الجمهوريـــــــــــــــــــــة الجزائريـــــــــــــــــــة الديموقراطيـــــــــــــــــــة الشعبيـــــــــــــــــــة

PEOPLE's DEMOCRATIC REPUBLIC OF ALGERIA

وزارة التــــــــعـــــليـــــم العــــــــــــالي والبـــــــحث العـــــــلـــــمــــــــي

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

جـــــــــامعة سعيدة – د. مولاي الطاهر

UNIVERSITY OF SAIDA - Dr. MOULAY TAHAR

Faculty of Mathematics, Computer Science and Telecommunications

MEMOIRE

Presented for the award of the **Master's Degree in Telecommunications**
Specialization: Telecommunication Systems

Presented by:

KHADIR Fatma Hadil

SADOUN Daouia

Machine Learning Techniques for THz Channel Estimation

Defended on June 17\06\2025 in front of the jury composed by:

Mr. CHETIOUI Mohammed	MCA	President
Mr. MANSOURI Boualem	Professor	Examiner
Mr. TAMI Abdelkader	MCA	Supervisor

2024/2025

Acknowledgments

*First and foremost, all thanks and praise to our God "**Allah**" for giving us the strength and patience to finish this task.*

*We would like to express our gratitude to our project supervisor, **Dr. TAMI Abdelkade**, for their valuable guidance, helpful feedback, and ongoing support and for his patience, motivation, enthusiasm, and immense knowledge Your expertise and encouragement have been instrumental in shaping the direction and quality of this project.*

Besides my advisor, I would like to thank the rest of my thesis committee, insightful comments, and hard questions.

A heartfelt thank you to our families, whose unwavering support and understanding have been a source of strength for us. Your belief in our abilities has always been a great motivator, and for their support, both moral and material, during the preparation phase.

DEDICATION

*To my beloved family: my **mother** and **father**, my sister **Ikram**, and my brother **krimo** and **Abdelkader**. Your unwavering support, endless love, and constant encouragement, Your sacrifices and teachings have shaped me into the person I am today. **Ikram** and **krimo** and **Abdelkader**., you have been my companions through thick and thin, sharing in my joys and comforting me in my sorrows.*

*To my best friend **hiba** and **bouchra**, your friendship is a precious gift that I cherish every day. Your loyalty, kindness, and understanding have made our bond unbreakable. You have been my confidante, my cheerleader, and my rock in completing our project.*

*And to my partner, **daouia**, your partnership and collaboration have been essential to our journey. Your dedication, hard work, and friendship have made this experience both successful and enjoyable. Thank you for being an exceptional partner.*

*To my supervisor, **Dr.TAMI Abdelkader**, your guidance and wisdom have been invaluable throughout this journey. To the jury, thank you for your time, your attention, and your valuable feedback.*

Thank you all for being an integral part of my journey. Your love and support have been my guiding light, and I am eternally grateful for each of you.

Khadir fatma Hadil

DEDICATION

Their last supplication will be, "Praise be to Allah, Lord of the worlds. "Allah Almighty has spoken the truth. .All praise is due to Allah, who made the beginnings easy, completed the endings, and helped us reach our goals. All praise is due to Allah; no effort is accomplished without His help, .All praise is due to Allah, who granted me this knowledge and supported me in completing it .All praise to Allah, in love, gratitude, and appreciation.

With all my love, I dedicate the fruit of my graduation:

- *To my great and resilient self, who persevered through all the setbacks despite the challenges. To the one whose name is forever linked to mine, that great man who supported me and encouraged me to achieve my ambitions, who gave his all and never held back .To the one who removed the thorns from my path and paved the way for me to learn: my rock, my safety, and the dearest person in my life:* ***my beloved father***

- *To the one who words cannot describe, she was a light in my darkness, her prayers were the secret to my success, and she was with me at all my times: my role model, my teacher, and my great lady:* ***my beloved mother***

- *To the springs of loyalty and sources of tenderness, to my dear sisters , my companions at work and through sleepless nights , and to my brothers* ***:Nasr Eddine ,Yacine*** *, and* ***mostapha allaoui*** *my contant support and the comfort of my days .To those who supported me and walked with me on every path ,I dedicate this work as a token of love and loyalty. And as a tribute to your pure hearts that have always embraced me .*

- ***To my dear friend and partner hadil,*** *having you by my side on this journey has been one of the greatest blessings, your support , effort ,and beautiful spirit made everything easier and more than just a partner you were a true sister in every sense .i'mtruly roud that you were part of this experience .*

- ***To Bouchra and Hiba,*** *You have been more than just friends you were my support in hard times and my laughter in difficult moments. With you, life felt wider, and I discovered that true love exists in sincere companionship. Thank you for your big hearts and for always being a steady presence in my world.*

- ***To Dr. Tami Abdelkader,***
  I sincerely thank you for your continuous support and encouragement throughout this journey. Your guidance and advice have had a great impact on achieving this accomplishment, and I am very grateful for everything you have done for me. May God reward you with the best

- ***To the jury****, thank you for your time, your attention, and your valuable feedback. Your insights have been crucial in shaping this work, and I am deeply grateful for your contributions.*

Sadoun Daouia

ABSTRACT

Terahertz (THz) communications is considered one of the most promising wireless technologies for the sixth generation (6G) and beyond. A fundamental challenge for the practical deployment of THz systems is accurate channel estimation, due to the unique propagation characteristics of THz frequencies. In this context, we address the problem of channel modeling and estimation by considering deterministic propagation and the physical characteristics specific to THz bands. We also explore the application of machine learning algorithms for THz channel estimation, including Neural Networks (NN), Logistic Regression (LR), and Projected Gradient Ascent (PGA).

we provide a clear explanation of machine learning and deep learning, introducing the three main types of machine learning: supervised, unsupervised. The objective is to offer a comprehensive understanding of what machine learning truly is and why it is essential. Furthermore, In the final chapter, we present simulation results showcasing a recent class of radio channel estimation based on Deep Neural Networks (DNN), which differs fundamentally from the classical channel estimation algorithms previously discussed.

Keywords: 6G wireless communication, Terahertz (THz) frequency band, channel estimation algorithms, Logistic Regression (LR), Projected Gradient Ascent (PGA), Deep Neural Networks (DNN)

RÉSUMÉ

Les Télécommunications basées sur le Terahertz (THz) sont considérées comme l'une des technologies sans fil les plus prometteuses pour la sixième génération (6G) et au-delà. Un défi fondamental pour le déploiement pratique des systèmes THz réside dans l'estimation précise du canal, en raison des caractéristiques de propagation uniques des fréquences THz. Dans ce contexte, nous abordons le problème de la modélisation et de l'estimation du canal en prenant en compte la propagation déterministe ainsi que les caractéristiques physiques spécifiques aux bandes THz. Nous explorons également l'application d'algorithmes d'apprentissage automatique pour l'estimation du canal THz, y compris les réseaux de neurones (NN), la régression logistique (LR) et l'ascension de gradient projetée (PGA).

Nous fournissons une explication claire de l'apprentissage automatique et de l'apprentissage profond, en introduisant les trois principaux types d'apprentissage automatique : supervisé, non supervisé. L'objectif est d'offrir une compréhension complète de ce qu'est réellement l'apprentissage automatique et pourquoi il est essentiel. De plus, dans le dernier chapitre, nous présentons des résultats de simulation mettant en évidence une nouvelle classe d'estimation de canal radio basée sur les réseaux de neurones profonds (DNN), qui diffère fondamentalement des algorithmes classiques d'estimation de canal précédemment abordés.

Mots-clés: Télécommunications sans fil 6G, Bande Terahertz (THz), Channel Estimations, Logistic Regression (LR), Projected Gradient Ascent (PGA), Deep Neural Networks (DNN)

LIST OF ABBREVIATIONS

5G	Fifth Generation	**NN**	neural networks
6G	Sixth Generation	**PGA**	projected gradient ascent
AI	Artificial Intelligence	**RIS**	Reconfigurable Intelligent Surfaces
		RT	Ray-Tracing
AoA	Angle of Arrival	**RX**	Receiver
AoSA	Array of Subarrays	**SVD**	Singular Value Decomposition
AWGN	additive white Gaussian Noise	**SVM**	Support Vector Machine
BER	Bit Error Rate	**SGD**	Stochastic gradient Descent
DCNN	Deep Convolutional Neural Network	**Tbps**	Terabit-per-second
FDTD	finite Difference Time Domain	**Tx**	Transmitter
FW	Frank-Wolfe	**THz**	Terahertz
HITRAN	HIgh-Resolution TRANsmission	**Tera-IoT**	Tbps Internet of Things
LOS	line-of-Sight	**UHSLLC**	Ultrahigh Speed with Low Latency Communications
LR	logistic regression	**UAVs**	Unmanned Aerial Vehicals
mmWave	Millimeter wave	**WiNoC**	Wireless Networks-On-Chip Communications
ML	Machine learning		
MSE	Mean Squared Error	**WPT**	Wireless Power Transmission

LIST OF CONTENT

LIST OF FIGURES

LIST OF TABELS

GENERAL INTRODUCTION

Communication methods have undergone a radical transformation since the dawn of humanity: over the past millennium, we shifted from relying on fire and animal messengers to using electrical signals.

In the last few decades, wireless systems have evolved at an astonishing pace today, terahertz (THz) communication systems (300 GHz–10 THz) are at the forefront of the transition toward sixth-generation (6G) networks, thanks to the extraordinarily high data rates they promise. However, operating at such high frequencies brings fundamentally new physical challenges that complicate channel modeling and estimation, including increased atmospheric absorption losses, near field multipath dispersion, elevated receiver noise, and diminished obstacle penetration compared to earlier mmWave bands (30 GHz–300 GHz). Against this backdrop, recent research has focused on developing mmWave MIMO channel estimation techniques as a stepping stone to THz applications. To overcome these obstacles, machine learning based solutions have emerged as a primary tool for improving channel estimation accuracy without relying on overly simplified statistical models. For example, some researchers have proposed using a Deep Convolutional Neural Network (DCNN) Channel estimation algorithms based on Deep Convolutional Neural Networks (DCNNs) are employed, where the terahertz (THz) channel is simulated using a channel model originally developed for millimeter-wave (mmWave) systems, with the frequency parameters adjusted accordingly, This approach leverages existing modeling techniques while tailoring them to the challenges of high-frequency communication. In addition, the search explores several machine learning algorithms applied in MIMO systems, including Projected Gradient Ascent (PGA), Frank-Wolfe techniques, and Logistic Regression.

Chapter I provides the introduction for Overview about 6G & Machine Learning. Chapter II describes the channel and system model. Chapter III explains the machine learning algorithms studied for channel estimation. Chapter IV concludes with the simulation results.

Chapter I :

Overview about 6G & Machine Learning

I.1. INTRODUCTION

As 5G communication networks are being deployed commercially, academia and industry started developing 6G wireless communication systems as we move beyond the capabilities of 5G the development of 6G (sixth Generation) wireless networks is set to revolutionize connectivity, offering ultra-fast speeds, near-zero latency, and intelligent automation [1] Commercial deployment of 6G is expected to begin around 2030. These networks will utlilize higher frequencies than 5G, offering significantly greater capacity and lower latency [2].

Terahertz (THz) propagation characteristics will play a key role in 6G mobile communication systems, integrating cuttingedge technologies such as Artificial Intelligence (AI), Machine Learning (ML), and the Internet of Everything (IoE) to create more autonomous, efficient, and intelligent communication networks. Machine Learning (ML) is crucial in the evolution of 6G and THz by enabling smart network management, predictive maintenance, and real time decision making. ML algorithms can optimize network performance and improve spectrum efficiency. In this context, we present a holistic, forward-looking vision that defines the core tenets of a 6G system. We believe that 6G will not just be an exploration of higher frequency spectrum bands combined with ML, but rather the convergence of emerging technological trends driven by exciting, underlying services [3]

I .2. Sixth-Generation Communication system

6G (sixth generation wireless) is the successor to 5G cellular technology. 6G networks will be able to use higher frequencies than 5G networks and provide substantially higher capacity and much lower latency. One of the goals of the 6G internet is to support one microsecond latency communications. This is 1,000 times faster (or 1/1000th the latency) than one millisecond throughput.

The 6G technology market is expected to facilitate large improvements in the areas of imaging, presence technology and location awareness. Working in conjunction with Artificial Intelligence (AI), the 6G computational infrastructure will be able to identify the best place for computing to occur; this includes decisions about data storage, processing and sharing [4]

6G is categorized into three main classes: Ultrahigh Data Density (UHDD), Ultrahigh Speed with Low Latency Communications (UHSLLC) and Ubiquitous Mobile Ultra Broadband (UMUB). Figure I.1 summarizes evolution from 1G to 6G. 6G is expected to fill gap of radio coverage limitation in previous generations [5]

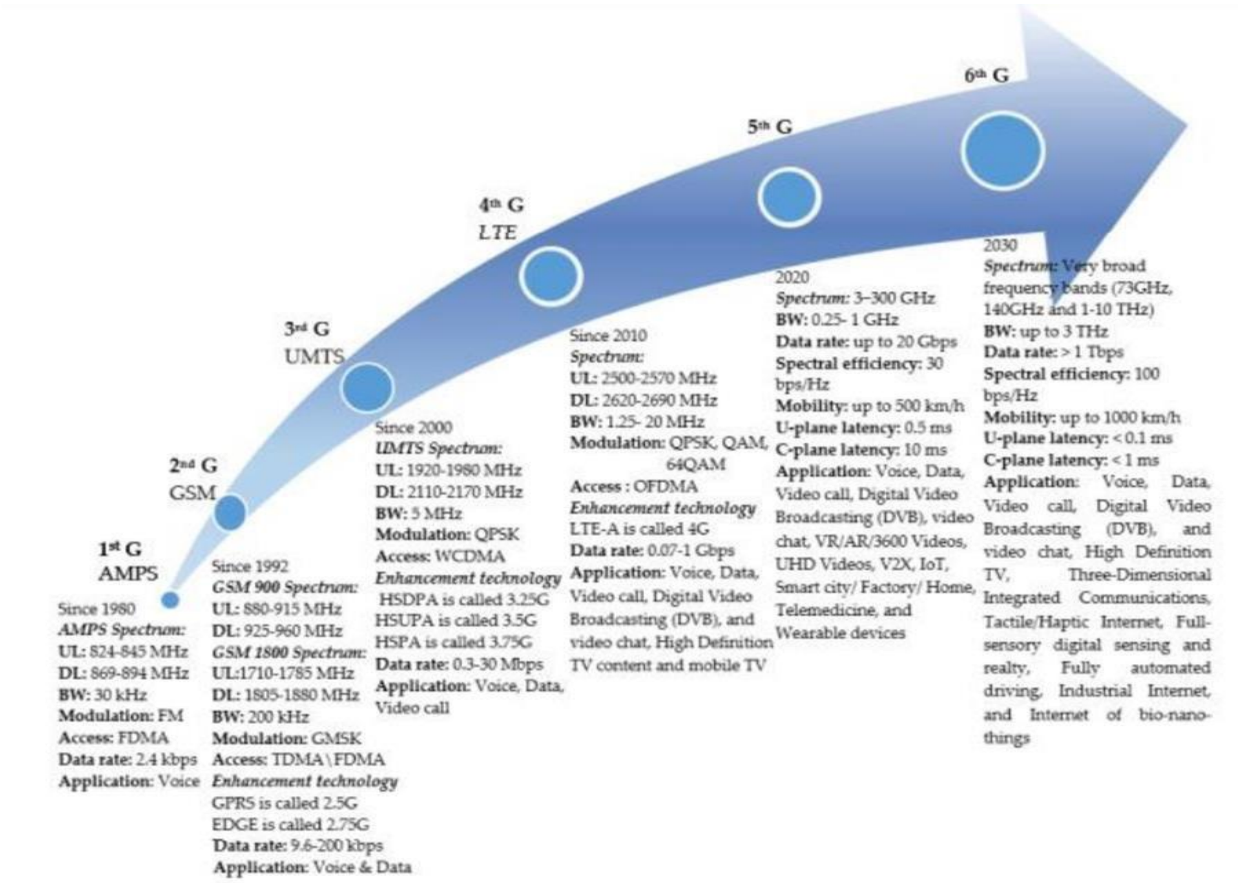


Figure I.1: Major Features for Different Generations of Wireless Cellular Communication (1-6G)

I.3. 6G NETWORK ARCHITECTURE

6G will be "digital, intelligent and ubiquitous". If the 5G era can realize the ubiquitous acquirement of information, then 6G should fully support the digitization of the world on the basis of 5G, and combine with the development of AI and other technologies to realize the ubiquitous acquirement of information and comprehensively enable everything.

In the future, it will move towards a "digital" world that combines virtual and reality. The world will generate a digital virtual world based on the physical world. Information and intelligence can be transmitted among persons, persons and things, and things and things in the physical world through the 6G. The twin virtual world is the simulation and prediction of the physical world, which will accurately reflect and predict the real state of the physical world, help human beings further emancipate themselves, improve the quality of life, and improve the efficiency of social production and governance, so as to achieve the goal of "creating a new world through digital innovation, and making all things intelligent"[6]

I.3.1. 6G Time-Space-Frequency Characteristics

6G requires to use a wide spectrum compared to previous generations to ensure high data speed. Some studies have suggested the multiplicity of frequency bands used in 6G, for example, millimeter wave band, terahertz band and visible light band to achieve transmission of hundreds of gigabytes per second. On the other hand, mobile phone networks will be combined with satellite systems and the Internet to build integrated networks. In the spatial dimension, a huge number of antennas will be used in transmitters regularly in the socalled ultra-huge MIMO(UM-MIMO) in the terahertz range. In the time dimension, there will be a clear improvement in response time, and there will be flexibility in the versatility of the systems as well, this facilitating their compatibility with 2G to 5G [7]

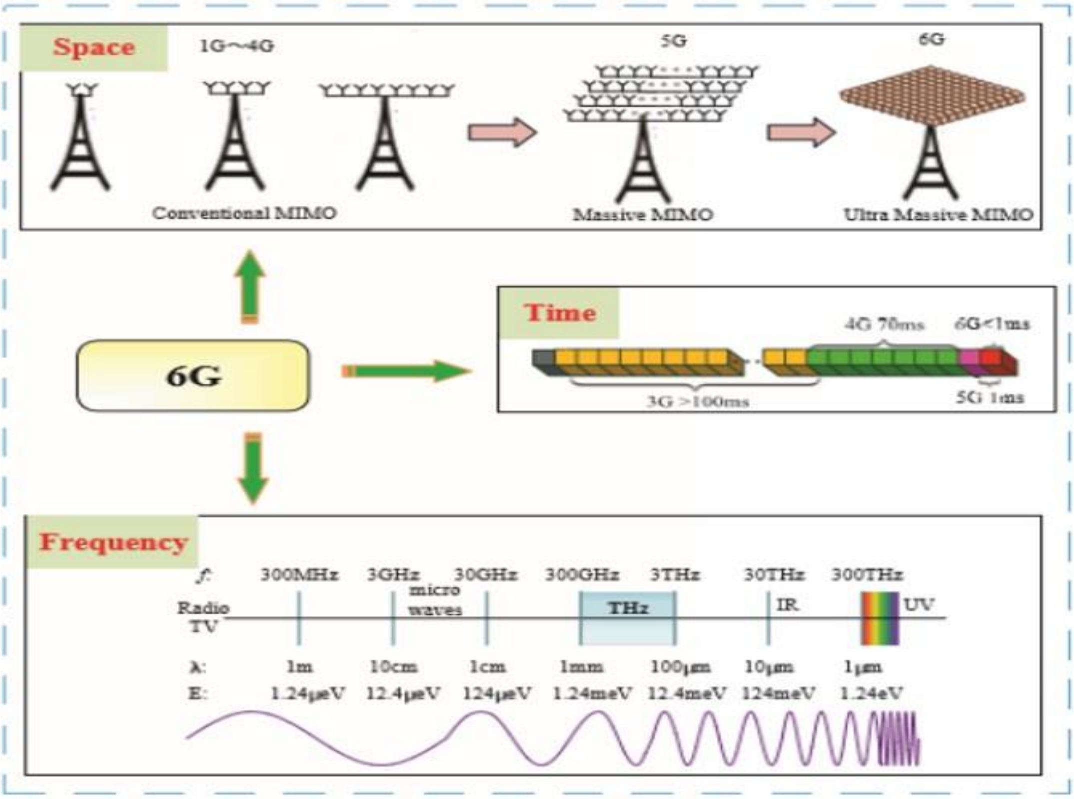


Figure I .2 : A Framework of 6G Founded On The Space Resource use, Frequency,Time

The specific characteristics are as follows:[6]

(a) In the time dimension, the basic time slot in 6G can be compressed to use the high frequency band more effectively and meet the demand of delay sensitive service. The flexibility and versatility of the network will be improved as the time slot becomes shorter.

(b) In the spatial dimension, ultra massive multiple input multiple output (UM-MIMO) for THz communication can support hundreds to thousands of transmit and receive antennas, and further utilize the "multipath" technology.

(c) In the frequency dimension, on the one hand, THz band and even visible light band will be used for 6G transmissions; on the other hand, in the future, mobile network can be integrated with satellite system and Internet to build space-ground integrated network. From the perspective of personal mobile communication, this will indeed increase the frequency range for services. As a result, 6G will use higher frequencies than previous generations of mobile communication systems to increase data rates.

I.4. 6G TECHNOLOGY CHARACTERISTICS

New technologies will be introduced into the 6G mobile communication system as follows: [6]

a) Novel discontinuous communication technology:

When new frequency bands such as millimeter wave and terahertz wave are added to the applied frequency band, 6G will adopt a very wide frequency band compared with the past. Therefore, it seems that there are many related research fields, such as optimizing the selection of multi band according to the application, studying the method of frequency reuse between cells, upgrading the duplex mode in uplink and downlink, and studying the utilization mode of low frequency band.

b) Ultra high rate, high reliability communication:

Wireless communication highly reliable control information is an important requirement of many industrial use cases (such as remote control and factory automation), and 6G is expected to achieve higher reliability and security than 5G. With the popularity of robots, unmanned aerial vehicles, and the expansion of radio coverage to the sky, highly reliable communication is required not only in limited areas such as factories, but also in wider areas, and it is possible to achieve highly reliable communication in various scenarios.

c) Network based positioning and sensing:

The 6G network will use a unified positioning and communication interface to improve control operations, which can rely on context information to form beam forming patterns, reduce interference and predict switching, and provide innovative user services, such as vehicle and electronic health services.

d) Terahertz communication:

Spectrum efficiency can be achieved by using THz communication (0.1-10THz) and using advanced UM-MIMO technology. RF frequency band has been almost exhausted, and it is not enough to meet the requirement of 6G. THz band will play an important role in 6G communication. THz band will become the next frontier of high data rate communication. The small wavelength of THz signal allows more antenna elements to be integrated into the equipment and base stations operating in frequency band.

The THz spectrum can resolve the spectrum scarcity problem and tremendously enhance current wireless system capacity. Various promising applications are envisaged, such as Tbps WLAN system (Tera-WiFi), Tbps Internet of Things (Tera-IoT) in wireless data center, Tbps Integrated Access Backhaul (Tera-IAB) wireless networks, and ultra-broadband THz space communications (Tera-SpaceCom), as illustrated. Besides these macro/micro-scale applications, the THz band can be utilized for wireless connections in nanomachine networks, to enable Wireless Networks-On-Chip Communications (WiNoC) and the Internet of Nano-Things (IoNT), motivated by the state of the art nanoscale transceivers and antennas that oscillate in the THz band [8]

Terahertz wave has many characteristics:

- Terahertz wave is easily absorbed by moisture in the air, which is more suitable for high-rate and short-range wireless communication;
- The wave beam is narrower and has better directionality, and has stronger anti interference ability;
- Terahertz wave has wide bandwidth, which can meet the demand of spectrum bandwidth in wireless broadband transmission.
- Terahertz wave can be widely used in space communication, especially for the communication between satellites or between satellite and ground;
- The propagation characteristics of electromagnetic wave show that the free space fading is proportional to the square of frequency, so terahertz has larger decline of free space compared with low frequency band.
- Terahertz signal is very sensitive to shadow and has great influence on coverage.
- At the moving rate, the channel coherence time is linearly related to the carrier frequency, which means that the coherent time of terahertz band is very small and the doppler spread is large, which is much faster than the frequency band used in the current cellular system.

Terahertz system is a highly spatially oriented signal transmission, which means that the path fading, service beam and cell correlation will change rapidly, and a fast adaptation mechanism is needed to overcome this fast changing intermittent connection problem [9][10]

e) Unmanned Aerial Vehicals (UAVs):

UAV will be an important part of 6G wireless communication. In many cases, UAV technology will be used to provide high data rate wireless connections. The base station entity will be installed on the UAV to provide cellular connectivity. UAVs have some characteristics that are not available in a fixed base station infrastructure, such as ease of deployment, strong LoS links, and degree of freedom with controllable mobility. In emergency situations such as natural disasters, it is not economically feasible to deploy ground communication infrastructure, and sometimes it is impossible to provide any services in unstable environments. UAVs can easily handle these situations. UAV will become a new mode in the field of wireless communication

f) Integration Of Sensing and Communication:

The key driver of autonomous wireless network is to be able to continuously sense the dynamic changes of the environment and exchange information among different nodes. In 6G, sensors will be tightly integrated with communications to support autonomous systems

g) Big data analysis*:*

Big data analysis is a complex process of analyzing various big data sets. This process discovers information, such as hidden patterns, unknown correlations, and customer preferences, to ensure perfect data management. Big data is collected from a variety of sources, such as videos, social networks, images and sensors. This technology will be widely used in the processing of massive data in 6G system.

h) WPT and energy harvesting:

Any IOT device in 6G will consume more power due to the huge computing demands of AI processing. WPT doesn't play a key role in 5G, but in 6G, it will eventually shine. First of all, because the density of wireless network continues to increase, the communication distance will be greatly shortened. In addition, the use of UAV as base station further shortens the distance, which makes WPT more meaningful. UAVs will benefit a lot from Wireless Power Transmission (WPT), which enables UAVs to move all the time. In addition, with the continuous progress of energy harvesting technology, energy harvesting from RF signals may become a feasible power supply for low-power applications.[10]

I.5. Machine Learning (ML)

I.5.1. Definition

Machine learning (ML) is a branch of Artificial Intelligence (AI) focused on enabling computers and machines to imitate the way that humans learn, to perform tasks autonomously, and to improve their performance and accuracy through experience and exposure to more data.

(UC Berkeley), breaks out the learning system of a machine learning algorithm into three main parts.

a. **A Decision Process**: In general, machine learning algorithms are used to make a prediction or classification. Based on some input data, which can be labeled or unlabeled, your algorithm will produce an estimate about a pattern in the data.

b. **An Error Function**: An error function evaluates the prediction of the model. If there are known examples, an error function can make a comparison to assess the accuracy of the model.

c. **A Model Optimization Process**: If the model can fit better to the data points in the training set, then weights are adjusted to reduce the discrepancy between the known example and the model estimate. The algorithm will repeat this iterative “evaluate and optimize” process, updating weights autonomously until a threshold of accuracy has been met [11]

I.5.2. Types of Machine learning

Machine learning models fall into three primary categories: [12]

a. Supervised learning :

Supervised learning, also known as supervised machine learning, is defined by its use of labeled datasets to train algorithms to classify data or predict outcomes accurately. As input data is fed into the model, the model adjusts its weights until it has been fitted appropriately. This occurs as part of the cross-validation process to ensure that the model avoids overfitting or underfitting. Supervised learning helps organizations solve a various real world problems at scale, such as filtering spam into a separate folder from in an inbox. Some methods used in supervised learning include neural networks, Naïve Bayes, linear regression, logistic regression, random forest, and Support Vector Machine (SVM).

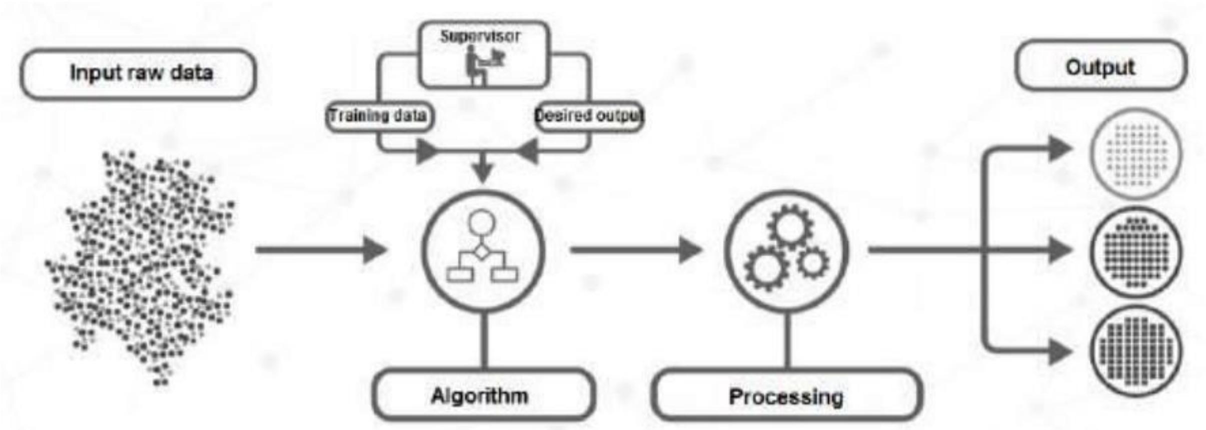


Figure I.3 : Supervised learning

b. Unsupervised learning :

Unsupervised learning, also known as unsupervised machine learning, uses machine learning algorithms to analyze and cluster unlabeled datasets (subsets called clusters). These algorithms discover hidden patterns or data groupings without the need for human intervention.

Unsupervised learning's ability to discover similarities and differences in information make it ideal for exploratory data analysis, cross-selling strategies, customer segmentation, and image and pattern recognition. It's also used to reduce the number of features in a model through the process of dimensionality reduction. Principal Component Analysis (PCA) and Singular Value Decomposition (SVD) are two common approaches for this. Other algorithms used in unsupervised learning include neural networks, k-means clustering, and probabilistic clustering methods.

c. Semi-supervised learning :

Semi-supervised learning offers a happy medium between supervised and unsupervised learning. During training, it uses a smaller labeled data set to guide classification and feature extraction from a larger, unlabeled data set. Semi-supervised learning can solve the problem of not having enough labeled data for a supervised learning algorithm. It also helps if it’s too costly to label enough data.

d. Reinforcement learning :

Reinforcement learning is a machine learning model that is similar to supervised learning, but the algorithm isn’t trained using sample data. This model learns as it goes by using trial and error. A sequence of successful outcomes will be reinforced to develop the best recommendation or policy for a given problem.

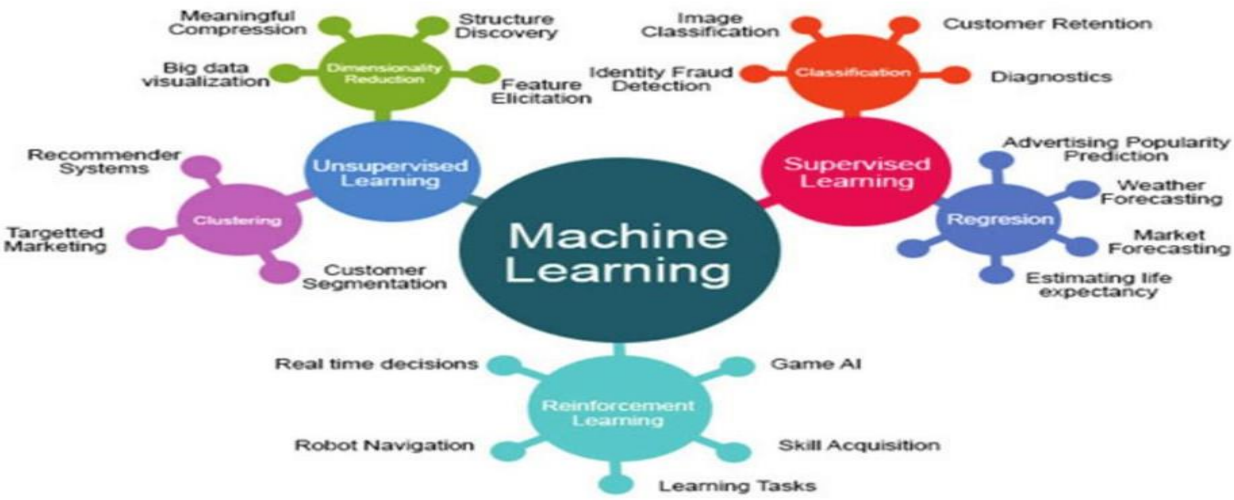


Figure I.4: Different Categories of Machine Learning and their Applications

I.5.3. Designing a machine Learning System

Designing a learning system in machine learning is a multi-step process that ensures the system is effective, efficient, and adaptable to real-world challenges. Each stage builds on the previous one to create a robust framework: [13]

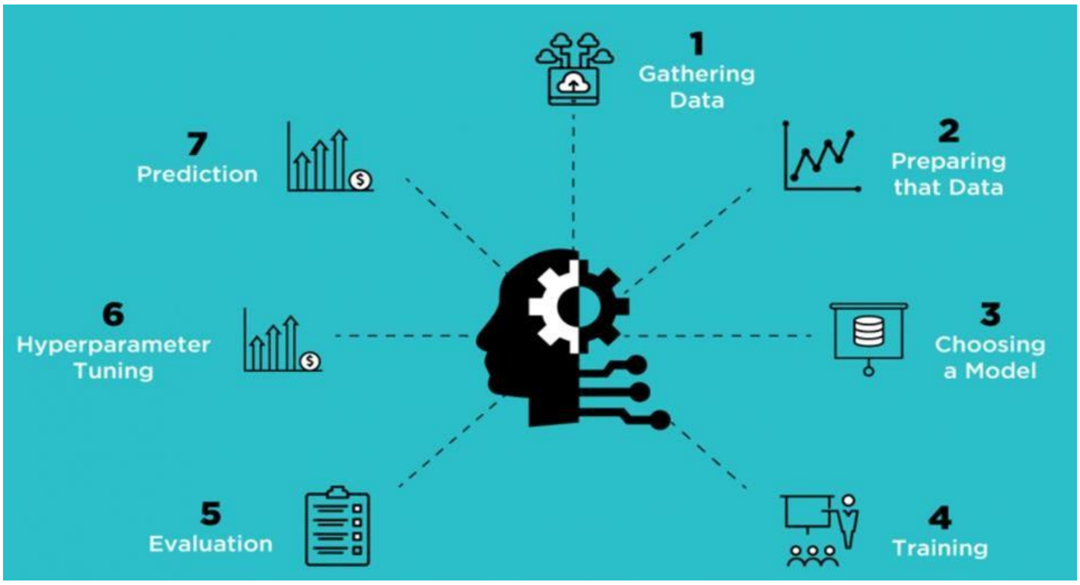


Figure I.5: Machine Learning Algorithms

a) Defining the Problem and Objectives:

The foundation of a learning system begins with a clear problem definition.it starts by identifying the type of task the system aims to solve such as classification, regression, or clustering Alongside the task, and it is measurable performance objectives using metrics like accuracy, precision, recall, or F1-score. These metrics act as benchmarks to assess the system's success and align its outcomes with organizational goals.

b) Data Collection and Preparation :

Data is the backbone of any learning system. Collecting and preprocessing high-quality, relevant data is critical for the system's success.

- Data Gathering: Data can come from databases, APIs, or real time sensors.

- Data Preprocessing: Raw data often contains inconsistencies, missing values, and outliers. Use techniques like imputation to handle missing values and normalization to scale features. Data cleaning ensures the dataset is ready for training, minimizing errors caused by poor-quality data.

c) Choosing the Training Experience :

Decide on the type of training experience based on the problem and the nature of the data:

- **Supervised Learning:** Best for tasks where labeled data is available

- **Unsupervised Learning**: Suitable for uncovering hidden patterns in unlabeled data
- **Reinforcement Learning:** Ideal for dynamic environments like robotics or gaming, where the system learns through trial and error.

d) Choosing a Representation for the Target Function:

Choosing the appropriate representation for the target function depends on the problem's complexity and data characteristics

- **Decision Trees**: Effective for hierarchical decision-making tasks.
- **Neural Networks**: Suitable for handling non-linear relationships in large, complex datasets.
- **Linear Models**: Best for interpretable, straightforward problems.

This step requires balancing model complexity with interpretability and computational efficiency.

e) Selecting a Function Approximation Algorithm:

To set a learning algorithm that effectively approximates the target function.

- **Gradient Descent**: Common in training neural networks for minimizing error.
- **Support Vector Machines (SVM)**: Used for classification tasks requiring clear decision boundaries.
- **K-Means Clustering**: Effective for grouping data points in unsupervised learning scenarios.

The algorithm choice must align with the problem type, data characteristics, and performance objectives.

f) Training the Model :

Training the chosen model using the prepared data. This involves

- Iteratively feeding the data into the algorithm.
- Adjusting model parameters to minimize errors using techniques like backpropagation in neural networks.

- Monitoring the training process to avoid issues like overfitting, which occurs when the model performs well on training data but poorly on unseen data.

g) Evaluating Model Performance :

Evaluating the model ensures it generalizes well to new data.

Analyze metrics such as :

- Accuracy for classification tasks.
- Mean Squared Error (MSE) for regression problems.
- SNR (Signal-To-Noise Ratio)

This evaluation process identifies weaknesses in the model, guiding further refinement.

h) Iterative Refinement :

Refining the model is a continuous process aimed at improving its performance:

- **Hyperparameter Tuning:** Adjust parameters like learning rate, depth of decision trees, or the number of layers in neural networks.
- **Retraining with Updated Data:** Incorporate new or additional data to enhance the model's understanding of the problem domain.

This iterative cycle ensures the learning system evolves, adapting to changing data and requirements while achieving optimal performance.

I.6. Conclusion

In conclusion, this chapiter examined the theoretical foundations and emerging technologies underpinning ML-driven networks. Although still in their infancy, these networks are poised to become a fundamental component of complex 6G systems. The envisioned 6G architecture will leverage distributed artificial intelligence to implement a fully user-centric network infrastructure. Consequently, terminal devices will be empowered to make autonomous network decisions based on prior operational outcomes, eliminating the need for continuous communication with a centralized controller. This distributed approach facilitates real-time ML processing with sub-millisecond latency, thereby meeting the stringent requirements of various 6G services and significantly enhancing network management responsiveness. The terahertz (THz) spectrum stands as a key enabler for 6G, offering ultra-wide bandwidth to meet the demands for extreme data rates and ultra-low latency. In the following chapter, we delve into THz channel models, their unique characteristics, and the challenges they pose, paving the way for optimized strategies to enhance future network performance

CHAPTER II:
THz CHANNEL MODELS

II.1. INTRODUCTION

In recent years, with the growing requirements of high data rate in various environments, the sixth generation (6G) wireless communication system has attracted increasing attention all over the world. It has been widely known that the objective of 6G is to realize the smart interconnection of everything, which requires ultra-high transmission data rate, ultra-high connection density and high reliability [14]

To target this, many studies have suggested that the terahertz (THz) technique is one of potential key enabling technologies in 6G. Indeed, the THz frequency band (0.1-10 THz) has the potential to carry information at a scale of terabit (Tbit, 1Tbit/s=1000Gbit/s). As we know, accurate and efficient wireless communication channel modeling plays an indispensable role in communication system design and performance optimization. This is due to the fact that we have to design the communication systems only by experience and experiment without channel models. Electro- magnetic radiation from a transmitter to a receiver is the typical work pattern of a wireless channel, and the change of channel strength with time or frequency is its major characteristic [15]

II.2. THz channel models

Terahertz (THz) communications, operating within the frequency range of 0.1-10 THz, are considered a crucial technology for the advancement of sixth generation (6G) wireless systems. The study of underlying THz wireless propagation channels provides the foundations for the development of reliable THz communication systems and their application. The wireless propagation channel serves as the conduit through which signals are transmitted from the transmitter (Tx) to the receiver (Rx), with channel properties playing a crucial role in determining the overall performance capabilities of wireless communications, as well as the effectiveness of specific transmission strategies and transceiver configurations. Given that wireless channels form the fundamental basis for constructing wireless communication systems in novel frequency spectrums and diverse environments, it is essential to investigate the propagation channels for Terahertz (THz) radio frequencies in anticipation of future 6G wireless communication technologies. The examination of wireless channel characteristics hinges on conducting physical channel measurements using channel sounders. Subsequent analysis of these measurements aids in the development of channel models, which aim to encapsulate the behavior of wave propagation with a suitable level of complexity, facilitating equitable comparisons of various algorithms, designs, and performance metrics within wireless networks [16]

II.2.1 Classification of THz channel models

This section provides a detailed discussion on THz channel modeling. While some channel models derived from the mmWave band can be applied to the THz band in practice, THz channel models have distinct classifications, as illustrated in Figure II. 1. Due to the varying parameters that must be carefully considered across different THz communication scenarios, most research over the past decade has focused on propagation models tailored to specific scenarios. Typically, THz channel modeling studies emphasize outdoor, indoor, and nanoscale channel models, employing both deterministic and stochastic approaches [15]

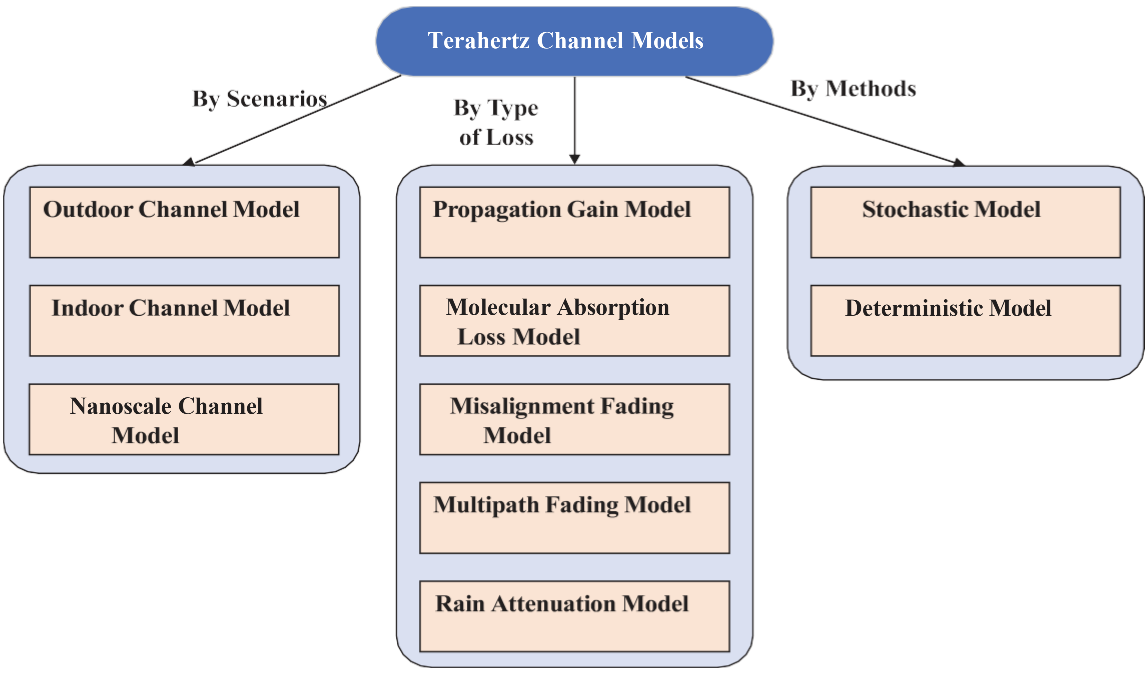


Figure II. 1: THz Channel Models Classification

II.2.1.1. THz channel models by scenarios:

a) Outdoor channel model:

Models that emulate THz channels in outdoor environments are scarce, primarily focusing on point-to-point links. This limitation arises due to the lack of reported experimental measurements in the literature. Notably, the first 120 GHz experimental radio station license has been issued, yet existing outdoor channel models still address only point-to-point scenarios. A key challenge in outdoor THz measurements is the interference from unintentional Non Line of Sight (NLOS) paths, which can significantly impact the Bit Error Rate (BER) [17].

The comprehensive mmWave channel modeling method from the previous chapter can be extended to outdoor environments by adjusting specific system parameters. In outdoor settings, differences in the positions of Reconfigurable Intelligent Surfaces (RIS) and terminals affect channel parameters while following the same fundamental modeling approach. the path loss exponent and shadow fading parameter in equation should be modified to match the outdoor propagation environment.

Regarding small scale fading, adjustments should be made for clusters whose angle of departure is directed toward the ground by reducing the maximum transmission distance based on terminal positions in the outdoor environment. The primary distinction from the indoor channel model lies in the channel characteristics between the RIS and the receiver (Rx). In outdoor settings with a random number of unique clusters, small-scale fading may influence this channel as well.

In cases where the distance between the RIS and Tx is relatively short and the Line of Sight (LOS) probability is high, the LOS dominated channel from equation remains useful. However, in more general outdoor scenarios, additional fading effects and random cluster distributions must be taken into account [18]

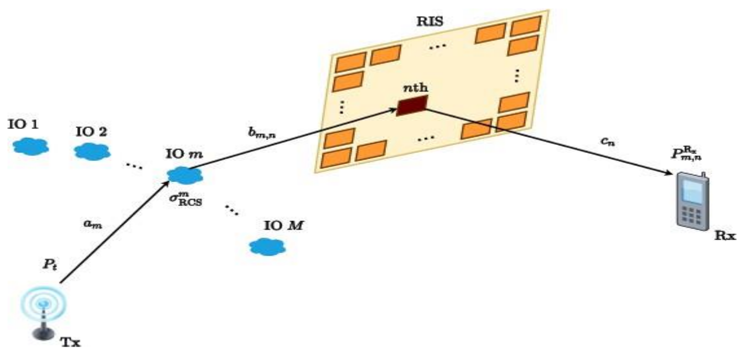


Figure II. 2: RIS-assisted communication between Tx-RIS and RIS-Rx and

Here, $d_{T-RIS}$is the distance between the Tx and the RIS, L_{T-RIS} is the path attenuation n is the path loss exponent, b is a system parameter, and f_0 is a fixed reference frequency (the centroid of all frequencies represented by the path loss model), and X_σ~ N (0,σ^2) is the shadow fading term in logarithmic units.

$$d_{T-RIS} = ((x^{RIS} - x^{Tx})^2 + (y^{RIS} - y^{Tx})^2 + (z^{RIS} - z^{RIS})^2)^{1/2} \quad \text{(II.1)}$$

$$L_{T-RIS} = -20log_{10}\left(\frac{4\pi}{\lambda}\right) - 10^n\left(1b\left(\frac{f-f_0}{f_0}\right)\right)log_{10}(d_{T-RIS})X_\sigma \quad \text{(II.2)}$$

$$\text{h}=\bar{\gamma}\sum_{c=1}^{\bar{C}}\sum_{s=1}^{\bar{S_c}}\overline{\beta_{c,s}}\sqrt{G_e(\theta_{c,s}^{Rx})L_{RIS-R}}\; a(\phi_{c,s}^{Rx},\theta_{c,s}^{Rx}) + g_{LOS} \quad \text{(II.3)}$$

where, $\bar{\gamma}$ is a normalization term, C and S_c stand for number of clusters and sub rays per cluster for the RIS-Rx link, $\beta_{c,s}$is the complex path gain, L_{RIS-R} is the path attenuation, $G_e(\theta_{c,s}^{Rx})$is the RIS element radiation pattern in the direction of the (c, s)th scatterer, $a(\phi_{c,s}^{Rx},\theta_{c,s}^{Rx})$ is the array response vector of the RIS for the given azimuth and elevation angles, and g_{LOS} is the LOS component [18]

b) Indoor channel model:

Several indoor channel models are available in the literature. Indoor channel models can be categorized into either analytical or stochastic models. In terms of deterministic channels, the ray-tracing model is usually applied. This technique is site- specific abiding with propagation theories and capturing the continuously adapting the model to a new environment, which can limit its time efficiency. From the communications perspective, it is fundamental to understand the large and small scale statistics of the channel including path loss, shadowing and multipath propagation. Hence, statistical methods arise as suitable options to model THz propagation based on empirical channel measurements. The first statistical model for THz channels, spanning the range between 275 and 325 GHz, the given model depends on extensive ray-tracing simulations to realize the channel statistical parameters. Yet, the information concerning the channel statistics such as the correlation function and power delay profile cannot be captured easily [17]

Indoor wireless communication systems require accurate channel models to describe signal propagation. Various models have been developed to account for multipath effects, scattering, and attenuation in indoor environments. One commonly used model is the Saleh-Valenzuela (S-V) model, which effectively captures multipath clustering and random variations in arrival and departure angles. The S-V model has been modified to suit Terahertz (THz) frequency bands by incorporating measurement based adjustments to reflect the distinctive propagation characteristics of high frequency signals. In this context, the model assumes that arrival paths consist of multiple clusters, each containing several rays, with the arrival time calculated relative to a reference point [18]

the arrival time of the first cluster is set as the reference time, $T_0 = 0$. The arrival time of the ith cluster and the arrival time of the lth ray in the ith cluster are denoted as T_i and τ_{il} , respectively. Then, the arrival time of each ray will be:

$$t_{il} = T_i + \tau_{il} \tag{II.4}$$

$$H(f,d) = \sum_{i=0}^{N_{clu-1}} \sum_{l=0}^{N_{ray}^{1}-1} \alpha_{il}(f,d) G_t(\phi_{il}^t, \theta_{il}^t) G_r(\phi_{il}^r, \theta_{il}^r) \times a_r(\phi_{il}^r, \theta_{il}^r) a_t^{\dagger}(\phi_{il}^t, \theta_{il}^t) \tag{II.5}$$

Where, $\alpha_{il}(f, d) = |\alpha_{il}(f, d)| e^{j\psi_{il}}$ denotes the path gain of the arrival ray, where ψ_{il} is the associated independent phase shift uniformly distributed over $(0,2\pi)$. $\phi_{il}^t/\theta_{il}^t$ and $\phi_{il}^r/\theta_{il}^r$ refer to the azimuth/elevation angles of departure an arrival, respectively. $G_t(\phi_{il}^t, \theta_{il}^t)$ and $G_r(\phi_{il}^r, \theta_{il}^r)$ represent the transmit and receive antenna gains, while vectors $a_t^{\dagger}(\phi_{il}^t, \theta_{il}^t)$ and $a_r(\phi_{il}^r, \theta_{il}^r) a_t^{\dagger}$ are the associated array steering.

Another approach to indoor channel modeling considers line of sight (LOS) and Non Line of Sight (NLoS) conditions. In certain scenarios, such as those involving Reconfigurable Intelligent Surfaces (RIS), the LOS component dominates due to the close proximity between the receiver (Rx) and the RIS. When the inter-terminal distance is less than 4.5 meters, the probability of LOS exceeds 50%, leading to stronger direct paths and reduced multipath interference. The channel gain in an RIS-assisted setup can be determined using azimuth and elevation angles, which influence the overall characteristics of signal attenuation and reflection.

$$g=\sqrt{G_e(\theta_{Rx}^{RIS})L_{RIS-R}e^{j\eta}a(\phi_{Rx}^{RIS},\theta_{Rx}^{RIS})} \tag{II.6}$$

Both models provide valuable insights into indoor Terahertz communication channels, with the S-V model focusing on multipath clustering, while the RIS-based model emphasizes LOS dominated conditions. The choice of an appropriate model depends on the specific indoor scenario and the desired propagation characteristics [18]

c) Nano-Scale channel model:

In the past few years, advancements in the field of nanotechnology have paved the way towards the development of miniaturized sensing devices which capitalize on the properties of novel nanomaterials. Such devices, denoted as nanodevices, can perform simple tasks including computing, data storing, sensing and actuation. As such, the formulation of nanonet works will allow various applications in the biomedical, industrial, and military fields. Based on radiative transfer theory and in light of molecular absorption, a physical channel model for wireless communication among nanodevices in the THz band. The provided model considers the contribution from the different types and concentrations of molecules, where the HITRAN database is used in order to compute the attenuation that a wave suffers from. The Beer Lambert law was used to compute the transmittance of the medium which relies on the medium absorption coefficient. The model provided was also utilized to compute the channel capacity of nanonetworks operating in the THz band, in which the authors deployed different power allocation schemes. The authors recommended using the lower end of the THz band which has lower absorption coefficients in order to ensure a strong received signal. Moreover, the sky noise model is the basis of the existing absorption noise models. The authors in elaborated on this topic by presenting different perspectives on how to model the molecular absorption noise. However, there is no real experiments conducted in order to validate the proposed models. Not only absorption, but also scattering of molecules and small particles affects the propagation of electromagnetic waves. Hence, a wideband multiple scattering channel model for THz frequencies. Further, the authors in presented an analytical model based on stochastic geometry for interference from omnidirectional nano-sensors. However, in their model, they disregarded interference arising due to the existence of base stations. The authors in tackled this issue where they studied interference from beamforming base stations. As such, it has been concluded that having a high density of base stations using beamforming with small beam- width antennas and deploying a low density of nano-sensors is recommended to improve the coverage probability [17].

where $\mathcal{L}^{mr}$ is the path loss that the main ray faces in its path to a point on the focal line. Note that $\vec{E}_0$ is initially considered to be polarized along

$$\vec{E}_F^{mr} = |\vec{E}_0| \mathcal{L}^{mr} \hat{a}_x \quad \text{(II.7)}$$

field coming through the main ray over the focal line can be also given, where γ(r) is the cell-size gain factor, $\mathcal{G}_{mp}$ is the multi-path gain caused by the reflected rays from adjacent cells

$$\mathrm{H}\,(f,d) = \gamma(r)\mathcal{G}_{mp}\,.\,(\,|\vec{E}_F^{mr}|e^{-j\omega\tau_{mr}} + |\vec{E}_F^{fr}|e^{-j\omega\tau_{fr}}) \quad \text{(II.8)}$$

II.2.1.2. THz channel models by type of loss:

a) Propagation gain model:

According to the Fris equation, the propagation gain in the free space can be given by:

$$h = \frac{c\sqrt{G_t G_r}}{4\pi d f} \quad \text{(II.9)}$$

Where c, G_t, G_r, d, and f respectively represent the velocity of light, the transmitting antenna gain, the receiving antenna gain, the transmission distance and frequency.

It is easy to check that the propagation gain in the THz band is much smaller than that in a radio band. Actually, such high spreading loss is a serious constraint to THz transmission. Since only G_t and G_r could be increased by improving hardware component performance, and hence the high gain antennas become essentially important in developing THz systems in order to overcome the high spreading loss [15]

b) Molecular absorption loss model:

The molecular absorption loss is one of the main fading sources in a THz channel. The energy of electromagnetic wave can be absorbed by the atmospheric molecules, such as water vapor and oxygen. When the operation frequency is higher than 200 GHz, the molecular absorption mainly originates from the water vapor, resulting in several absorption peaks at resonant frequencies, which makes the THz frequency band split into several transmission windows

yAccording to the Beer Lambert Law, the molecular absorption gain is given by:

$$\mathrm{h} = e^{\frac{-1}{2}k(f)d} \tag{II.10}$$

where absorption coefficient k (f) can be calculated by:

$$\mathrm{k(f)}=\sum_{i,g} \frac{p}{p0} \frac{Tst}{T} Q^{i,g} \sigma^{i,g}(f) \tag{II.11}$$

The total absorption coefficient k(f) is the sum of those for isotopologue $i(i \in \{1,2, \ldots \ldots N_i\}$ $i(i \in \{1,2, \ldots \ldots N_g\}$ of ges Besides, T and p are the

temperature and the pressure respectively, while p Tst and $p0$ denote the standard temperature and the standard atmospheric pressure respectively. Furthermore, $Q^{i,g}$and $\sigma^{i,g}$stand for the total number of molecules per volume unit and the absorption cross section of the given gas mixture. The computing methods or definition of these parameters are detailed.

In actual calculation, we can usually extract these parameters from some specific database. In most studies of THz channel modeling, the HIgh-Resolution TRANsmission molecular absorption database (HITRAN) is adopted.

However, the accurate computation of above models is quite complex and tedious. Recently, a simplified but fairly accurate model has been proposed to compute the molecular absorption loss in the 275-400 GHz band. In this model, the absorption coefficient is approximately written as

$$\mathrm{k(f)}=y_1(f,u) + y_2(f,u) + g(f) \tag{II.12}$$

where:

$$y_1(f,u) = \frac{A(u)}{B(u)+\left(\frac{f}{100c}-c1\right)^2} \tag{II.13}$$

$$y_2(f,u) = \frac{c(u)}{D(u)+(\frac{f}{100c}-c1)^2} \tag{II.14}$$

$$G(f) = p_1 f^3 + p_2 f^2 + p_3 f^1 + p_4 \qquad \text{(II.15)}$$

The evaluation expressions of $A(u)$, $B(u)$, $c(u)$, $D(u)$and the values of parameters c_1, c_2, p_1, p_2, p_3, p_4 refer to Besides, u stands for the volume mixing ratio of water vapor

$$u=\frac{\emptyset}{100}\frac{p_w(T,p)}{p} \qquad \text{(II.16)}$$

where p_w denotes the saturated water vapor partial pressure, which is given by

$$p_w = 6.1121(1.0007 + 3.46 \times 10^{-6} p) \times \exp\left(\frac{17.502T}{240.97 + T}\right) \qquad \text{(II.17)}$$

according to the Buck equation. Besides, φ stands for the relative humidity.

On this basis, the approximate formulas of transmission windows. Actually, this simplified model has been used in lots of THz studies, such as channel modeling performance evaluation and reconfigurable intelligent surface (RIS). Furthermore, the state of the art work in presented a novel simplified model for molecular absorption loss in the 100-450 GHz band, and exhibited six absorption peaks in this band, including 119 GHz, 183 GHz, 325 GHz, 380 GHz, 439 GHz and 448 GHz

In fact, the atmosphere can affect THz communications. The first is to cause the atmospheric attenuation, or molecular absorption loss, which has been introduced detailedly in this part. Secondly, since the phase velocities and refractive indexes are different in different frequencies, the arrival time of waves in different frequencies will be different, which leads to the waveform broadening. This phenomenon is also known as atmospheric dispersion. Furthermore, the atmospheric turbulence refers to effects of the air motion and the inhomogeneous distribution of air parameters, which typically brings random fluctuations for THz signals [15].

c) Misalignment fading model:

As aforementioned, high gain antennas are required and indeed widely adopted in THz systems owing to the presence of severe path loss. For instance, the corrugated conical horn antennas with gain of 55 dBi are adopted to achieve a 850 m link at 240 GHz. Since the antenna beam width is inversely proportional to the antenna gain, the THz systems usually need highly directional and narrow beam width antennas. Nevertheless, the highly directional antennas are extremely susceptible to the random motion of antennas, which originates in traffic, wind, and so on. As a consequence, it easily causes the problem of antenna alignment, and hence the pointing error is one of the paramount factors influencing the performance of THz wireless communications.

The relationship between reflectarray antenna gain and beam width was discussed. Moreover, suggested that the misalignments induced by the mobility userend devices may result in severe deterioration of channel capacity and outages in THz wireless communications with highly directional antennas, and presented a mathematical framework to analyze this issue, the effect of antenna directivities on the THz links in indoor scenarios is examined. Be sides, the authors put forward a misalignment model based on statistics, and suggest an optimum antenna configuration to minimize the transmission attenuation. Since the THz wave is quasi optical, the study argued that the misalignment fading can be modeled from a pointing error model in free- space optical communications. After this statistical misalignment fading model is proposed, it has been used in many subsequent studies [15]

the receiver's effective area is assumed to be a circle with a radius of a. Also, the beam of a transmitter is circle, and its radius changes with the distance. The radius at the distance d is denoted as ρ, which meets $0 \le \rho \le wd$ where wd denotes the maximum radius of the transmission beam. Moreover, the antenna movements from building sways give rise to pointing errors between transmission and receiving beams. The radial distance between such beams is denoted as r. On this basis, a statistical model can be utilized to characterize the misalignment fading coefficient, which is approximated as

$$\mathrm{h}(r,d) \approx A_0 \exp\left(-\frac{2r^2}{w_{eq}^2}\right) \tag{II.18}$$

where A_0 is the fraction of received power, which is determined by the radius of the receiver's effective area and the maximum radius of the transmission beam at the distance. Besides, weq denotes the equivalent beam width. A_0 and weq are studied in details.

As stated the random motion of antennas can be modeled as the Gaussian movement. Furthermore, assuming the elevation and the horizontal sways follow independent identical Gaussian distributions, the radial distance obeys a Rayleigh distribution. Therefore, the probability density function of the misalignment fading can be evaluated as

$$f_{h_{mis}}(x) = \frac{\gamma^2}{A_0^{\gamma^2}} x^{\gamma^2-1} \quad , 0 \leq x \leq A_0 \tag{II.19}$$

where the parameter γ is the ratio between the receiver's equivalent beam width and the radial displacement standard deviation.

Although the misalignment fading induced by highly-directional antennas triggers a drastic decrease of the received power, such as up to 13 dB power decrease for 20◦ tilts, it is still worth of rewarding for THz wireless networks with particular respects.

the propagation gain and the molecular absorption loss are both deterministic variables, while the above misalignment fading is a random variable. However, if antennas move regularly rather than meeting random motion, the performance of stochastic misalignment fading models will be degraded. For example, in an eavesdropping scenario, the angle be- tween the antenna of a legitimate user and that of an eavesdropper may be constant. In this case, an exact characterization of the misalignment fading is urgently needed. The impact of antenna directivity at 300 GHz is studied by means of a Gaussian beam model based on ray tracing (RT). So far, there is limited re- search investigating deterministic models of misalignment fading, and it is extremely meaningful to further construct a more comprehensive and accurate deterministic misalignment fading model in the future [15]

d) Multipath fading model:

The multipath fading is definitely one of the key aspects, and it is of great significance to develop a novel multipath channel models for THz communications. In fact, the power difference between the Line of Sight (LoS) and Non Line of Sight (NLoS) is larger in the THz band than that in the mmWave bands. On average, when compared to the LoS path, the attenuation of the power of the firstorder reflection path is larger than 10 dB, and that of the secondorder reflection path is larger than 20 dB in the 275-325 GHz band in an indoor environment. Furthermore, according to the previous study, there are an exceedingly limited number of NLoS paths. Hence, THz channels are generally considered to be LoS dominant and NLoS assisted, and sensitive to obstacles. A lot of research on THz channels so far focuses on the LoS path, and neglects NLoS paths since the LoS path usually plays a decisive role in THz propagation, while a multipath channel should be considered in several specific scenarios, the indoor

THz communication just as an example, since the multipath effects are often obvious in such limited spaces [15]

A lot of studies of THz multipath channels especially indoor channels, have been indeed done, and suggest that the S-V channel model can be adopted for THz communications. Moreover, a 2-D geometrical propagation model for THz indoor scenarios. Based on RT, the authors in discussed multipath channels in details, including the LoS, reflected, scattered and diffracted ray propagation schemes. To be more specific, the Kirchhoff theory, the modified Beckmann Kirchhoff theory and the Fresnel Knife Edge Diffraction theory are used to characterize the reflected, scattered and diffracted paths, respectively. a new two-path channel model was proposed in the 275-400 GHz band, which is composed of the LoS path and one reflected path. In addition, the work presented an analytical model for a THz multipath channel, which analytically derives the number of multipath components and the probability of the LoS.

On the other hand, quite a few studies characterize the THz multipath fading by means of stochastic models. For instance, the authors presented a THz multipath channel model. According to, the first- order statistics of its attenuation factor can be characterized by the Nakagamim or Rician distribution in the LoS dominant scenarios, while they should be described by the Rayleigh or Nakagamim distribution when there is no LoS path. the $\alpha - \mu$ distribution is suggested to model the THz multipath channel, and the channel capacity and outage probability are presented based on it. Very recently, the work stated that the fluctuating Two Ray (FTR) distribution can characterize the small-scale amplitude fading well, since the FTR distribution fits the measurement data much better than the Rician, Gaussian, Nakagamim distributions in the train-to-train scenarios at 304 GHz [15].

$$y(t) = g_1 \times s(t) + 0.5[\tau \times g_2 \times s(t)] + 0.25[\tau \times g_2 \times s(t)] + n(t) \tag{II.20}$$

Where, y(t) is output signal, s (t) is input signal, τ is delay or phase shift, $g_1$is fixed gain, g_2 is variable gain and n (t) is noise.

e) Rain attenuation model:

Rain attenuation is one of important obstacle to over-come for imaging and sensing system to detect the hazardous things using Terahertz waves above 300 GHz because of its masking action. Raindrop-size distribution has been found to play an important role in monitoring rainfall and in predicting the rain attenuation. The rain attenuation is particularly severe and greatly dependent on various models of raindrop-size distribution in a Millimeter and Terahertz wave system.

Rain attenuation using three types of raindrop-size distributions and a specific attenuation model for use in prediction method recommended by ITU-R. For calculations using by raindrop-size distributions, rain specific attenuation A in dB/km is calculated by integrating all of the drop sizes as [20]

$$A = 4.343 \int Q(D, \lambda, m) N(D) dD \tag{II.21}$$

where Q is the attenuation cross section that is a function of the drop diameter

D, the wavelength of the radio wave λ, and the complex refractive index of the water drop m, which is a function of the frequency and the temperature, and

N(D) is the drop-size distribution. The attenuation cross section Q is found by applying the classical scattering theory of Mie for a plane wave radiation to an absorbing sphere particle. According to Hulst, the cross-section Q is expanded as

$$Q(\mathrm{D}, \lambda, m) = \frac{\lambda^2}{2\pi} \sum_{n=1}^{\infty} (2n+1) Re[\, a_n + b_n] \tag{II.22}$$

where a_n and b_n are the Mie scattering coefficients, which are complex functions of m, D, and λ. The complex refractive index of liquid water

m was taken from. The “Mie scattering coefficients” a_n and b_n in Equation represent a contribution to the scattered field from the multi poles induced in the sphere, such as raindrops. For calculation by using the recommended prediction methods by ITU-R, rain specific attenuation R γ dB/ km is obtained from the rain rate R mm/hr using the power-law relationship:

$$\gamma_{R=KR^{\alpha}} \tag{II.23}$$

Values for the constants for the coefficients k and α are determined as functions of frequency, f GHz, in the range from 1 to 1000 GHz, from the equations which have been developed from curve-fitting to power-law coefficients derived from scattering calculations. It is shown in ITU-R P.838-3 [20]

II.2.1.3. THz channel models by methods:

a) Stochastic model:

A stochastic model has the capacity to handle uncertainties in the inputs applied, making it useful for representing real-world systems where variability is present. These models possess inherent randomness, meaning that even when provided with the same set of parameter values and initial conditions, they generate an ensemble of different outputs. This characteristic distinguishes stochastic models from deterministic ones, as the latter always produce the same outcome under identical conditions. By incorporating randomness, stochastic models better capture uncertainty and fluctuations in complex systems. [21] [22]

The following theorems return the probability density function (PDF) and cumulative density function (CDF) of the random process that is used to model the stochastic behavior of the channel,

$$|h_{fp}| = |h_f||h_p| \quad \text{(II.24)}$$

Theorem 1: The PDF of |hfp | can be analytically evaluated as

$$f_{|h_{fp}|} = \gamma^2 A_0^{-\gamma^2} \frac{\mu^{\frac{\gamma^2}{\alpha}}}{\hat{h}_f^{\alpha}\Gamma(\mu)} x^{\gamma^2-1} \times \Gamma\left(\frac{\alpha\mu-\gamma^2}{\alpha}, \mu\frac{x^{\alpha}}{\hat{h}_f^{\alpha}} A_0^{-\alpha}\right) \quad \text{(II.25) (II.22) (II.22) (II.22)}$$

Proof: Please see Appendix A.

Theorem 2: The CDF of |hfp | can be obtained as

$$F_{|h_{fp}|}(x) = 1 - \frac{1}{\alpha}\frac{x^{\gamma^2}}{\hat{h}_f^{\gamma^2}}\frac{\gamma^2}{A_0^{\gamma^2}} \times \sum_{k=0}^{\mu-1}\frac{\mu^{\frac{\gamma^2}{\alpha}}}{k!}\Gamma\left(\frac{\alpha k-\gamma^2}{\alpha}, \mu\frac{x^{\alpha}}{\hat{h}_f^{\alpha}} A_0^{-\alpha}\right) \quad \text{(II.26)}$$

Proof: Please see Appendix B.

it is evident that the presented distribution depends on the multipath fading channels characteristics, which are modeled through the parameters α and μ, as well as the level of misalignment fading that is described via the parameters γ^2 and A_0[23].

b) Deterministic model:

Deterministic channel models accurately model the wave propagation based on the theory of electromagnetic (EM) wave propagation. The approach is site-specific, and requires detailed geometric information of the propagation environment, dielectric properties of materials and spatial positions of the Tx and the Rx. Therefore, a deterministic approach provides a good agreement between the simulation results and the measurements in general, though the accuracy varies based on the specific method, the accuracy of the environmental information, and the analyzed frequency band. The results from the deterministic modeling can be useful by themselves (e.g., for deployment planning), to provide statistical channel information by applying Monte Carlo analysis on many random transmit/receive locations, and/or as input for statistical channel modeling. In particular, RT, and finite Difference Time Domain (FDTD) are two representative methods of deterministic channel modeling, while the use of measured, stored impulse responses (or equivalent) is another possible deterministic approach [24].

1) Ray-Tracing: RT has emerged as a popular technique for the analysis of site-specific scenarios, due to its ability to analyze very large structures with reasonable computational resources. The ray-tracing algorithm models the propagation of electromagnetic waves based on the high-frequency approximation of Maxwell's equations, geometrical optics. The locations of the Tx and the Rx are first specified, followed by determining all possible routes between the transceivers, based on high-frequency-approximation rules like Geometric Optic (GO), Geometric Theory of Diffraction (GTD), Uniform Theory of Diffraction (UTD), and Kirchhoff theory. The technique is especially suitable for THz channels due to the fact that these approximations become more accurate due to the stronger corpuscular property in the THz band, which is associated with the wave-particle (wave corpuscle) duality of light.

2) Finite-Domain Time-Domain: FDTD is also known as Yee's method named after the Chinese American applied mathematician Kane S. Yee. It is a numerical analysis technique that directly solves Maxwell's equations. FDTD can resolve the impact of small and complex scatterers, and rough surfaces in the THz band, but suffers from very high computational complexity when applied to an environment that has large dimensions in units of wavelength, as often occurs for THz channels. Furthermore, a database of the environment with sufficient resolution, e.g., a point cloud from laser scanning (see above) is required [24]

3) Measurement-based: The measurement-based approach relies on channel measurement along with data storage. The concept of "stored measurements" has been used at least since the 1990s, when channel sounder measurements started to be stored digitally. Various projects, such as the Metamorp project, attempted to standardize formats for data storage both in the time domain (as impulse response) and frequency domain (transfer function). Major challenges revolve in particular around unified formats of metadata such as calibration data of the channel sounders, and descriptions of the measurement parameters and environments. More recently, the principle of "open source" data has motivated many researchers to place measurement results online for download. Various standardization groups, including the NextG Channel alliance aim to facilitate data exchange. The challenges in the context of THz channels revolve around the size of the measured data, both due to the large bandwidth, and large antenna arrays [24]

II.3. Characteristics Thz channel model:

- **Path loss models:**

The path loss elements include free-space loss, atmospheric losses due to gaseous and water vapor absorption, precipitation, fading loss due to multipath, and other miscellaneous effects based on frequency and the environment [25]

Path loss models are used to compute the decrease in the power of a radio signal as it propagates away from the transmitter. Path loss models are implemented by path loss modules, which are submodules of the radio medium module. The default path loss model is most often free space path loss, which computes attenuation according to the inverse square law along a single line-of-sight propagation path. This is a simple model, and realistic only in certain cases. Because of its low computational cost, it is also useful if the emphasis of the simulation is not on the accuracy of radio propagation (e.g. for testing protocols.) However, there are several more path loss models available in INET, suitable for various other scenarios. Here is a list of the path loss module types featured in this showcase example:

$$PL = Alog_{10}(d) + B + X_{\sigma} \tag{II.27}$$

where PL is short for path loss, d is the distance between the Tx and the Rx, A is the slope, B is the intercept, and Xσ is the shadow fading, which can be expressed as a Gaussian variable with zero mean value and a standard deviation of σ [25]

- **Free Space Path Loss** computes the loss of signal power in a single line-of-sight propagation path, without any reflections or shadowing.
- **Two Ray Ground Reflection** computes the loss of signal power by assuming a line-of-sight wave interfering with another wave reflected from the ground between the transmitter and the receiver. This model computes interference in the far-field only and is the same as free space path loss up until a certain crossover distance [26]
- **Two Ray Interference** is the same as the two-ray ground reflection model in the far-field, but it models the interference of the two waves in the near-field as well.
- **Rician Fading** is a stochastic path loss model that assumes a dominant line-of-sight signal and multiple reflected signals between the transmitter and the receiver. It is useful for modeling radio propagation in an urban environment.
- **Log Normal Shadowing** is a stochastic path loss model where power levels follow a lognormal distribution. It is useful for modeling shadowing caused by objects such as trees [26]

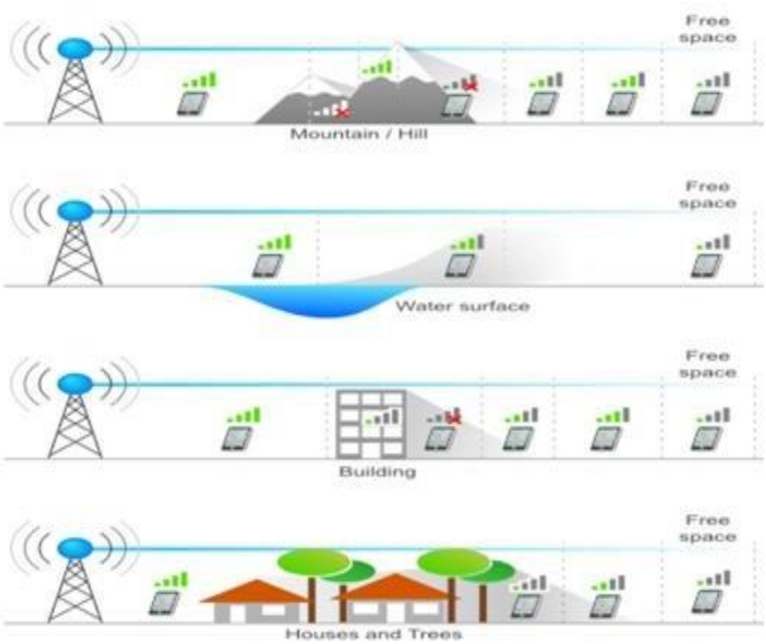


Figure II.4: Examples of Path Loss Models

II .4. CONCLUSION

Based on the discussions presented above, the conclusions are obtained. Terahertz (THz) communication, operating in the 0.1–10 THz frequency range, is a key technology for 6G wireless systems, enabling ultra-high data rates in the terabit-per-second (Tbps) range. However, THz signals suffer from high path loss, molecular absorption, and limited transmission range, making line-of-sight (LoS) propagation, high-gain directional antennas, and beamforming essential. THz channel models are classified based on outdoor, indoor, and nanoscale environments, each addressing unique propagation challenges such as multipath fading, misalignment fading, and rain attenuation. While deterministic models like ray-tracing (RT) and finite-difference time-domain (FDTD) provide precise predictions, stochastic models are used for practical deployment scenarios. Despite these challenges, THz technology holds great promise for 6G, high-speed indoor networks, nanoscale communication, and advanced imaging applications. Future research must focus on optimizing hardware, signal processing, and propagation models to overcome current limitations and fully harness THz potential

CHAPTER III:
Algorithms of THz Channel Estimation

III.1. INTRODUCTION

Terahertz communication is one of the most promising wireless communication technologies for 6G generation and beyond. and it is based on measurements specifically made for an intended communication system. Propagation models are the base for channel modelling, as they try to describe signal changes during its travel from the transmitter to the receiver [27]

For THz systems to be practically adopted, As the key to wireless communication, channel estimation has become a hot research topic in recent years. We consider the problem of channel modeling and estimation with deterministic channel propagation and the related physical characteristics of THz bands, and benchmark various machine learning algorithms to estimate THz channel, including neural networks (NN), logistic regression (LR), and projected gradient ascent (PGA) [28]

The aim to introduce wireless channel models by providing a selection of the most popular ones. The types of fading typical of the wireless environment are also presented, together with the relevant propagation models. carrying out vast indoor and outdoor measurement campaigns at the packet level, in order to define channel models derived [27]

III.2. Channel Estimation

In all communication systems, data is transferred from source to the destination in form of signals. These signals traverse different medium which can be wired or wireless. Copper wires or fibre cables are two examples of wired medium while air is a wireless medium. These mediums are also called channel. When a signal passes from channel, it is distorted from the noise or from other signals traversing that same medium. This means that when signal is received at its destination, it could have errors. So, in order to remove the noise and distortion effects of channel from the received signal, channel's properties have to be found out. The process of figuring out channel characteristics is called Channel Estimation [29].

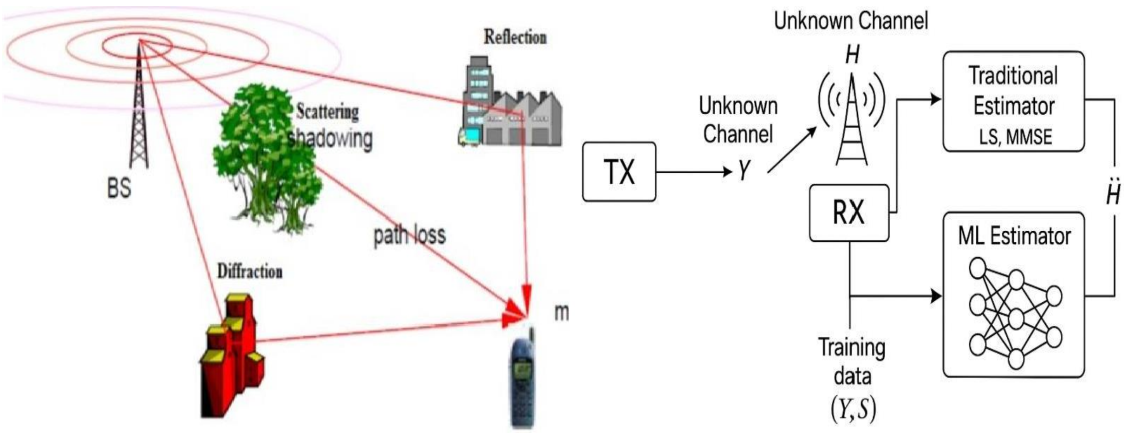


Figure III.1: Learning Based Channel Estimation

Channel estimation process consists of multiple steps. First a mathematical model is created of the channel. Then a signal which is known by both sender and receiver is transmitted over the channel. When the receiver receives the signal, it is of course distorted and contains noise from the channel, but the receiver also knows the original signal, thus it can compare the original signal and received signal to extract the properties of channel and the noises added to the sent signal in the channel.

To put is in 3 main steps:

1. Mathematical model for channel is created. This model correlates sent and received signal using channel matrix.

2. A signal known by both sender and receiver is sent by sender over the channel.

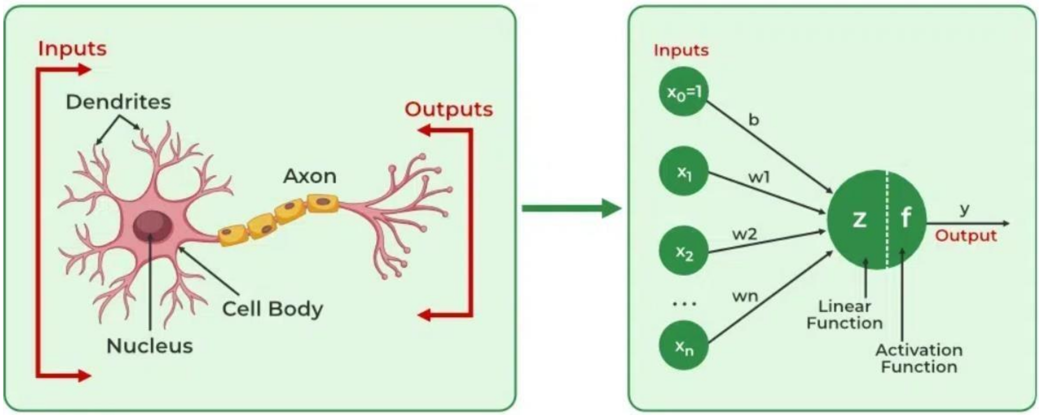
3. Receiver compares the received signal with original signal and figures III.1 out the values in channel matrix [29]

III.3. Algorithms of THz channel estimation

III.3.1. Neural Network Algorithm (NN)

Neural networks are machine learning models that mimic the complex functions of the human brain. These models consist of interconnected nodes or neurons that process data, learn patterns, and enable tasks such as pattern recognition and decision making.

In this article, we will explore the fundamentals of neural networks, their architecture, how they work, and their applications in various fields. Understanding neural networks is essential for anyone interested in the advancements of artificial intelligence [30]



FigureIII.2: Analogy Between a Biological Neuron and an Artificial Neuron

III.3.1.1. Neural Network Architecture:

Input Layer: This is where the network receives its input data. Each input neuron in the layer corresponds to a feature in the input data.

Hidden Layers: These layers perform most of the computational heavy lifting. A neural network can have one or multiple hidden layers. Each layer consists of units (neurons) that transform the inputs into something that the output layer can use.

Output Layer: The final layer produces the output of the model. The format of these outputs varies depending on the specific task (e.g., classification, regression) [30]

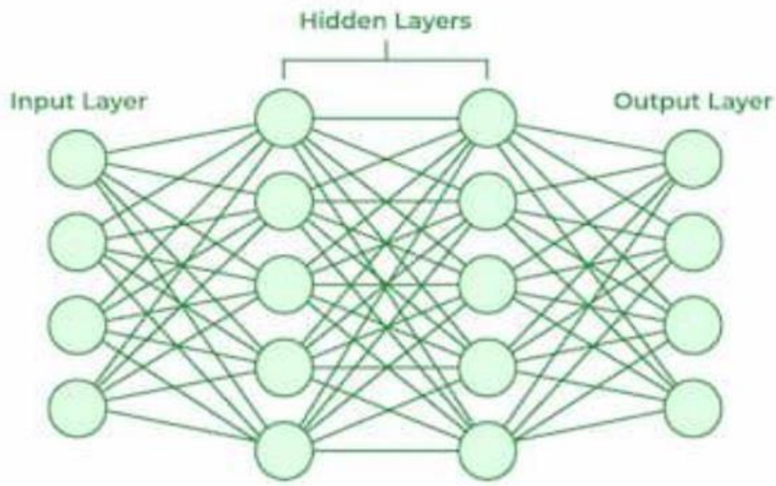


Figure III.3: Neural Network Architecture

III.3.1.2. Principal of Neural Networks Working:

1. Forward Propagation :

When data is input into the network, it passes throughh the network in the forward direction, from the input layer through the hidden layers to the output layer. This process is known as forward propagation.

Here's what happens during this phase:

- **Linear Transformation:**

Each neuron in a layer receives inputs, which are multiplied by the weights associated with the connections. These products are summed together, and a bias is added to the sum. This can be represented mathematically as:

$$Z = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + b \quad \text{(III.1)}$$

Where w represents the weights, x represents the inputs, and b is the bias.

- **Activation:**

The result of the linear transformation (denoted as z) is then passed through an activation function. The activation function is crucial because it introduces non-linearity into the system, enabling the network to learn more complex patterns. Popular activation functions include ReLU, sigmoid, and tanh.

In the input layer, tanh function is selected as activation function since the input data contains negative values, which can be expressed as

$$f_{tanh}(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{III.7}$$

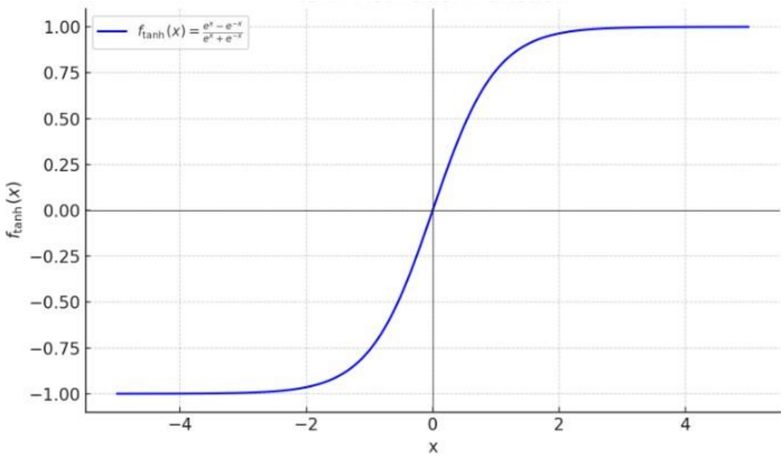


Figure III.4: tanh Function

Where x stands for the input of the function. The rectified linear unit (ReLU) function is explored at the CV layers for its fast computation speed, which is represented as:

$$f_{ReLU}(x) = \max(0, x) \tag{III.8}$$

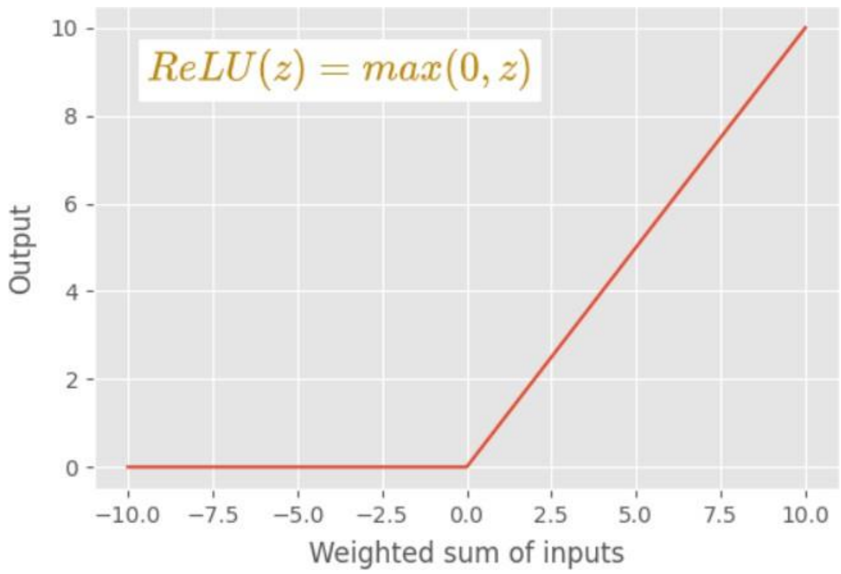


Figure III.5: ReLU Function

We also choose the sigmoid function at the output layer as

$$f_{sigmoid}(x) = \frac{1}{1+e^{-x}} \quad \text{(III.9)}$$

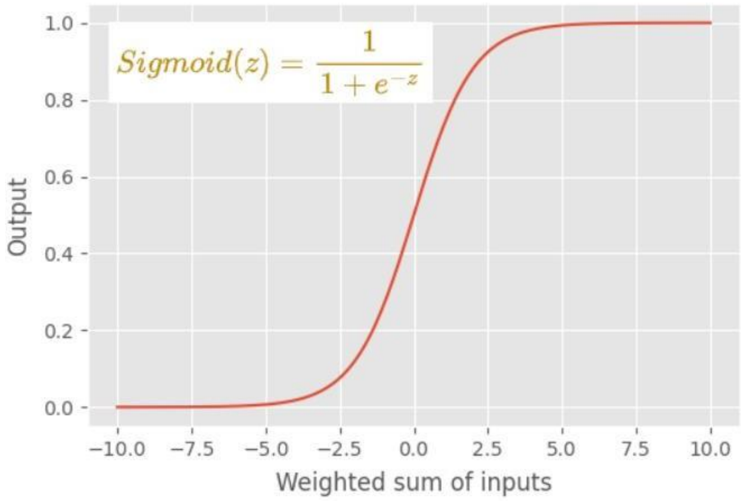


Figure III.6: Sigmoid Function

- **Backpropagation :**

After forward propagation, the network evaluates its performance using a loss function, which measures the difference between the actual output and the predicted output. The goal of training is to minimize this loss. This is where backpropagation comes into play:

- **Loss Calculation:** The network calculates the loss, which provides a measure of error in the predictions. The loss function could vary; common choices are mean squared error for regression tasks or cross-entropy loss for classification.
- **Gradient Calculation:** The network computes the gradients of the loss function with respect to each weight and bias in the network. This involves applying the chain rule of calculus to find out how much each part of the output error can be attributed to each weight and bias.
- **Weight Update:** Once the gradients are calculated, the weights and biases are updated using an optimization algorithm like stochastic gradient descent (SGD). The weights are adjusted in the opposite direction of the gradient to minimize the loss. The size of the step taken in each update is determined by the learning rate [30]

III.3.1.3. Types of Neural Networks

There are five types of neural networks that can be used.

Feedforward Networks: A feedforward neural network is a simple artificial neural network architecture in which data moves from input to output in a single direction.

Multilayer Perceptron (MLP): MLP is a type of feedforward neural network with three or more layers, including an input layer, one or more hidden layers, and an output layer. It uses nonlinear activation functions.

Deep Convolutional Neural Network (DCNN): A Convolutional Neural Network (DCNN) is a specialized artificial neural network designed for image processing. It employs convolutional layers to automatically learn hierarchical features from input images, enabling effective image recognition and classification [30]

where the THz Array of Subarrays AoSA architecture is equipped at both transmitter (Tx) and receiver (Rx). In the AoSA, there are N_{RF} RF- chains with N_{RF}<<N where N represents the number of antennas, leading that each RF-chain connects to

$N_a = N/N_{RF}$ antennas that form one subarray. Among the subarrays, analog beamforming and combining are conducted. In this case, the corresponding analog beamforming matrix F and combining matrix W hold the same block diagonal structure. In particular, the form of W can be expressed as [31]

$$\mathrm{W} = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{N_{RF}} \end{bmatrix} \qquad \text{(III.2)}$$

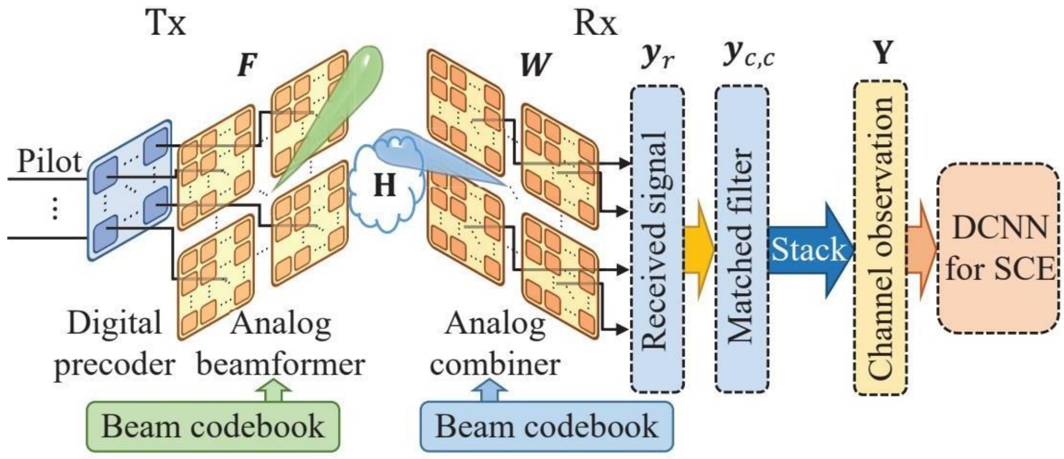


Figure III.7: THz AoSA system model for SCE

where $W_k = [w_{1,k}, \dots, w_{Na,k}]^T$represents the analog combining vector of the K^{th} RF-chain, with k = 1, ..., N_{RF}. Since the analog combining is implemented by phase shifters, each element in wk can be expressed as

$$w_{na,k} = \left({}^{1}\!/_{\sqrt{N}}\right) e^{j2\pi\widetilde{w}_{n_a,k}} \quad \text{(III.3)}$$

Where $n_a = 1, \dots, N_a$ stands for the index of the antenna on the subarray, the phase shift coefficient satisfies $0 \leq \widetilde{w}_{n_a,k} \leq 1$. In addition, during the SCE stage,

At the receiver side, by denoting the c^{th} codeword at Tx and Rx as F_c and W_c , respectively, with c = 1, ..., C is the codeword index, the received signal can be represented as

$$y_r = W_c^H H F_C S + W_C^H N \quad \text{(III.4)}$$

where H stands for the THz channel matrix. The transmitted pilot symbol vector is denoted by s. Since the pilot symbol vectors are orthogonal to each other, we have $SS^H = I_{Na}$. In addition, n refers to the complex additive white Gaussian noise (AWGN), with zero mean and variance σ_n^2,which follows the distribution as n$\sim CN(0, \sigma_N^2 In)$. Furthermore, the digital precoding matrix is chosen as an identity matrix during the pilot transmission stage, without loss of generality.

After that, the received signal y_r is sent to the matched filter, which can be represented by multiplying s^H to y_r as

$$y_{c,c} = W_C^H H F_C + N_C \quad \text{(III.5)}$$

Where $N_C = W_C^H ns^H$ describes the modified noise. Following the same procedure, C^2matrices like can be acquired by varying the beam codewords at both Tx and Rx sides, to traverse all codebook combinations. Stacking them together, we can construct the channel observation matrix as

$$\mathrm{Y}=\overline{W}^H H\overline{F} + N \qquad \text{(III.6)}$$

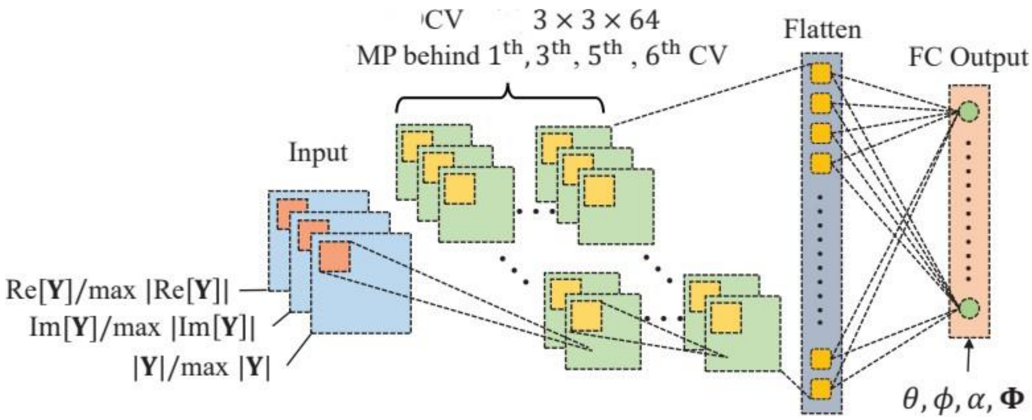


Figure III. 8: The Structure of The DCNN Framework Proposed in [31]

The structure of the proposed DCNN network is illustrated in Fig III. 8, which learns the estimation of channel parameters in light of the channel observation matrix. In total, the DCNN architecture contains several layers, including one input layer, many convolution (CV) layers, max-pooling (MP) layers, one flattening layer and one fully-connected (FC) output layer. In each layer, an activation function describing the non- linear mapping relationship is tailored to the neurons, which is basic unit of the DCNN network

after the last CV layer, a flatting layer rearranges the neurons into one dimension and connects to the FC output layer for parameter estimation results. Distinguished from CV layer, the neurons between neighboring layers in FC are fully connected. the output z (M) based on the proposed method is represented as [31]

$$Z^{(M)} = F^{(M)}(F^{(M-1)}(\dots F^{(1)}(x))) \qquad \text{(III.10)}$$

Recurrent Neural Network (RNN): An artificial neural network type intended for sequential data processing is called a Recurrent Neural Network (RNN). It is appropriate for applications where contextual dependencies are critical, such as time series prediction and natural language processing, since it makes use of feedback loops, which enable information to survive within the network.

$$h_t = \sigma_h(w_{xhxt} + w_{hh}h_{t-1} + b_h) \tag{III.11}$$

where w_{xh} is the weight matrix between the input and hidden layer, w_{hh} is the weight matrix for the recurrent connection, b_h is the bias vector, and σ_h is the activation function, typically the hyperbolic tangent function (tanh) or the rectified linear unit. The output at each time step, t, is given by the following:

$$y_t = \sigma_y(w_{hy}h_t + b_y) \tag{III.12}$$

where w_{hy} is the weight matrix between the hidden and output layers, b_y is the bias vector, and σ_y is the activation function for the output layer.[32]

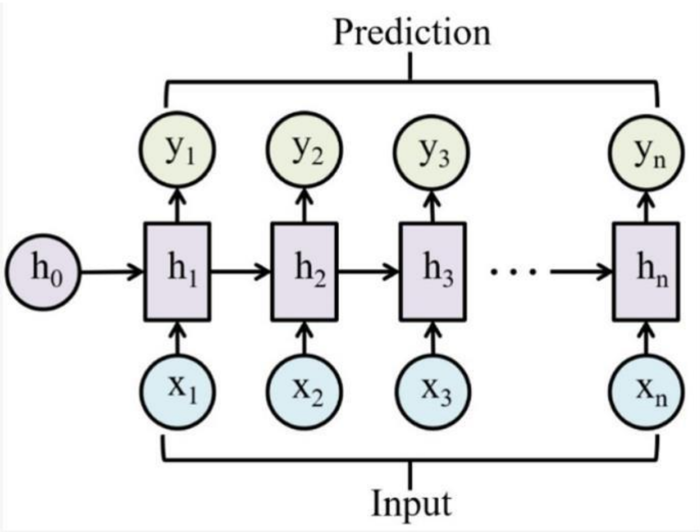


Figure III. 9: Basic RNN Architecture

Long Short-Term Memory (LSTM): LSTM is a type of RNN that is designed to overcome the vanishing gradient problem in training RNNs. It uses memory cells and gates to selectively read, write, and erase information [30]

III.3.1.4. Advantages of Neural Networks

Neural networks are widely used in many different applications because of their many benefits:

Adaptability: Neural networks are useful for activities where the link between inputs and outputs is complex or not well defined because they can adapt to new situations and learn from data.

Pattern Recognition: Their proficiency in pattern recognition renders them efficacious in tasks like as audio and image identification, natural language processing, and other intricate data patterns.

Parallel Processing: Because neural networks are capable of parallel processing by nature, they can process numerous jobs at once, which speeds up and improves the efficiency of computations.

Non-Linearity: Neural networks are able to model and comprehend complicated relationships in data by virtue of the non-linear activation functions found in neurons, which overcome the drawbacks of linear models [30]

III.3.1.5. Disadvantages of Neural Networks:

Neural networks, while powerful, are not without drawbacks and difficulties:

Computational Intensity: Large neural network training can be a laborious and computationally demanding process that demands a lot of computing power.

Black box Nature: As “black box” models, neural networks pose a problem in important applications since it is difficult to understand how they make decisions.

Overfitting: Overfitting is a phenomenon in which neural networks commit training material to memory rather than identifying patterns in the data. Although regularization approaches help to alleviate this, the problem still exists.

Need for Large datasets: For efficient training, neural networks frequently need sizable, labeled datasets; otherwise, their performance may suffer from incomplete or skewed data [30]

III.3.1.6. Applications of Neural Networks

Neural networks have numerous applications across various fields:

Image and Video Recognition: CNNs are extensively used in applications such as facial recognition, autonomous driving, and medical image analysis.

Natural Language Processing (NLP): RNNs and transformers power language translation, chatbots, and sentiment analysis.

Finance: Predicting stock prices, fraud detection, and risk management.

Healthcare: Neural networks assist in diagnosing diseases, analyzing medical images, and personalizing treatment plans.

Gaming and Autonomous Systems: Neural networks enable real-time decision-making, enhancing user experience in video games and enabling autonomous systems like self-driving cars [30]

III.3.2. Logistic Regression Algorithm (LR)

Logistic regression is a statistical method for developing machine learning models with binary dependent variables, i.e. binary. Logistic regression is a statistical technique used to describe data and the relationship between one dependent variable and one or more independent variables. used for classification tasks where the goal is to predict the probability that an instance belongs to a given class or not. Logistic regression is a statistical algorithm which analyze the relationship between two data factors. The article explores the fundamentals of logistic regression, it's types and implementations.

Logistic regression is used for binary classification where we use sigmoid function, that takes input as independent variables and produces a probability value between 0 and 1.

For example, we have two classes Class 0 and Class 1 if the value of the logistic function for an input is greater than 0.5 (threshold value) then it belongs to Class 1 otherwise it belongs to Class 0. It's referred to as regression because it is the extension of linear regression but is mainly used for classification problems.

Key Points :

- Logistic regression predicts the output of a categorical dependent variable. Therefore, the outcome must be a categorical or discrete value.
- It can be either Yes or No, 0 or 1, true or False, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lie between 0 and 1.
- In Logistic regression, instead of fitting a regression line, we fit an "S" shaped logistic function, which predicts two maximum values (0 or 1) [33]

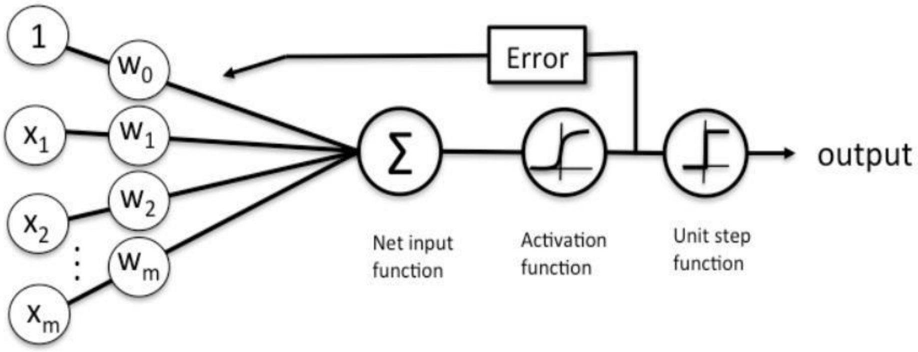


Figure III. 10: Logistic Regression Algorithm

III.3.2.1. Types of Logistic Regression

On the basis of the categories, Logistic Regression can be classified into three types:

Binomial: In binomial Logistic regression, there can be only two possible types of the dependent variables, such as 0 or 1, Pass or Fail, etc.

Multinomial: In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable.

Ordinal: In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as "low", "Medium", or "High" [33]

III.3.2.2. Assumptions of Logistic Regression

We will explore the assumptions of logistic regression as understanding these assumptions is important to ensure that we are using appropriate application of the model. The assumption include:

Independent observations: Each observation is independent of the other. meaning there is no correlation between any input variables.

Binary dependent variables: It takes the assumption that the dependent variable must be binary or dichotomous, meaning it can take only two values. For more than two categories SoftMax functions are used.

Linearity relationship between independent variables and log odds: The relationship between the independent variables and the log odds of the dependent variable should be linear.

No outliers: There should be no outliers in the dataset.

Large sample size: The sample size is sufficiently large [33]

III.3.2.3. Logistic Regression working

The logistic regression model transforms the linear regression function continuous value output into categorical value output using a sigmoid function, which maps any real-valued set of independent variables input into a value between 0 and 1. This function is known as the logistic function. [33]

Let the independent input features be:

$$X = \begin{bmatrix} x_{11} & \dots & x_{1m} \\ x_{21} & \dots & x_{2m} \\ x_{n1} & \dots & x_{nm} \end{bmatrix} \tag{III.13}$$

and the dependent variable is Y having only binary value i.e. 0 or 1.

$$Y=\begin{cases} 0 & if\ class\ 1 \\ 1 & if\ class\ 2 \end{cases} \tag{III.14}$$

then, apply the multi-linear function to the input variables X.

$$Z= \left(\sum_{i=1}^{n} w_i x_i\right) + b \tag{III.15}$$

Here x_i is the ith observation of X, $\omega_i = [\omega_1, \omega_2, \omega_3, \dots, \omega_m]$ is the weights or Coefficient, and b is the bias term also known as intercept. simply this can be represented as the dot product of weight and bias

$$Z = \omega \cdot X + b \tag{III.16}$$

- **Equation of Logistic Regression :**

The odd is the ratio of something occurring to something not occurring. it is different from probability as the probability is the ratio of something occurring to everything that could possibly occur. so odd will be:[33]

$$\frac{p(x)}{1-p(x)} = e^{z} \tag{III.18}$$

Applying natural log on odd. then log odd will be:

$$\log\left[\frac{p(x)}{1-p(x)}\right] = z \tag{III.19}$$

$$\frac{p(x)}{1-p(x)} = e^{\omega.X+b} \tag{III.20}$$

Exponentiate both sides

$$P(x)=e^{\omega.X+b}.\,(1-p(x)) \tag{III.21}$$

$$P(x)=\frac{e^{\omega.X+b}}{1+e^{\omega.X+b}} \tag{III.22}$$

then the final logistic regression equation will be:

$$p(X\,;b,\omega)=\frac{e^{\omega.X+b}}{1+e^{\omega.X+b}}=\frac{1}{1+e^{-\omega.X+b}} \tag{III.23}$$

- **Likelihood Function for Logistic Regression**

The predicted probabilities will be:

- for y=1 The predicted probabilities will be: p(X;b, ω) = p(x)
- for y = 0 The predicted probabilities will be: 1-p (X;b, ω) = 1-p(x)

$$L(b,\omega)=\prod_{i=1}^{n} p(x_i)^{y_i}(1-p(x_i))^{1-y_i} \tag{III.24}$$

Taking natural logs on both sides;

$$\text{Log}\,(L(b,\omega)) = \sum_{i=1}^{n} y_i logp(x_i) + (1-y_i)\log(1-p(x_i))$$

$$=\sum_{i=1}^{n} -log1 + e^{\omega.x_i+b} + \sum_{i=1}^{n} y_i(\omega.x_i + b) \quad \text{(III.25)}$$

- **Gradient of the log-likelihood function**

To find the maximum likelihood estimates, we differentiate w.r.t ω

$$\frac{\partial J(l(b,\omega)}{\partial \omega_i} = \sum_{i=n}^{n}(y_i - p(x_i; b, \omega))x_{ij} \quad \text{(III.26)}$$

III.3.2.4. Advantages of logistic regression in ML

Simplicity and Interpretability: Easy to implement and understand; useful when quick results and insights are needed.

Efficiency: Performs well even when data isn't perfect; the underlying math is simple and fast to optimize.

Insightful: Suitable for binary classification; helps identify variable importance and their impact direction (positive or negative).

III.3.2.5. Disadvantages of logistic regression in ML

Limited to Discrete Outcomes: Cannot handle continuous target variables.

Assumes Linearity: Assumes linear relationships between variables, which may not hold in real-world data.

Poor at Modeling Complex Relationships: Struggles with nonlinear or highly interdependent features.

Risk of Overfitting: May overfit large or complex datasets without proper regularization [34]

III.3.3. Frank-Wolfe Algorithm (FW)

The Frank-Wolfe (FW) method, also known as the conditional gradient method, is a well-studied first-order algorithm for smooth convex optimization with a bounded feasible region. Compared to other first-order methods, such as projected gradient methods and proximal type methods, where a projection operation onto the feasible set is required at every iteration, the Frank-Wolfe method avoids projection by minimizing a linear objective over the feasible set.

Solving this problem is often computationally more attractive than a projection step in several large-scale problems arising in machine learning [35]

$$H = \min_{x \in D} f(x) \qquad \text{(III.27)}$$

III.3.3.1. Types of frank-wolfe:

- **Unbounded Frank-Wolfe Algorithms**

In this section, we present two new Frank-Wolfe algorithms designed to solve with an unbounded feasible region. As mentioned above, we alternate between performing a Frank-Wolfe step on the bounded set S and a gradient descent step along the subspace T. Since the gradient descent direction can also be viewed as an extreme ray to the solution of the Frank-Wolfe linear subproblem along the linear subspace T, we call this the “Unbounded Frank-Wolfe Method”.

Algorithm 1 presents our first algorithm: the unbounded Frank-Wolfe method, for solving. In each iteration, we first perform a gradient descent step along the unbounded subspace T. To do so, we first compute the negative projected gradient onto T, namely PT f(xk). Unlike gradient descent, the traditional Frank-Wolfe method does not have linear convergence even if the objective function is strongly convex. An intuitive explanation is that the solutions to the linear subproblems in the Frank-Wolfe algorithm may alternate between two extreme points of the constraint set, the iterate solutions zigzag, slowing down the convergence of the algorithm. Awaystep Frank-Wolfe method allows moving in a direction opposite to the maximal solution of the linear model.

Algorithm 2 adapts the away-step Frank-Wolfe method to solve with the unbounded feasible region T S. The major difference of Algorithm 2 with Algorithm 1 is that the former allows performing an away step for the update along S [35]

III.3.3.2 Frank-Wolfe method

The Frank-Wolfe method, also called conditional gradient method, uses a local linear expansion of:

$$s^{(k-1)} \in \underset{s \in c}{argmin} \nabla f(x^{(k-1)})^T \qquad \text{(III.28)}$$

$$x^{(k)} = (1-\gamma_k)x^{(k-1)} + \gamma_{k}s^{(k-1)} \qquad \text{(III.29)}$$

Note that there is no projection; update is solved directly over C

Default step sizes: $\gamma_k = 2/(k+1)$, $k = 1, 2, 3, \ldots\ldots$ Note for any $0 \le \gamma_k \le 1$, we have $x^{(k)} \in C$ by convexity. Can rewrite update as

$$x^{(k)} = x^{(k-1)} + \gamma_k(s^{(k-1)} - x^{(k-1)}) \qquad \text{(III.30)}$$

i.e., we are moving less and less in the direction of the linearization minimizer as the algorithm proceeds [36]

- **Algorithm Franke-Wolfe method to estimate:**

Instead of estimating the original mmWave channel matrix H, the authors propose estimating a pseudo-channel defined as:

$$G = HS \qquad \text{(III.31)}$$

where S is a known training matrix. Since H is assumed to be low-rank, the matrix G also retains this property. Estimating G simplifies the optimization process.

Low-Rank Constraint: The rank constraint rank(X) = r is non-convex and difficult to handle directly. Instead, it is relaxed using the nuclear norm:

This creates a convex constraint set, enabling efficient optimization. The parameter β can be chosen based on prior knowledge of the channel statistics.

Log-Likelihood Function: The log-likelihood function measures how well a candidate matrix X matches the 1-bit quantized measurements Y, accounting for Gaussian noise.

For the real part of the received signal, the log-likelihood is:[37]

$$\mathcal{L}_{Y^R}(X^R) = \sum_{k=1}^{N}\sum_{l=1}^{N}[\mathbb{1}_{[Y^R_{k,l}=1}\log\left(\Phi(X^R_{k,l}/\sigma)\right) + \mathbb{1}_{[Y^r_{k,l}=1}\log\left(1-\Phi(X^R_{k,l}/\sigma)\right)]$$

(III.32)

where $\Phi(X^R_{k,l}/\sigma)$ is the cumulative distribution function (CDF) of the standard normal distribution.

The full log-likelihood includes both real and imaginary parts:

$$\mathcal{L}_Y(X) = \mathcal{L}_{Y^R}(X^R) + \mathcal{L}_{Y^1}(X^1) \quad \text{(III.33)}$$

Because both components are concave, the overall function $\mathcal{L}_Y$(X) is also concave, making it well-suited for convex optimization methods.

approximation is determined by the gradient, this method selects a matrix D_t within $\|X\|* \leq \beta$ that maximizes the inner product $\ll D_t, \nabla\mathcal{L}_Y(X_t) \gg$. The matrix D_t is simply the rank-one approxi-mation of the gradient $\nabla\mathcal{L}(X_t)$. For a step size of γ_t , the optimization variable X_t is incremented by $\gamma_t(D_t - X_t)$ to obtain X_{t+1}. A summary of the Franke-Wolfe technique to estimate G is given in Algorithm To achieve a low complexity implementation of Algorithm, we use the power method to compute the rank one- approximation of $\nabla\mathcal{L}_Y(X_t)$. Each iteration of the power method requires multiplying an N × N matrix with an N × 1 vector. With the power method-based implementation, the complexity of a single iteration of the Franke-Wolfe method is O (N^2) which is lower than that of the PGA algorithm [37]

for t= 1 to T_{max} do

$D_t \leftarrow$ Rank 1 approx. of $\nabla\mathcal{L}_Y(X_t)$ by power method

$\gamma_t \leftarrow 2/(t+2)$

$$X_{t+1} = X_t + \gamma_t(D_t - X_t) \quad \text{(III.34)}$$

Stop if 0< $\mathcal{L}_Y(X_{t+1}) - \mathcal{L}_Y(X_t) < \epsilon \left|\mathcal{L}_Y(X_t)\right|$

End for

$$\hat{G} = X_{t+1} \quad \text{(III.35)}$$

III.3.3.3. Applications:

The Frank-Wolfe algorithm appears in many different contexts. Here are some examples [38]

- **Structured SVM:**

Given n samples x = (x_1, .., x_n) and their corresponding labels y = (y_1, . ., y_n). Given a weight vector w, we would like to minimize

$$\underset{y\in\{-1,+1\}^n}{\text{Min}} <\omega, \Phi_x(y)> + \frac{\lambda}{2}\|\omega\|^2 \quad \text{(III.36)}$$

This objective function is a support function (of the convex hull conv $\{\Phi_x(y) \mid y \in \{-1,1\}^m\}$) plus a squared norm. The dual of it can be derived analogously to that of the Lovász ex- tension plus squared norm, and looks similar to the min-norm problem for submodular optimization. Applying the Frank-Wolfe algorithm to the dual is, according to our above reasoning, equivalent to applying a subgradient method to the primal (non-smooth) SVM problem.

Frank-Wolfe method for the structured SVM, and derive a stochastic block coordinate descent method. This can be related to a stochastic gradient method in the primal.

- **Herding Problem:**

In the herding problem, we are are given a set of samples x_1 ,. ., x_n and are trying to ap-proximate a given mean (expectation of a feature function or sufficient statistic)

$$\mu = \mathbb{E}_{p(x)}\Phi(x) \quad \text{(III.37)}$$

by the average of a few sample points. The original Herding method picks those greedily. This method can be viewed as a Frank-Wolfe method applied to the objective

$$\min_{\omega \in conv(\{x\}_{j=1}^{n})} \|\omega - \mu\|^2 \quad \text{(III.38)}$$

With an appropriately chosen step size, we get $\omega = \frac{1}{t}\sum_{j=1}^{t}\Phi(x_j)$and hence the differ-ence between the empirical and the population mean

$$\| \frac{1}{t}\sum_{j=1}^{t}\Phi(x_j) - \mu\|^2 \quad \text{(III.39)}$$

that is being minimized.

The equivalence between Herding and Frank-Wolfe

- **Boosting:**

Boosting too can be viewed as a Frank-Wolfe method. Details are discussed in . Suppose B is the convex hull of the set of all hypotheses. We aim to choose a weight function ω(x) that minimizes

$$\underset{\omega(x) \in \mathcal{B}}{Min} \quad \mathbb{E}_{x,y} loss(\omega(x), y). \tag{III.40}$$

III.3.4. Projected Gradient Ascent (PGA)

Gradient ascent is an ubiquitous optimization algorithm used to train machine learning (ML) algorithms from simple linear regression models to sophisticated transformer architectures. ML researchers are typically introduced to unconstrained optimization problems.

The key idea behind PGD is to project the current solution onto the feasible region at each iteration. This projection step involves mapping the updated parameter onto the closest point within the feasible region in case if lands outside of it, effectively "projecting" it onto the constraint boundaries. By doing so, PGD guarantees that the resulting parameters remain feasible throughout the optimization process [39]

III.3.4.1. Projected Gradient Ascent working:

The starting point is just an arbitrary point for us to evaluate the performance. From that starting point, we will find the derivative (or slope), and from there, we can use a tangent line to observe the steepness of the slope. The slope will inform the updates to the parameters. the weights and bias. The slope at the starting point will be steeper, but as new parameters are generated, the steepness should gradually reduce until it reaches the lowest point on the curve, known as the point of convergence.

Similar to finding the line of best fit in linear regression, the goal of gradient ascent is to minimize the cost function, or the error between predicted and actual y. In order to do this, it requires two data pointsa direction and a learning rate. These factors determine the partial derivative calculations of future iterations, allowing it to gradually arrive at the local or global minimum (i.e. point of convergence.

Learning rate (also referred to as step size or the alpha): is the size of the steps that are taken to reach the minimum. This is typically a small value, and it is evaluated and updated based on the behavior of the cost function. High learning rates result in larger steps but risks overshooting the minimum. Conversely, a low learning rate has small step sizes. While it has the advantage of more precision, the number of iterations compromises overall efficiency as this takes more time and computations to reach the minimum [40]

The cost (or loss): function measures the difference, or error, between actual y and predicted y at its current position. This improves the machine learning model's efficacy by providing feedback to the model so that it can adjust the parameters to minimize the error and find the local or global minimum. It continuously iterates, moving along the direction of steepest descent (or the negative gradient) until the cost function is close to or at zero. At this point, the model will stop learning. Additionally, while the terms, cost function and loss function, are considered synonymous, there is a slight difference between them. It’s worth noting that a loss function refers to the error of one training example, while a cost function calculates the average error across an entire training set [40]

The “training” phase in machine learning usually involves numerical optimization. Minimizing a function f depending on d parameters w

$$H = \underset{\omega \in \mathbb{R}^d}{Min}\, f(\omega) \quad \text{(III.41)}$$

For differentiable f, a prototypical method is gradient ascent

$$\omega^{k+1} = \omega^k - \alpha_k \nabla f(\omega^k) \quad \text{(III.42)}$$

Cost of update is O(d) in terms. Guaranteed to decrease f for small enough step size α_k [41]

- **Projected gradient ascent method to estimate**

For PGA algorithm, the learning process is based on gradient ascent, a similar learning algorithm to gradient descent with a positive addition of the gradient at each update. PGA includes a projection step at each iteration, assuming that the matrix H has lower rank r than N, r ≪ N, where N = M_t = M_r . Singular Value Decomposition (SVD) and simplex projection are used to find the closest low rank matrix to the updated estimation of H. For Franke-Wolfe algorithm, an additional step is included to the basic gradient ascent learning, where we update the channel matrix differently: instead of adding a weighted gradient matrix, we compute the top singular vector of the gradient, then we soustract it from the gradient and update the channel matrix [28]

We now explain PGA-based estimation of G from Y. For a step size of η, the ascent step in PGA shifts X_t $\eta\nabla\mathcal{L}(X_t)$ The matrix obtained after shifting $X_t$is defined as Z_{t+1}It is important to note that Z_{t+1}may not belong to the constraint set,$\| X \|*\leq \beta, even\ when\ X_t$satisfies the constraint. The projection step in PGA finds a matrix within the constraint set that is closest to Z_{t+1} . The projection, defined as X_{t+1} , is derived using the singular value decomposition (SVD) of $Z_{t+1}$and a simplex projection. The PGA algorithm to estimate G is summarized in Algorithm. It can be noticed that the complexity of a gradient step, computing $\nabla\mathcal{L}_Y(X), is\ O(N^2)$. The complexity of the SVD step in PGA, however, is$O(N^3)$. Therefore, every iteration of the PGA algorithm has a complexity of O(N^3).[37]

For t = 1 to T_{max} do

$$Z_{t+1}= X_t + \eta\nabla\mathcal{L}_Y(X_t) \tag{III.43}$$

Compute the SVD: $Z_{t+1}diag(d_{t+1})V^*_{t+1}$

$\pi_{t+1} \leftarrow projection\ of\ d_{t+1}on\ \{d: 1^T d = \beta, d \geq 0\}$

$$X_{t+1} = U_{t+1}diag(d_{\pi+1})V^*_{t+1} \tag{III.44}$$

Stop if 0 < $\mathcal{L}_Y(X_{t+1}) - \mathcal{L}_Y(X_t) < \epsilon \left|\mathcal{L}_Y(X_t)\right|$

End for: $\hat{G} = X_{t+1}$

III.3.4.2. Types of Projected Gradient Ascent

There are three types of gradient descent learning algorithms: batch gradient descent, stochastic gradient descent and mini-batch gradient descent [40]

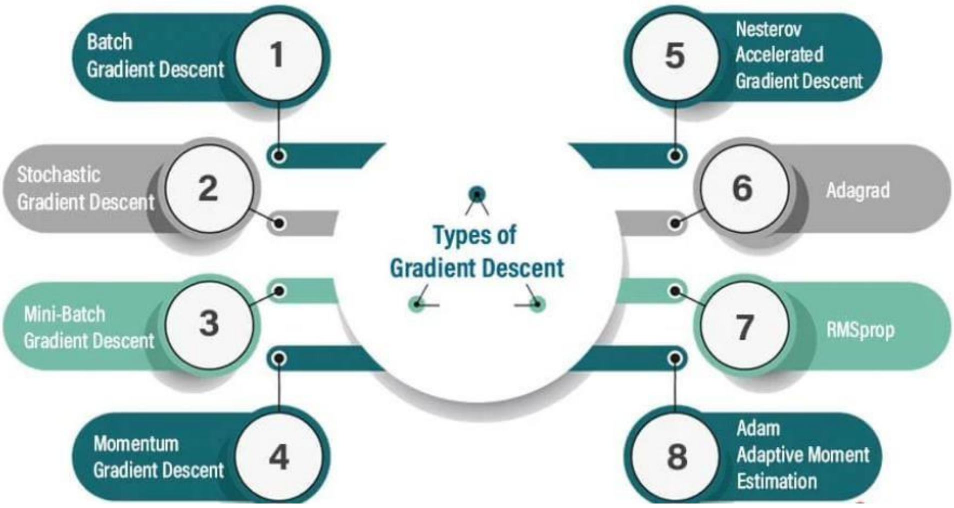


Figure III. 11: Types of Projected Gradient Ascent

- **Batch gradient descent :**

Batch gradient descent sums the error for each point in a training set, updating the model only after all training examples have been evaluated. This process referred to as a training epoch.

While this batching provides computation efficiency, it can still have a long processing time for large training datasets as it still needs to store all of the data into memory. Batch gradient descent also usually produces a stable error gradient and convergence, but sometimes that convergence point isn't the most ideal, finding the local minimum versus the global one.

- **Stochastic gradient descent :**

Stochastic gradient descent (SGD) runs a training epoch for each example within the dataset and it updates each training example's parameters one at a time. Since you only need to hold one training example, they are easier to store in memory. While these frequent updates can offer more detail and speed, it can result in losses in computational efficiency when compared to batch gradient descent. Its frequent updates can result in noisy gradients, but this can also be helpful in escaping the local minimum and finding the global one.

- **Mini-batch gradient descent :**

Mini-batch gradient descent combines concepts from both batch gradient descent and stochastic gradient descent. It splits the training dataset into small batch sizes and performs updates on each of those batches. This approach strikes a balance between the computational efficiency of batch gradient descent and the speed of stochastic gradient descent.

- **Momentum Gradient Descent:**

Momentum GD adds a momentum term to the update rule, which helps to smooth out oscillations in the gradient descent path and accelerate convergence. The momentum term accumulates the gradient updates over time and helps the algorithm to move faster in the correct direction and avoid local optima.

- **Nesterov Accelerated Gradient Descent (NAG):**

Nesterov's Accelerated Gradient Descent is a modification of momentum Gradient Descent that accounts for the next step's momentum and improves convergence speed. NAG computes the gradient of the cost function after adding the momentum term to the current position, which provides a better estimate of the gradient and improves convergence speed.

- **Adagrad:**

Adagrad adapts the learning rate for each model parameter based on the history of gradients for that parameter. This approach helps to converge quickly on sparse features, making it useful for natural language processing and computer vision applications.

- **RMSprop:**

RMSprop (Root Mean Square Propagation) is a modification of Adagrad that normalizes the historical gradient sum. It helps to adapt more robustly to the changing gradient surface and prevents the learning rate from becoming too small.

- **Adam (Adaptive Moment Estimation):**

Adam combines the concepts of momentum and adaptive learning rates. It uses the first and second moments of the gradient to calculate the adaptive learning rate and momentum term. Adam is a popular optimization algorithm for training deep neural networks due to its effectiveness in converging quickly to the minimum.

In general, selecting the appropriate Gradient Descent algorithm depends on the specific problem, dataset size, and model complexity. It is often beneficial to experiment with multiple algorithms and tune their hyperparameters to optimize convergence speed and accuracy [42]

III.3.4.3. Advantages of Projected Gradient ascent

Projected Gradient ascent (PGA) offers several advantages in the realm of adversarial attacks:

Robust Adversarial Examples: PGA is known for generating adversarial examples that are robust across various models, making it a potent tool for evaluating and enhancing model robustness.

Transferability: Adversarial examples crafted using PGA on one model often transfer well to other models, demonstrating its effectiveness in generating universal perturbations.

Stability: PGA attacks are less sensitive to the choice of hyperparameters, providing a stable and reliable method for crafting adversarial examples [43]

III.3.4.4. Disadvantages of Projected Gradient Ascent (PGA)

While PGD is a powerful technique, it comes with its own set of challenges and limitations: [42]

Increased Computational Cost: PGA attacks involve multiple iterations of gradient ascent, leading to increased computational cost compared to single-step methods.

Limited Understanding of Robustness: Despite its success, PGA does not necessarily provide a complete understanding of a model's robustness, as it might not cover all possible types of adversarial attacks.

Hyperparameter Sensitivity: Although less sensitive than some other methods, PGD's performance can still be influenced by the choice of hyperparameters, requiring careful tuning [43]

III.4. Conclution

Based on the discussions presented above, the conclusions, has been developed to enhance the wireless communications domain by offering a precise and adaptable platform various machine learning algorithms to estimate THz channel, including neural networks (NN), logistic regression (LR), and projected gradient ascent (PGA), Frank-Wolfe. We chose these specific algorithms as they proved good performance for similar problems

CHAPTER IV: *SIMULATIONS AND RESULTS*

IV.1. Introduction:

Simulation plays a vital role in understanding, analyzing, and optimizing complex systems across various scientific and engineering disciplines. As the development of sixth-generation (6G) wireless networks and terahertz (THz) communication systems accelerates, accurate channel modeling and estimation have become critical challenges in the field.

In this context, this study focuses on the deterministic propagation of signals and the physical characteristics inherent to THz frequency bands, which are essential for enabling high-capacity and reliable wireless communication. The aim is to provide a deeper understanding of channel behavior in THz environments through precise and efficient simulation models [28]

In this section we will use python as a main program for simulation. We use it to create a simple DNN model to use it as a tool for channel estimation

IV.2. Python

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

Often, programmers fall in love with Python because of the increased productivity it provides. Since there is no compilation step, the edit-test-debug cycle is incredibly fast. Debugging Python programs is easy: a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it raises an exception. When the program doesn't catch the exception, the interpreter prints a stack trace. A source level debugger allows inspection of local and global variables, evaluation of arbitrary expressions, setting breakpoints, stepping through the code a line at a time, and so on. The debugger is written in Python itself, testifying to Python's introspective power. On the other hand, often the quickest way to debug a program is to add a few print statements to the source: the fast edit-test-debug cycle makes this simple approach very effective.

IV.3. System Model:

We create a model channel that pass by the following main steps:

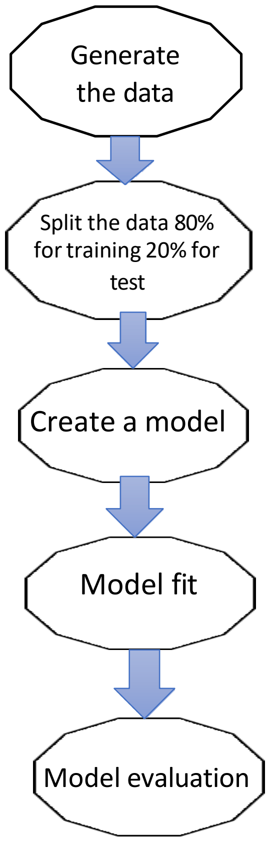


Figure IV. 1: Steps of our Model Diagram.

IV.4. Simulation

in the context of terahertz (THz) communication channels. This setup typically leverages Multiple Input Multiple Output (MIMO) techniques to enhance signal quality and data rates of the environment subject to it with creation of Thz wireless channel

Where K is the Rician K-factor the channel matrix H follows:

$$H = \sqrt{PL}.(K/K+1).H_{LOS} + 1/K+1 \; .H_{NLOS} \tag{IV.1}$$

The aim of simulation is estimate a THz wireless channel based on ML technique. In the first we need to generate a realistic Terahertz (THz) wireless channel matrix for a 2×2 Multiple-Input Multiple-Output (MIMO) system operating at 300 GHz in a short-range indoor environment. The channel matrix must be stored in a 3D array of dimensions (200,000, 2, 2), where each of the 200,000 instances represents a different channel realization (e.g., time samples, spatial positions, or independent channel realizations). Each 2×2 matrix represents the complex channel gains between 2 transmit and 2 receive antennas.

Parameter	Values
Frequency	300.0 GHz
Wavelength	1.00 mm
Antenna spacing	0.50 mm
Rician K-factor	15 dB
Distance range	1-5 m
Data Set length	2 .000.000
Mean channel gain	-15.13 dB
Std channel gain	-15.14 dB
Mean condition number	4.66
Spatial correlation (TX)	0.001
Spatial correlation (RX)	0.001
Temperature	20°C = 293.15 K
Humidity	0.45 = 45%
Pressure	101325 Pa
Environment	Los
Data type	complex128
Memory usage	122.1 MB

Table IV.1: Parameters of Thz channel Model.

We used a model Deep Neural Networks (DNN) functional with different values of noise to evaluate our model, in parallel to the change

With 5 hidden layer:

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	(None, 4)	0	-
input_2 (InputLayer)	(None, 4)	0	-
dense (Dense)	(None, 64)	320	input_1[0][0]
dense_1 (Dense)	(None, 64)	320	input_2[0][0]
concatenate (Concatenate)	(None, 128)	0	dense[0][0], dense_1[0][0]
dense_2 (Dense)	(None, 128)	16,512	concatenate[0][0]
dense_3 (Dense)	(None, 128)	16,512	dense_2[0][0]
dense_4 (Dense)	(None, 128)	16,512	dense_3[0][0]
dense_5 (Dense)	(None, 128)	16,512	dense_4[0][0]
dense_6 (Dense)	(None, 128)	16,512	dense_5[0][0]
dense_7 (Dense)	(None, 4)	516	dense_6[0][0]

Total params: 83,716 (327.02 KB)
Trainable params: 83,716 (327.02 KB)
Non-trainable params: 0 (0.00 B)

Figure IV. 2: DNN Model With 5 Hidden Layers.

IV.4.1. THz Channel Modeling Approach

The THz channel model incorporates three main propagation effects:

- **Path Loss Model**: Free-space path loss combined with atmospheric absorption
  - Free-space path loss: $L_{f_s} = (4\pi df/c)^2$

    - Molecular absorption: Exponential decay based on ITU-R P.676 recommendations

- **Molecular Absorption:** Dominated by water vapor at 300 GHz
    - Absorption coefficient calculated using Van Vleck-Weisskopf line shape
    - Simplified model: $\gamma \approx 0.1 dB/m$ at 300 GHz for standard conditions
- **Small-Scale Fading:** Combination of LOS and NLOS components
    - Rician fading for dominant LOS path
    - Rayleigh fading for scattered components

Spatial correlation based on antenna spacing

IV.4.2. Explanation of Model Compilation

This method configures the model for training with consideration parameters must be considered:

- **SGD (Stochastic Gradient Descent):** The optimizer used in this model, which is a classical method for optimizing neural networks. It updates the model's weights during training to minimize the loss function**.**
- **The learning rate is set to 0.001:** which controls the size of the steps taken during optimization. A smaller learning rate results in slower but more precise learning.
- **The momentum parameter:** is set to 0.9, which helps accelerate the learning process in the correct direction and reduces oscillations by accumulating the effects of previous updates.
- **MeanSquaredError:** is the loss function, which measures how close the model's predictions are to the actual values by computing the average of the squared differences. This type of loss is commonly used in regression problems where the goal is to predict continuous numerical values such as signal strength, speed, or price.
- **Accuracy:** is the metric used to evaluate the model's performance is accuracy, which calculates the proportion of correct predictions. However, it is important to note that accuracy is typically suited for classification tasks.

For SNR=30db:

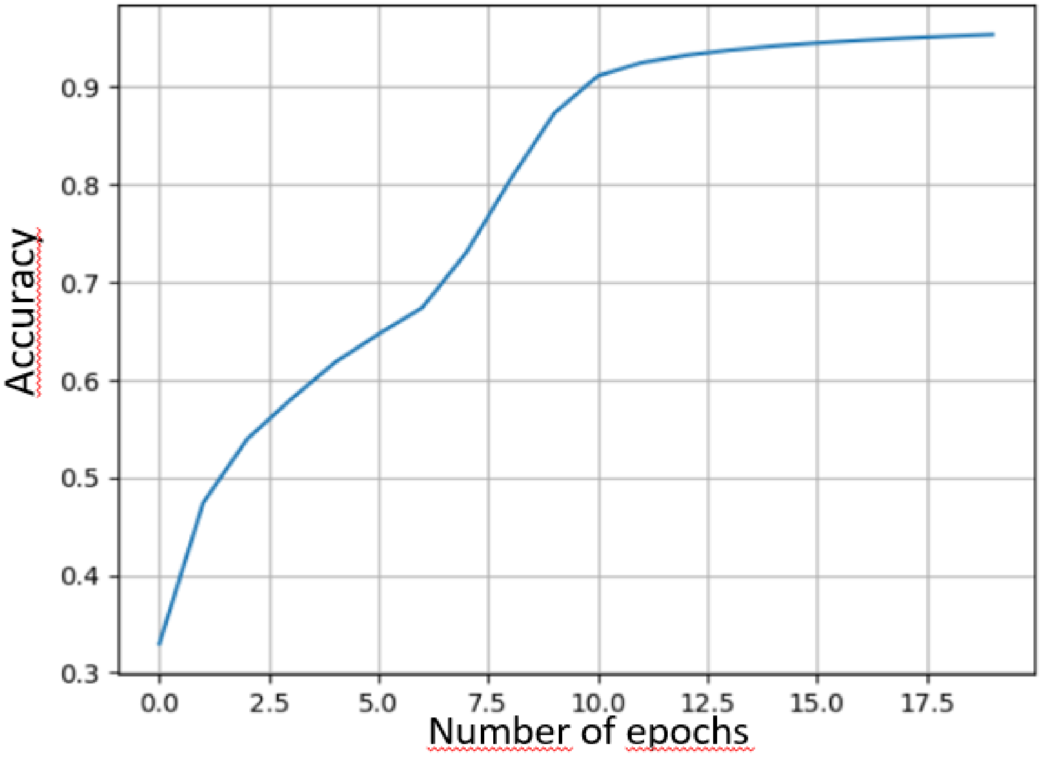


Figure IV. 3: Development of the Accuracy During Training.

The graph shows in the Figure IV. 3 the model's training accuracy increases significantly during the initial epochs, starting at around 33% and reaching approximately 95%, indicating that the model is learning well from the training data. After the tenth epoch, the improvement begins to slow down, and the performance reaches a state of stability. Although the result indicates successful training

With epochs 20 :

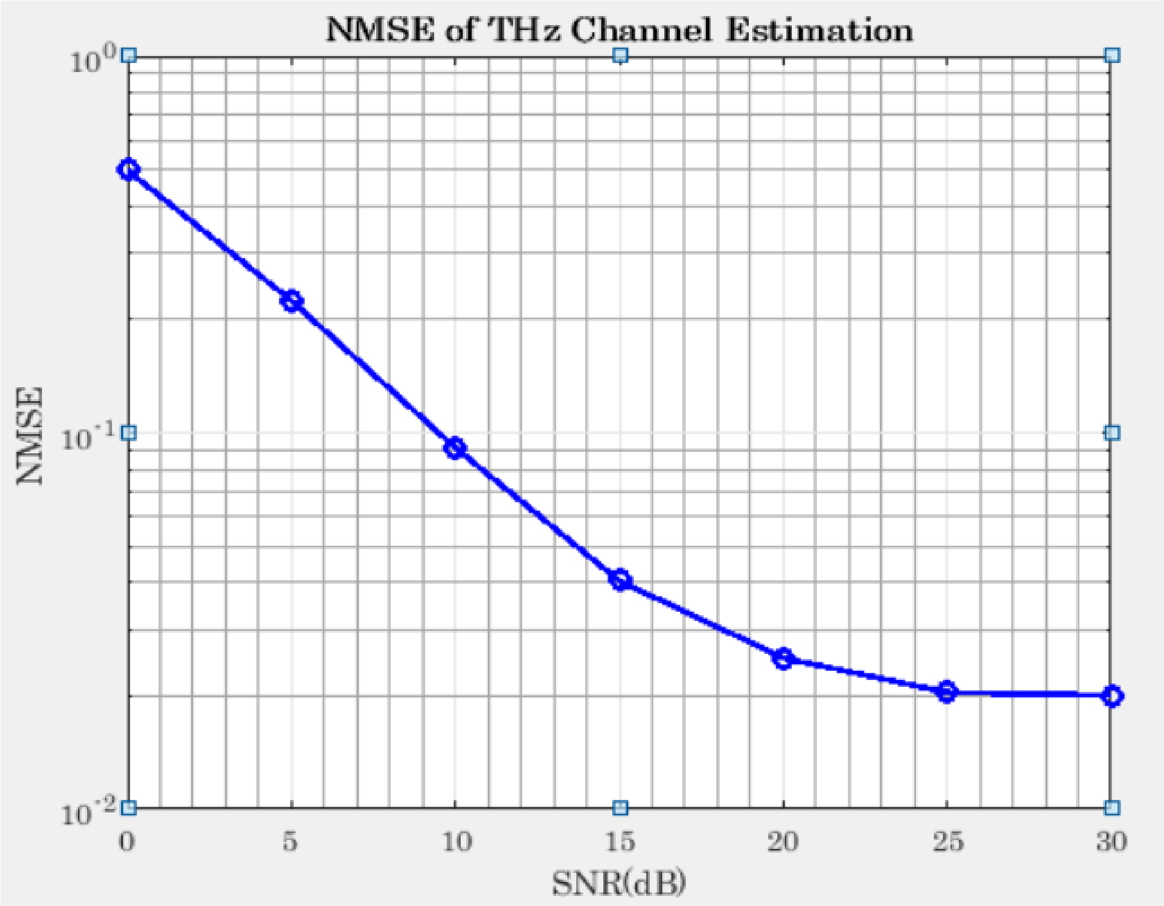


Figure IV. 4: NMSE and SNR for wireless THz channel estimation

The graph shows the relationship between the Normalized Mean Squared Error (NMSE) and the Signal-to-Noise Ratio (SNR) for THz channel estimation, where the horizontal axis represents the SNR in decibels and the vertical axis shows the NMSE on a logarithmic scale. It is observed that the NMSE decreases significantly as the SNR increases from 0 to around 20 dB, indicating a noticeable improvement in estimation accuracy as noise impact is reduced. However, after 20 dB, this improvement slows down and reaches a saturation point around an NMSE value of 10^{-2}, suggesting that the limiting factors are no longer just noise but could include algorithmic constraints, modeling limitations, or numerical precision issues. This indicates that the method used for channel estimation is effective at medium to high SNR values.

With SNR 30db:

With epochs 30:

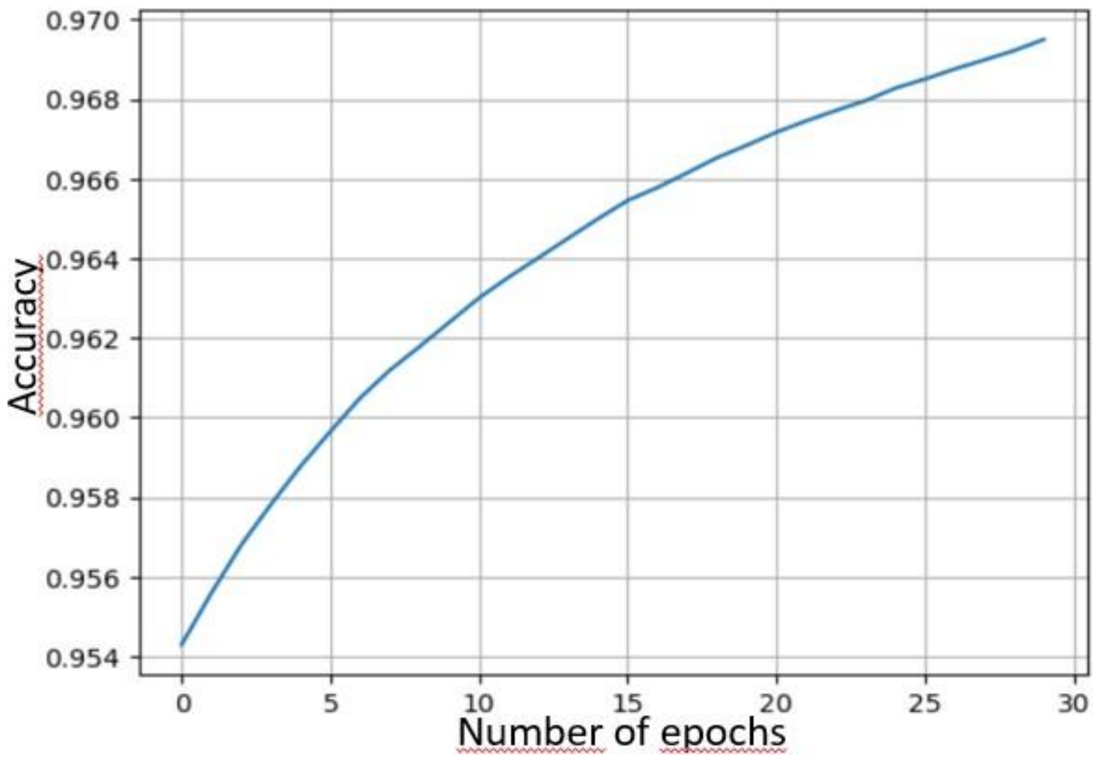


Figure IV. 5: Development of The Accuracy During Training.

This graph shows the change in accuracy with the number of epochs during model training. As shown, the accuracy gradually improves from around 0.954 to approximately 0.970 as the number of epochs increases from 0 to 30, From the figure IV.5, we can easily notice that the performance of the channel estimation based on DNN is related to the number of epochs, i.e. when the number of epochs is more than 10 epochs.

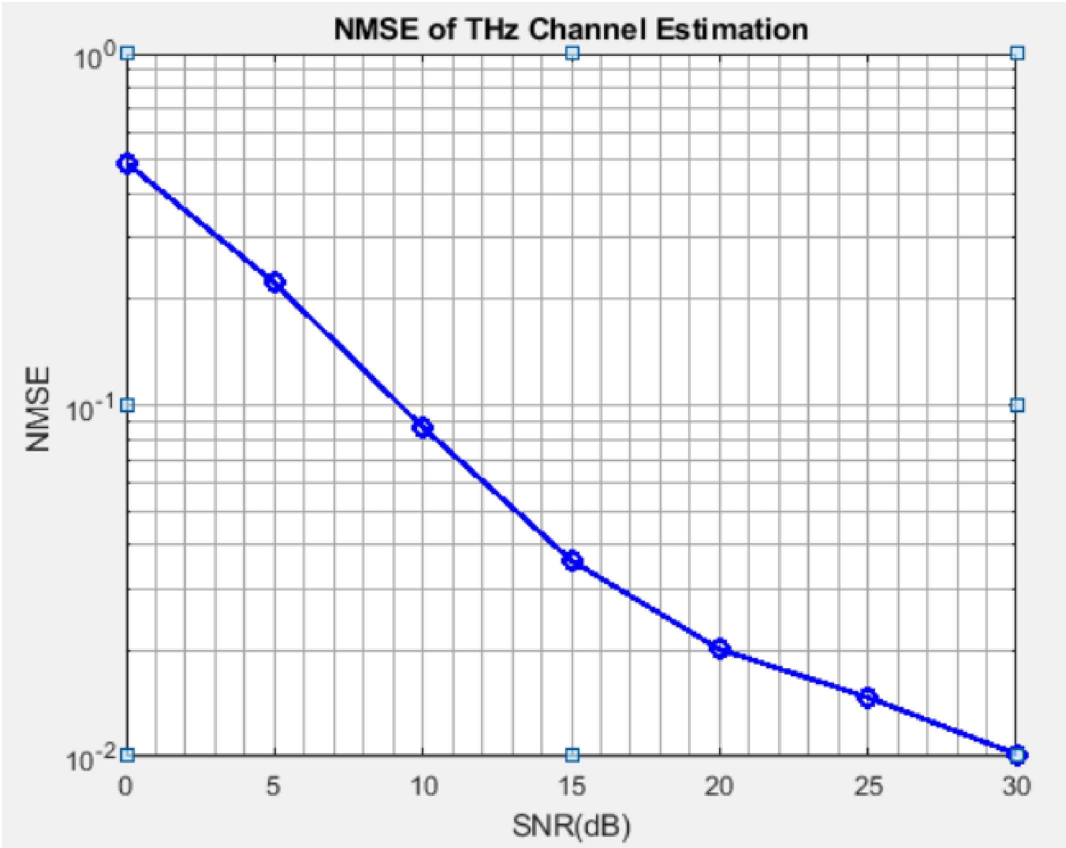


Figure IV. 6: NMSE and SNR for wireless THz channel estimation

The plot clearly shows that the THz channel estimation model heavily depends on the signal quality (SNR), where an increase in SNR leads to reduced error and improved estimation accuracy. This behavior is expected and desirable in estimation models for communication systems, reflecting the efficiency of the employed deep neural network.

IV.5 CONCLUSION

In this study, we presented an estimation of wireless THz channel response to overcome the technical challenges and enhance the performance of communication systems, components using a dual-antenna system by creating a model Thz according to the appropriate environment. As newcomers to machine learning and deep learning, we used a Deep Neural Network (DNN) model for estimating a 2×2 MIMO flat-fading channel. From the resulting graphs, it can be observed that the model's performance begins to degrade as the Signal-to-Noise Ratio (SNR) decreases. Furthermore, the accuracy of the channel estimation is closely linked to the number of training epochs, with noticeable improvements occurring when the number of epochs exceeds 10. This highlights the importance of sufficient training time to achieve reliable estimation results.

GENERAL CONCLUSION

Terahertz (THz) channel models serve as a fundamental pillar in driving the advancement of wireless communication networks, particularly by offering deep insights into the unique physical characteristics of signal propagation at high frequencies in sixth-generation (6G) systems. These models take into account critical factors such as severe power loss and atmospheric absorption, enabling accurate simulations and the optimization of THz wireless links. As research and technologies for 6G continue to evolve, these models will become indispensable tools for ensuring ultra-fast, reliable connectivity and enabling innovative applications across various domains from terabit-per-second data transmission to advanced sensing systems.

To overcome the technical challenges and enhance the performance of communication systems, a set of machine learning algorithms has been proposed, specifically designed to handle the complex nature of THz channels. These include Neural Networks (NN), Logistic Regression (LR), Projected Gradient Ascent (PGA), and Frank-Wolfe techniques. The selection of these algorithms is based on their proven efficiency in similar channel estimation tasks. They are expected to play a significant role in developing intelligent and highly accurate solutions for THz channel estimation, making them well-suited to meet the advanced technical requirements of 6G systems.

However, Deep Learning has recently gained prominence as a result of its superior accuracy when trained with large amounts of data, In this work, we presented the THz channel response, focusing on the components, using a simple Deep Neural Network (DNN) model developed for estimating a 2×2 MIMO flat-fading channel. It is important to note that the model's performance is closely related to both the accuracy of the channel estimation and the signal-to-noise ratio (SNR). As the number of training epochs increases, the accuracy of the model improves accordingly, highlighting the importance of sufficient training to achieve reliable channel estimation.

Because of the short time that we had and the lack of a powerful calculator machine, we suggest as a future work to make comparison between PGA ,LR ,and NN THz channel estimation algorithms

BIBLIOGRAPHICAL REFERENCES

[1] : Inomata Minoru, ; Wataru Yamada; Nobuaki Kuno ,« *Terahertz Propagation Characteristics for 6G Mobile Communication Systems »*. 2021.

[2] : Dua, T. R. « *Overview of 6G Network & Technology »*. Digital Infrastructure Providers Association (DIPA), 7 Bhai Vir Singh Marg, Gole Market, New Delhi 110001, India.

[3] : Saad, Walid, Mehdi Bennis, and Mingzhe Che. « *A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems »*.

[4] : **"6G."** *TechTarget SearchNetworking*, https://www.techtarget.com/searchnetworking/definition/6G.

[5] : Mohsan, Syed Agha Hassnain, and Yanlong Li. « *A Contemporary Survey on 6G Wireless Networks: Potentials, Recent Advances, Technical Challenges, and Future Trends"*.

[6] : Xu, Guangbin. *"Research on 6G Mobile Communication System." Journal of Physics: Conference Series*, vol. 1693, 2020, article 012101, IOP Publishing.

[7] : Jazea, Najim Abdallah, and Abdullsalam M. Saeed. *"Characteristics Analysis of 6G Wireless Networks: Review, Vision, Challenges." American Scientific Research Journal for Engineering, Technology, and Sciences*, vol. 87, no. 1, 2022, pp. 218–229.

[8] : Chen, Z., et al. *"A Survey on Terahertz Communications." China Communications*, vol. 16, no. 2, Feb. 2019, pp. 1–35.

[9] : He, H., et al. *"Model-Driven Deep Learning for Physical Layer Communications." IEEE Wireless Communications*, vol. 26, no. 5, 2019, pp. 77–83.

[10] : Letaief, K. B., et al. *"The Roadmap to 6G: AI-Empowered Wireless Networks." IEEE Communications Magazine*, vol. 57, no. 8, 2019, pp. 84–90.

[11] : "The IBM Watson® System That Won the Jeopardy! Challenge in 2011**."** *IBM Think*, IBM, 2011 , https://www.ibm.com/think/topics/machine-learning

[12] : Gupta, Mayank. *"Designing a Learning System in Machine Learning." Applied AI Course*, 30 Dec. 2024,

https://www.appliedaicourse.com/blog/designing-a-learning-system-in-machine-learning/.

[13] : Khan, M. Qurratulain, Abdo Gaber, Philipp Schulz, and Gerhard Fettweis. *"Machine Learning for mmWave and THz Beam Management: A Survey and Open Challenges." [*Member, IEEE*]*, 2023.

[14] : Yuan, Yuan, et al. "A 3D Geometry-Based THz Channel Model for 6G Ultra Massive MIMO Systems." *IEEE Transactions on Vehicular Technology*, vol. 71, no. 3, Mar. 2022, pp. 2251-2266. IEEE, doi:10.1109/TVT.2022.3147796.

[15] : S. Liu, X. Yu, R. Guo, Y. Tang, and Z. Zhao, "THz Channel Modeling: Consolidating the Road to THz Communications," Zhejiang Laboratory, Hangzhou, China.

[16] : Han, C., and Y. Chen. "Propagation Modeling for Wireless Communications in the Terahertz Band." *IEEE Communications Magazine*, vol. 56, no. 6, 2018, pp. 96–101

[17] : Elayan, Hadeel, et al. "Terahertz Band: The Last Piece of RF Spectrum Puzzle for Communication Systems." *IEEE Open Journal of the Communications Society*, vol. 1, 2019, pp. 1–20, doi:10.1109/OJCOMS.2019.2953633

[18] : Lin, Cen, and Geoffrey Ye Li. "Indoor Terahertz Communications: How Many Antenna Arrays Are Needed?" *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, June 2015, pp. 3097–3107

[19] : Başar, Ertuğrul, et al. "Indoor and Outdoor Physical Channel Modeling and Efficient Positioning for Reconfigurable Intelligent Surfaces in mmWave Bands." *IEEE Transactions on Communications*, vol. 69, no. 12, Dec. 2021, pp. 8600–8611

[20] : shii, S., S. Sayama, and K. Mizutani. "Rain Attenuation at Terahertz." *Wireless Engineering and Technology*, vol. 1, no. 2, 2010, pp. 92–95, doi:10.4236/wet.2010.12014.

[21] : Robins, Adam. "Stochastic vs. Deterministic Models: Understand the Pros and Cons." *EV Blog*, 23 June 2020, blog.ev.uk/stochastic-vs-deterministic-models-understand-the-pros-and-cons.

[22] : Kenton, Will. "Stochastic Modeling: Definition, Advantage, and Who Uses It." *Investopedia*, updated 15 Aug. 2024,

[23] : Boulogeorgos, Alexandros-Apostolos A., Evangelos N. Papasotiriou, and Angeliki Alexiou. "Analytical Performance Assessment of THz Wireless Systems." *IEEE Access*, vol. 7, 2019, pp. 11436–11453, doi:10.1109/ACCESS.2019.2892198

[24] : Han, Chong, et al. "Terahertz Wireless Channels: A Holistic Survey on Measurement, Modeling, and Analysis." *IEEE Communications Surveys & Tutorials*, vol. 24, no. 3, 2022, pp. 1670–1707, doi:10.1109/COMST.2022.3182539.

[25] : Yi, Haofan, et al. "Characterization for the Vehicle-to-Infrastructure Channel in Urban and Highway Scenarios at the Terahertz Band." *IEEE Access*, vol. 7, 2019, pp. 166984–166996, doi:10.1109/ACCESS.2019.2953890.

[26] : "Path Loss Showcase – INET Framework." *INET Framework Documentation*, https://inet.omnetpp.org/docs/showcases/wireless/pathloss/doc/. Accessed 13 Mar. 2025.

[27] : Barsocchi, Paolo. *Channel Models for Terrestrial Wireless Communications: A Survey. National Research Council – ISTI Institute, 2009. Web. https://iris.cnr.it/handle/20.500.14243/148770*

[28] : Bensalem, Mounir, and Admela Jukan. Benchmarking Machine Learning Techniques for THz Channel Estimation Problems. Proceedings of the 45th International Convention on Information, Communication and Electronic Technology (MIPRO), 2022, pp. 459–464. IEEE, https://doi.org/10.23919/MIPRO55190.2022.9803610.

[29] : Govil, Rakshit. "Different Types of Channel Estimation Techniques Used in MIMO-OFDM for Effective communication Systems." International Journal of Engineering Research and Technology, vol. 7, no. 7, 2018, pp. 1–6.

[30] : GeeksforGeeks. "What is a Neural Network?" GeeksforGeeks, 3 Apr. 2025, www.geeksforgeeks.org/neural-networks-a-beginners-guide/.

[31] : Chen,Y.& Han,C.(2020,December).Deep CNN-based spherical -wave channel estimation for terahertz ultra-massive MIMO systems. In 2020 IEEE Global communication conference(GLOBECOM)(PP.1-6). IEEE.

https://doi.org/10.1109/GLOBECOM42002.2020.9322411

[32] : Lee, Y., & Shin, S. (2024). *Deep Learning-Based Music Emotion Recognition:A Review*. Information, *15*(9), 517.

https://doi.org/10.3390/info15090517

[33] : GeeksforGeeks. "Understanding Logistic Regression." GeeksforGeeks, 7 Feb. 2025, www.geeksforgeeks.org/understanding-logistic-regression/.

www.stat.cmu.edu/~ryantibs/convexopt/lectures/frank-wolfe.pdf

[34] : Grammarly. (n.d.). *What is logistic regression in machine learning?* Grammarly. Retrieved May 4, 2025, from

https://www.grammarly.com/blog/ai/what-is-logistic-regression/

[35] : Wang, Haoyue, et al. "Frank-Wolfe Methods with an Unbounded Feasible Region and Applications to Structured Learning." arXiv preprint, arXiv:2012.15361, 30 Dec. 2020, https://arxiv.org/abs/2012.15361.

[36] : Tibshirani, Ryan. "Frank-Wolfe Method." Convex Optimization 10-725, Carnegie Mellon University, Fall 2019,

[37] : Myers, N. J., Tran, K. N., & Heath, R. W., Jr. (2020). Low-rank mmWave MIMO channel estimation in one-bit receivers. In *Proceedings of the 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5005–5009). IEEE.

https://doi.org/10.1109/ICASSP40776.2020.9054266

[38] : Nguyen, Linh V. "Frank-Wolfe Algorithm." 6.883 Learning with Combinatorial Structure, Massachusetts Institute of Technology, 2015, people.csail.mit.edu/linhvn/6.883/lectures/lecture14.pdf.

[39] : Han, Byeolyi, et al. "Bootcamp Summer 2023 Week 1 – Projected & Mirror Gradient Descent." CORE Robotics Lab, 30 July 2023, core-robotics.gatech.edu/2023/07/30/bootcamp-summer-2023-week-1-mirror-descent/.

[40] : " IBM Think, 2025, www.ibm.com/think/topics/gradient-descent ‘’.

[41] : Schmidt, Mark. "Numerical Optimization for Machine Learning: Projected-Gradient, Projected-Newton, and Frank-Wolfe." University of British Columbia, Summer 2022–Summer 2023, www.cs.ubc.ca/~schmidtm/Courses/5XX-S22/.

[42] : Pedamkar, Priya. "Gradient Descent in Machine Learning: Optimized Algorithm." *EDUCBA*, 11 May 2023, www.educba.com/gradient-descent-in-machine-learning/.

[43] : Zachariah, Arun George. "Unveiling the Power of Projected Gradient Descent in Adversarial Attacks." *Medium*, 26 Dec. 2023, https://medium.com/@zachariaharungeorge/unveiling-the-power-of-projected-gradient-descent-in-adversarial-attacks-2f92509dde3c.