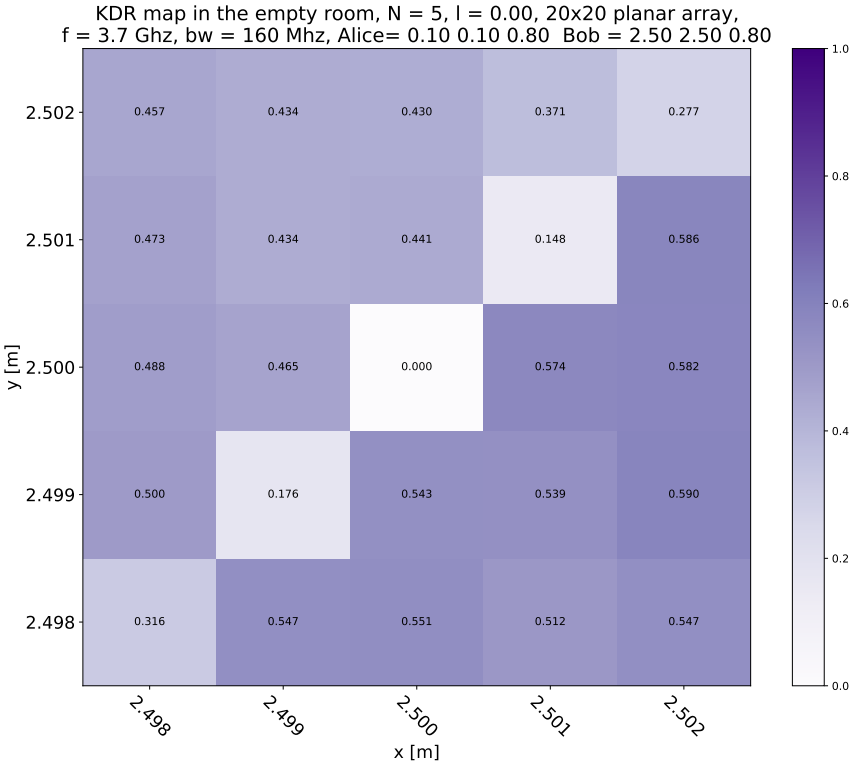


(a) *Root Mean Square Error map with a bandwidth of* 160 MHz

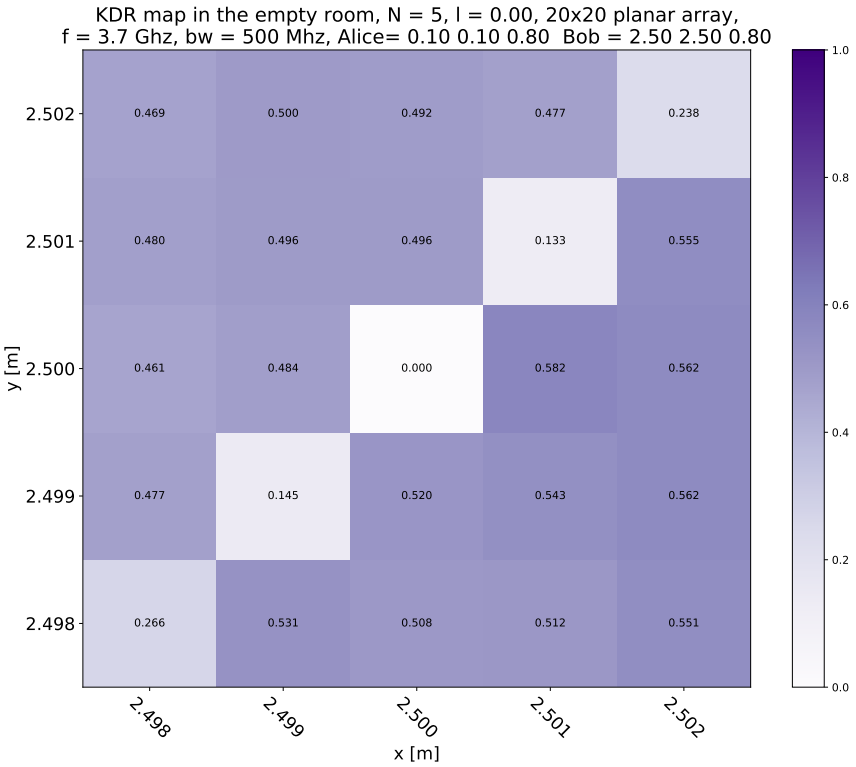


(b) *Root Mean Square Error map with a bandwidth of* 500 MHz

Figure 5.22: *Spatial map of the RMSE of the* ***empty room*** *with the* ***dipole****, distance of* 1 mm. *Abscissa and ordinate are the positions of Eves.*

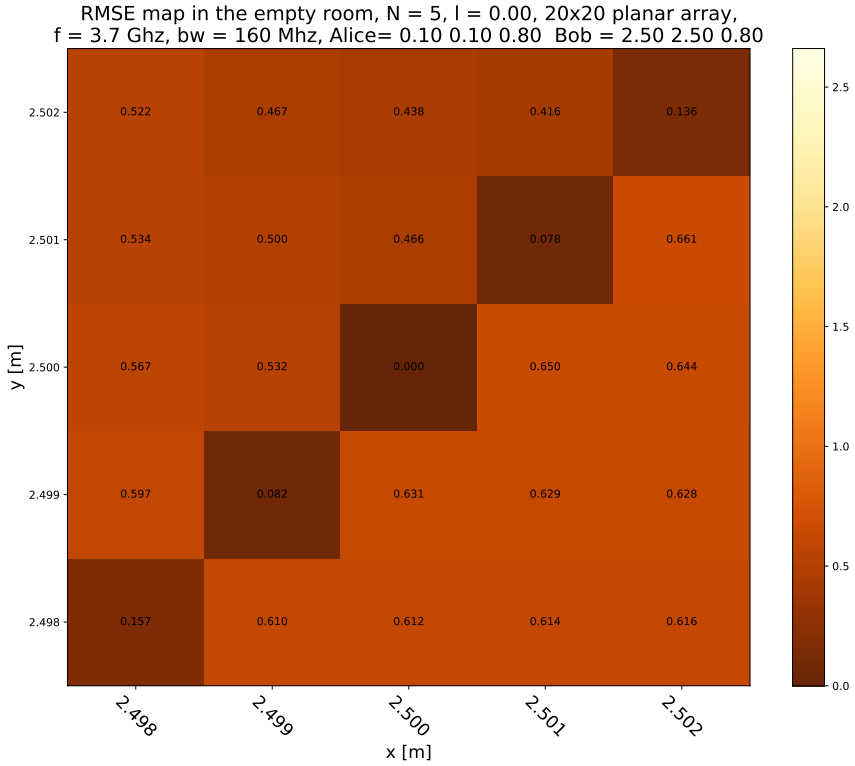


(a) *Key Disagreement Rate map with a bandwidth of* $160\,\mathrm{MHz}$

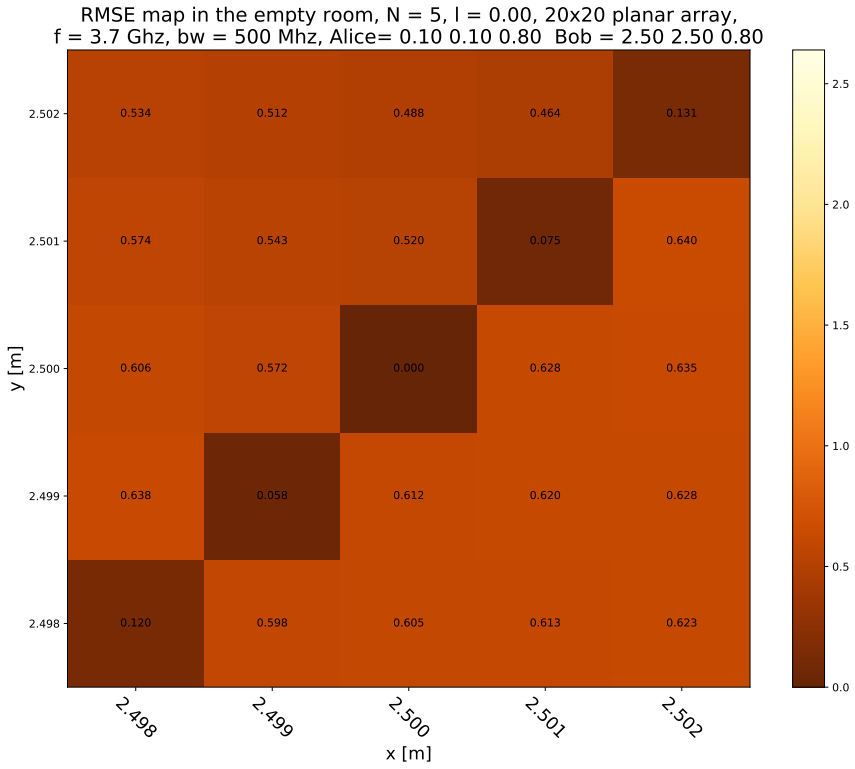


(b) *Key Disagreement Rate map with a bandwidth of* $500\,\mathrm{MHz}$

Figure 5.23: *Spatial map of the KDR of the* ***empty room*** *with the* 20×20 ***array****, distance of* $1\,\mathrm{mm}$*. Abscissa and ordinate are the positions of Eves.*

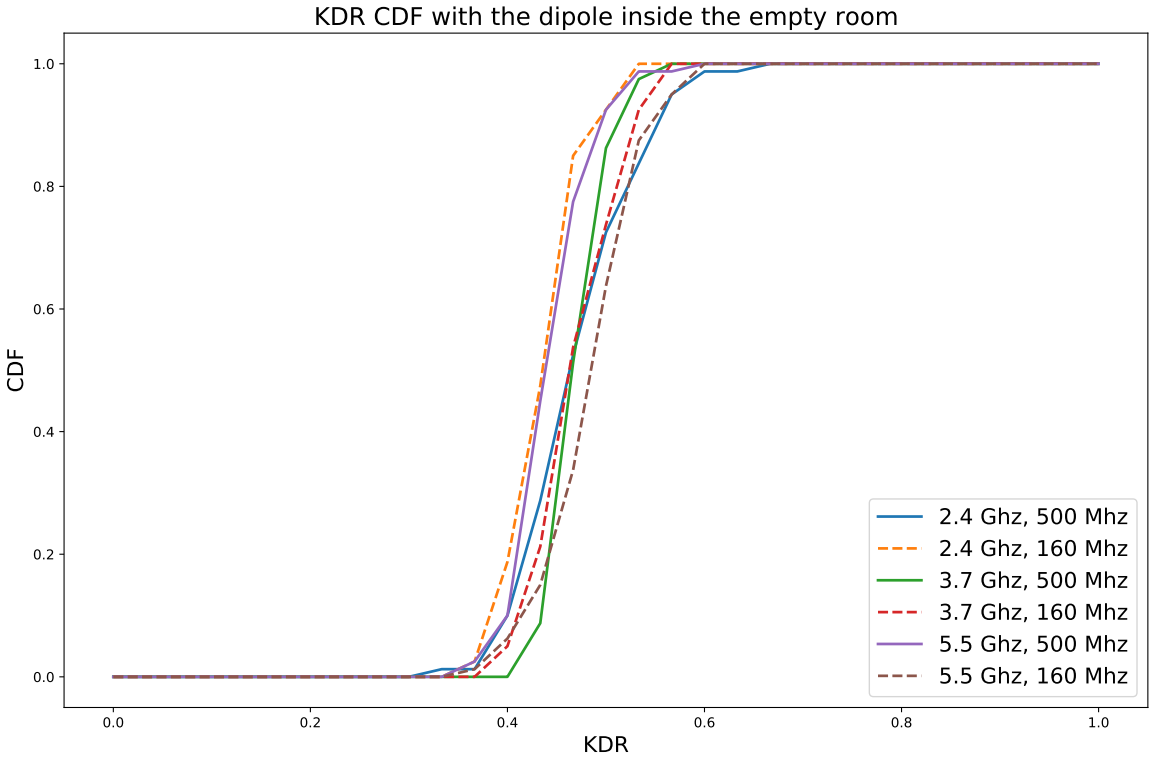


(a) *Root Mean Square Error map with a bandwidth of* 160 MHz

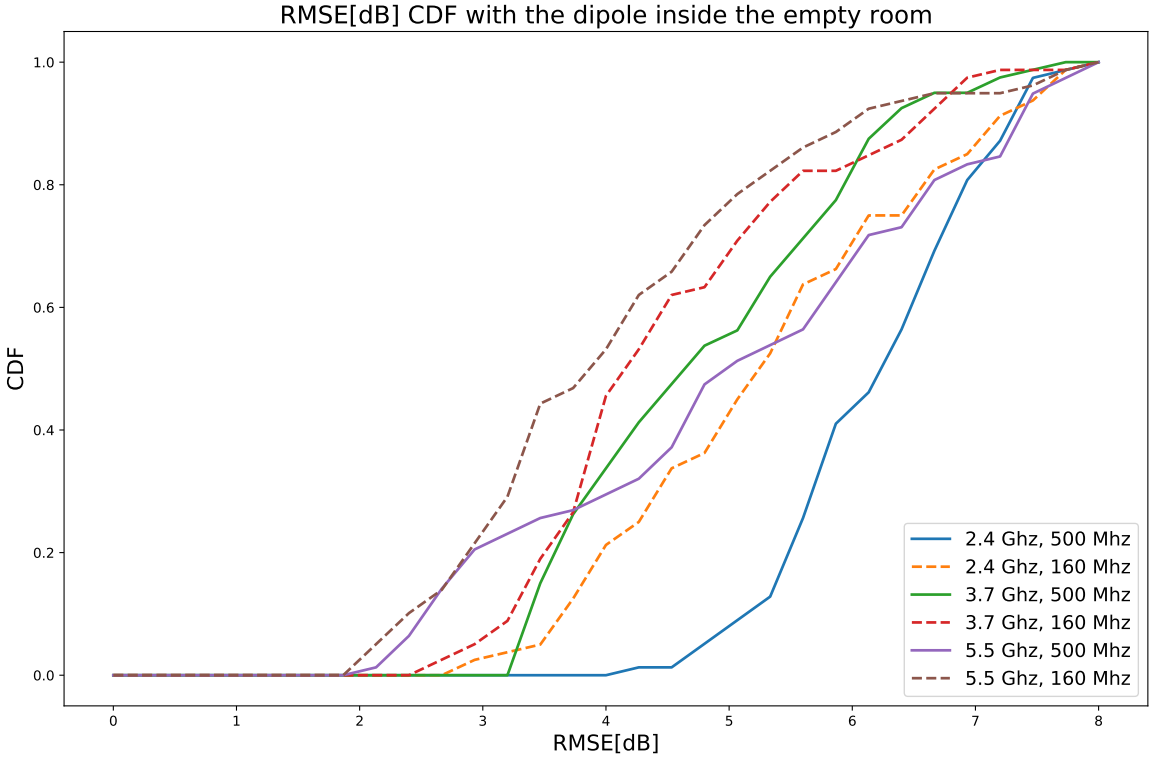


(b) *Root Mean Square Error map with a bandwidth of* 500 MHz

Figure 5.24: *Spatial map of the RMSE of the* ***empty room*** *with the* 20×20 ***array****, distance of* 1 mm*. Abscissa and ordinate are the positions of Eves.*

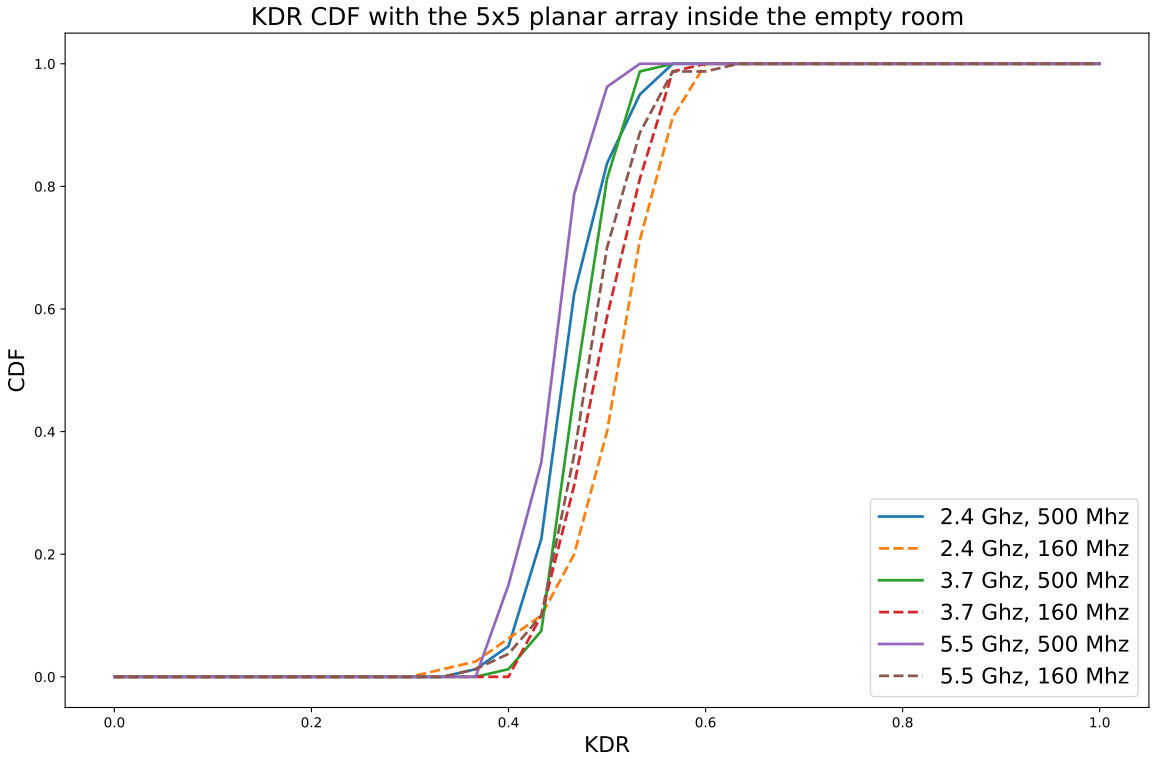


(a) *CDF of the KDR*

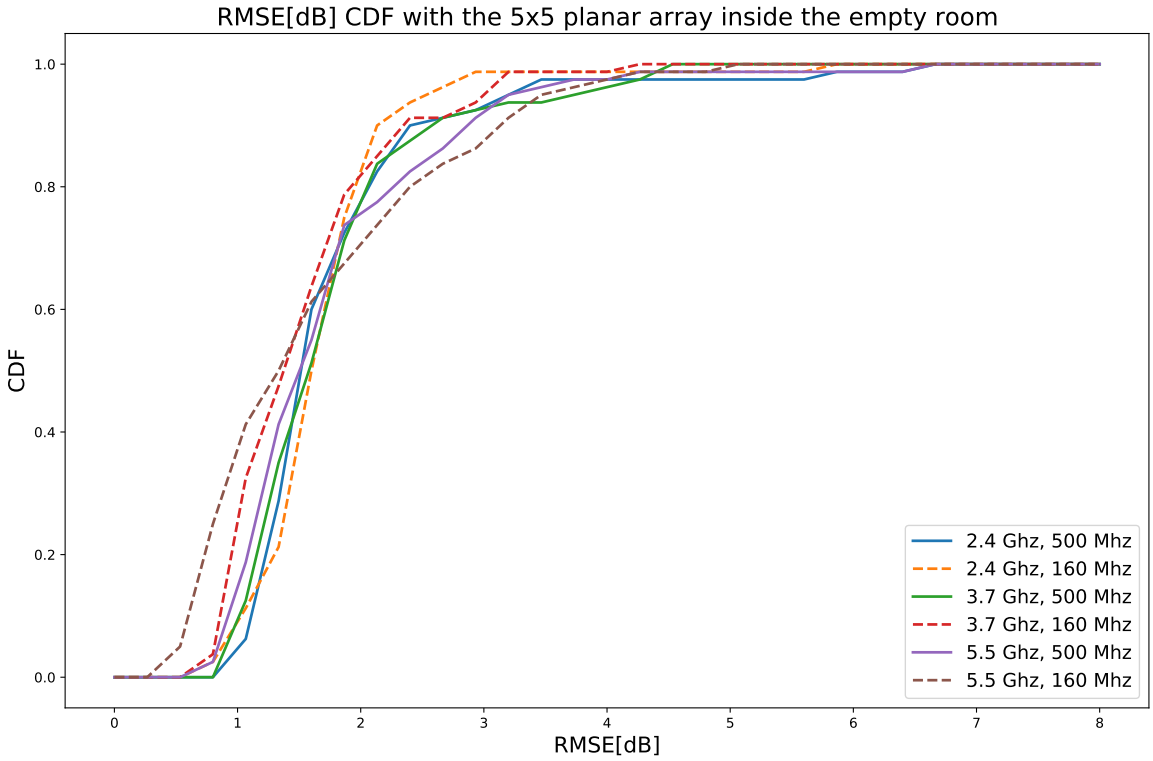


(b) *CDF of the RMSE*

Figure 5.25: *Cumulative Distribution Function of the KDR and RMSE map, in the* ***empty room*** *with the* ***dipole****.*

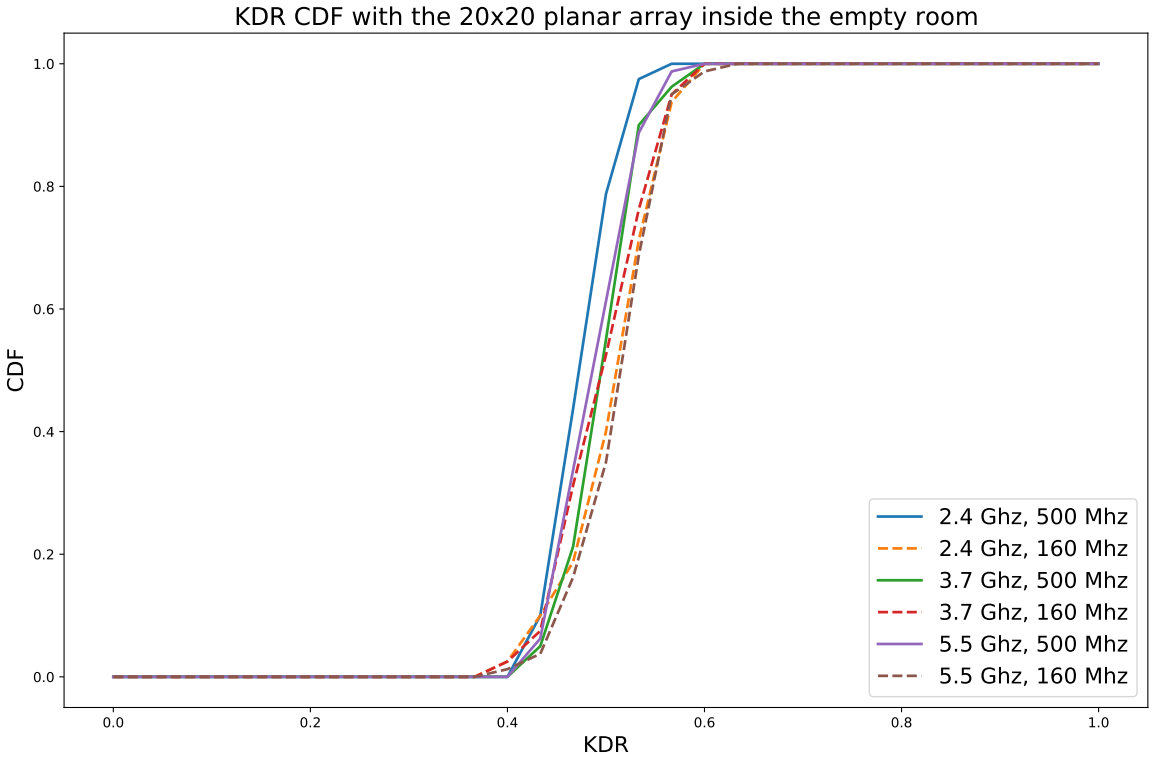


(a) *CDF of the KDR*

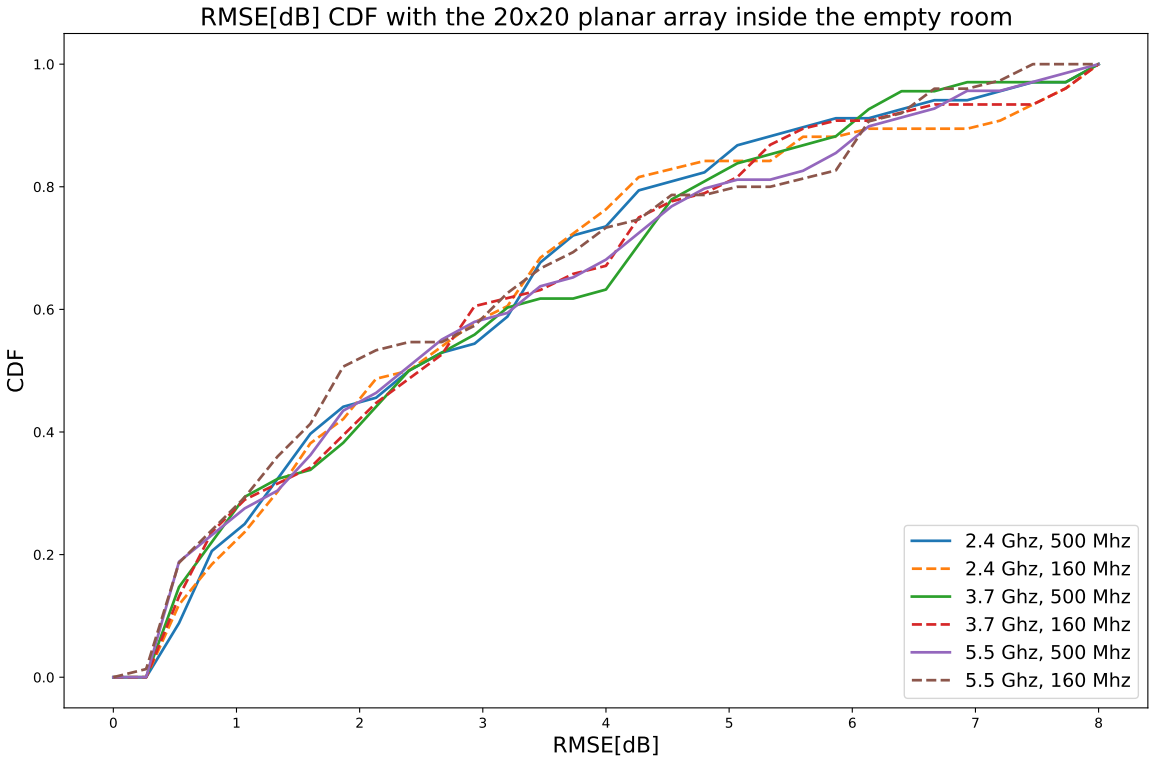


(b) *CDF of the RMSE*

Figure 5.26: *Cumulative Distribution Function of the KDR and RMSE map, in the* ***empty room*** *with the* 5×5 ***array****.*



(a) *CDF of the KDR*



(b) *CDF of the RMSE*

Figure 5.27: *Cumulative Distribution Function of the KDR and RMSE map, in the* ***empty room*** *with the* 20×20 ***array****.*

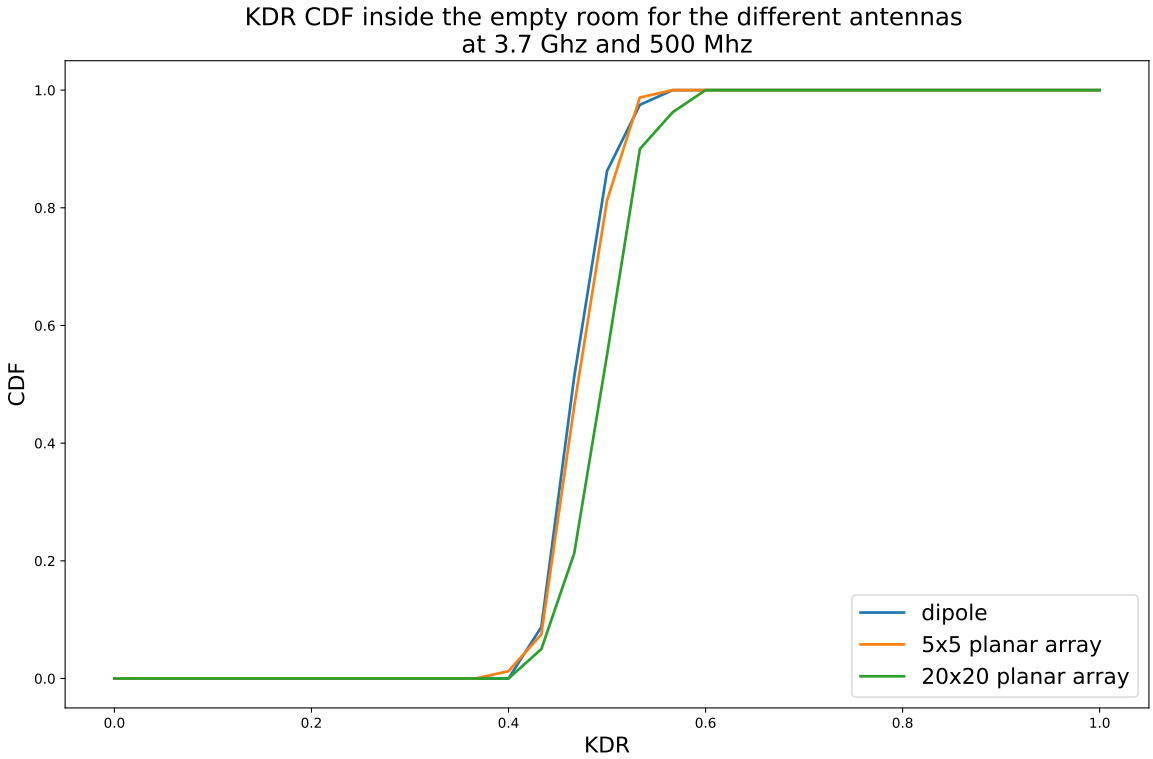


Figure 5.28: *CDF of the KDR in the* ***empty room*** *with respect to the three antennas, at* $3.7\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$

5.6.2 Spatial assessment in the complex room

In Fig. 5.29 there is the scheme of the users: Alice (red dot) is at the end of the room, Bob (green dot) is in front of her behind the wall with Eves (blue dots) around. Things are similar also in the complex room: in Fig. 5.30 and Fig. 5.31 there are the maps of KDR and RMSE with the dipole, in Fig. 5.32 and Fig. 5.33 the maps with the planar array. The channels with the dipole present an high value of the RMSE, the KDR is around 50% and it is uniform inside the environment. In case the directive antenna is considered, the distribution changes a bit. Looking at the RMSE, for both the bandwidths, it is possible to see that the Eves in the rectangle between $y = 3.125$ and $y = 3.875$: these positions are all behind the wall and perhaps the direction still shows an high gain, therefore the channel are easily similar. Despite the RMSE, the KDR does not present the same behaviour. Instead, the positions with the same x of Bob and closer to the bed (higher y) show very similar channels, looking both at the RMSE and at the KDR. In addition, in Fig. 5.32a there is a singular spot in $x = 3.062, y = 3.312$ that shows a low KDR, which might be caused by a similar multipath distribution. However, this situation happens for channels with a low dynamic, so situations in which Physical Layer based-key generation does not work fine.

The KDR CDF with the dipole (Fig. 5.38) shows that regardless of the link features

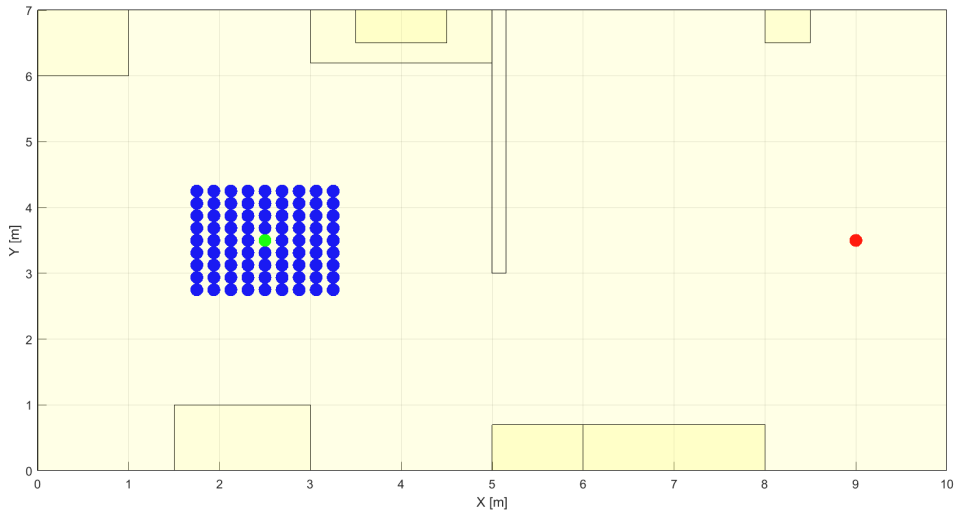
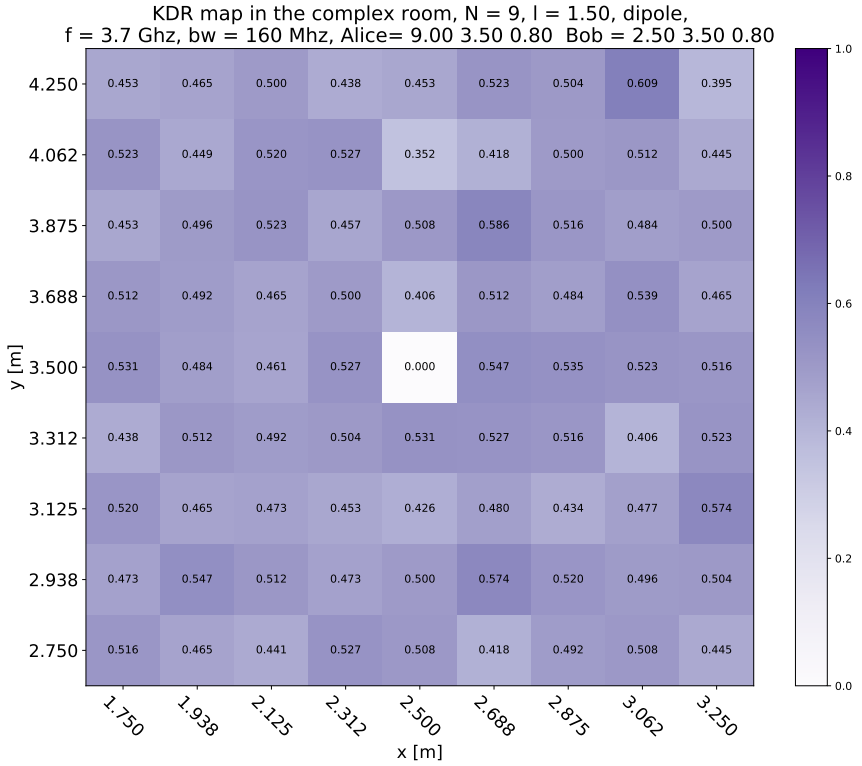


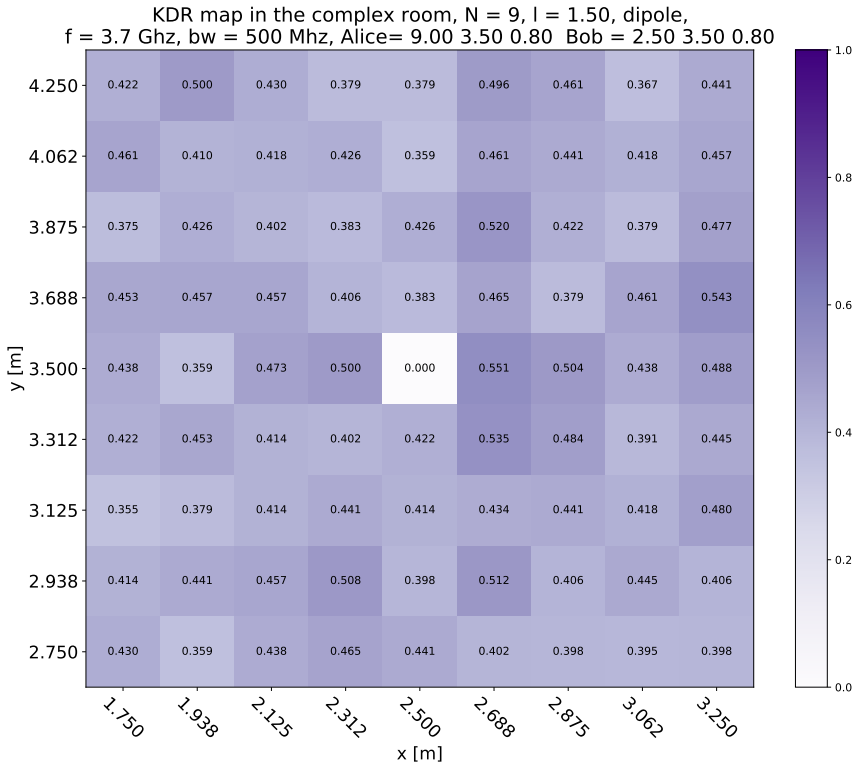
Figure 5.29: *Scheme of the grid of receivers (Eves) in the* ***complex room****.*

the KDR remains around 0.4 or 0.5. Instead, for the RMSE the channels at $5.5\,\mathrm{GHz}$ present the highest mean value of RMSE. The CDFs of the 5×5 array are not compressed as the previous environment: now, the situation with $3.7\,\mathrm{GHz}$ shows higher RMSE with respect to the other frequencies. As for the case with the 20×20 array, the situation is similar with respect to the previous case. Different from the previous case, the KDR with respect to the antennas (Fig. 5.41) shows that when the dipole is employed the KDR has the lowest mean value, whereas as the directivity increases the average value increases.

In this environment, when Eve is very close to Bob and Alice employes the dipole, the KDR (Fig. 5.34) is more or less uniform around 0.25 in both bandwidths, even though some points behind and in front of Bob shows a smaller value. In addition, by looking at the RMSE (Fig. 5.35) the values remains quite low: the channels are indeed similar, but still the quantization makes the keys different. In case the 20×20 array is employed the KDR (Fig. 5.36) shows a pattern made of lines perpendicular to the line connecting Alice and Bob, as the RMSE does (Fig. 5.37). This evaluation is the results of simulations, in a real scenario these trends might not hold anymore, since the quantization is not as sensible as with the computer. In addition, the channel estimation can have some errors and might not be able to recognize small channel variations as the simulation does.

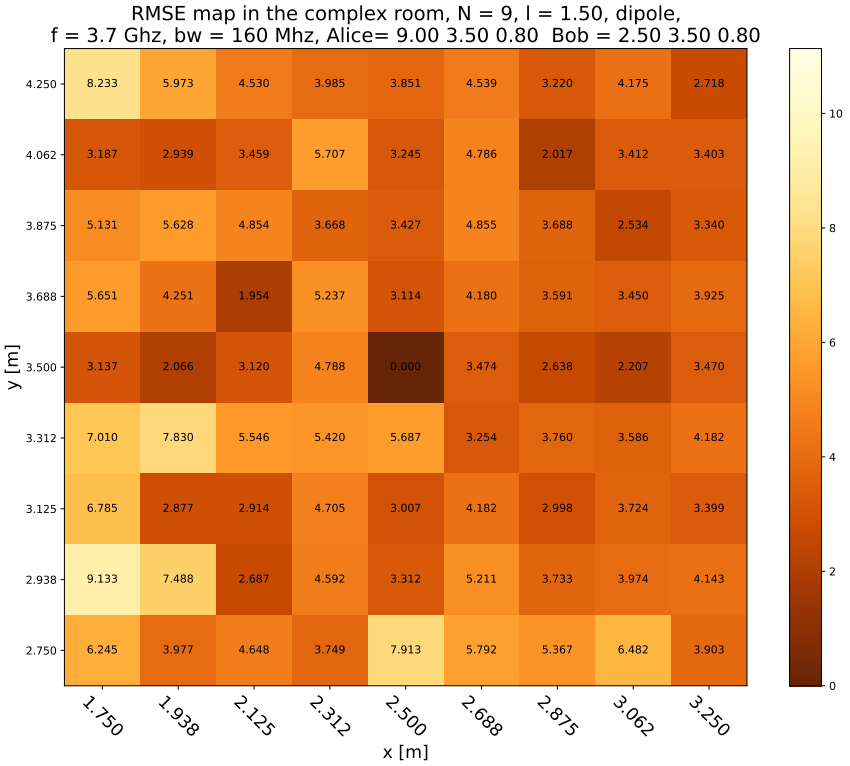


(a) *Key Disagreement Rate map with a bandwidth of* 160 MHz

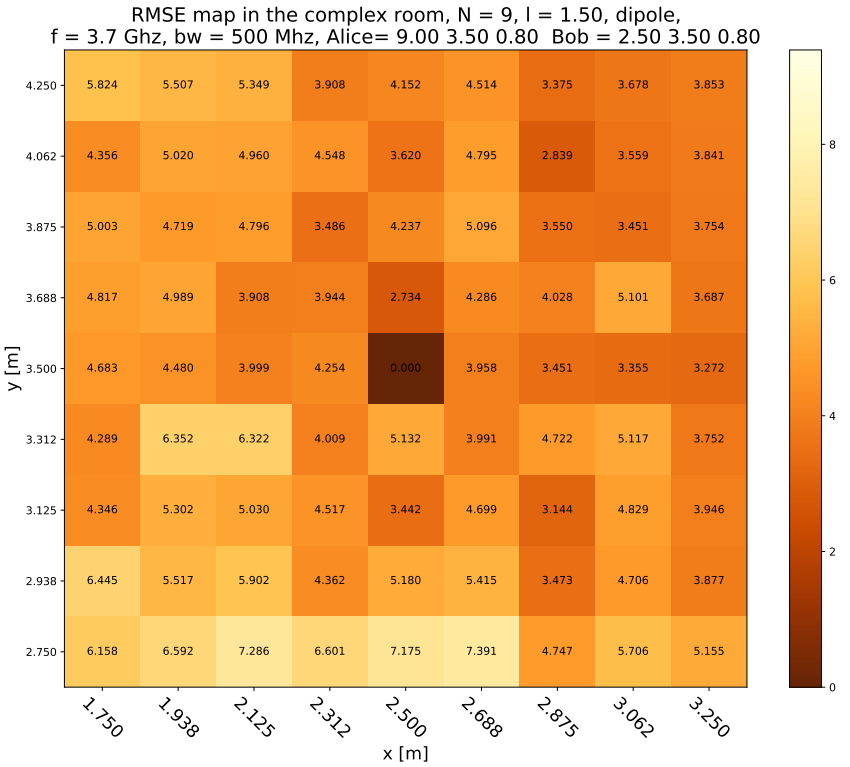


(b) *Key Disagreement Rate map with a bandwidth of* 500 MHz

Figure 5.30: *Spatial map of the KDR of the* ***complex room*** *with the* ***dipole****. Abscissa and ordinate are the positions of Eves.*

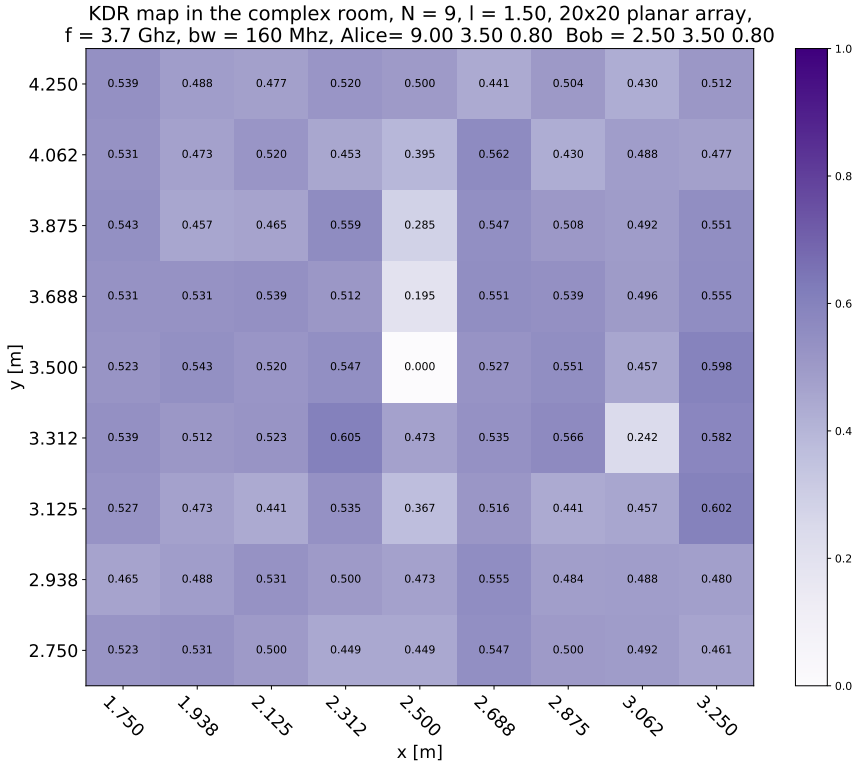


(a) *Root Mean Square Error map with a bandwidth of* 160 MHz

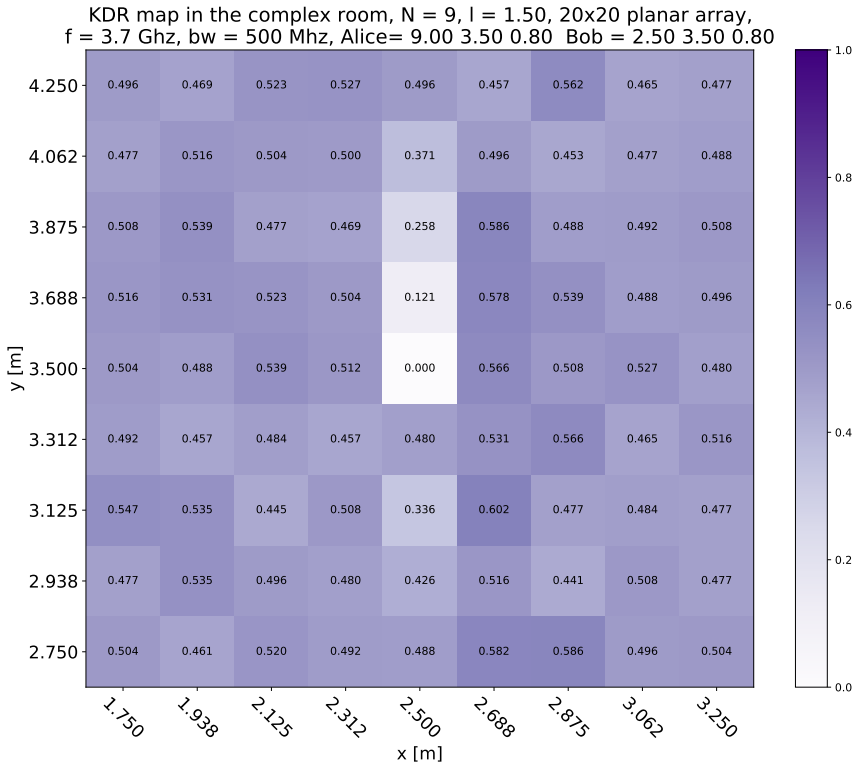


(b) *Root Mean Square Error map with a bandwidth of* 500 MHz

Figure 5.31: *Spatial map of the RMSE of the* ***complex room*** *with the* ***dipole****. Abscissa and ordinate are the positions of Eves.*

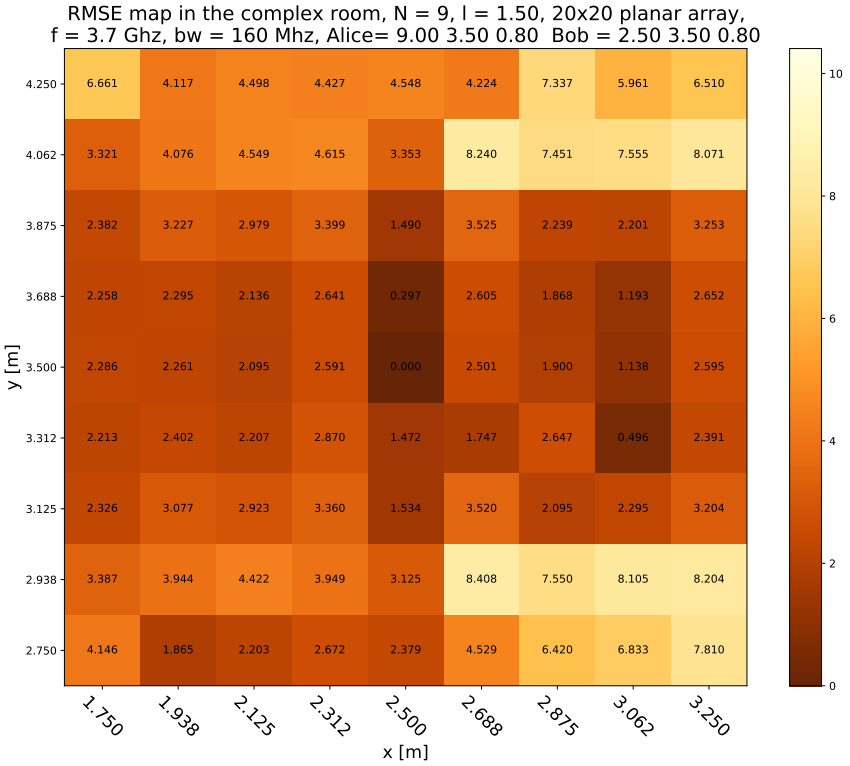


(a) *Key Disagreement Rate map with a bandwidth of* 160 MHz

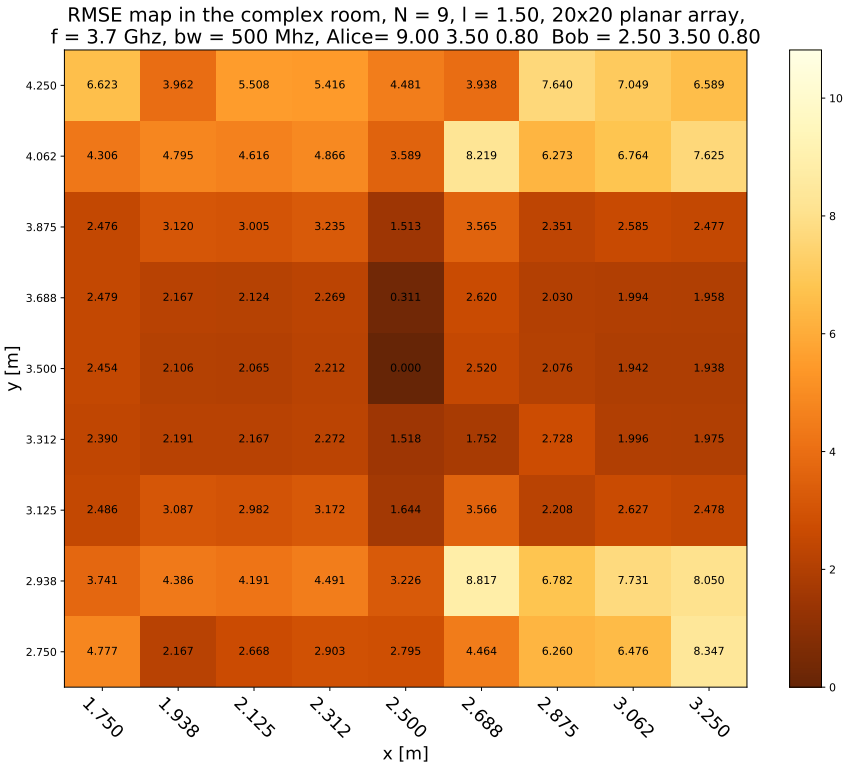


(b) *Key Disagreement Rate map with a bandwidth of* 500 MHz

Figure 5.32: *Spatial map of the KDR of the* ***complex room*** *with the* ***20x20 array****. Abscissa and ordinate are the positions of Eves.*

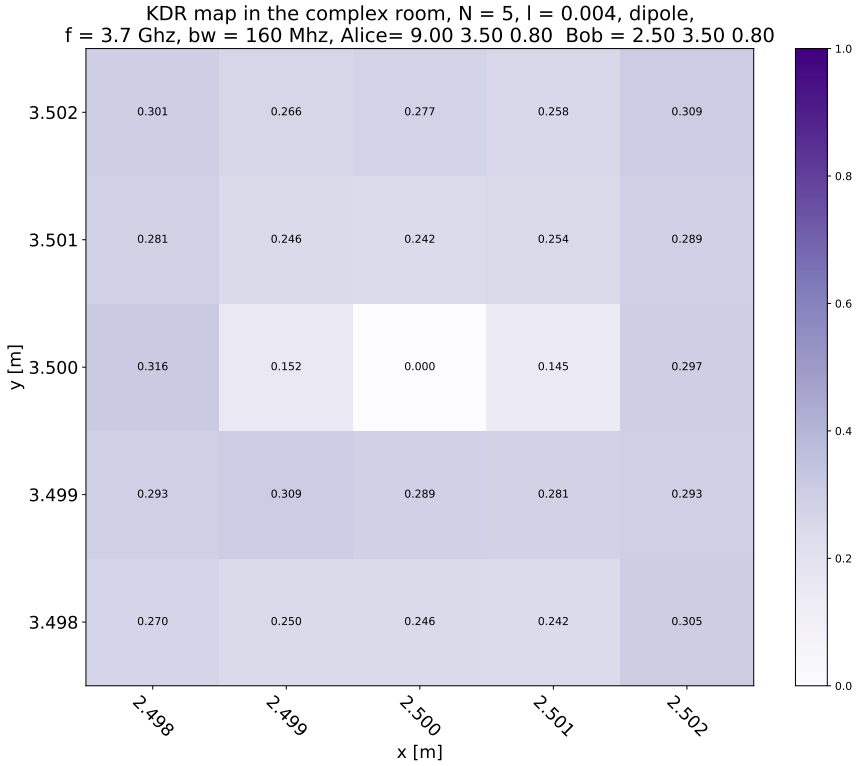


(a) *Root Mean Square Error map with a bandwidth of* 160 MHz

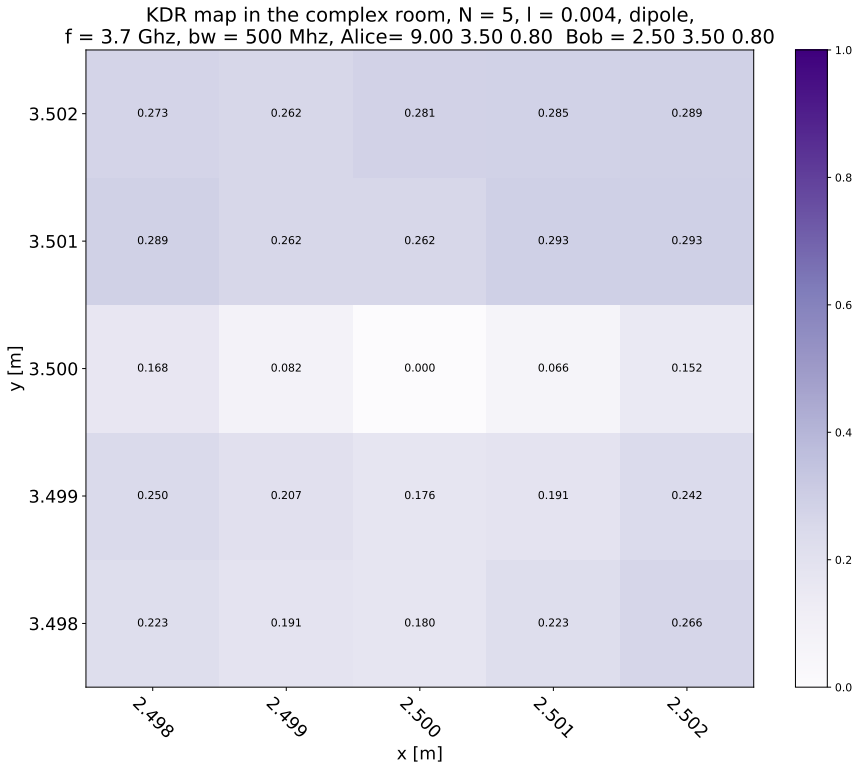


(b) *Root Mean Square Error map with a bandwidth of* 500 MHz

Figure 5.33: *Spatial map of the RMSE of the* ***complex room*** *with the* ***20x20 array****. Abscissa and ordinate are the positions of Eves.*

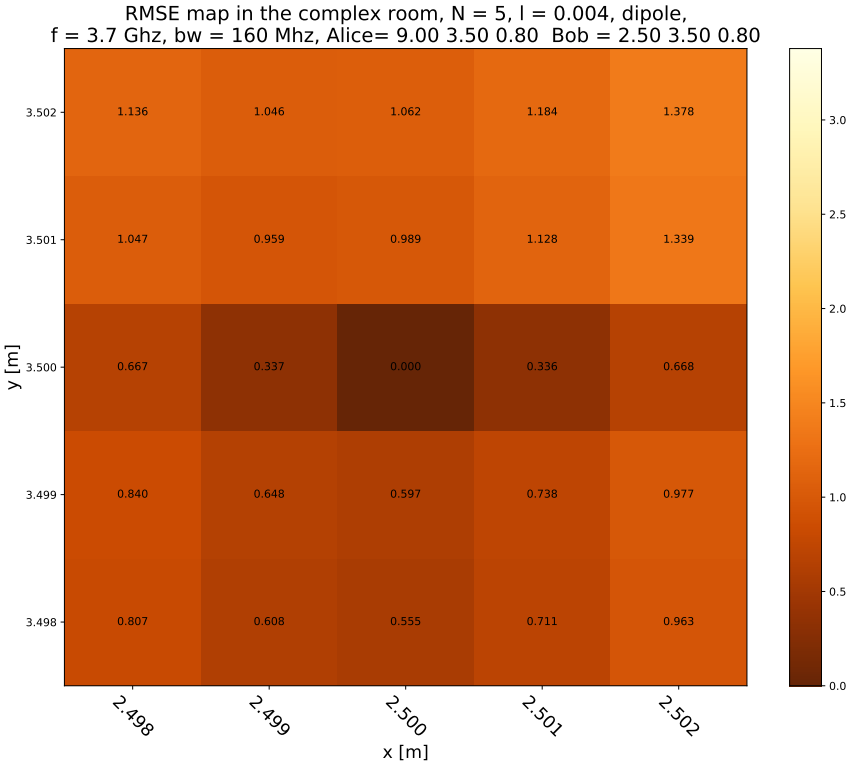


(a) *Key Disagreement Rate map with a bandwidth of* 160 MHz

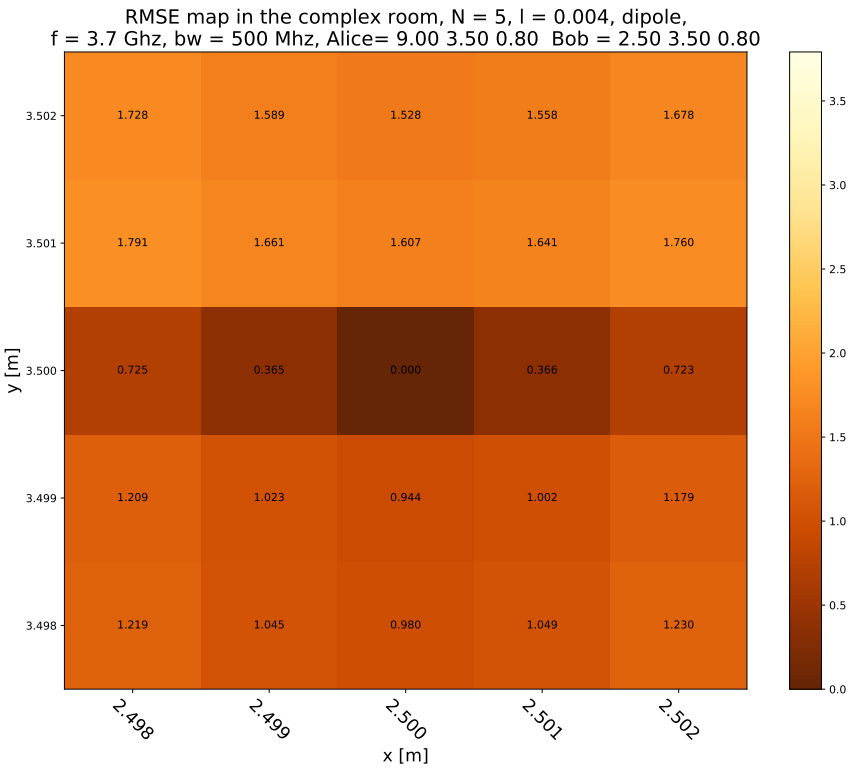


(b) *Key Disagreement Rate map with a bandwidth of* 500 MHz

Figure 5.34: *Spatial map of the KDR of the* ***complex room*** *with the* ***dipole****, distance of* 1 mm*. Abscissa and ordinate are the positions of Eves.*

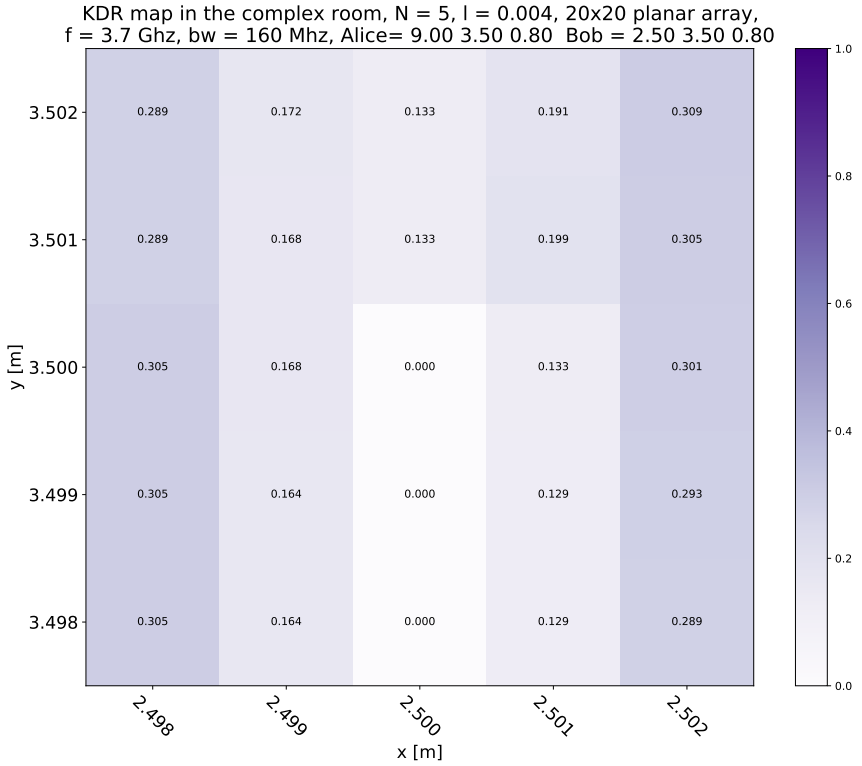


(a) *Root Mean Square Error map with a bandwidth of* 160 MHz

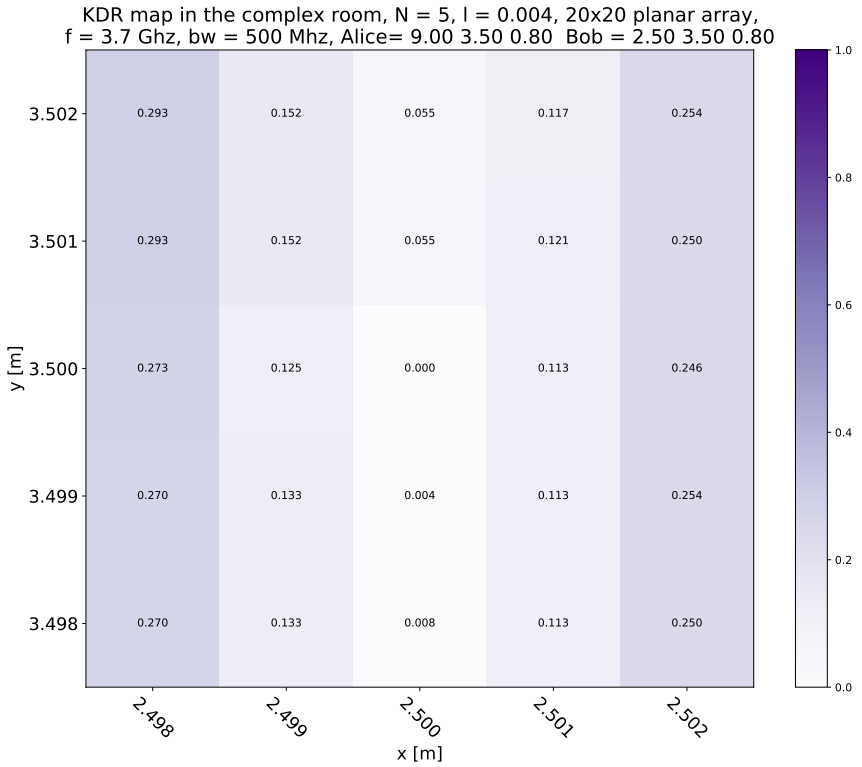


(b) *Root Mean Square Error map with a bandwidth of* 500 MHz

Figure 5.35: *Spatial map of the RMSE of the* ***complex room*** *with the* ***dipole****, distance of* 1 mm*. Abscissa and ordinate are the positions of Eves.*

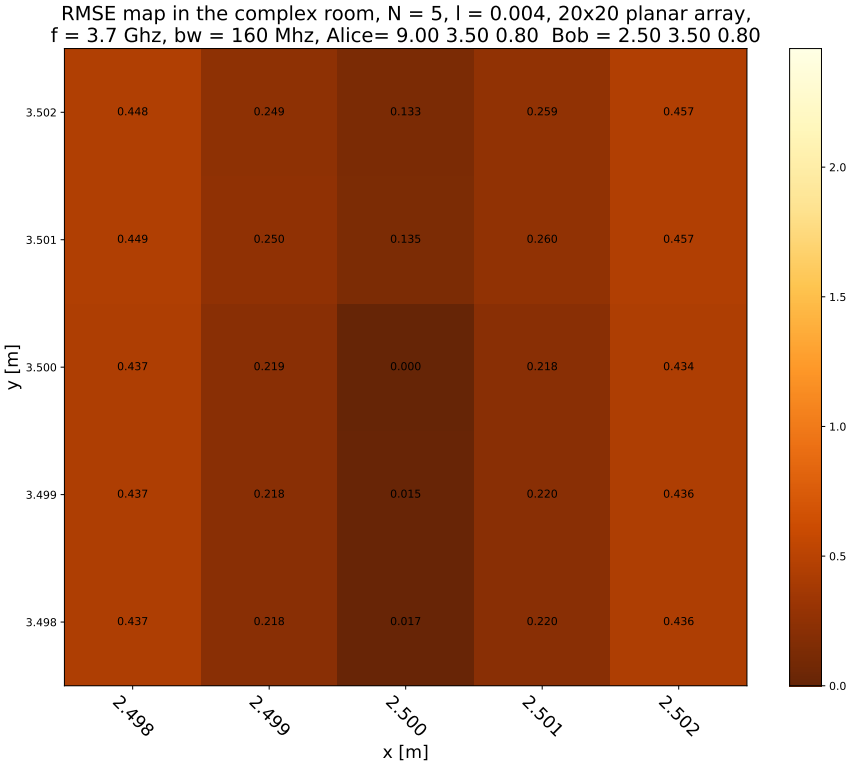


(a) *Key Disagreement Rate map with a bandwidth of* $160\,\mathrm{MHz}$

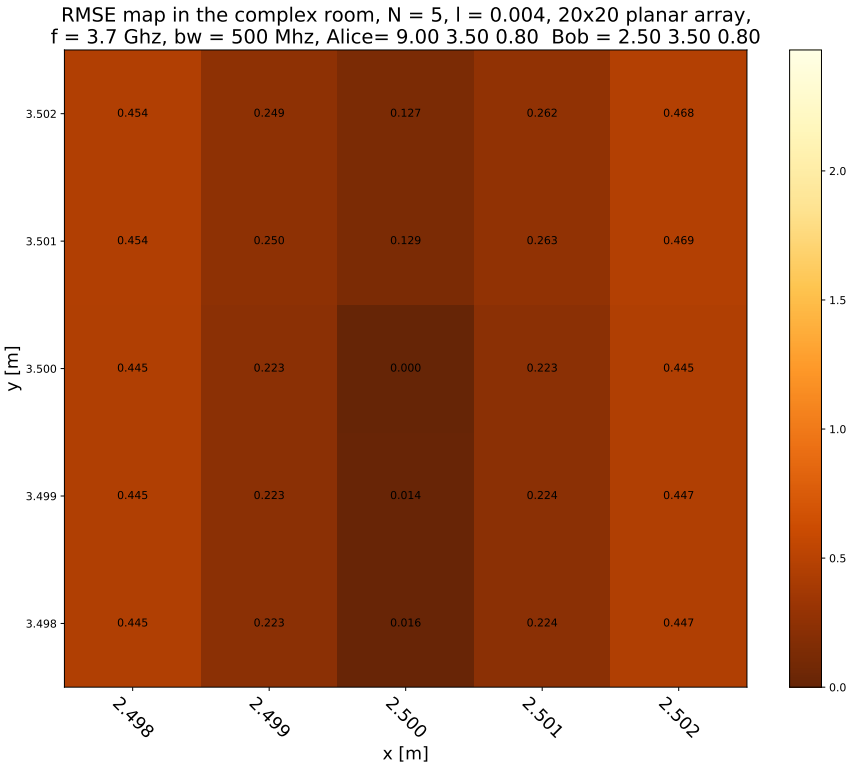


(b) *Key Disagreement Rate map with a bandwidth of* $500\,\mathrm{MHz}$

Figure 5.36: *Spatial map of the KDR of the* ***complex room*** *with the* 20×20 ***array****, distance of* $1\,\mathrm{mm}$*. Abscissa and ordinate are the positions of Eves.*

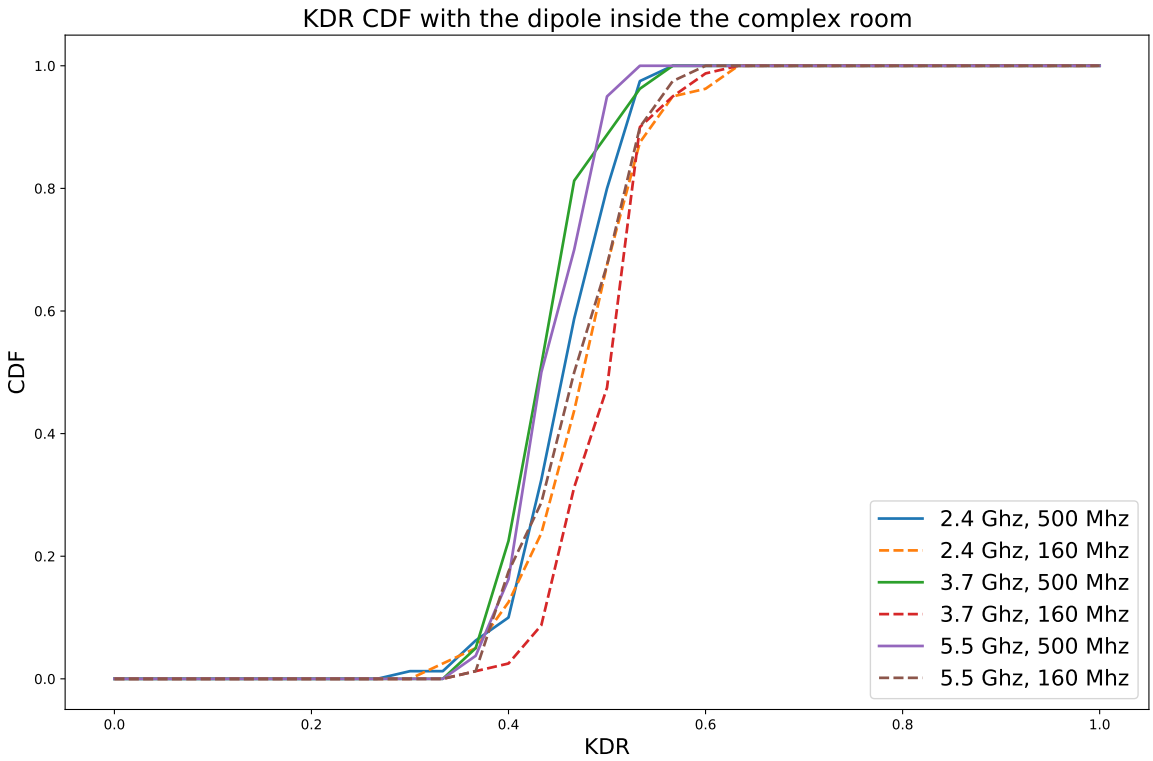


(a) *Root Mean Square Error map with a bandwidth of* 160 MHz

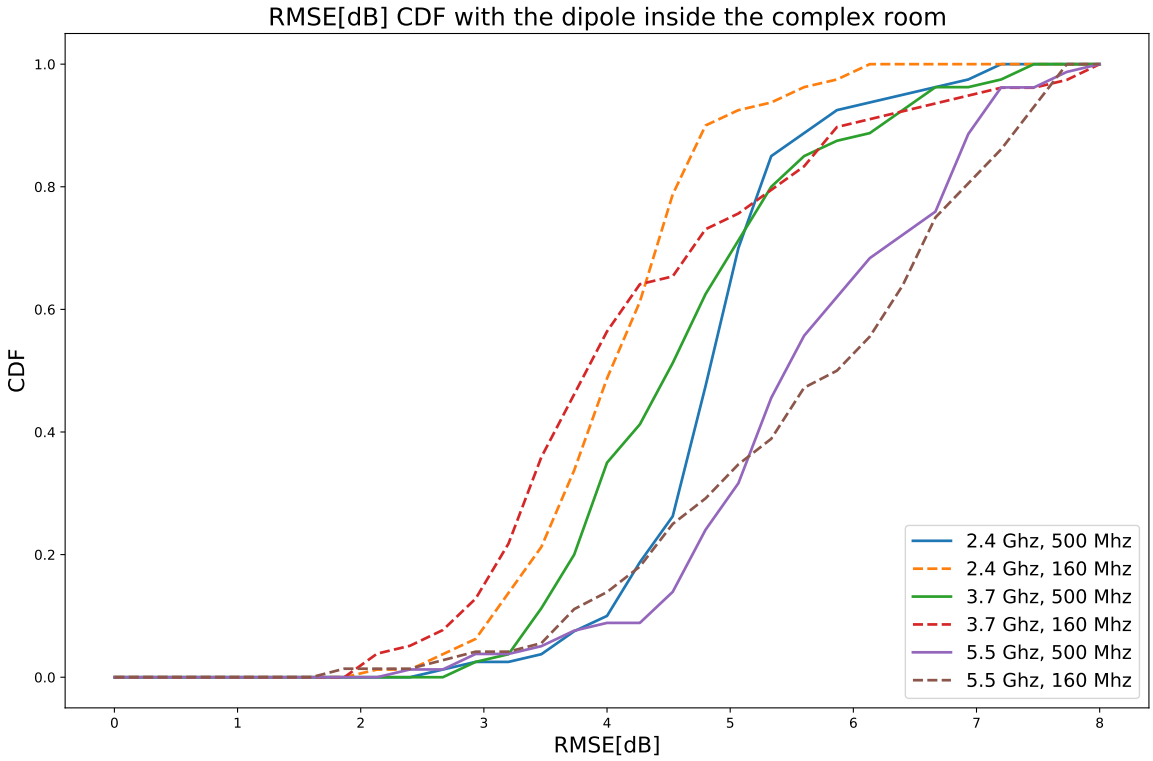


(b) *Root Mean Square Error map with a bandwidth of* 500 MHz

Figure 5.37: *Spatial map of the RMSE of the* ***complex room*** *with the* 20×20 ***array****, distance of* 1 mm*. Abscissa and ordinate are the positions of Eves.*

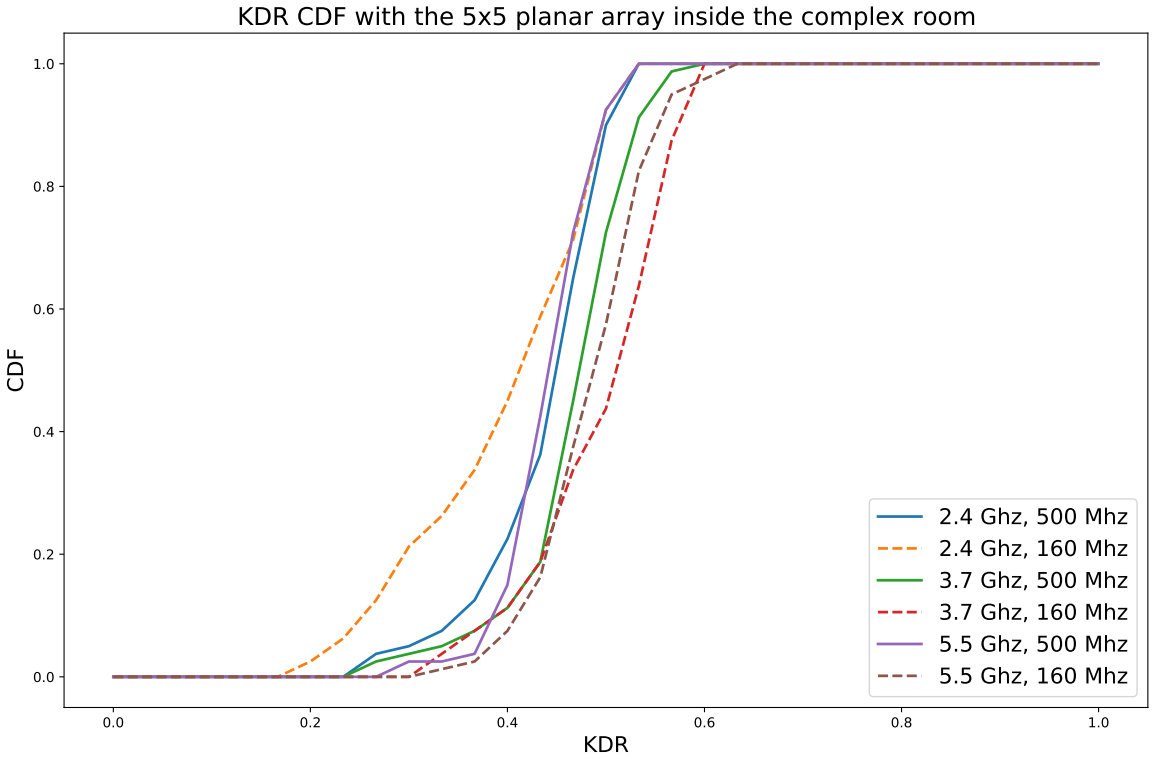


(a) *CDF of the KDR*

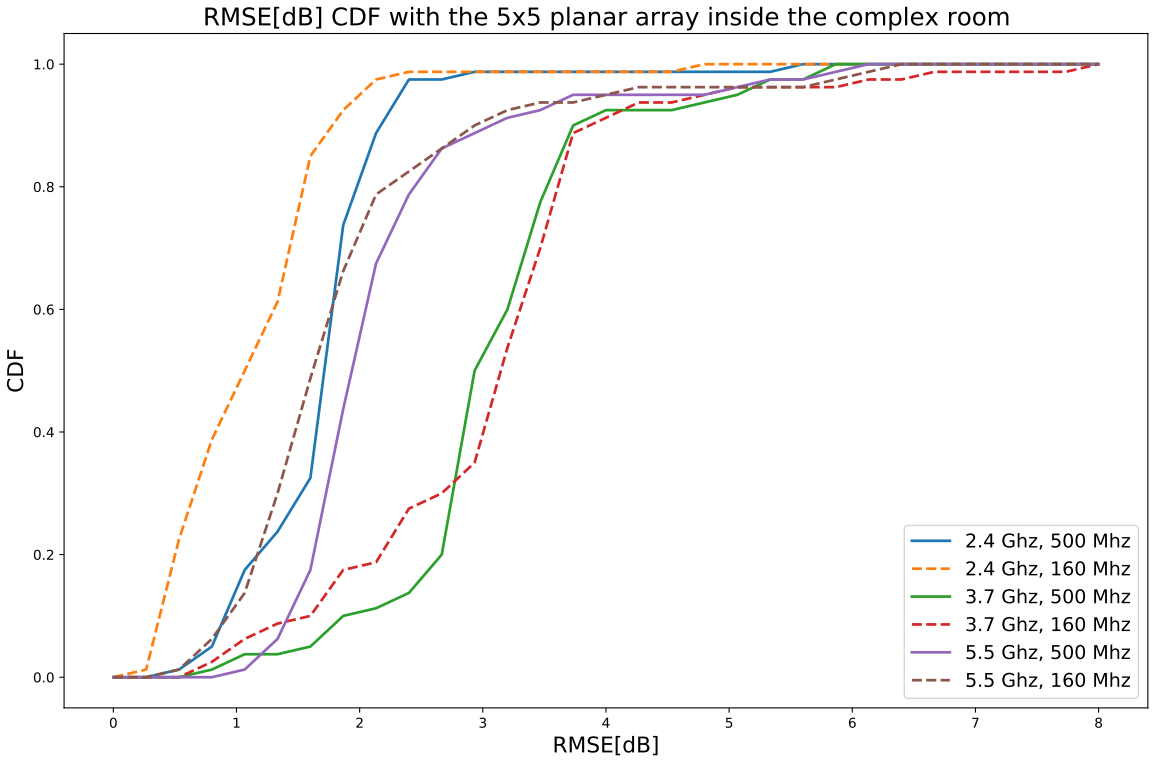


(b) *CDF of the RMSE*

Figure 5.38: *Cumulative Distribution Function of the KDR and RMSE map, in the* ***complex room*** *with the* ***dipole****.*

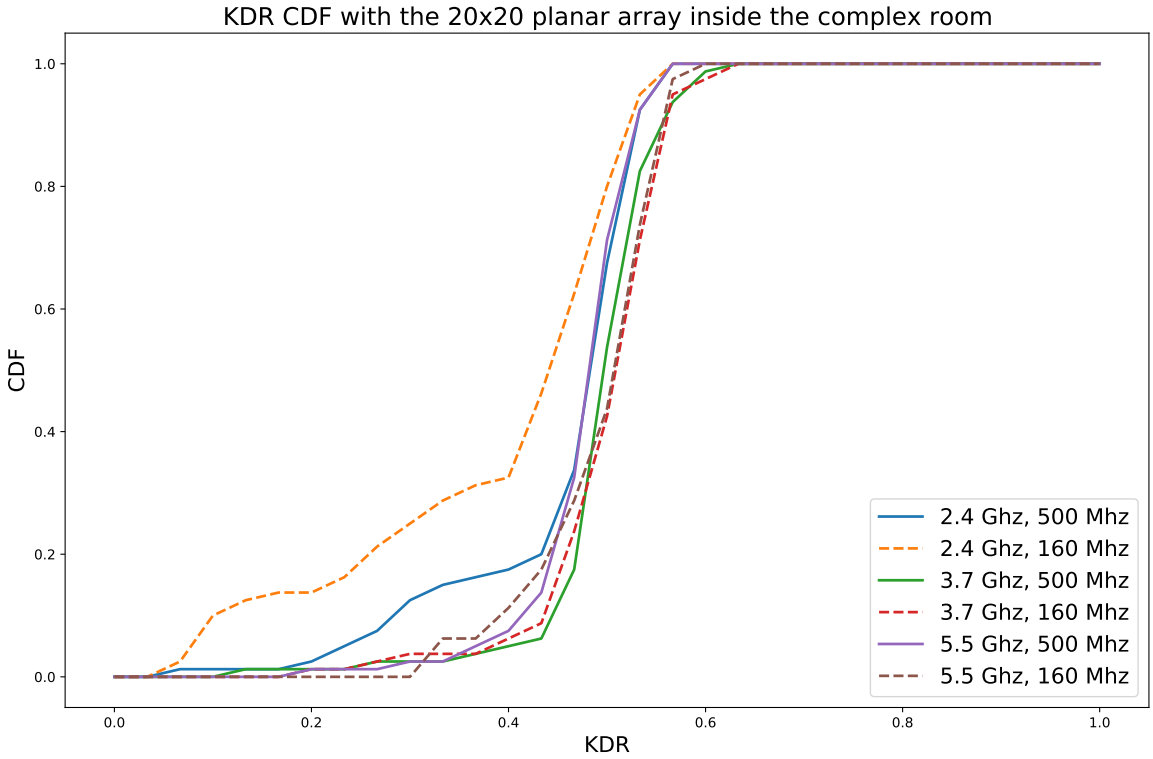


(a) *CDF of the KDR*

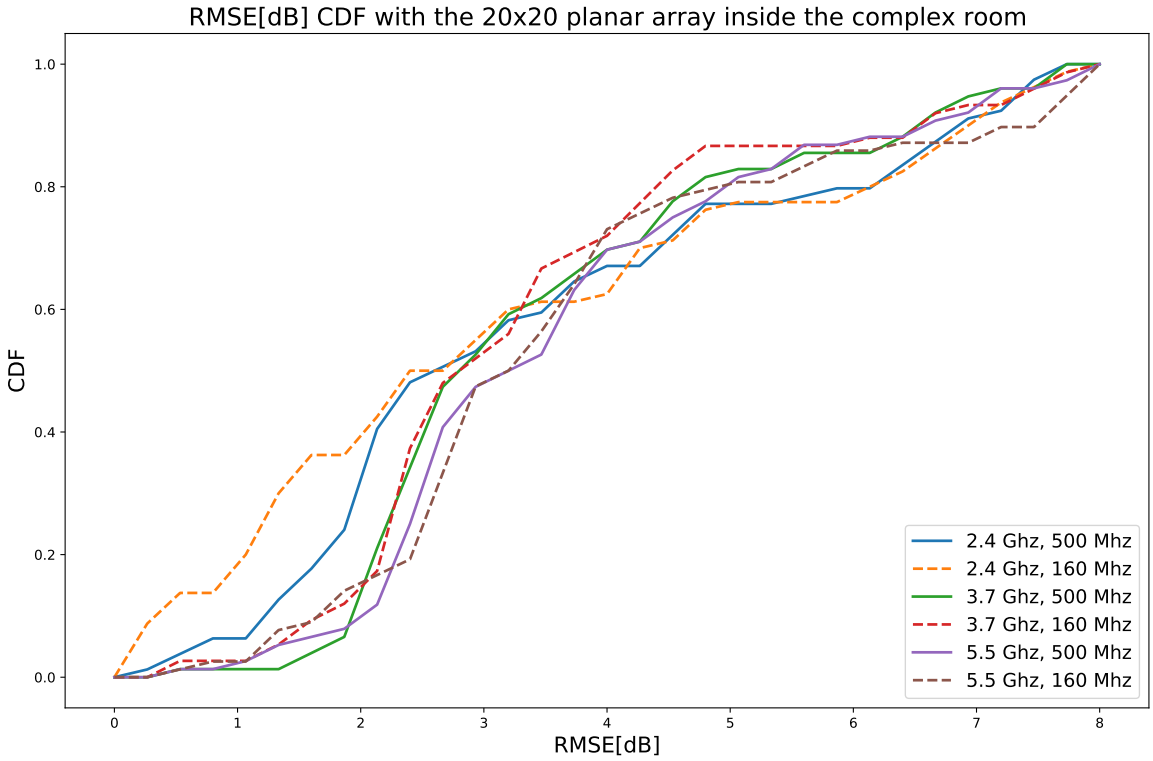


(b) *CDF of the RMSE*

Figure 5.39: *Cumulative Distribution Function of the KDR and RMSE map, in the* ***complex room*** *with the* ***5x5 array****.*



(a) *CDF of the KDR*



(b) *CDF of the RMSE*

Figure 5.40: *Cumulative Distribution Function of the KDR and RMSE map, in the* ***complex room*** *with the* ***20x20 array****.*

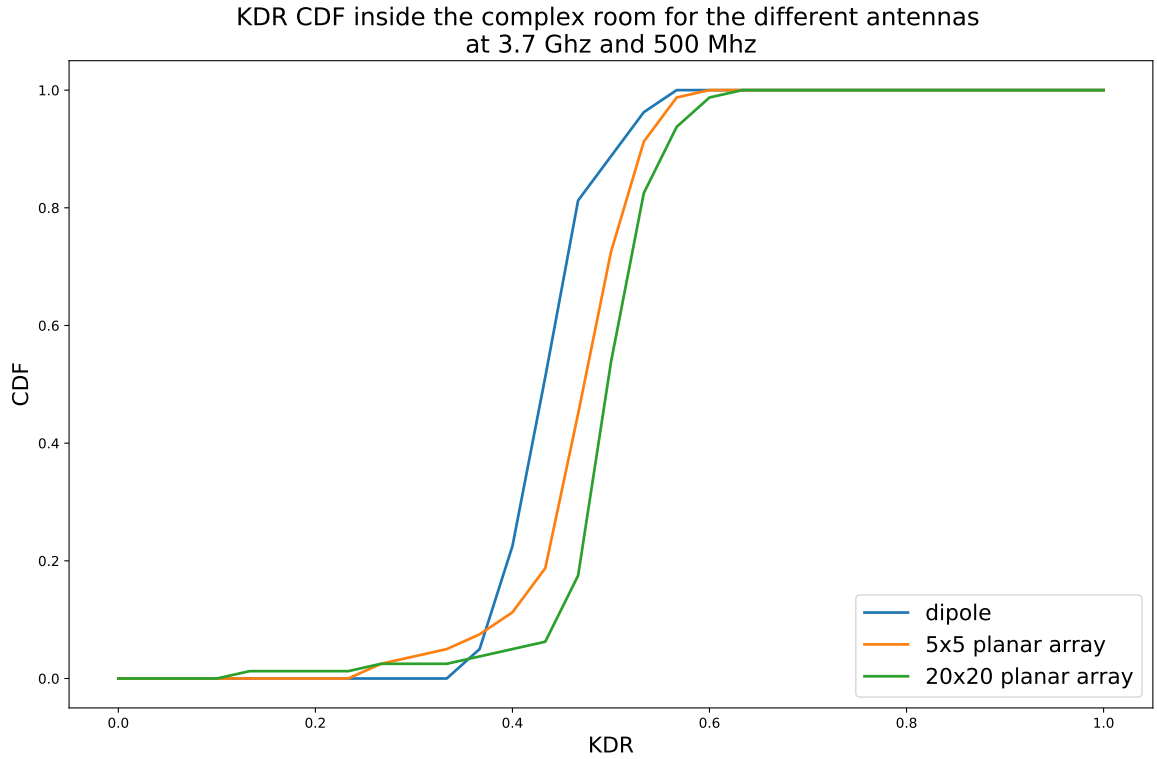


Figure 5.41: *CDF of the KDR in the* ***complex room*** *with respect to the three antennas, at* $3.7\,\mathrm{GHz}$ *and a bandwidth of* $500\,\mathrm{MHz}$

5.7 Summary

The possibility of using the Ray Tracing as a tool to evaluate the impact of the channel on the Physical Layer based-Key generation has been evaluated in this chapter. In particular, it has been noticed that with directive antennas, even though the simulations show a good situation for the key generation, the dynamic may be comparable with the noise which disrupt the possibility to generate equal keys. Moreover, the possibility to use the Ray Tracing to evaluate the spatial decorrelation has been presented.

Conclusions

This work was aimed at a general assessment of the performance of Physical Layer Security protocols for wireless communications by means of Ray tracing simulations. In this framework, Chapters 1 and 2 have been devoted to a survey on cryptography and fundamentals of Physical Layer Security. Chapter 3 has presented the methodologies and the tools considered to tackle the Ray Tracing attack issue, which has been then assessed in Chapter 4. In particular, Ray Tracing has turned out unfit to deterministically track the small fading characteristics of the channel. This is mainly due to the fact that Ray Tracing is not able is not able to properly compute the phase of the rays, which makes the simulated transfer function of the channel greatly different from the real one. However, the Ray Tracing is still able to reliably model the statistics of fast fading of the channel and it is able to follow the entropy of the channel. Therefore, Chapter 5 has reported reported some considerations about using the Ray Tracing as a design and evaluation tool for PLS solutions. In particular, the Physical Layer based-Key generation with the Filter-bank model can be used as long as the channel has enough bandwidth and dynamic. When the users employ directive antennas, even though the simulation shows an high percentage of random keys, the entropy is small and the dynamic is comparable to the noise fluctuations, which brings to the impossibility for the users of generating the same key. In the end, even though the Ray Tracing is not able to predict the exact multipath pattern, it is still a good tool to evaluate the impact of the channel parameters on the Key generation, taking into account also the specific characteristic of the environment. Finally, since the fading distribution is reliably modelled by the Ray Tracing (provided that that the environment and antennas are correctly modelled), it is also possible to perform some spatial correlation assessments, to show that in a specific environment a possible eavesdropper can hardly steal the secret if they are far from legitimate users, regardless of any other link parameters.

Bibliography

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[2] "2019 cyber safety insights report global results," March 2020. https://now.symassets.com/content/dam/norton/campaign/NortonReport/2020/2019_NortonLifeLock_Cyber_Safety_Insights_Report_Global_Results.pdf.

[3] J. Zhang, G. Li, A. Marshall, A. Hu, and L. Hanzo, "A New Frontier for IoT Security Emerging From Three Decades of Key Generation Relying on Wireless Channels," *IEEE Access*, vol. 8, pp. 138406–138446, 2020.

[4] A. Grau, "How to Build a Safer Internet of Things," 2015. https://spectrum.ieee.org/telecom/security/how-to-build-a-safer-internet-of-things.

[5] E. Marin, D. Singelée, F. Garcia, T. Chothia, R. Willems, and B. Preneel, "On the (in)security of the latest generation implantable cardiac defibrillators and how to secure them," pp. 226–236, 12 2016.

[6] E. Marin, D. Singelée, F. D. Garcia, T. Chothia, R. Willems, and B. Preneel, "On the (in)Security of the Latest Generation Implantable Cardiac Defibrillators and How to Secure Them," p. 226–236, 2016.

[7] Y. Zou, J. Zhu, X. Wang, and L. Hanzo, "A Survey on Wireless Security: Technical Challenges, Recent Advances, and Future Trends," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1727–1765, 2016.

[8] J. Zhang, S. Rajendran, Z. Sun, R. Woods, and L. Hanzo, "Physical layer security for the internet of things: Authentication and key generation," *IEEE Wireless Communications*, vol. 26, no. 5, pp. 92–98, 2019.

[9] Q. Xu, R. Zheng, W. Saad, and Z. Han, "Device fingerprinting in wireless networks: Challenges and opportunities," *IEEE Communications Surveys & Tutorials*, vol. 18, 01 2015.

[10] A. Mpitziopoulos, D. Gavalas, G. Pantziou, and C. Konstantopoulos, "Defending Wireless Sensor Networks from Jamming Attacks," pp. 1 – 5, 10 2007.

[11] K. Pelechrinis, C. Koufogiannakis, and S. V. Krishnamurthy, "On the Efficacy of Frequency Hopping in Coping with Jamming Attacks in 802.11 Networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 10, pp. 3258–3271, 2010.

[12] X. Lin, "CAT: Building Couples to Early Detect Node Compromise Attack in Wireless Sensor Networks," pp. 1–6, 2009.

[13] *Fundamentals of Cryptology: A Professional Reference and Interactive Tutorial*. Spinger Nature.

[14] *Practical Cryptography*. Wiley Publishing.

[15] NIST. `www.nist.gov`.

[16] S. Rao, D. Mahto, D. YADAV, and D. Khan, "The AES-256 Cryptosystem Resists Quantum Attacks," *International Journal of Advanced Research in Computer Science*, vol. 8, pp. 404–408, 04 2017.

[17] G. S. Vernam, "Secret signaling system," 1919.

[18] V. Bhatia and K. Ramkumar, "An Efficient Quantum Computing technique for cracking RSA using Shor's Algorithm," pp. 89–94, 2020.

[19] C. E. Shannon, "Communication theory of secrecy systems," *The Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.

[20] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," *Theoretical Computer Science*, vol. 560, p. 7–11, Dec 2014.

[21] R. Hughes, W. Buttler, P. Kwiat, S. Lamoreaux, G. Morgan, J. Nordholt, and C. Peterson, "Quantum cryptography for secure free-space communications," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 3615, 06 1999.

[22] A. D. Wyner, "The wire-tap channel," *The Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[23] M. Bloch and J. Barros, ch. 1. Cambridge University Press, 2011.

[24] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian wire-tap channel," *IEEE Transactions on Information Theory*, vol. 24, no. 4, pp. 451–456, 1978.

[25] H. Qin, X. Chen, Y. Sun, M. Zhao, and J. Wang, "Optimal Power Allocation for Joint Beamforming and Artificial Noise Design in Secure Wireless Communications," pp. 1 – 5, 07 2011.

[26] R. Ahlswede and I. Csiszar, "Common randomness in information theory and cryptography. i. secret sharing," *Information Theory, IEEE Transactions on*, vol. 39, pp. 1121 – 1132, 08 1993.

[27] U. Maurer, "Secret key agreement by public discussion from common information," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 733–742, 1993.

[28] J. Zhang, T. Q. Duong, A. Marshall, and R. Woods, "Key generation from wireless channels: A review," *IEEE Access*, vol. 4, pp. 614–626, 2016.

[29] S. Ben Hamida, J.-B. Pierrot, B. Denis, C. Castelluccia, and B. Uguen, "On the security of uwb secret key generation methods against deterministic channel prediction attacks," 09 2012.

[30] O. A. Topal, G. K. Kurt, and H. Yanikomeroglu, "Securing the inter-spacecraft links: Physical layer key generation from doppler frequency shift," *IEEE Journal of Radio Frequency Identification*, pp. 1–1, 2021.

[31] C. Huth, R. Guillaume, T. Strohm, P. Duplys, I. A. Samuel, and T. Güneysu, "Information reconciliation schemes in physical-layer security: A survey," *Computer Networks*, vol. 109, pp. 84–104, 2016. Special issue on Recent Advances in Physical-Layer Security.

[32] N. Hosseinidehaj, Z. Babar, R. Malaney, S. X. Ng, and L. Hanzo, "Satellite-based continuous-variable quantum communications: State-of-the-art and a predictive outlook," *IEEE Communications Surveys Tutorials*, vol. 21, no. 1, pp. 881–919, 2019.

[33] S. Jana, S. N. Premnath, M. Clark, S. K. Kasera, N. Patwari, and S. V. Krishnamurthy, "On the effectiveness of secret key extraction from wireless signal strength in real environments," in *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking*, MobiCom '09, (New York, NY, USA), p. 321–332, Association for Computing Machinery, 2009.

[34] J. Zhang, S. K. Kasera, and N. Patwari, "Mobility assisted secret key generation using wireless link signatures," in *in INFOCOM miniconference*, 2009.

[35] Y. Wei, K. Zeng, and P. Mohapatra, "Adaptive wireless channel probing for shared key generation based on pid controller," *IEEE Transactions on Mobile Computing*, vol. 12, no. 9, pp. 1842–1852, 2013.

[36] M. Zoli, A. N. Barreto, S. Köpsell, P. Sen, and G. Fettweis, "Physical-layer-security box: a concept for time-frequency channel-reciprocity key generation," *EURASIP Journal on Wireless Communications and Networking*, vol. 2020, 06 2020.

[37] F. Fuschini, E. M. Vitucci, M. Barbiroli, G. Falciasecca, and V. Degli-Esposti, "Ray tracing propagation modeling for future small-cell and indoor applications: A review of current techniques," *Radio Science*, vol. 50, no. 6, pp. 469–485, 2015.

[38] F. Gil, A. Claro, J. Ferreira, C. Pardelinha, and L. Correia, "A 3-d extrapolation model for base station antennas' radiation patterns," vol. 3, pp. 1341–1345 vol.3, 1999.

[39] "R&S®FSH handled spectrum analyzer-specificaitons." `https://scdn.rohde-schwarz.com/ur/pws/dl_downloads/dl_common_library/dl_brochures_and_datasheets/pdf_1/FSH_dat-sw_en_5214-0482-22_v2900.pdf`.

[40] "Documentation of the Fast Fourier Transform implemented in the package `scipy`." `https://docs.scipy.org/doc/scipy/reference/generated/scipy.fft.fft.html`.

[41] "Documentation of `numpy.histogram`." `https://numpy.org/doc/stable/reference/generated/numpy.histogram.html`.

[42] G. V. Steeg, "Non Parametric Entropy Estimator Toolbox." `https://github.com/gregversteeg/NPEET`.

[43] "Documentation of `numpy.histogram2d`." `https://numpy.org/doc/stable/reference/generated/numpy.histogram2d.html`.

[44] E. M. Vitucci, F. Mani, T. Mazloum, A. Sibille, and V. Degli Esposti, "Ray tracing simulations of indoor channel spatial correlation for physical layer security," pp. 1–5, 2015.

[45] Y. Kim, Y. Kim, J. Oh, H. Ji, J. Yeo, S. Choi, H. Ryu, H. Noh, T. Kim, F. Sun, Y. Wang, Y. Qi, and J. Lee, "New radio (nr) and its evolution toward 5g-advanced," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 2–7, 2019.