

國立臺灣大學電機資訊學院資訊網路與多媒體研究所

博士論文

Graduate Institute of Networking and Multimedia

College of Electrical Engineering and Computer Science

National Taiwan University

Doctoral Dissertation

智慧反射表面輔助毫米波通信：基於加性注意力輔助
對抗自編碼器的波束成型設計

Intelligent Reflecting Surface-Assisted Millimeter Wave
Communications: Additive Attention-aided Adversarial
Autoencoder-based Beamforming Design

陳弘運

Hong-Yun Chen

指導教授：周承復 博士

Advisor: Cheng-Fu Chou, Ph.D.

中華民國 112 年 11 月

November 2023



國立臺灣大學博士學位論文
口試委員會審定書

智慧反射表面輔助毫米波通信：基於加性注意力輔助對抗自編碼器的波束成型設計

Intelligent Reflecting Surface-Assisted Millimeter Wave
Communications: Additive Attention-aided Adversarial
Autoencoder-based Beamforming Design

本論文係陳弘運君（學號 D05944018）在國立臺灣大學資訊網路與多媒體研究所完成之博士學位論文，於民國一十二年十一月一日承下列考試委員審查通過及口試及格，特此證明

口試委員：

周承復

(指導教授)

江英超

蔡子偉

吳曉文

呂政修

黃志偉

廖婉君

曾士青

黃瑞光

鄭卜壬

所長：





Acknowledgements

我要感謝周承復教授，感謝他的親切關懷、悉心指導和淵博學識。從他深入淺出、生動活潑的講解中，我對無線通訊有了更深入的瞭解。感謝他在繁忙的教學研究工作之餘，他仍對我的論文指導和閱讀，感謝他不斷的鼓勵和打氣。還有感謝口試委員廖婉君教授，呂政修教授，吳曉光教授，黃志煒教授，蔡子傑教授，張英超教授，鄭瑞光教授，葉士青教授的悉心指導與建議。最後，我要感謝我的爸媽的養育之恩，感謝我的爸媽對我沒有止境的關心和支持，我內心有著最真誠的感謝。





摘要

6G 技術在速度、延遲和容量方面都超過了 5G，並引入了至關重要的智慧反射面 (IRS)。在微控制器的管理下，這種高性價比的無源元件陣列可通過精確操縱傳入的無線電波來優化無線通訊，從而提高網路覆蓋、容量和能效。然而，現實世界中 IRS 的混合波束成形面臨著雜訊和干擾的困難。為了處理這個困難，我們提出了 AAE-AATT-波束成形 (Adversarial AutoEncoder with Additive ATTention Beamforming)。加性注意力是一種強大的機制，用於在時域中對輸入序列中元素全域依賴性。AAE 學習到潛在空間在捕捉頻域和空間域通道內波束成形的基本特徵扮演重要角色。在類比預編碼模組中，自動編碼器被用來優化潛空間，並忠實地重建潛空間以匹配原始輸入通道信號資料，從而顯著提高捕捉全域特徵的準確性。數位預編碼模組利用具有平移不變性等特徵的 2D-CNN 捕獲通道預編碼的基本頻率和空間特徵，同時最大限度地減少干擾。在類比波束成形模組中，使用了門控遞迴單元 (GRU)，其重定和更新門控制單元內的資訊流。這提高了特徵捕捉的準確性。數位波束成形模組採用 1D-CNN 技術，善於捕捉連續資料中的局部模式和依賴關係，因此適用於時間序列分析等任務。該模組能有效捕捉關鍵通道波束成形特徵，同時減少干擾。實驗數值顯示和之前的研究對照，MSE、可實現速率、泛化性和強健性都有大幅提高。

關鍵字：第六代 (6G)、加性注意力、對抗自動編碼器、波束成形、毫米波、智

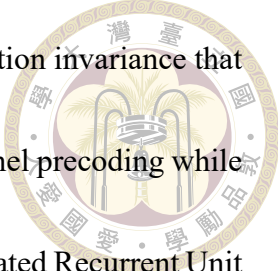
慧反射面 (IRS)。





Abstract

The 6G technology, surpassing 5G in speed, latency, and capacity, introduces the crucial intelligent reflecting surface (IRS). Managed by a microcontroller, this cost-effective array of passive elements optimizes wireless communication by precisely manipulating incoming radio waves, enhancing network coverage, capacity, and energy efficiency. However, real-world hybrid beamforming in the IRS faces challenges from noise and interference. To tackle this issue, we present AAE-AATT-Beamforming (Adversarial AutoEncoder with Additive ATTention Beamforming). Additive attention is a powerful mechanism for modeling global relationships among elements within an input sequence in the time domain. AAE learned latent space plays a pivotal role in capturing essential features for intra-channel beamforming in both frequency and spatial domains in frequency and spatial domains. In the analog precoding module, an autoencoder is utilized to optimize the latent space and faithfully reconstruct it to match the original input channel signal data that significantly enhancing the accuracy of capturing global domain features. The



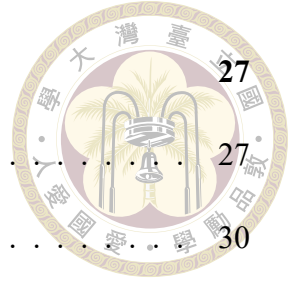
digital precoding module utilizes a 2D-CNN with features like translation invariance that enabling it to capture essential frequency and spatial features for channel precoding while minimizing interference. Within the analog beamforming module, a Gated Recurrent Unit (GRU) is used, featuring reset and update gates that control information flow within the cell. This enhances feature capture accuracy. The digital beamforming module employs a 1D-CNN, adept at capturing local patterns and dependencies in sequential data, making it suitable for tasks like time series analysis. This module effectively captures key channel beamforming features while reducing interference. Numerical results demonstrate substantial improvements in MSE, achievable rate, generalizability, and robustness compared to prior research.

Keywords: Sixth-generation (6G), additive attention, adversarial autoencoder, beamforming, millimeter-wave, Intelligent reflecting surface (IRS).



Contents

	Page
Verification Letter from the Oral Examination Committee	i
Acknowledgements	iii
摘要	v
Abstract	vii
Contents	ix
List of Figures	xi
List of Tables	xiii
Chapter 1 Introduction	1
1.1 Introduction	1
1.2 Related Work	2
1.3 Contribution	5
Chapter 2 Preliminary Study	11
2.1 6G	11
2.2 IRS	13
2.3 Channel estimation	18
2.4 Precoding	21
2.5 Beamforming	22



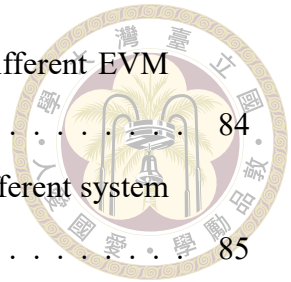
Chapter 3	System Model	27
3.1	Channel Model	27
3.2	Signal Model	30
3.3	LOS and NLOS	35
Chapter 4	METHOD	39
4.1	Additive Attention Mechanism	42
4.2	Adversarial Autoencoder	49
4.3	Analog precoding	56
4.4	Digital precoding	58
4.5	Analog beamforming	59
4.6	Digital beamforming	65
4.7	Computational Complexity Analysis	67
Chapter 5	Results	69
5.1	The Investigation of System Parameters	71
5.2	The Investigation of Loss Functions	75
5.3	The Investigation of Spectral Analysis	77
5.4	The Investigation of Interference Analysis	78
5.5	The Investigation of Generalization	79
5.6	The Investigation of Robustness	82
5.7	The Investigation of System Architecture Ablation	84
Chapter 6	Conclusions	89
6.1	Conclusions	89
References		91



List of Figures

3.1	Scenario diagram of IRS-assisted MIMO OFDM system.	31
3.2	Schematic diagram of the IRS-assisted beamforming design architecture with mmwave MIMO OFDM system.	32
3.3	Illustration of the structure of the proposed additive attention-aided adversarial autoencoder beamforming design.	35
4.1	Scenario of the proposed beamforming design.	41
4.2	The architecture of the proposed adversarial autoencoder design.	49
4.3	The architecture of GRU and LSTM.	59
5.1	Comparison of MSE versus SNR between different methods without beamforming.	71
5.2	Comparison of MSE versus SNR between different methods with beamforming.	72
5.3	Comparison of achievable rate versus SNR between different methods without beamforming.	72
5.4	Comparison of achievable rate versus SNR between different methods with beamforming.	75
5.5	Comparison of loss functions versus SNR between different methods. . .	76
5.6	Comparison of attention map between different methods.	76
5.7	Comparison of subcarrier channel prediction for different methods.	76
5.8	Comparison of SNR versus SINR between different methods.	80
5.9	The generalization evaluation of different maximum Doppler frequencies between different methods.	82

5.10 Comparison of MSE performance against the impact of different EVM between different methods.	84
5.11 Comparison of MSE performance against the impact of different system architecture ablation.	85
5.12 Comparison of achievable rate performance against the impact of different system architecture ablation.	88





List of Tables

1.1	NOTATIONS AND THEIR DEFINITIONS	9
4.1	COMPUTATIONAL COMPLEXITY OF THE DIFFERENT RESEARCH METHODS	55
5.1	SIMULATION PARAMETERS	71





Chapter 1 Introduction

1.1 Introduction

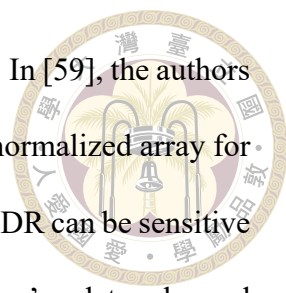
The demand for mobile data rates has surged, driven by advancements in wireless communication technologies in recent years. Looking forward to the sixth generation (6G) cellular system, there's a strong expectation of even greater data capacity[2, 3, 10, 12, 58]. Meeting the immense data flow and connectivity needs of 6G necessitates solutions like increased transmission bandwidth and enhanced spectral efficiency. However, the proliferation of RF chains in 6G wireless communication systems not only escalates RF circuit costs but also consumes additional energy.

In response to these challenges and the quest to boost spectrum and energy efficiency while augmenting data rates, researchers are investigating the potential of Intelligent Reflecting Surfaces (IRS). These surfaces consist of two-dimensional metasurface arrays, featuring multiple passive and reconfigurable elements. What sets these elements apart is their capacity to dynamically redirect incident signals without any additional energy consumption. By deploying IRS, we can significantly enhance spectrum efficiency in a cost-effective manner, reducing power consumption and network complexity by capitalizing on the passive nature of these elements. IRS opens up opportunities to boost wireless communication system performance while mitigating associated drawbacks[11, 21, 39].

1.2 Related Work

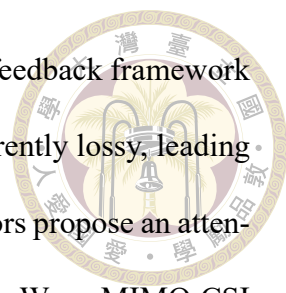


Hybrid beamforming in wireless communication systems have recently been extensively explored using both traditional methods and deep neural network (DNN) approaches. In traditional methods, the following techniques were commonly employed. In [44], the proposed approach incorporates the application of the successive convex approximation (SCA) technique, a prominent iterative optimization method frequently employed in solving complex non-convex optimization problems for IRS-aided multi-cell MISO system. However, each iteration of SCA involves solving a convex subproblem that approximates the original non-convex problem. Depending on the problem's size and complexity, solving these subproblems may become a bottleneck, especially in massive scale applications. In [50], authors introduced a novel methodology that utilizes a semidefinite programming (SDP) approach to effectively address the optimization challenges associated with the spatial channel covariance matrix (CCM) estimation and beamforming for IRS-assisted mmWave communication systems. However, SDP may have a duality gap, meaning that the optimal dual solution may not always be equal to the optimal primal solution. In [60], the authors undertook a comprehensive study focusing on the simultaneous design of active and passive beamforming techniques to facilitate power-efficient communication in the RIS-aided SWIPT-enabled MIMO downlink communication network. Leveraging the methodology of alternating optimization (AO), they proposed an effective framework aimed at achieving enhanced power efficiency and robust communication performance within the considered MIMO downlink communication network incorporating RIS technology. Nonetheless, the order in which variables are updated in AO can significantly affect its convergence rate and final solution. Weak choices of variable ordering



can lead to slow convergence or getting stuck in suboptimal solutions. In [59], the authors propose the Lagrangian dual reformulation (LDR) to investigate the normalized array for wideband THz RIS communication system. But the effectiveness of LDR can be sensitive to algorithm parameters, such as the choice of the Lagrangian multipliers' update rules and convergence criteria. Poor choices can lead to slow convergence or numerical instability. In [36], the authors considered the joint BS transmit beamforming designs for RIS assisted MU-MISO systems with fractional programming (FP) is proposed. However, FP problems are generally non-convex, which means they may have multiple local optimum. This can make it challenging to find the global optimum, and solutions obtained may not be guaranteed to be the best possible.

As communication systems grow in complexity, traditional physical layer techniques struggle to meet performance demands. Deep learning, particularly DNN, has emerged as a promising approach to address these challenges. DNN have been applied to various communication problems, including signal detection[30, 47, 63], beamforming[7, 15, 16, 23, 31, 37, 46], channel estimation[20, 26, 27, 32], and precoding [18, 25, 49, 54]. In [46], the authors developed and formulated an advanced Long Short-Term Memory (LSTM)-based algorithm with the capability to intelligently and dynamically track link connections within the context of both fixed and adaptive power constraints, focusing specifically on the integration of analog and digital beamforming techniques. The algorithm's design was geared towards enhancing the adaptability and robustness of the overall system while optimizing power consumption and ensuring efficient link tracking performance. However, LSTM is sensitive to random weight initializations. In [23], the authors consider the use of convolutional neural network (CNN) for channel prediction and beamforming in TDD systems. Yet, CNNs struggle to capture global features effectively due to their limited



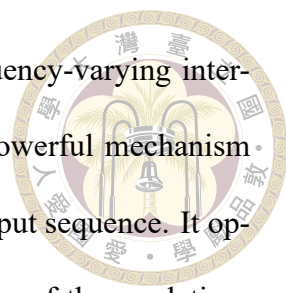
perceptual domain. In [16], the authors propose an autoencoder CSI feedback framework for beamforming design. However, autoencoder compression is inherently lossy, leading to information loss during dimensionality reduction. In [37], the authors propose an attention mechanisms to obtain higher gains through beamforming for mmWave MIMO CSI feedback which achieving improved performance. However, this approach hasn't been explored for IRS-aided MIMO systems. In our prior research[5], we addressed the issue of insufficient attention to information in distinct representation subspaces across various locations. Our research methodology encompassed the strategic implementation of a hybrid approach incorporating Adversarial Autoencoder (AAE) techniques applied directly to raw data, combined with the incorporation of additive attention mechanisms. This integrated strategy facilitated the extraction of multifaceted mapping information from various data subspaces, enabling the revelation of intricate and concealed patterns within the complex dataset. Leveraging the principles of adversarial encoding within the latent space, we meticulously preserved critical information necessary for the comprehensive capture and analysis of global domain features across the temporal, spatial, and frequency domains. Notably, this approach stands out significantly from existing studies[5] due to its innovative amalgamation of AAE methodologies and additive attention techniques, fostering a deeper understanding of the data's underlying structure and interrelationships. The transmission of electromagnetic waves over radio channels is influenced by obstacles, and our AAE is designed to capture the relationship between channel precoding in the time, space, and frequency domains. While CNNs, as mentioned in [23], struggle with global information perception and can lose positional information in input data due to convolutional operations, we chose to use a denoising autoencoder instead of a hybrid CNN-based precoding approach to achieve higher performance.

1.3 Contribution



In this paper, we introduce AAE-AATT-Beamforming (Adversarial AutoEncoder with Additive ATTention Beamforming) and delve into its beamforming design. In our innovative approach, we tackle the intricate issue of signal interference often neglected in conventional hybrid beamforming designs. Specifically, we address the complexities arising from the interaction of subcarriers induced by diffuse scattering during signal propagation, which is further exacerbated by the presence of noise and distortions prevalent in real-world environments. To combat these challenges effectively, we leverage the power of an adversarial autoencoder (AAE), an advanced neural network architecture that harnesses the potential of learning a highly structured latent space. Within the AAE model, the intricate process of encoding the channel data facilitates the transformation of raw input data into a comprehensive latent space representation. This latent space, acquired through the AAE's powerful learning capabilities, empowers the model to generate realistic data samples that closely resemble the characteristics of the original dataset. By leveraging the latent space knowledge gleaned from the AAE, our methodology seeks to establish a robust and effective framework for mitigating the impact of signal interference and enhancing the overall performance of hybrid beamforming techniques. This versatility is advantageous in applications like signal processing and the creation of new data points resembling the training data. Significantly, the learned latent space assumes a crucial role in capturing vital features for intra-channel beamforming across both frequency and spatial domains.

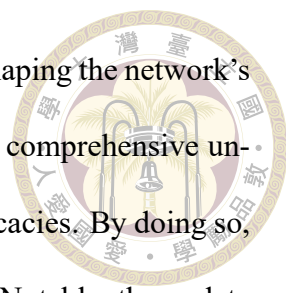
The primary contributions of this thesis are as follows:

- 
- First, our AAE-AATT-Beamforming addresses time and frequency-varying interference between different channels. Additive attention is a powerful mechanism for modeling global relationships among elements within an input sequence. It operates without imposing rigid assumptions regarding the nature of these relationships, making it highly adaptable for capturing diverse dependencies. In the context of channel data analysis, this layer assumes a crucial role by effectively capturing inter-channel features in the time domain characteristics, with a specific emphasis on the intricate and complex interdependencies present in the channel data.
 - Second, we propose AAE-AATT-Beamforming to tackle subcarrier channel interference. In the sophisticated AAE model, the input channel data is seamlessly integrated and meticulously processed through the intricate workings of an encoder, meticulously mapping the intricate nuances of the input data into a finely tuned latent space representation. This transformative process serves as the cornerstone for the AAE's exceptional capabilities, empowering the model with the distinctive ability to generate highly realistic data samples, each meticulously crafted based on the profound insights gleaned from the learned latent space. Leveraging this sophisticated latent space understanding, the AAE transcends conventional limitations, allowing for the creation of data samples that seamlessly encapsulate the intricate characteristics and underlying patterns inherent within the original dataset. This process not only enables the AAE to serve as a powerful generative model but also equips it with the unique potential to facilitate a wide array of applications across various domains, including signal processing and the seamless generation of data points that closely mirror the characteristics and properties of the training data. This capability proves valuable in various applications such as signal processing

and the generation of new data points resembling the training data. Importantly, the learned latent space plays a pivotal role in capturing essential features for intra-channel beamforming in both frequency and spatial domains.



- Third, within the analog precoding module, an autoencoder is employed to optimize the latent space and reconstruct it as faithfully as possible to the original input channel signal data. This intricate and carefully orchestrated process encompasses a sophisticated transformation of feature vectors, expertly navigating the shift from a lower-dimensional space to a meticulously crafted higher-dimensional realm. This deliberate and strategic maneuver serves as a catalyst, sparking a remarkable surge in the model's capability to adeptly capture and encapsulate a diverse array of global domain features. By harnessing the power of this meticulous transformation, the model gains a profound understanding of the intricate nuances and complexities inherent within the global domain, thus laying the foundation for a profound and comprehensive analysis of the underlying patterns and intricacies within the dataset. The digital precoding module employs a 2D-CNN with characteristics like local neuron connectivity, translation invariance, and weight sharing. This configuration is adept at capturing crucial frequency and spatial features necessary for channel precoding while simultaneously reducing interference. This carefully designed analog beamforming module incorporates a sophisticated Gated Recurrent Unit (GRU), an advanced architectural component fortified with two intricately interwoven gating mechanisms: the reset gate and the update gate. Operating in unison, these gates assume a critical role in regulating the intricate flow of information within the cell, orchestrating a delicate interplay between the retention of valuable insights from the previous time step and the assimilation of novel and pertinent information. This in-

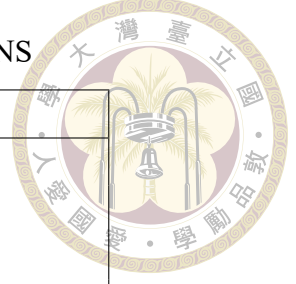
The logo of National Taiwan University (NTU) is located in the upper right quadrant of the page. It is a circular emblem with a gold border. Inside the circle, there is a central figure of a person holding a torch, surrounded by the university's name in Chinese characters: '國立台灣大學' (National Taiwan University) and '敦品勵學' (Cultivate Virtue, Encourage Learning).

tricate dance of retention and incorporation is instrumental in shaping the network's decision-making process, enabling it to achieve a nuanced and comprehensive understanding of the underlying data dynamics and temporal intricacies. By doing so, the GRU effectively models complex temporal dependencies. Notably, the update gate addresses the vanishing gradient problem by regulating the extent to which the previous hidden state should be retained versus updated with new information. This mechanism significantly enhances the accuracy of feature capture within the analog beamforming module. And, the digital beamforming module utilizes a 1D-CNN. 1D-CNN, renowned for their remarkable ability to capture intricate local patterns and dependencies embedded within sequential data, have emerged as a pivotal tool for a diverse array of data analysis tasks, particularly finding their niche in the challenging domain of time series analysis. Leveraging their inherent capacity to discern subtle nuances and extract meaningful insights from temporal sequences, these specialized networks stand as robust tools capable of unraveling complex temporal relationships and uncovering underlying trends that might otherwise remain obscured. By tapping into their profound expertise in local feature extraction, 1D-CNNs offer a versatile and powerful solution to the intricate challenges posed by the analysis of time-varying data, enabling a comprehensive understanding of the dynamic interplay between various data points and the temporal evolution of critical phenomena. This configuration proves advantageous in capturing the necessary features for channel beamforming while minimizing interference.

- Our proposed AAE-AATT-Beamforming model undergoes a thorough assessment of diverse system parameters, generalizability, and robustness through comparisons with current technological methods [16, 23, 37, 46]. Numerical results exhibit

Table 1.1: NOTATIONS AND THEIR DEFINITIONS

Notation	Description
N_b	The transmitter antennas
\mathbf{n}	Gaussian noise
N_b^{RF}	The RF chains
\mathbf{x}	Transmitted signals for user equipment
N_u	The receiver antennas
\mathbf{y}_k	The k -th OFDM subcarrier received signal
$\mathbf{H}_{\text{eff},k}$	The k -th OFDM subcarrier cascade channel
ρ	The base station transmission power
\mathbf{G}	The base station-IRS signals channels
\mathbf{T}	The IRS-user equipment signals channels
\mathbf{s}	The base station transmitted signals
\mathbf{W}_{BB}	The digital baseband combiner
\mathbf{W}_{RF}	The RF combiner
M	IRS passive elements
\mathbf{F}_{RF}	The analog RF precoder
\mathbf{F}_{BB}	The digital baseband precoder



substantial improvements across multiple metrics, including mean squared error (MSE), achievable rate, generalizability, and robustness. These advancements are particularly notable when contrasted with outcomes from prior studies [16, 23, 37, 46].

Notation: $(\cdot)^*$ denote conjugate transpose operators, $E(\cdot)$ denote the statistical expectation operation, $(\cdot)^T$ denote transpose operators, $(\cdot)^{-1}$ denote inverse operators, \mathbf{I} represent the identity matrix, and $\text{diag}\{\cdot\}$ denote the construction of a diagonal matrix.






Chapter 2 Preliminary Study

2.1 6G

The development of each new generation of wireless communication technology, including 6G, is driven by several key factors, with the growing number of internet-connected devices, higher resolution content (e.g., 4K and 8K video streaming), and emerging applications like augmented reality (AR), virtual reality (VR), and holography, there is a significant increase in data demand. 6G aims to provide the capacity and data rates to support these requirements. This stringent demand arises from the necessity to facilitate swift and seamless communication and data processing, allowing for instantaneous decision-making and response mechanisms in high-stakes scenarios. By adhering to these stringent latency constraints, these advanced systems can ensure real-time responsiveness and operational precision, thereby fostering enhanced safety, efficiency, and reliability in mission-critical operations[1, 34, 40, 51, 66]. 6G is expected to provide ultra-low latency to enable these applications. The Internet of Things (IoT) is expected to continue growing with billions of connected devices. 6G is designed to support massive machine-type communication, where devices may need to transmit small amounts of data intermittently over long periods. To reduce the environmental impact and extend the battery life of devices, 6G is expected to focus on energy efficiency in both device operations and network

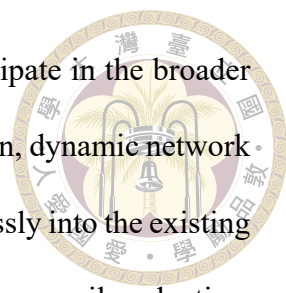
The logo of National Taiwan University (NTU) is located in the upper right quadrant of the page. It is a circular emblem with a gold border. Inside the circle, there is a central figure of a person holding a book, surrounded by the university's name in Chinese characters: "國立台灣大學" (National Taiwan University) at the top and "崇德廣業" (Chongde Guangye) at the bottom.

infrastructure[8, 9, 38, 56, 61]. As the radio spectrum becomes increasingly congested, 6G aims to use the spectrum more efficiently, potentially utilizing higher frequencies and advanced spectrum sharing techniques. 6G is expected to facilitate global connectivity with consistent services, potentially bridging the digital divide by connecting remote and underserved areas[14, 17, 35, 45, 53]. Recognizing the evolving threat landscape and the growing concerns surrounding data privacy and cyber threats, 6G technology is designed to incorporate resilient security protocols, multifaceted encryption mechanisms, and sophisticated privacy safeguards, ensuring the protection of sensitive data and the preservation of user confidentiality. By leveraging cutting-edge cryptographic techniques, secure transmission protocols, and proactive threat detection systems, 6G aims to establish a robust security framework that can effectively mitigate emerging cyber risks and safeguard the integrity of critical communication networks and data transmissions[19, 57, 64]. 6G is expected to support sustainability goals, such as reducing carbon emissions through more efficient communication networks and enabling applications that promote environmental monitoring and protection. Development of new wireless technologies and standards drives innovation, creates economic opportunities, and supports job growth in technology-related industries. Nations and companies that lead in 6G technology development and deployment can gain a competitive advantage in the global technology landscape[42, 43]. In summary, the need for 6G wireless communications arises from the increasing demands of new applications, the desire for improved performance and efficiency, and the potential for technological innovation and economic growth. While 5G is currently being deployed and expanded, research and development efforts for 6G are already underway to address the future communication requirements of society and industry[13, 52].

2.2 IRS

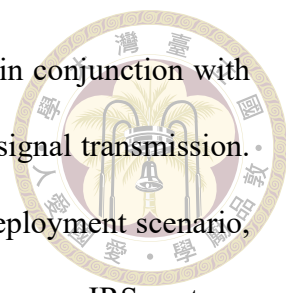


The architecture of an IRS typically consists of the following key components, the core of an IRS is its reflecting elements, often called "metasurfaces" or "metastructures." These elements are composed of materials with unique electromagnetic properties that allow them to manipulate the phase, amplitude, and polarization of incident electromagnetic waves. They can be made up of various types of materials, including metamaterials or metamaterial-inspired structures. IRS may be equipped with sensors to gather information about the environment, such as the location of users or the quality of incoming signals. These sensors provide data that can be used to optimize the reflection patterns. The controller is responsible for processing sensor data and making decisions about how to configure the reflecting elements to achieve specific communication objectives. IRS represent a transformative technology in the realm of wireless communication systems, offering unparalleled capabilities in the manipulation and enhancement of radio wave propagation. These sophisticated surfaces leverage a combination of advanced algorithms and optimization techniques to dynamically adjust the phase shifts and angles of the individual reflecting elements, enabling precise control over signal propagation and reception. By intelligently manipulating the reflected signals, IRS devices can effectively steer the direction of electromagnetic waves, optimize signal paths, and minimize signal attenuation and interference, thereby significantly improving the overall signal quality, coverage, and network performance. Moreover, IRS devices are designed with integrated communication interfaces that facilitate seamless integration with the broader network infrastructure. These interfaces enable IRS devices to establish robust connections with various network components, including base stations, access points, and other wireless devices.


The logo of National Taiwan University (NTU) is located in the upper right quadrant of the page. It is a circular emblem with a gold border. Inside the circle, there is a central figure of a person standing under a tree, with a bell above their head. The Chinese characters "國立台灣大學" (National Taiwan University) are written around the inner edge of the circle.

This interconnectedness empowers IRS technology to actively participate in the broader communication ecosystem, facilitating coordinated signal optimization, dynamic network management, and efficient resource allocation. By integrating seamlessly into the existing communication framework, IRS devices contribute to the creation of more agile, adaptive, and high-performance wireless networks capable of meeting the escalating demands for data capacity, coverage, and reliability. This enables them to receive instructions from network controllers and exchange information about their status and capabilities[4, 33, 48].

IRS represent a groundbreaking advancement in wireless communication technology, offering a wealth of transformative capabilities that can be seamlessly integrated into existing and future wireless communication networks. With the advent of 5G and the relentless pursuit of even more sophisticated technologies in the future, the integration of IRS into the existing infrastructure holds tremendous promise for revolutionizing the way we transmit and receive data. By integrating IRS into the fabric of contemporary communication networks, we can leverage the advanced capabilities of these intelligent surfaces to enhance the overall network performance, extend coverage, and optimize the utilization of available resources. The incorporation of IRS technology into the existing framework can enable the development of highly efficient and adaptable communication ecosystems that can dynamically adapt to changing network demands, environmental conditions, and user requirements. Furthermore, the integration of IRS into 5G and upcoming wireless communication networks can significantly augment the network's capacity, improve spectral efficiency, and reduce signal interference and propagation losses. The intelligent manipulation of signal paths and the precise control over signal reflection and propagation facilitated by IRS technology can pave the way for more reliable, secure, and high-speed data transmission, fulfilling the mounting demands for ultra-fast and seamless connectiv-

The logo of National Taiwan University (NTU) is located in the upper right quadrant of the page. It is a circular emblem with a gold border. Inside the circle, there is a central figure of a person holding a torch, surrounded by the university's name in Chinese characters: '國立台灣大學' (National Taiwan University) and '崇德廣業' (Chongde Guangye).

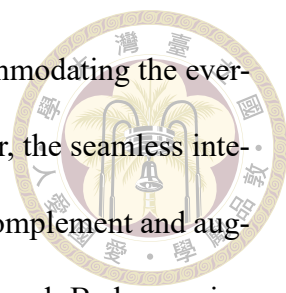
ity across a diverse array of applications and use cases. They work in conjunction with base stations, access points, and other network elements to enhance signal transmission. IRS devices require a power supply to operate. Depending on the deployment scenario, they may be powered by batteries, solar panels, or other energy sources. IRS systems rely on specialized software to manage the configuration of the reflecting elements and optimize signal propagation. This software can incorporate machine learning algorithms, optimization routines, and real-time feedback mechanisms. Some IRS implementations may include antenna arrays to enhance their functionality. These arrays can help focus signals in specific directions. The architecture of an IRS can vary depending on its intended application and deployment scenario. For example, IRS used in indoor environments may have different characteristics than those deployed in outdoor urban settings or for satellite communication. The key is that IRS devices are programmable and can adapt to changing communication needs dynamically, making them a versatile tool for optimizing wireless communication[22, 29, 65]. IRS are a promising technology in the field of wireless communication, and they are envisioned to serve several essential purposes, IRS can enhance the strength of wireless signals by reflecting them in a desired direction. This is particularly useful in scenarios with obstacles or signal blockages. By intelligently redirecting signals, IRS can improve signal quality and coverage. IRS can extend the coverage of wireless networks, especially in challenging environments such as urban areas with high-rise buildings or rural areas with sparse infrastructure. It can effectively bring wireless connectivity to areas that were previously underserved. IRS can increase the capacity of wireless networks by focusing signals where they are needed most. This is especially valuable in crowded environments like stadiums, airports, and urban centers where many users are trying to connect simultaneously. By optimizing signal paths, IRS can reduce

The logo of National Taiwan University (NTU) is located in the upper right quadrant of the page. It is a circular emblem with a gold border. Inside the circle, there is a central figure of a person holding a torch, surrounded by the university's name in Chinese characters: '國立台灣大學' (National Taiwan University) at the top and '崇德 尚學 博愛' (Chongde, Shangxue, Boai) at the bottom.

the power requirements of communication devices. This can lead to improved energy efficiency, longer battery life for mobile devices, and reduced environmental impact. IRS can be used to create controlled communication zones. By directing signals only to specific areas or devices, it can enhance privacy and security by limiting the risk of eavesdropping or interference. IRS represent a pioneering technology that can significantly contribute to the reduction of communication latency, a critical factor in the successful deployment of a wide array of cutting-edge applications. This reduction in latency is particularly crucial for mission-critical applications, including autonomous vehicles, remote surgery, real-time industrial automation, and other time-sensitive operations that rely on instantaneous data transmission and response. By integrating IRS into the communication infrastructure, it becomes possible to strategically manipulate and optimize signal propagation, thereby minimizing the delays typically encountered in data transmission. The dynamic control and precise adjustment of signal paths facilitated by IRS technology enable the swift and efficient redirection of signals, effectively reducing the time required for data to travel between the source and the destination. This capability is especially beneficial in scenarios where even the slightest latency can have far-reaching consequences. Additionally, the integration of IRS in communication networks allows for the implementation of advanced signal processing and beamforming techniques, which can substantially enhance the overall speed and efficiency of data transmission. This, in turn, translates into improved responsiveness, increased reliability, and heightened levels of operational efficiency in critical applications that demand instantaneous and real-time data communication. As a result, the incorporation of IRS into the communication infrastructure promises to usher in a new era of ultra-low latency communication, revolutionizing various sectors and applications that rely on swift and dependable data exchange. By optimizing signal paths, IRS can help

ensure that data reaches its destination quickly. IRS can enable more efficient use of the radio spectrum. By focusing signals precisely, it reduces interference and allows for the reuse of spectrum in different spatial regions, effectively increasing spectral efficiency.

In some cases, IRS can be a cost-effective solution for improving wireless connectivity compared to deploying additional infrastructure or more powerful transmitters. IRS systems can be reconfigured dynamically to adapt to changing communication needs or to address specific challenges, making them a versatile solution for various scenarios. IRS is seen as a complementary technology to 6G and future wireless standards. It can help overcome some of the limitations of traditional network architecture. In summary, IRS technology has emerged as a transformative innovation with the capacity to revolutionize the landscape of wireless communication systems. Its potential to significantly enhance the performance, expand the coverage, and optimize the overall efficiency of communication networks has attracted substantial attention from both the academic and industrial communities. One of the key advantages offered by IRS lies in its ability to dynamically manipulate and control the propagation of electromagnetic waves, allowing for the precise shaping and steering of signal beams. This capability contributes to the establishment of robust and reliable communication links, resulting in improved signal strength, enhanced coverage, and minimized signal degradation across diverse environments. Furthermore, the integration of IRS into wireless communication systems enables the implementation of advanced beamforming techniques, thereby facilitating the directed transmission of signals to specific users or devices. This targeted signal transmission enhances the overall efficiency of the network by reducing unnecessary signal interference and maximizing the utilization of available bandwidth. As a result, IRS technology not only bolsters the performance of wireless communication networks but also lays the groundwork for the

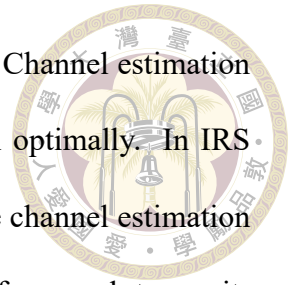


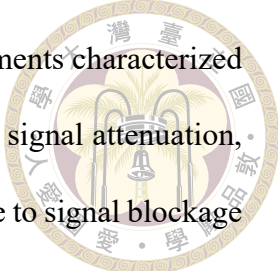
development of high-speed, high-capacity networks capable of accommodating the ever-increasing demands for data transmission and connectivity. Moreover, the seamless integration of IRS into existing communication infrastructure serves to complement and augment the capabilities of conventional technologies, such as 5G and beyond. By harnessing the full potential of IRS, wireless communication systems can achieve unprecedented levels of reliability, speed, and connectivity, leading to improved user experiences, enhanced network performance, and the establishment of a solid foundation for the realization of the full potential of future communication technologies. It is particularly valuable in scenarios with complex propagation environments, high user density, or stringent requirements for privacy and security[24, 41, 55].

2.3 Channel estimation

Channel estimation is a vital component in IRS communication systems for several critical reasons. And, to realize the full potential of spatial multiplexing, accurate knowledge of the channel is required. Channel estimation enables the receiver to differentiate between the various transmitted streams, separating and decoding them effectively. In IRS systems, signals from different transmit antennas may experience different fading and attenuation effects as they propagate through the wireless channel. Accurate channel estimation helps the receiver mitigate these effects by compensating for the variations in signal amplitude and phase. This compensation is crucial for accurate signal detection and decoding, especially in the presence of channel impairments like fading and multipath propagation. In IRS systems, both precoding and beamforming techniques rely on knowledge of the channel. Precoding involves shaping the transmitted signal based on the channel characteristics to maximize data rates or minimize interference. Beamforming, on

the other hand, focuses the reception of signals in desired directions. Channel estimation provides the necessary information for these techniques to function optimally. In IRS scenarios, where multiple transmitters and receivers coexist, accurate channel estimation is crucial for interference management. By estimating the channels from each transmitter to the receiver, the system can apply interference mitigation techniques, such as spatial nulling or interference cancellation, to improve communication quality and system performance. Channel estimation allows the system to adapt the modulation and coding schemes dynamically based on the channel conditions. Adapting modulation and coding schemes based on the channel quality is a crucial strategy employed in wireless communication systems to ensure optimal data transmission performance across varying environmental conditions. This adaptive approach facilitates the dynamic adjustment of the data rate and error correction capabilities in response to the fluctuating quality of the communication channel. When the channel quality is deemed favorable, the implementation of higher-order modulation techniques, combined with sophisticated coding schemes, becomes instrumental in maximizing the data transmission capacity. This optimization strategy is particularly beneficial in scenarios where high-speed data transfer and efficient bandwidth utilization are paramount, such as multimedia streaming, high-definition video conferencing, or large-scale data transfer applications. Conversely, when the channel conditions deteriorate, necessitating the mitigation of potential signal impairments, lower-order modulation and coding schemes are preferred. By employing simpler modulation formats and error correction techniques, the system can ensure reliable data transmission even in the presence of signal fading, noise, or interference. This adaptive adjustment helps to maintain a robust and stable link between the transmitter and receiver, minimizing the risk of data loss or transmission errors, and ensuring the delivery of consistent and dependable



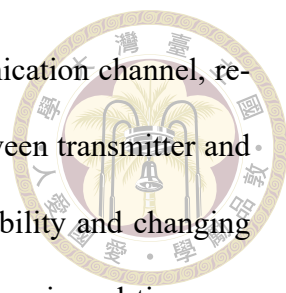


communication services. Such an approach proves crucial in environments characterized by challenging propagation conditions, such as urban areas with high signal attenuation, remote regions with limited infrastructure, or environments susceptible to signal blockage and multipath fading. The dynamic modulation and coding scheme adaptation mechanism serve as a cornerstone in the design and operation of modern wireless communication systems, enabling the optimization of data transmission performance, the efficient utilization of available resources, and the provision of reliable and seamless connectivity across diverse and challenging environments. Accurate channel estimation at the user devices provides feedback that helps the base station optimize beamforming for each user, enhancing spectral efficiency and overall system performance. In IRS systems, diversity gain is achieved by exploiting multiple paths and spatial dimensions. Accurate channel estimation allows the system to realize diversity gains by combining signals from different antennas coherently. This diversity improves the reliability of communication in challenging wireless environments. Channel estimation assists in intelligent resource allocation, such as power allocation and bandwidth allocation. For applications with stringent QoS requirements, such as teleconferencing or internet relay chat, accurate channel estimation is essential. It helps ensure that the required level of service quality is maintained, even in dynamic and changing wireless environments. In summary, channel estimation in MIMO systems is fundamental for optimizing data rates, managing interference, enabling advanced techniques like beamforming, and adapting to varying channel conditions. It plays a central role in achieving the performance gains and spectral efficiency that IRS technology promises in wireless communication.

2.4 Precoding



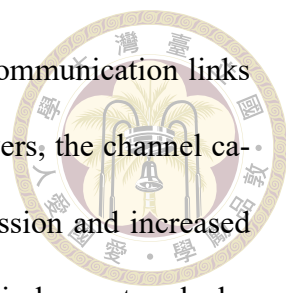
Precoding is an essential technique in IRS systems for several important reasons, one is that to simultaneously transmit multiple data streams over the same frequency band. Precoding helps in the efficient utilization of the available spatial dimensions by shaping and optimizing the transmitted signals. This enables spatial multiplexing, where multiple independent data streams can be sent and received in parallel, significantly increasing the data rate and system capacity. By carefully designing the precoding matrix, it is possible to mitigate interference and noise, leading to improved signal-to-noise ratio (SNR) and reduced bit error rates (BER). This is especially valuable in scenarios with high interference or when signals must travel through challenging environments. In IRS systems, signals transmitted from multiple antennas may interfere with each other at the receiver. Precoding can be employed to reduce or eliminate interference between the transmitted signals, allowing multiple users or devices to communicate simultaneously in the same frequency band without significant degradation in performance. Precoding can be used to focus transmitted energy in specific directions, effectively extending the coverage and range of wireless communication systems. By directing signals towards the intended receivers, it is possible to achieve better coverage in certain areas, especially in outdoor and long-range communication scenarios. Precoding techniques can exploit the spatial diversity provided by multiple antennas to improve link reliability. By transmitting redundant information across multiple antennas, it becomes more likely that at least one of the received signals will be of sufficient quality, reducing the likelihood of communication failures, particularly in fading or noisy channels. Precoding helps maximize the achievable capacity of IRS systems. By optimizing the transmitted signals based on the channel conditions, it



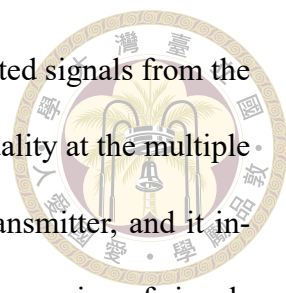
is possible to approach the theoretical capacity limits of the communication channel, resulting in higher data rates and spectral efficiency. The channel between transmitter and receiver in wireless communication can vary due to factors like mobility and changing environments. Precoding can adapt to these changing channel conditions in real-time, optimizing signal transmission for current channel characteristics. In summary, precoding plays a crucial role in IRS systems by enhancing data rates, signal quality, interference management, coverage, and reliability. It is a key technology that allows MIMO systems to fully leverage their spatial diversity and multiplexing capabilities, making it an integral part of modern wireless communication systems.

2.5 Beamforming

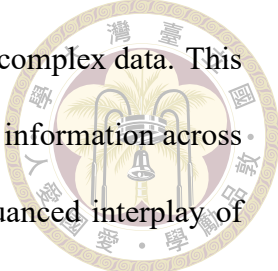
Beamforming is an essential technique in IRS systems for several important reasons, beamforming allows the transmitter to concentrate signal energy in specific directions or toward particular receivers. This spatial focus enhances the signal strength and quality at the intended destinations, effectively increasing the coverage area and range of wireless communication systems. It ensures that transmitted signals reach their intended targets with minimal wastage of energy in other directions. In IRS wireless communication scenarios with multiple transmitters and receivers, interference between signals can be a significant challenge. Beamforming can be used to minimize interference by directing the transmitted signal away from interfering sources or nulling them out. This interference mitigation improves overall system performance, especially in crowded spectral environments. By focusing the transmitted energy in the direction of the receiver, beamforming increases the received signal power while minimizing noise and interference from other directions. This results in a higher SNR, leading to improved communication reliability

The logo of National Taiwan University (NTU) is located in the upper right quadrant of the page. It is a circular emblem with a gold border. Inside the circle, there is a central figure of a person holding a torch, surrounded by the university's name in Chinese characters (國立台灣大學) and English (NATIONAL TAIWAN UNIVERSITY).

and reduced BER. Beamforming can be used to create directional communication links with higher data rates. By steering transmit beams toward the receivers, the channel capacity can be more effectively utilized, enabling faster data transmission and increased spectral efficiency. Beamforming can extend the coverage area of wireless networks by directing signals to reach locations that may be challenging to access with traditional omnidirectional transmission. This is especially valuable in outdoor and long-range communication scenarios. Focusing the transmitted signal in a specific direction reduces energy wastage in unwanted directions. This energy-efficient operation is particularly crucial in battery-powered and energy-constrained devices, such as mobile phones and IoT sensors. Beamforming can adapt to changing channel conditions in real-time. By continuously monitoring the channel state and adjusting the beam direction, the system can maintain optimal communication performance even in dynamic environments. Directional transmission reduces the risk of eavesdropping or interception of signals by unintended receivers. Beamforming can enhance the security of wireless communication by limiting signal exposure to potential attackers. In IRS communication systems, beamforming can be used to establish reliable links over longer distances without the need for additional infrastructure, such as relay stations or additional base stations. Beamforming can help increase the capacity of cellular networks by allowing more users to share the same spectrum concurrently. By directing signals to specific users or sectors, the system can support a larger number of simultaneous connections. In summary, beamforming in IRS systems is crucial for optimizing signal transmission, improving link quality, reducing interference, extending coverage, conserving energy, and adapting to varying channel conditions. In the IRS wireless communication systems, precoding and beamforming are two distinct techniques used to improve the performance of data transmission. Here's the difference

The logo of National Taiwan University (NTU) is located in the upper right quadrant of the page. It is a circular emblem with a gold border. Inside the circle, there is a central figure of a person holding a torch, surrounded by the university's name in Chinese characters: "國立台灣大學" (National Taiwan University) and "崇德廣業" (Chongde Guangye).

between them, precoding is a technique used to optimize the transmitted signals from the multiple antennas at the transmitter to improve the received signal quality at the multiple antennas at the receiver. Precoding is typically performed at the transmitter, and it involves complex signal processing techniques. It focuses on spatial processing of signals before transmission to exploit the spatial characteristics of the IRS channel, such as reducing interference or increasing the SNR. Precoding can be used to transmit multiple data streams simultaneously, improving the system's throughput. Beamforming is a subset of precoding and specifically refers to the process of shaping and directing the transmitted signal in a specific direction or towards a specific receiver or set of receivers. Beamforming can be relatively simpler than general precoding techniques and is often used for specific applications like directional transmission. It primarily focuses on steering the transmitted signal in a particular direction to maximize the signal power at the intended receiver while minimizing interference to other receivers. For single-stream beamforming, it is often referred to as "single-user beamforming." In summary, precoding is a broader term that encompasses various techniques to optimize signal transmission in IRS systems. Beamforming, on the other hand, is a specific type of precoding that emphasizes directing signals in a particular spatial direction to maximize reception at the intended receivers. Both techniques aim to improve wireless communication performance by exploiting the spatial characteristics of the channel. In our prior research[5, 6], we addressed the issue of insufficient attention to information in distinct representation subspaces across various locations. In the pursuit of advancing the frontiers of research in this domain, our study has strategically leveraged a synergistic combination of advanced AAE methodologies on raw data, with a novel emphasis on harnessing the power of additive attention. By synergistically integrating these sophisticated techniques, we were able to unravel and illuminate



previously concealed patterns embedded within the intricate layers of complex data. This transformative approach facilitated the generation of diverse mapping information across distinct subspaces, enabling us to glean valuable insights into the nuanced interplay of hidden patterns and correlations that often evade conventional analysis techniques. One of the defining features of our work lies in the integration of adversarial encoding within the latent space, an innovative strategy aimed at meticulously preserving essential information critical for capturing comprehensive global domain features spanning space, time, and frequency domains. This distinctive approach sets our research apart from earlier studies[5, 6], establishing a novel paradigm that is poised to make significant contributions to the overarching field of data analysis and pattern recognition. In the rigorous experimental phase of our study, we meticulously curated and executed an extensive comparison of various modules, rigorously examining and dissecting their performance in detail. This systematic evaluation revealed compelling evidence and compellingly demonstrated the superior efficacy and robustness of our proposed methodology. The results, meticulously analyzed and meticulously presented, not only underscore the tangible impact of our work but also serve as a testament to the transformative potential of our innovative framework in reshaping and redefining the existing landscape of data analysis and interpretation. Drawing from these intricate analyses and results, our study lays the groundwork for a new paradigm of data analysis that promises to unlock previously untapped insights and pave the way for a more nuanced and comprehensive understanding of complex data dynamics. Through this work, we aim to contribute significantly to the advancement of the field and pave the way for further explorations and applications in diverse domains.

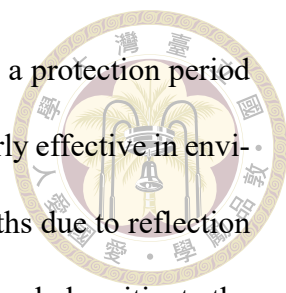




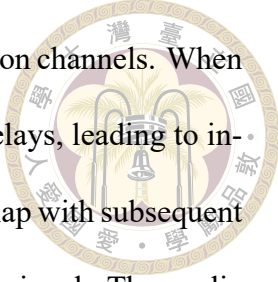
Chapter 3 System Model

3.1 Channel Model

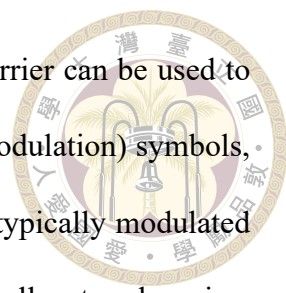
We study a IRS-assisted mm-wave MIMO OFDM system comprising of a base station with N_t^{RF} transmit radio frequency (RF) chains and N_t transmitting antennas, also a user with N_r^{RF} receive RF chains and N_r receiving antennas, assisted by an IRS with M reflecting elements as shown in Fig. 3.1. The OFDM is a modulation and multiplexing approach utilized in digital communication systems. It is broadly adopted in various technologies, including 4G LTE, 5G, 6G, and digital television broadcasting. OFDM is known for its efficiency in handling high data rates and mitigating issues related to multipath interference. OFDM divides the available bandwidth into multiple subcarriers, also known as tones or bins. These subcarriers are closely spaced in frequency and are orthogonal to each other, meaning they do not interfere with each other. Each subcarrier can be thought of as a separate narrowband channel. The orthogonality of subcarriers is a critical feature of OFDM. Because the subcarriers do not overlap in frequency and are orthogonal, they can be transmitted simultaneously without causing interference between them. OFDM transmits data in parallel by modulating each subcarrier with its own data symbol. This parallel transmission increases the overall data throughput. OFDM includes guard intervals between symbols or subcarriers to deal with multipath interference, which

The logo of National Taiwan University (NTU) is located in the upper right quadrant of the page. It is a circular emblem with a purple and gold color scheme. The outer ring contains the university's name in Chinese characters: '國立台灣大學' at the top and 'NTU' at the bottom. The inner circle features a central figure, possibly a scholar or a historical figure, surrounded by a laurel wreath. The text '崇德 尚學 博愛' is inscribed around the inner circle.

can cause intersymbol interference (ISI). The guard intervals provide a protection period against delayed versions of the transmitted signal. OFDM is particularly effective in environments with multipath propagation, where signals take multiple paths due to reflection and scattering. The guard intervals and the orthogonality of subcarriers help mitigate the effects of multipath fading. OFDM is known for its spectral efficiency, as it can efficiently use available frequency bands. The closely spaced subcarriers allow for efficient utilization of the frequency spectrum. OFDM systems can adapt the modulation and coding scheme (MCS) for each subcarrier independently, based on channel conditions. This adaptability ensures that OFDM can provide high data rates while maintaining robustness in challenging channel conditions. OFDM is used in 5G NR (New Radio), 6G, digital television broadcasting (DVB-T, ATSC), and many other wireless communication technologies. Overall, OFDM has become a fundamental technology for high-speed wireless communication due to high spectral efficiency, and adaptability to changing environments. In OFDM, the IFFT and the addition of the cyclic prefix (CP) are essential components of the modulation scheme. They play crucial roles in the transmission and reception of OFDM signals. In the OFDM system, data symbols are typically modulated onto multiple subcarriers, each representing a specific frequency. These subcarriers are orthogonal to each other, meaning they don't interfere with each other. However, to transmit data over a channel, we need to convert these frequency-domain symbols into a time-domain signal. The IFFT performs this conversion. It takes the data symbols on each subcarrier and combines them to create a complex time-domain signal. The result is a time-domain waveform that represents the composite OFDM symbol. The IFFT operation ensures that the subcarriers' frequencies are orthogonal in the time domain, allowing for simultaneous transmission and avoiding interference. The addition of a cyclic prefix is a technique used



to combat the effects of multipath interference in wireless communication channels. When a signal travels through a channel, it can experience reflections and delays, leading to intersymbol interference (ISI). ISI occurs when parts of one symbol overlap with subsequent symbols, making it challenging to correctly demodulate the received signal. The cyclic prefix is a copy of the end of the OFDM symbol that is appended to the beginning. It's a guard interval, consisting of the last part of the symbol, copied and added to the front. By adding the cyclic prefix, the OFDM symbol's cyclically extended version contains enough information to fill in the delayed or distorted portions caused by multipath propagation. This mitigates ISI, making it easier for the receiver to recover the original data. The cyclic prefix simplifies the equalization process at the receiver. Overall, IFFT and cyclic prefix work together to convert data symbols from the frequency domain to the time domain and protect the transmitted signal against the adverse effects of multipath propagation in the channel. This combination of techniques makes OFDM a robust and efficient modulation scheme for high-speed data transmission in various wireless communication systems, such as Wi-Fi and 4G/5G/6G cellular networks. After IFFT and adding cyclic prefix, the BS uses an analog RF precoder $\mathbf{F}_{\text{RF}} \in \mathbb{C}^{N_t \times N_t^{\text{RF}}}$, analog beamforming $\mathbf{B}_{\text{RF}} \in \mathbb{C}^{N_t \times N_t^{\text{RF}}}$, and each pilot frame with OFDM symbols of K subcarriers. In OFDM, subcarriers are individual carriers or tones within the total bandwidth allocated for data transmission. OFDM divides the available frequency spectrum into numerous subcarriers, each of which can be thought of as a narrowband signal. These subcarriers are spaced at specific frequencies, and they collectively carry data. Subcarriers are orthogonal to each other. This orthogonality is achieved by ensuring that the frequency spacing between subcarriers is such that the sinc function of one subcarrier is zero at the frequency position of another subcarrier. This orthogonality minimizes interference between subcarriers, allowing them to be



transmitted simultaneously without mutual interference. Each subcarrier can be used to carry data, such as binary symbols, QAM (Quadrature Amplitude Modulation) symbols, or other digital modulation formats. The data on each subcarrier is typically modulated separately before being transmitted. OFDM systems can dynamically allocate subcarriers to different users, services, or purposes, making it a flexible modulation scheme. For example, in a multi-user environment, different subcarriers can be assigned to different users based on their channel conditions and data rate requirements. In some OFDM systems, certain subcarriers at the edges of the frequency spectrum are reserved as guard bands. These guard bands are not used for data transmission but serve as a buffer to prevent interference with adjacent communication systems or to provide protection against spectral leakage. OFDM's use of subcarriers is particularly effective in combating the effects of frequency-selective fading in wireless channels. When some subcarriers experience fading or interference, others may remain unaffected, allowing for the reconstruction of the transmitted data. OFDM systems can adapt the number of subcarriers based on the available bandwidth and the specific application. For example, Wi-Fi standards like 802.11a/g/n/ac/ax use varying numbers of subcarriers depending on the channel bandwidth. Overall, subcarriers are a fundamental component of OFDM modulation, enabling the efficient transmission of data over a wide range of communication channels, including wired and wireless systems.

3.2 Signal Model

The transmitted signal from the BS is $\mathbf{s}[k]$, where $\mathbf{s}[k] \in \mathbb{C}^{N_s}$ represents OFDM symbols, $\mathbf{F}_{\text{BB}}[k] \in \mathbb{C}^{N_t^{\text{RF}} \times N_s}$ represents digital baseband precoder, $\mathbf{B}_{\text{BB}}[k] \in \mathbb{C}^{N_t^{\text{RF}} \times N_s}$ represents digital beamforming, and it is assumed that power constraint is satisfied by

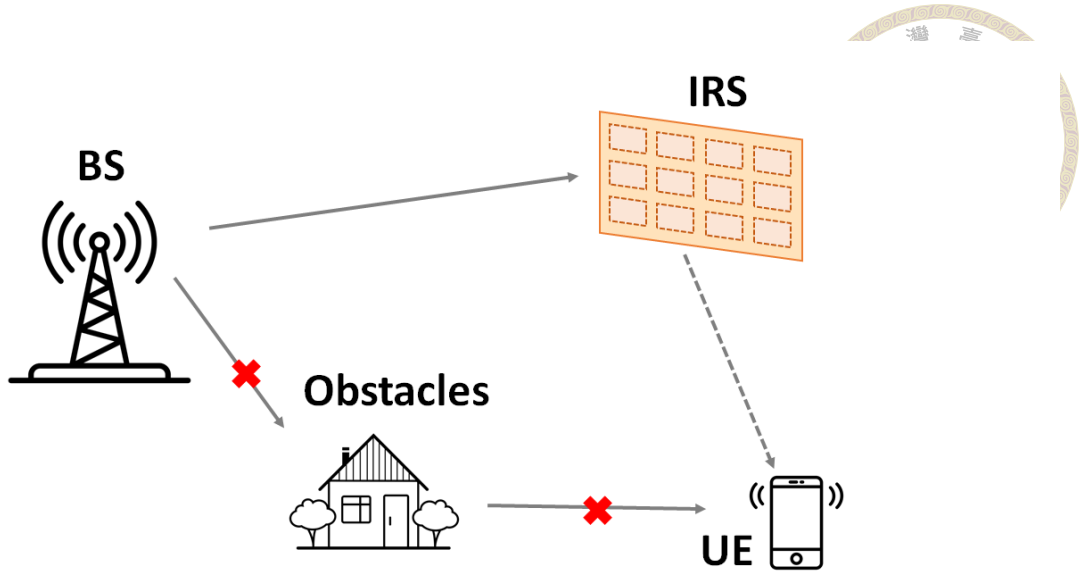


Figure 3.1: Scenario diagram of IRS-assisted MIMO OFDM system.

$\mathbb{E}[\mathbf{s}[k]\mathbf{s}^H[k]] = KN_s\mathbf{I}_{N_s}$. The UE removes the cyclic prefix and performs an N -point FFT frequency transformation, the UE applies an analog combiner $\mathbf{W}_{\text{RF}} \in \mathbb{C}^{N_r \times N_r^{\text{RF}}}$ and digital combiner $\mathbf{W}_{\text{BB}} \in \mathbb{C}^{N_r^{\text{RF}} \times N_s}$ in Fig. 3.2. Then, the received signals at the UE at the k -th subcarrier can be written as

$$\mathbf{y}[k] = \mathbf{s}[k]\mathbf{F}_{\text{BB}}[k]\mathbf{B}_{\text{BB}}[k]\mathbf{F}_{\text{RF}}\mathbf{B}_{\text{RF}}\mathbf{H}_{\text{eff}}[k]\mathbf{W}_{\text{RF}}\mathbf{W}_{\text{BB}}[k] + \mathbf{n}[k]. \quad (3.1)$$

where $\mathbf{H}_{\text{eff}}[k] \in \mathbb{C}^{N_r \times N_s}$ denote the signals vector, the BS communicates N_s sequential signals to the UE through k subcarriers, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_{N_t})$ and $\mathbf{n}[k] \in \mathbb{C}^{N_r}$ are the Gaussian noise. The channel model used in this paper is the Saleh-Valenzuela model as the channel model for the BS-IRS and IRS-User links. $\mathbf{H}_{\text{eff}}[k] \triangleq \mathbf{G}[k]\mathbf{\Theta}\mathbf{T}[k]$ represent the equivalent channels from the BS to the UE, and $\mathbf{\Theta} = \text{diag}(e^{j\theta_1}, e^{j\theta_2}, \dots, e^{j\theta_{N_s}})$ denotes a phase shifter and θ_n is the n -th element of IRS, and the channel matrix $\mathbf{G}[k]$ is modeled as

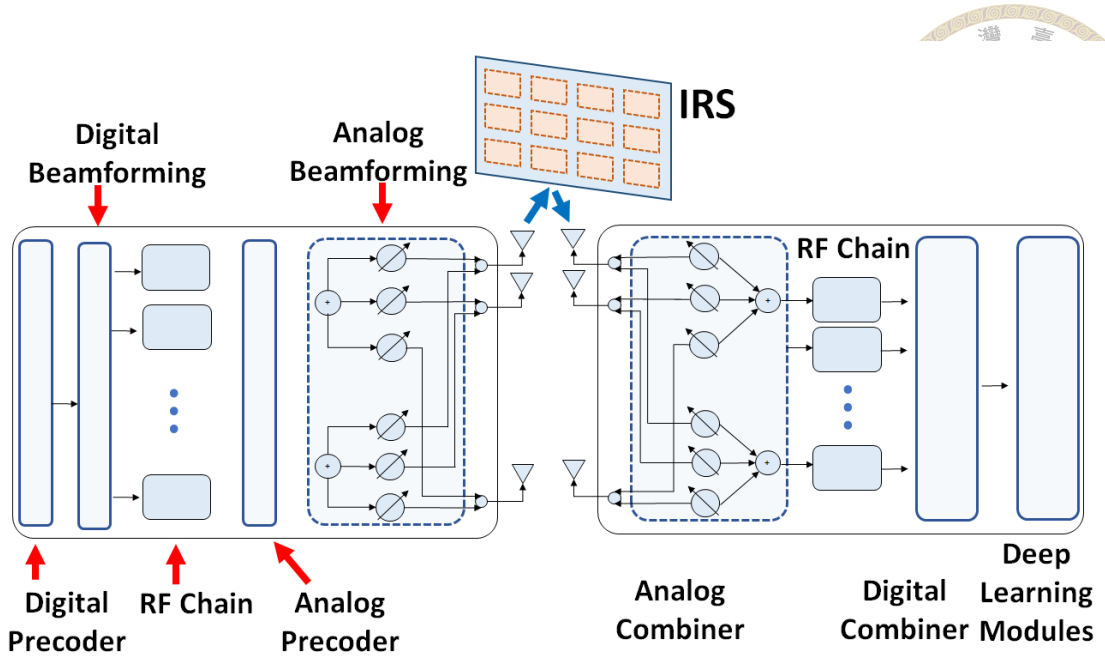


Figure 3.2: Schematic diagram of the IRS-assisted beamforming design architecture with mmwave MIMO OFDM system.

$$\mathbf{G}[k] = \sqrt{\frac{N_t L_x \times L_z}{M N_t}} \sum_{l_g=1}^{L_g} \mathbf{g}_{l_g}^{\text{BI}} \mathbf{a}_{l_g}(\theta_{\text{BI},l_g}, \phi_{\text{BI},l_g}) \mathbf{b}_{l_g}^H(\theta_{\text{BI},l_g}), \quad (3.2)$$

where $\mathbf{a}_{l_g}(\theta_{\text{BI},l_g}, \phi_{\text{BI},l_g})$ and $\mathbf{b}_{l_g}(\theta_{\text{BI},l_g})$ represents the array response vectors of the l_g -th propagation paths at the BS and IRS, $(\theta_{\text{BI},l_g}, \phi_{\text{BI},l_g})$ represents the azimuth and elevation of AoA, $\mathbf{g}_{l_g}^{\text{BI}}$ represents the complex channel gain, the array response vectors can be written as

$$\mathbf{a}_{l_g}(\theta_{l_g}, \phi_{l_g}) = \frac{1}{\sqrt{L_x \times L_z}} \left[1 \dots e^{j\lambda 2\pi f (l_x \sin(\theta_{l_g}) \sin(\phi_{l_g}) + l_z \cos(\phi_{l_g}))} \dots \right]^T, \quad (3.3)$$

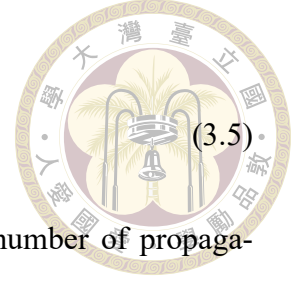
$$\mathbf{b}_{l_g}(\theta_{l_g}) = \frac{1}{\sqrt{N_T}} \left[1 e^{j\lambda 2\pi f \sin(\theta_{l_g})} \dots e^{j\lambda(N_T-1) \sin(\theta_{l_g})} \right]^T, \quad (3.4)$$

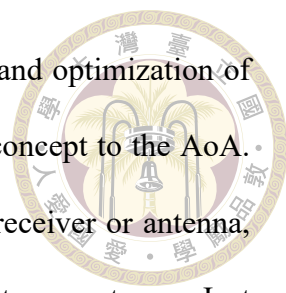
where f represents the carrier frequency, L_x represents the number of reflecting elements of the IRS on horizontal plane, λ represents the antenna spacing, and L_z represents the number of reflecting elements of the IRS on vertical plane. Similarly, the channel vector

$\mathbf{T}[k]$ is modeled as

$$\mathbf{T}[k] = \sqrt{\frac{L_{\text{IRS}}}{MN_r}} \sum_{l_t=1}^{L_t} \mathbf{t}_{l_t}^{\text{IU}} \mathbf{a}_{l_t}(\theta_{\text{IU},k}, \phi_{\text{IU},k}), \quad (3.5)$$

where \mathbf{t}_k^{IU} represents the complex channel gain, L_t represents the number of propagation paths, $\theta_{\text{IU},k}$ and $\phi_{\text{IU},k}$ represents the azimuth and elevation of AoD, respectively, and $\mathbf{a}_{l_t}(\theta_{\text{IU},k}^{\text{AoD}}, \phi_{\text{IU},k}^{\text{AoD}})$ represents the array response vector of the l_t -th propagation paths at the IRS and UE. The AoA refers to the direction from which a signal, such as an electromagnetic wave or sound wave, arrives at a receiver or antenna. In various fields, including telecommunications, radar, acoustics, and astronomy, AoA is typically measured in degrees or radians. In degrees, it often ranges from 0° to 360° , where 0° represents a signal coming from the north direction, 90° from the east, 180° from the south, and 270° from the west. In some cases, it may be measured relative to other reference points. AoA estimation is commonly used in antenna arrays, where multiple antennas are used to receive signals. By analyzing the phase differences of signals received at different antennas, it's possible to estimate the angle of arrival of a signal source. This technique is known as spatial signal processing. In wireless communication, it can help optimize beamforming to direct signals toward specific users or avoid interference. In acoustics, it aids in locating sound sources. When there are multiple sources of signals arriving at a receiver, AoA estimation can help distinguish between them. This is crucial for scenarios like multi-user wireless communication or radar systems tracking multiple targets. The geometry and arrangement of antennas in an array play a significant role in AoA estimation. Different array configurations, such as uniform linear arrays (ULA), uniform planar array (UPA), or more complex geometries, can provide varying levels of AoA accuracy. Advanced algorithms and signal processing methods are often employed to mitigate these challenges. Overall, AoA estimation is a valuable tool in many fields where the direction of signal



The logo of National Taiwan University (NTU) is located in the upper right quadrant of the page. It is a circular emblem with a gold border. Inside the circle, there is a central figure of a person holding a torch, surrounded by the university's name in Chinese characters: '國立台灣大學' (National Taiwan University) and '崇德廣業' (Chongde Guangye).

sources needs to be determined. It enables better spatial awareness and optimization of various systems and technologies. The AoD is the complementary concept to the AoA. While AoA refers to the direction from which a signal arrives at a receiver or antenna, AoD refers to the direction in which a signal departs from a transmitter or antenna. Just as AoA helps determine the direction from which a signal arrives at a receiver, AoD helps determine the direction in which a signal is transmitted from a source or antenna. AoA and AoD are often used together to describe the full spatial characteristics of signal propagation. AoD is typically measured in degrees or radians, similar to AoA. AoD estimation is important in various applications, including wireless communication systems. Knowing the AoD is crucial for beamforming, which involves directing signals toward specific angles or directions to optimize communication links. In beamforming, the transmitter adjusts the phase and amplitude of signals transmitted from an antenna array to steer the signal in a specific direction. The AoD is a key parameter used to control the direction of the transmitted beam. In systems with multiple transmit antennas, such as IRS systems, AoD estimation helps determine the angles at which signals are transmitted from each antenna. This information is used to achieve spatial diversity and multiplexing in wireless communication. By estimating the AoD of signals received at multiple receiver locations, it's possible to triangulate the source's position. AoD estimation often involves signal processing techniques similar to those used in AoA estimation, such as beamforming, and other spatial signal processing methods. Estimating AoD accurately can be challenging due to factors like noise, multipath propagation, and interference. Advanced algorithms and techniques are used to enhance AoD estimation performance. And, AoD is an important parameter in wireless communication, radar, and localization systems, as it describes the direction in which signals are transmitted from a source or antenna. Accurate AoD

estimation is crucial for optimizing communication links and spatial awareness in various applications.

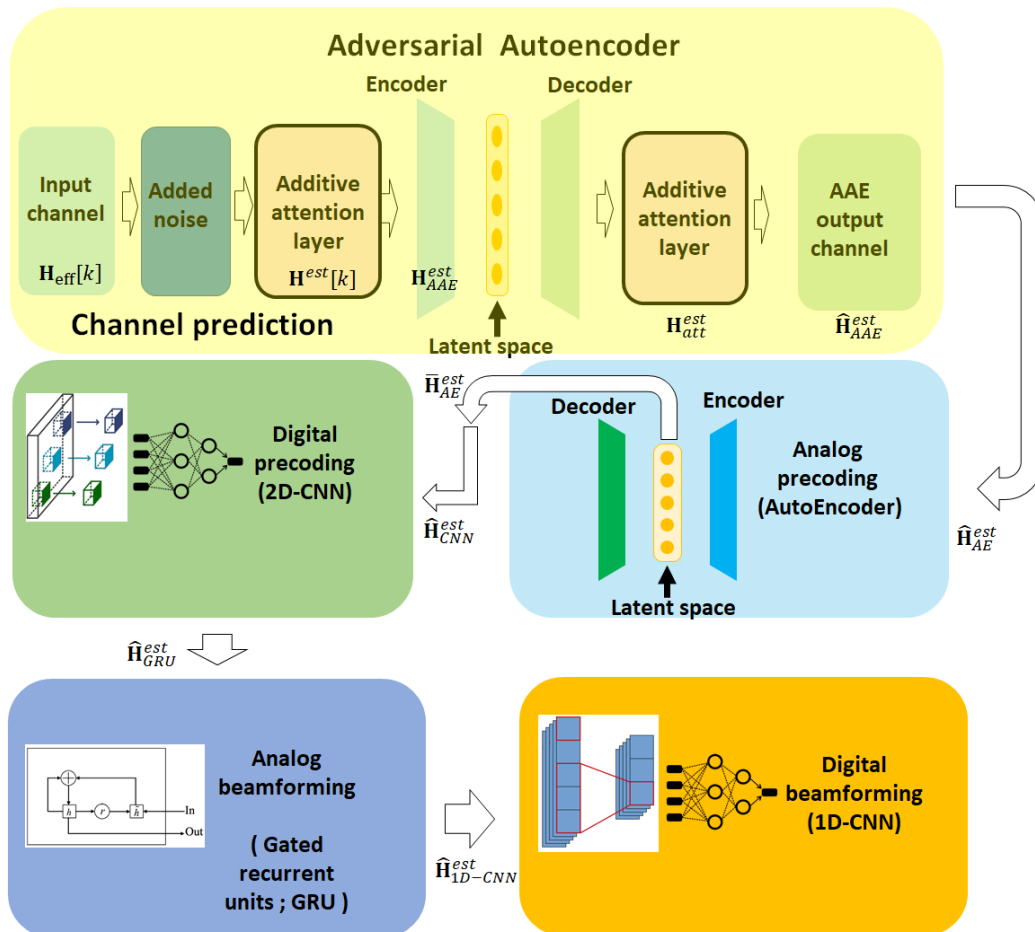
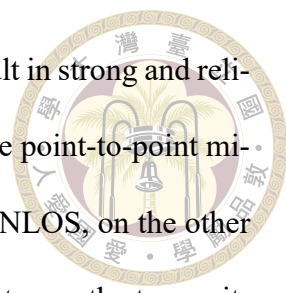


Figure 3.3: Illustration of the structure of the proposed additive attention-aided adversarial autoencoder beamforming design.

3.3 LOS and NLOS

In the practical network deployment of mobile communication, LOS and NLOS are terms commonly used in the context of wireless communication and radio propagation to describe whether there is a direct, unobstructed path between a transmitter and a receiver. LOS conditions are often characterized by minimal signal attenuation and are desirable

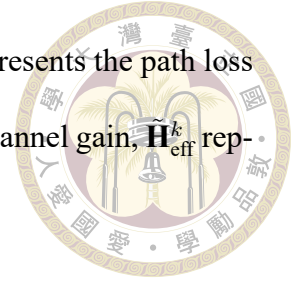


for many wireless communication systems because they typically result in strong and reliable signal reception. LOS is commonly encountered in scenarios like point-to-point microwave links, satellite communication, and optical communication. NLOS, on the other hand, refers to scenarios where there is no clear, unobstructed path between the transmitter and the receiver. In NLOS conditions, obstacles, such as buildings, terrain features, trees, or other structures, obstruct or scatter the signal path. NLOS conditions often lead to signal reflection, diffraction, and multiple signal paths (multipath propagation) due to the scattering of signals off surfaces. NLOS conditions can result in signal attenuation, signal fading, and increased susceptibility to interference, making wireless communication more challenging. NLOS conditions are commonly encountered in urban environments, indoor wireless communication, and scenarios where the direct line of sight is blocked. Understanding whether a wireless communication link is LOS or NLOS is crucial for designing and optimizing wireless networks and selecting appropriate technologies. In NLOS scenarios, additional techniques like beamforming, signal processing, and the use of reflectors (e.g., IRS) may be employed to improve signal quality and reliability. When LOS condition exists, the wireless signals are transmitted in a straight line between the transmitter and receiver without any obstruction, the channel coefficients are composed of LOS and NLOS, the NLOS is still modeled as a complex Gaussian distribution, and the LOS is modeled as a channel gain which can be modeled as

$$\mathbf{G}[k] = \sqrt{\kappa d_{B,m}^{-\alpha}[k]} \sqrt{\frac{\hat{R}}{1 + \hat{R}}} \mathbf{h}_{B,m}^{\text{LoS}}[k], \quad (3.6)$$

$$\mathbf{T}[k] = \sqrt{\kappa d_{m,k}^{-\alpha}[k]} \sqrt{\frac{\hat{R}}{1 + \hat{R}}} \mathbf{h}_{m,k}^{\text{LoS}}[k] + \sqrt{\frac{1}{1 + \hat{R}}} \mathbf{h}_{m,k}^{\text{NLoS}}[k], \quad (3.7)$$

where d represents the distance, \hat{R} represents the Rician factor, α represents the path loss exponent, $\hat{\mathbf{H}}_{\text{eff}}[k]$ represents the NLOS component, κ represents the channel gain, $\hat{\mathbf{H}}_{\text{eff}}^k$ represents the LOS component.

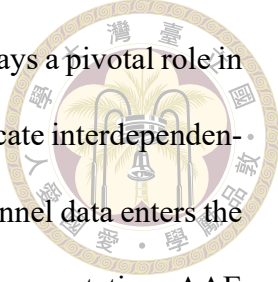






Chapter 4 METHOD

Our proposed AAE-AATT-Beamforming model is a comprehensive system consisting of several key components: additive attention, AAE (Adversarial Autoencoder), analog precoding module, digital precoding module, analog beamforming module, and digital beamforming module. This system operates as follows: The channel signal originates at the base station and is transmitted to the IRS, and then reflect to the user equipment (UE). The UE acts as the receiver, receiving the transmitted signal, and it serves as the input channel for further processing in Fig. 4.1. A central processing center gathers power measurements from BS and employs them for two primary purposes: firstly, to determine which BS is hosting the UE and roughly estimate its distance. Once the hosting cell is identified, the cell is further divided into three-sector sites, each offering 120-degree coverage. Secondly, these power measurements, primarily from the downlink signal, assist in identifying the sector to which the UE belongs. This Sector ID identification significantly reduces the search space for the Angle of Arrival (AoA), allowing the base station to efficiently utilize beamforming techniques. By steering the array's beam in the specific direction corresponding to the UE's sector, the base station optimizes the achievable data rate performance for that UE. The input channel data first passes through a additive attention layer. Additive attention is capable of modeling global relationships between elements in the input sequence. It doesn't make strong assumptions about the type of relationships that



exist, allowing it to capture a wide range of dependencies. This layer plays a pivotal role in capturing important channel features, particularly focusing on the intricate interdependencies in the channel data. Following the additive attention layer, the channel data enters the AAE model. Here, an encoder maps the input data to a latent space representation. AAE can generate realistic data samples from a learned latent space. This makes them useful for tasks like signal processing, and generating new data points that are similar to the training data. to optimize the model parameters. The learned latent space is instrumental in capturing crucial features for the subsequent beamforming processes. The analog precoding module employs an autoencoder to minimize the latent space and reconstruct it as closely as possible back to the original input channel signal data. This process entails the transformation of bottleneck features, leading to a substantial enhancement in the accuracy of capturing global domain features. The digital precoding module utilizes 2D-CNN equipped with local neuron connectivity, translation invariance, and weight sharing. This configuration aids in capturing frequency and spatial features essential for channel precoding while minimizing interference. Analog beamforming module use Gated Recurrent Unit (GRU) that have two gating mechanisms, namely the reset gate and the update gate, serve as crucial elements for governing the information flow within the cell. These gates provide the network with the capability to determine which information from the previous time step to forget, and which to update and retain. The analog beamforming module employs an GRU to helps in modeling complex temporal dependencies effectively. The update gate in GRU allows the model to decide how much should be updated with new information. This helps mitigate the vanishing gradient problem by preventing gradients from becoming too small during backpropagation, significantly enhancing analog beamforming module feature capture accuracy. The digital beamforming module utilizes 1D-CNN equipped with

local neuron connectivity. 1D-CNN excel at capturing local patterns and dependencies within sequential data. They can automatically learn features such as time series analysis. This configuration aids in capturing frequency and spatial features essential for channel beamforming while minimizing interference. By knowing which sector the UE belongs to, the base station can employ beamforming techniques to steer the antenna array's beam in the precise direction of the UE. This optimization enhances the achievable data rate and performance for the UE, providing a more reliable and efficient wireless communication experience.

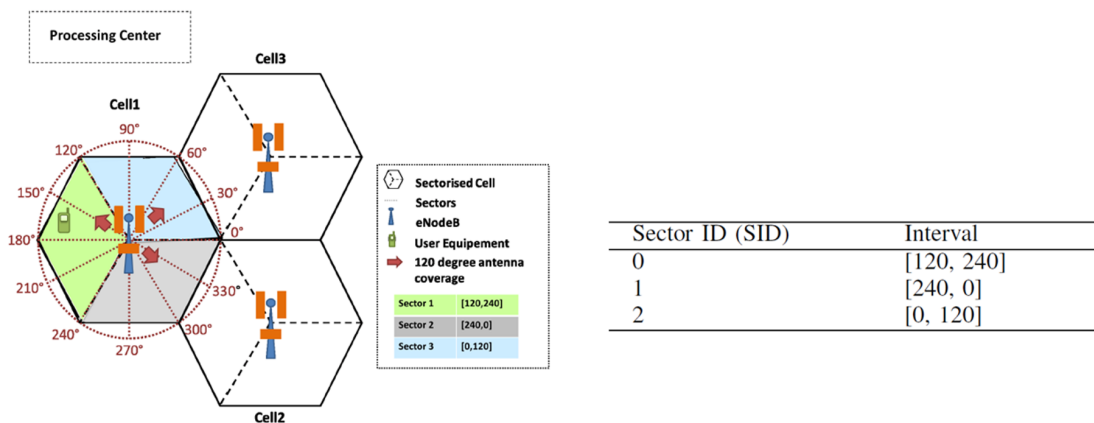
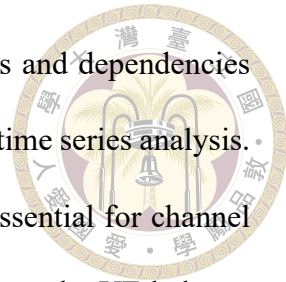


Figure 4.1: Scenario of the proposed beamforming design.

Through this hybrid beamforming process in the proposed AAE-AATT-Beamforming model, we aim to identify shared relationships and informative patterns within wireless channel data. The model incorporates additive attention and AAE techniques to learn these valuable patterns, capturing spatial, temporal, and structural information. Consequently, the model can efficiently optimize precoders, beamformers, and combiners to maximize the achievable data rate within the constraints of the hardware. The primary goal behind IRS beamforming design is to optimize the achievable rate R_a at the user over the IRS

channel, R_a is given by

$$R_a = \frac{1}{K} \sum_{k=1}^K \log_2 \left(1 + \frac{P}{\sigma_n^2} \|\mathbf{F}_{\text{BB}}[k] \mathbf{F}_{\text{BB}}^{\text{H}}[k] \mathbf{B}_{\text{BB}}[k] \times$$



$$\mathbf{B}_{\text{BB}}^{\text{H}}[k] \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{RF}}^{\text{H}} \mathbf{B}_{\text{RF}} \mathbf{B}_{\text{RF}}^{\text{H}} \mathbf{H}_{\text{eff}}[k] \mathbf{H}_{\text{eff}}^{\text{H}}[k] \mathbf{\Lambda}[k]^{-1} \|\right), \quad (4.1)$$

where $\mathbf{\Lambda}[k] = \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{RF}}^{\text{H}} \mathbf{W}_{\text{BB}}[k] \mathbf{W}_{\text{BB}}^{\text{H}}[k] \in \mathbb{C}^{N_s \times N_s}$ corresponds to the noise term in (4.1).

The design of hybrid beamforming can be expressed as an optimization problem given below

$$\text{P1: maximize } \sum_{k=1}^K \log_2(1 + \text{SINR}_k) \quad (4.2a)$$

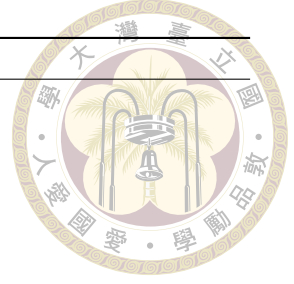
$$\text{s.t. } \|\mathbf{B}_{\text{RF}} \mathbf{B}_{\text{BB}}[\mathbf{k}]\|_{\mathcal{B}}^2 \leq P_{\text{max}}(\text{SID} : 1, 2, 3) \quad (4.2b)$$

$$|\Theta_m| = 1, \forall m = 1, 2, \dots, M. \quad (4.2c)$$

where \mathcal{B}_{RF} and \mathcal{W}_{RF} are the sets of feasible beamformers and combiners that follow the restriction of having unit-modulus constraints. Our main goal is to build a more accurate and efficient beamforming design model by extracting and obtaining multi-dimensional relationships and features.

4.1 Additive Attention Mechanism

The model's primary operations commence with the adept utilization of an additive attention mechanism, strategically employed to efficiently condense the intricate input query sequence into a comprehensive global query vector. This critical initial step sets the

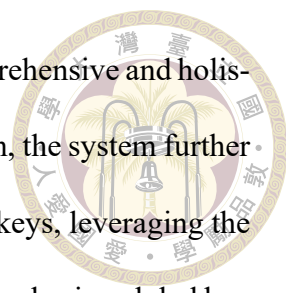


Algorithm 1 Proposed AAE-AATT-Beamforming model

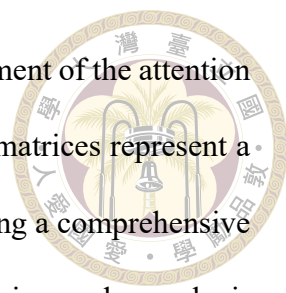
Input: $x, \mathbf{G}[k], \mathbf{T}[k], \mathbf{H}_{\text{eff}}[k], \Theta$.

- 1: Determines which specific BS is hosting the UE.
 - 2: Provides an estimate of the user's distance from BS.
 - 3: Cell is further subdivided into three-sector sites.
 - 4: Each sector offering 120-degree coverage.
 - 5: Identifying the Sector ID.
 - 6: Initialize the channel complex value input $\mathbf{H}_{\text{eff}}[k]$.
 - 7: Convert the $\mathbf{H}_{\text{eff}}[k]$ into a real and imaginary value of $\mathbf{H}^{\text{est}}[k]$.
 - 8: **Repeat :**
 - 9: **Channel prediction :**
 - 10: Calculate the proper query \mathbf{Q} , key \mathbf{K} , and value \mathbf{V} .
 - 11: Obtain the output $\mathbf{H}_{\text{att}}^{\text{est}}[k]$ of the additive attention by using (4.4), (4.6).
 - 12: Compute AAE encoder and decoder parameters.
 - 13: Calculate the $\hat{\mathbf{H}}_{\text{AAE}}^{\text{est}}$ of AAE is then constructed from $\mathbf{H}_{\text{AAE}}^{\text{est}}$ by using (4.10), (4.11).
 - 14: Attain the best channel prediction matrix $\mathbf{H}_{\text{att}}^{\text{est}}[k]$ by using (4.10).
 - 15: **Analog precoding :**
 - 16: Input the AAE $\mathbf{H}_{\text{AAE}}^{\text{est}}$ into the $\hat{\mathbf{H}}_{\text{AE}}^{\text{est}}$ by the encoder.
 - 17: Iteratively update the weight \mathbf{W}, \mathbf{W}' and bias values \mathbf{b}, \mathbf{b}' by using (4.14), (4.15).
 - 18: Calculate the AE difference between $\hat{\mathbf{H}}_{\text{AE}}^{\text{est}}$ and $\tilde{\mathbf{H}}_{\text{AE}}^{\text{est}}$ by using (4.15).
 - 19: Obtain the optimal $\tilde{\mathbf{H}}_{\text{AE}}^{\text{est}}$ by using (4.15).
 - 20: Retrieve the best analog precoding matrix \mathbf{F}_{RF} by using (4.15).
 - 21: **Digital precoding :**
 - 22: Input the AE's latent space output $\bar{\mathbf{H}}_{\text{AE}}^{\text{est}}$ into 2D-CNN by using (4.15).
 - 23: Calculate the 2D-CNN difference between $\tilde{\mathbf{H}}_{\text{2D-CNN}}^{\text{est}}$ and the $\hat{\mathbf{H}}_{\text{2D-CNN}}^{\text{est}}$ until convergence. Seize the best digital precoding matrix \mathbf{F}_{BB} .
 - 24: **Analog beamforming :**
 - 25: Input the 2D-CNN output channel signals $\bar{\mathbf{H}}_{\text{2D-CNN}}^{\text{est}}$ into the GRU by using (4.20).
 - 26: Calculate the GRU difference between $\tilde{\mathbf{H}}_{\text{GRU}}^{\text{est}}$ and $\hat{\mathbf{H}}_{\text{GRU}}^{\text{est}}$ until convergence.
 - 27: **Digital beamforming :**
 - 28: Input the GRU output channel signals $\bar{\mathbf{H}}_{\text{GRU}}^{\text{est}}$ into the 1D-CNN by using (4.22).
 - 29: Calculate the 1D-CNN difference between the $\tilde{\mathbf{H}}_{\text{1D-CNN}}^{\text{est}}$ and the $\hat{\mathbf{H}}_{\text{1D-CNN}}^{\text{est}}$ until convergence by using (4.19), (4.20), (4.21), (4.22), (4.23).
 - 30: BS apply beamforming techniques to steer the array's beam.
- Output:** $y[k], \mathbf{F}_{\text{BB}}[k], \mathbf{F}_{\text{RF}}, \mathbf{B}_{\text{RF}}, \mathbf{B}_{\text{BB}}[k], \mathbf{W}_{\text{RF}},$ and $\mathbf{W}_{\text{BB}}[k]$
-

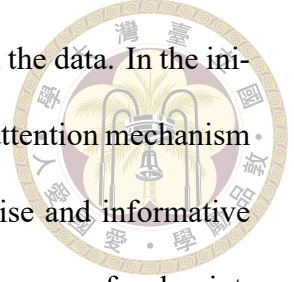
foundation for the subsequent intricate interactions that follow, facilitating a sophisticated analysis of the underlying data. With the groundwork laid by the global query vector, the system seamlessly transitions into capturing the intricate and nuanced interplay between the global query vector and attention keys. Through a meticulously orchestrated process involving element-wise multiplication, the system successfully encapsulates the intricate



relationships and dependencies within the data, contributing to a comprehensive and holistic understanding of the input sequence. Building upon this foundation, the system further advances its analytical capabilities by meticulously summarizing the keys, leveraging the powerful additive attention mechanism to derive a comprehensive and cohesive global key vector. This critical step marks a significant milestone in the process, allowing the system to holistically comprehend the intricate network of dependencies and interactions embedded within the data. Following the comprehensive analysis of the global key vector, the system seamlessly progresses into an advanced stage where it further explores the multifaceted relationships between the global key and attention values. Through a meticulous interplay of element-wise multiplication and a carefully tailored linear transformation, the system adeptly cultivates a comprehensive set of global context-aware attention values, each imbued with a nuanced understanding of the intricate contextual nuances embedded within the data. Finally, armed with this extensive reservoir of context-aware values, the system synthesizes a final output by carefully integrating these values with the attention query, allowing for a comprehensive and nuanced understanding of the underlying data. This intricate process, designed to streamline computational complexity while effectively capturing the nuanced contextual intricacies within the input sequence, represents a significant breakthrough in the field of data analysis and processing. Furthermore, the initial phase of our comprehensive model involves a meticulously designed and orchestrated conversion of the input embedding matrix into a sophisticated set of query, key, and value sequences, establishing a strong and robust foundation for the subsequent stages of analysis and comprehension. Their meticulous orchestration and arrangement within the broader framework of the input matrix play a pivotal role in shaping the subsequent stages of data processing and analysis. As the system progresses through its intricate operations, the



data undergoes a transformative process, culminating in the establishment of the attention query, key, and value matrices, denoted as $\mathbf{Q}, \mathbf{K}, \mathbf{V} \in \mathbb{R}^{N \times d}$. These matrices represent a significant advancement in the system's analytical capabilities, offering a comprehensive and nuanced understanding of the data's underlying nuances, intricacies, and complexities. Their meticulous construction and careful arrangement of elements within the broader framework of the data enable the system to delve deeper into the data's intricate relationships, dependencies, and patterns, laying the groundwork for more sophisticated analyses and insights. Diving into the specific characteristics of these matrices, we encounter the comprehensive composition of the attention query matrix \mathbf{Q} , meticulously crafted to encapsulate the critical elements and aspects of the data that require further analysis and examination. Complementing this, the key matrix \mathbf{K} plays a pivotal role in shaping the system's understanding of the data's underlying relationships and dependencies, contributing to a comprehensive and nuanced understanding of the data's intricacies. Simultaneously, the value matrix \mathbf{V} offers a comprehensive and nuanced perspective, encapsulating the critical elements and nuances embedded within the data, facilitating a comprehensive and holistic understanding of the data's underlying patterns, trends, and dependencies. Collectively, the attention query, key, and value matrices play a critical role in shaping the system's understanding of the data, laying the groundwork for more sophisticated analyses and insights in the subsequent stages of the data processing pipeline. Embarking on an intricate journey through the intricate mechanisms of data processing, we encounter the concept of additive attention—an instrumental mechanism renowned for its unparalleled ability to efficiently summarize critical information within a sequence, all while maintaining linear computational complexity. This pivotal mechanism serves as a cornerstone for the subsequent stages of data analysis, offering a comprehensive and nuanced perspective

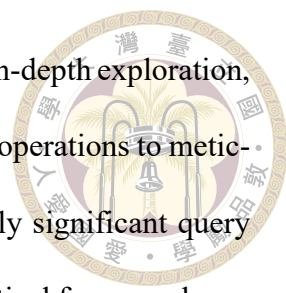


on the underlying patterns, trends, and relationships embedded within the data. In the initial phases of data processing, the system deftly employs the additive attention mechanism to condense the complex and multifaceted query matrix into a concise and informative global query vector, denoted as $\mathbf{q} \in \mathbb{R}^d$. This global query vector serves as a focal point, encapsulating the comprehensive contextual information embedded within the broader attention query. By carefully summarizing the critical components of the attention query, the global query vector lays the groundwork for more sophisticated analyses and insights, enabling the system to glean valuable and actionable information from the intricate data at hand. As we delve deeper into the intricacies of the system's operations, we uncover the intricate computational processes that underlie the determination of the attention weight α_i for the i -th query vector. This intricate computation involves a meticulous analysis of the contextual relevance and significance of each component within the attention query, culminating in the determination of the attention weight —a critical metric that signifies the relative importance and relevance of each component within the broader context of the data. This process of determining the attention weight serves as a pivotal step in the system's quest to distill critical insights and glean valuable information from the complex and multifaceted data. And, the attention weight α_i can be written as:

$$\alpha_i = \frac{\exp(\mathbf{w}_q^T \mathbf{q}_i / \sqrt{d})}{\sum_{j=1}^N \exp(\mathbf{w}_q^T \mathbf{q}_j / \sqrt{d})}, \quad (4.3)$$

where N denote the sequence length, $\mathbf{w}_q \in \mathbb{R}^d$ is a parameter vector and d denote the hidden dimension. The global attention query can be written as:

$$\mathbf{q} = \sum_{i=1}^N \alpha_i \mathbf{q}_i. \quad (4.4)$$

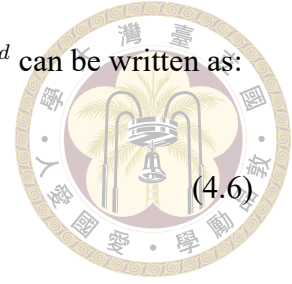


Subsequently, our analytical journey propels us into the realm of in-depth exploration, where we systematically leverage the power of element-wise product operations to meticulously dissect and model the intricate interplay between the globally significant query vector and individual key vectors. Harnessing the power of this analytical framework, we embark on a meticulous exploration of the data, carefully applying mathematical principles and methodologies to scrutinize the interplay between the global query vector and individual key vectors. This exploration culminates in the formation of a global context-aware key matrix, a critical artifact that encapsulates and synthesizes the dynamic and multifaceted relationships embedded within the data. Denoted as $\mathbf{p}_i = \mathbf{q}\mathbf{k}_i$, this matrix serves as a cornerstone artifact, offering profound insights into the data's intricate nuances and facilitating a more comprehensive and nuanced understanding of the complex relationships and dependencies that underlie the data. Moreover, to streamline and optimize our analytical process, we judiciously leverage the efficiency and computational advantages afforded by the additive attention mechanism. This mechanism plays a crucial role in distilling and summarizing the key insights embedded within the global context-aware key matrix, offering a concise and comprehensive overview of the data's overarching patterns and relationships. The process of deriving the additive attention weight for the i -th vector involves a meticulous and calibrated calculation, characterized by a profound understanding of the data's underlying patterns and dependencies. This intricate computation culminates in the determination of the additive attention weight, serving as a critical artifact that underpins the data's overarching narrative and offers profound insights into its intricate tapestry. The additive attention weight β_i can be written as:

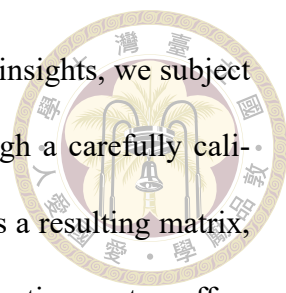
$$\beta_i = \frac{\exp(\mathbf{w}_k^T \mathbf{p}_i / \sqrt{d})}{\sum_{j=1}^N \exp(\mathbf{w}_k^T \mathbf{p}_j / \sqrt{d})}, \quad (4.5)$$

where $\mathbf{w}_k \in \mathbb{R}^d$ is the parameter vector. The global key vector $\mathbf{k} \in \mathbb{R}^d$ can be written as:

$$\mathbf{k} = \sum_{i=1}^N \beta_i \mathbf{p}_i. \quad (4.6)$$



In the final phase of our data processing journey, we delve into the intricate nuances of modeling the dynamic and multifaceted interaction between the attention value matrix and the globally significant key vector—an endeavor that significantly bolsters and amplifies our context modeling efforts, enabling us to gain a comprehensive dependencies embedded within the data. To facilitate a meticulous exploration of this dynamic interaction, we adopt a strategic and methodical approach, mirroring the principles and methodologies employed during the query-key interaction modeling phase. Leveraging the profound insights and valuable lessons gleaned from our previous endeavors, we embark on a journey through the complex web of relationships and dependencies that underscore the data, unveiling critical patterns and trends that underlie the data's intricate tapestry. As we traverse this intricate web, we meticulously engage in a series of element-wise product operations, meticulously applying mathematical principles and methodologies to dissect and analyze the complex interplay between the global key vector and individual value vectors. Through this meticulous analysis, we uncover a wealth of information and insights, each contributing to a more comprehensive and nuanced understanding of the data's underlying patterns and relationships. The culmination of this meticulous exploration yields the key-value interaction vector, denoted as $\mathbf{u}_i = \mathbf{k}\mathbf{v}_i$, a critical artifact that encapsulates the dynamic and multifaceted relationship between the globally significant key vector and the individual value vectors. This interaction vector serves as a key artifact, offering profound insights into the data's intricate nuances and facilitating a more comprehensive and nuanced understanding of the complex relationships and dependencies that underlie the data.



Moreover, to further enhance our understanding and uncover deeper insights, we subject each key-value interaction vector to a transformative journey through a carefully calibrated linear transformation layer. This transformative journey yields a resulting matrix, which encapsulates the hidden representation of each key-value interaction vector, offering profound insights and nuanced perspectives that contribute to a more comprehensive and nuanced understanding of the intricate relationships embedded within the data.

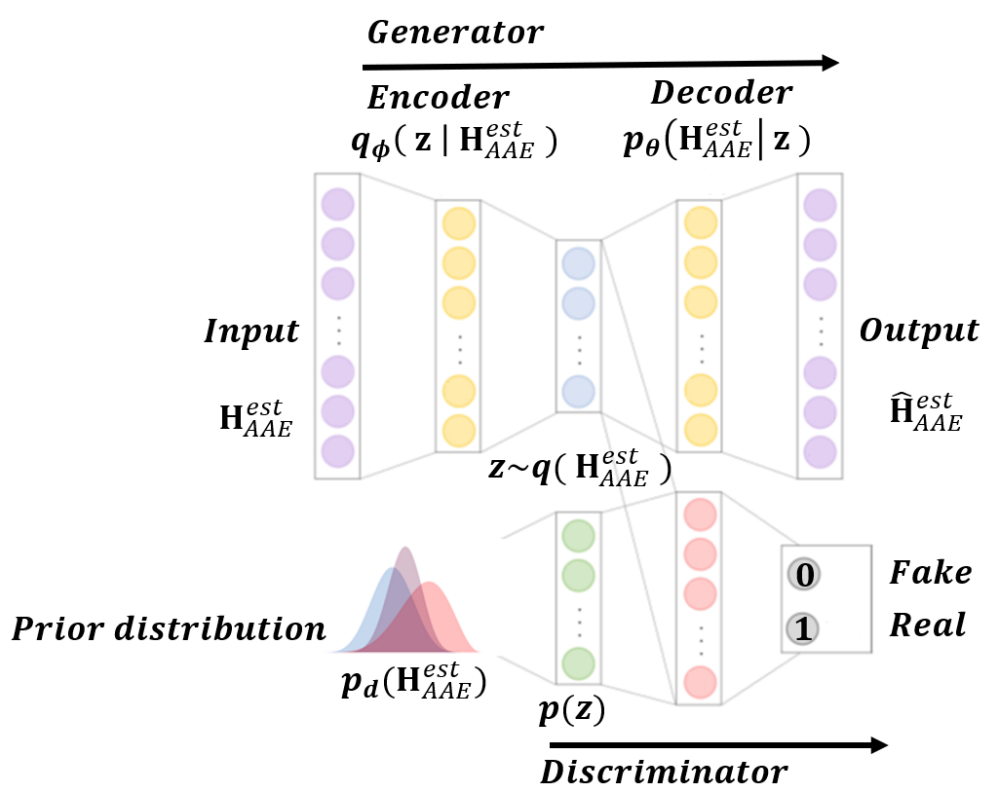
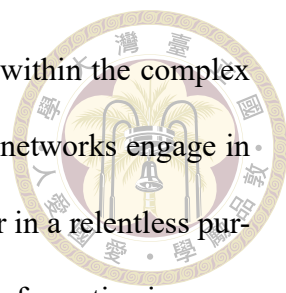


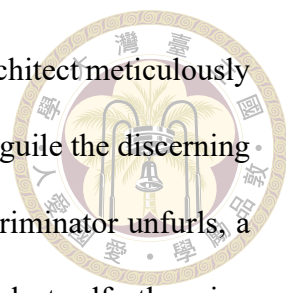
Figure 4.2: The architecture of the proposed adversarial autoencoder design.

4.2 Adversarial Autoencoder

Within the intricate framework of adversarial autoencoders, the fundamental challenge of data generation is elegantly reframed as an intricate, high-stakes minimax adversarial game. Central to this conceptualization are two critical players: the discriminator



D and the generator G , each wielding unique capabilities and roles within the complex framework of data generation and synthesis. Together, these pivotal networks engage in a dynamic and nuanced dance, each striving to outmaneuver the other in a relentless pursuit of data generation excellence and fidelity. Embarking on this transformative journey, the generator network assumes the mantle of an innovative architect, drawing upon a sophisticated function $G(\mathbf{z})$ to seamlessly bridge the gap between input samples $\mathbf{H}_{AAE}^{\text{est}}$ and a carefully curated ensemble of samples sourced from a carefully calibrated prior distribution $p(\mathbf{z})$. In doing so, the generator deftly orchestrates a delicate symphony, leveraging its strategic understanding of data structures and distributions to craft an intricate and nuanced tapestry of data points that mirror and echo the complex dynamics embedded within the original data space. In stark contrast, the discriminator $D(\mathbf{z})$ assumes a more discerning role, adroitly sifting through a diverse array of data samples to discern the subtle, yet crucial, nuances that demarcate the authentic data distribution from the meticulously crafted ensemble generated by our ingenious generative model. Armed with a sophisticated understanding of the data's underlying intricacies and patterns, the discriminator brings to bear its discerning gaze, meticulously scrutinizing each data point and discerning the minute markers and distinctive traits that set the actual data distribution apart from the simulated ensemble of data samples meticulously constructed by the generator. In this high-stakes, cutthroat adversarial game, the two networks engage in a relentless duel of wits and strategies, each seeking to outmaneuver and outsmart the other in a relentless pursuit of data generation excellence and fidelity. As the game unfolds, the generator and discriminator deftly navigate the complex and multidimensional data landscape, each striving to assert its dominance and prowess in the pursuit of data generation perfection. In our AAE model, the overarching objective of the enigmatic function $G(\mathbf{z})$



assumes a crucial and pivotal role, serving as a cunning and shrewd architect meticulously orchestrating a symphony of data points designed to confound and beguile the discerning discriminator. As this intricate interplay between generator and discriminator unfurls, a mesmerizing adversarial game of epic proportions ensues, each network steadfastly vying for the upper hand and seeking to assert its dominance in the realm of data generation and synthesis. Central to the foundation of our AAE model lie the twin pillars of the encoder and decoder network, intricately woven together to form the backbone of our sophisticated autoencoder framework. Within this meticulously crafted structure, the encoder operates as a perceptive and discerning observer, deftly extracting the subtle nuances and underlying patterns embedded within the data fabric and distilling them into a refined and compact representation. On the other hand, the decoder assumes the role of a skilled artisan, adeptly leveraging its strategic understanding of the latent space to reconstruct and reimagine the data points in a faithful and accurate manner, meticulously breathing life into the latent representations and weaving them into a coherent and immersive narrative that closely mirrors the original data distribution. In a masterful stroke of innovation and ingenuity, we seamlessly integrate the output \mathbf{H}_{att}^{est} from the additive attention module into the input \mathbf{H}_{AAE}^{est} of the AAE module, forging a seamless and cohesive bridge between the two pivotal components. This strategic integration ensures a harmonious synergy between the various elements of our model, facilitating a smooth and fluid transition of data points across the intricate and dynamic landscape of the autoencoder framework. As a result, the AAE model emerges as a powerful and versatile tool, capable of generating data samples that seamlessly mimic and reflect the complex dynamics and underlying patterns embedded within the original data distribution. For the input \mathbf{H}_{AAE}^{est} , and the encoder can be written

as



$$\mathbf{C} = F_{enc}(\mathbf{H}_{AAE}^{est}), \quad (4.7)$$

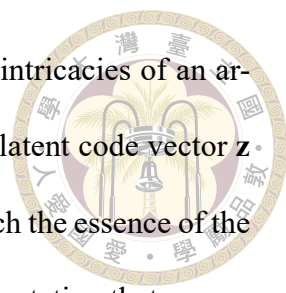
where \mathbf{C} is the codeword. Then, the decoder transforms the code back into the raw input data, and the decoder of a autoencoder can be written as

$$\hat{\mathbf{H}}_{AAE}^{est} = F_{dec}(\mathbf{C}). \quad (4.8)$$

Further analysis of the solutions to the min-max adversarial game reveals insights into the dynamics of the training process, shedding light on the convergence properties of GAN and their stability during training is represented as:

$$\min_G \max_D E_{\mathbf{H}_{AAE}^{est} \sim p_d} [\log D(\mathbf{H}_{AAE}^{est})] + E_{\mathbf{z} \sim p(\mathbf{z})} [\log(1 - D(G(\mathbf{z})))] \quad (4.9)$$

Our AAE network unfolds as a captivating tale of interwoven complexities, where the very essence of an autoencoder undergoes a transformative metamorphosis, emerging as a powerful and dynamic generative model that transcends the boundaries of traditional data synthesis and reconstruction. At the heart of this metamorphosis lies the ingenious integration of an adversarial process seamlessly interwoven into the intricate fabric of the autoencoder framework, breathing life into a novel and unprecedented paradigm of data generation and representation. Through the dexterous and artful orchestration of a carefully orchestrated interplay between the traditional reconstruction error criterion and an adversarial training criterion inspired by the groundbreaking concepts of GAN, the AAE model becomes imbued with an unparalleled capacity to not only faithfully reconstruct the input data but also to imbue it with a unique and compelling perspective, seamlessly



fusing the essence of the original data distribution with the nuanced intricacies of an arbitrary prior distribution. Within this transformative framework, the latent code vector \mathbf{z} assumes a pivotal and central role, serving as the conduit through which the essence of the input data is distilled and channeled into a refined and distilled representation that encapsulates the very essence and intricacies of the underlying data distribution. Additionally, the encoder $q(\mathbf{z}|\mathbf{H}_{AAE}^{est})$, traditionally entrusted with the solemn duty of encoding input data within the confines of an autoencoder framework, now assumes a dual and multifaceted role, deftly navigating the intricate interplay between data representation and generation within the broader context of the adversarial framework. It is within this intricate interplay and seamless fusion of diverse and multifaceted elements that the AAE model finds its true essence and purpose, transcending the boundaries of traditional data synthesis and representation, and ushering in a new era of data generation and reconstruction that seamlessly marries the nuances of the original data distribution with the boundless creativity and imagination of the generative model. And, the latent code vector \mathbf{z} assumes a pivotal role as it serves as the intermediary representation that encapsulates the essential characteristics of the input \mathbf{H}_{AAE}^{est} . This latent space representation acts as the bridge between the input data domain and the generative model's output, enabling the AAE to effectively reconstruct the input data while maintaining the underlying data distribution $p_d(\mathbf{H}_{AAE}^{est})$. Furthermore, the encoder $q(\mathbf{z}|\mathbf{H}_{AAE}^{est})$, which typically plays the role of encoding input data in an AE framework, takes on a dual role here. The encoder's critical role within the AAE is reflected in its ability to facilitate the establishment of the aggregate posterior distribution $q(\mathbf{z})$, which represents the synthesized latent code vectors corresponding to the input data distribution. This aggregate posterior distribution embodies the amalgamation of the encoded latent space representations and serves as the basis for guiding the genera-

tive process toward the creation of synthetic data samples that align with the desired data distribution properties can be written as:

$$q(\mathbf{z}) = \int_{\mathbf{H}_{AAE}^{est}} q(\mathbf{z}|\mathbf{H}_{AAE}^{est})p_d(\mathbf{H}_{AAE}^{est})d(\mathbf{H}_{AAE}^{est}). \quad (4.10)$$

Within the intricate framework of the AAE, the prior distribution $p(\mathbf{z})$ assumes a pivotal role as a guiding principle for the latent code vector \mathbf{z} . This prior distribution serves as a critical anchor, imparting structure and coherence to the generative process by delineating the overarching characteristics that the synthesized latent code vectors should embody. By enforcing this prior distribution on the latent space, the AAE framework ensures that the generated data samples maintain fidelity to the predefined statistical properties and structural nuances, thereby facilitating the creation of synthetic data that accurately represents the underlying data distribution. The role of the encoder function $q(\mathbf{z}|\mathbf{H}_{AAE}^{est})$ within the generator G is multifaceted and multifunctional. Acting as the cornerstone of the generative model, the encoder function assumes the critical responsibility of extracting and condensing the essential information from the input data, thereby enabling the accurate representation and synthesis of the latent code vectors. Through its intricate encoding process, the encoder function captures the salient features and key attributes of the input data, facilitating the translation of this information into the latent space representation. By establishing a robust and coherent connection between the input data distribution and the synthesized latent code vectors, the encoder function ensures that the generative process remains aligned with the desired data distribution properties, thereby enabling the AAE to produce synthetic data samples that faithfully mirror the statistical and structural essence of the original data. The dynamic interplay between the prior distribution, the latent code vector, and the encoder function within the AAE framework underscores the intricate na-



Table 4.1: COMPUTATIONAL COMPLEXITY OF THE DIFFERENT RESEARCH METHODS

Algorithm	Computational complexity
Proposed AAE-AATT-Beamforming	$\mathcal{O}(4d_i(N_t N_r(4(N_t N_r + 1)(4(N_t N_r \cdot 4(N_t N_r))) + (d_i((N_t N_r)N_t N_r))N_t N_r \sum_{\epsilon=1}^{L_c} N_{\epsilon-1} N_{\epsilon}))$
DAE-CNN-ATT-Beamforming	$\mathcal{O}(4d_i^2(N_t N_r(4(N_t N_r + 1)(4(N_t N_r \cdot 4(N_t N_r))) + (d_i^2(4N_t N_r)N_t N_r) \sum_{\epsilon=1}^{L_c} N_{\epsilon-1} N_{\epsilon}))$
DAE-CNN-Beamforming	$\mathcal{O}(4(N_t N_r(4(N_t N_r + 1)(4(N_t N_r \cdot 4(N_t N_r))) + 4(N_t N_r) + N_t N_r \sum_{\epsilon=1}^{L_c} N_{\epsilon-1} N_{\epsilon}))$
DAE-Beamforming	$\mathcal{O}(((4N_t N_r)N_t N_r((4N_t N_r + 1)((4N_t N_r \cdot (4N_t N_r)))$
LSTM-DAE-Beamforming	$\mathcal{O}(4(N_t N_r(4(N_t N_r + 1)(4(N_t N_r \cdot 4(N_t N_r))) + 4(N_t N_r) + (T d_h^2(N_t N_r) + T d_h d_i(N_t N_r)))$

T is sequence length, $N_{\epsilon-1}$ represent the numbers of input feature of the ϵ th layer, N_{ϵ} represent the numbers of output feature of the ϵ th layer, d_i is input dimensionality, L_c is number of convolutional layers, and d_h is hidden state.

ture of the generative process, highlighting the significance of maintaining fidelity to the underlying data distribution properties while synthesizing coherent and realistic data samples. Hence, the intricate optimization process of the encoder revolves around the careful calibration and fine-tuning of the latent code vector \mathbf{z} to align with the distribution $q(\mathbf{z})$, strategically confounding the discriminating capabilities of the discriminator D . Through this deliberate and nuanced interplay, the encoder endeavors to synthesize latent code vectors that strategically blur the distinction between the desired distribution $q(\mathbf{z})$ and the expected distribution $p(\mathbf{z})$ enforced by the discriminator D . This adversarial interplay forms the crux of the AAE's optimization process, ultimately manifesting as a min-max game of optimization objectives can be written as:

$$\min_G \max_D \mathbb{E}_{\mathbf{z} \sim q(\mathbf{z} | \mathbf{H}_{\text{AAE}}^{\text{est}})} [\log D(\mathbf{z})] + \mathbb{E}_{\mathbf{z}_{\text{prior}} \sim p(\mathbf{z}_{\text{prior}})} [\log(1 - D(G(\mathbf{z})))] \quad (4.11)$$



4.3 Analog precoding

The analog precoding in our previous study used a general neural network model[62] and successive convex approximation (SCA)[28] that is more difficult to capture whether the features reconstruct the original signals. Therefore, we use autoencoder to reconstruct the latent space back to the input channel signal data as much as possible, and extract from bottleneck features, so that our analog precoding module can improve the accuracy of capturing the features. We can use this upper bound to optimize by using the following Jensen's inequality, and the achievable rate is represented by

$$\begin{aligned}
R_a &\stackrel{(a)}{\leq} \frac{1}{K} \sum_{k=1}^K \log_2 \left(\left| 1 + \frac{P}{\sigma_n^2} \mathbf{F}_{\text{BB}}[k] \mathbf{F}_{\text{BB}}^H[k] \mathbf{B}_{\text{BB}}[k] \mathbf{B}_{\text{BB}}^H[k] \times \right. \right. \\
&\quad \left. \left. \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{RF}}^H \mathbf{B}_{\text{RF}} \mathbf{B}_{\text{RF}}^H \mathbf{H}_{\text{eff}}[k] \mathbf{H}_{\text{eff}}^H[k] \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{BB}}[k] \mathbf{W}_{\text{BB}}^H[k] \right| \right) \\
&\stackrel{(b)}{=} \frac{1}{K} \sum_{k=1}^K \log_2 \left(\left| 1 + \frac{PN}{\sigma_n^2} \mathbf{F}_{\text{BB}}[k] \mathbf{F}_{\text{BB}}^H[k] \mathbf{B}_{\text{BB}}[k] \mathbf{B}_{\text{BB}}^H[k] \times \right. \right. \\
&\quad \left. \left. \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{RF}}^H \mathbf{B}_{\text{RF}} \mathbf{B}_{\text{RF}}^H \mathbf{H}_{\text{eff}}[k] \mathbf{H}_{\text{eff}}^H[k] \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{BB}}[k] \mathbf{W}_{\text{BB}}^H[k] \right| \right) \\
&\stackrel{(c)}{\leq} \log_2 \left(\left| \mathbf{I}_{N_{\text{RF}}} + \frac{PN}{\sigma_n^2} \mathbf{F}_{\text{BB}}[k] \mathbf{F}_{\text{BB}}^H[k] \mathbf{B}_{\text{BB}}[k] \mathbf{B}_{\text{BB}}^H[k] \times \right. \right. \\
&\quad \left. \left. \mathbf{F}_{\text{RF}} \mathbf{F}_{\text{RF}}^H \mathbf{B}_{\text{RF}} \mathbf{B}_{\text{RF}}^H \sum_{k=1}^K \mathbf{H}_{\text{eff}}[k] \mathbf{H}_{\text{eff}}^H[k] \mathbf{W}_{\text{RF}} \mathbf{W}_{\text{RF}}^H \mathbf{W}_{\text{BB}}[k] \mathbf{W}_{\text{BB}}^H[k] \right| \right), \tag{4.12}
\end{aligned}$$

Then, the analog precoding matrix can be written as

$$\mathbf{F}_{\text{RF}} = \frac{1}{\sqrt{N_k}} \text{diag} (e^{j\Theta}), \tag{4.13}$$

Training on the reconstruction loss allows learning useful properties of the data in an unsupervised manner. We put the output $\hat{\mathbf{H}}_{AAE}^{est}$ of the AAE module into the input $\hat{\mathbf{H}}_{AE}^{est}$ of the AE module. Source signals $\hat{\mathbf{H}}_{AE}^{est} \in \mathbb{R}^n$ and maps it to latent space $\mathbf{Z}_{AE} \in \mathbb{R}^m$, and the encoder function generates a latent space containing the code for the source signal. The decoder reconstructs the input according to the latent space. The encoder transforms the input data into a lower dimensional representation can be written as

$$\mathbf{Z}_{AE} = \sigma \left(\mathbf{W} \hat{\mathbf{H}}_{AE}^{est} + \mathbf{b} \right), \quad (4.14)$$

where \mathbf{b} represent the bias of encoder, σ represent the activation function of encoder, and \mathbf{W} represent the weight of encoder. The decoder converts the representation of the code back to the original input data can be written as

$$\tilde{\mathbf{H}}_{AE}^{est} = \sigma' \left(\mathbf{W}' \mathbf{Z}_{AE} + \mathbf{b}' \right), \quad (4.15)$$

where \mathbf{b}' represent the bias of decoder, σ' represent the activation function of decoder, and \mathbf{W}' represent the weight of decoder. The training process of an autoencoder involves minimizing the reconstruction error, and it is common to use the squared difference of the reconstruction error (e.g., MSE) as the loss function $\left\| \hat{\mathbf{H}}_{AE}^{est} - \tilde{\mathbf{H}}_{AE}^{est} \right\|^2$ and optimize it using a backpropagation algorithm. Then, we put the output $\tilde{\mathbf{H}}_{AE}^{est}$ of the AE module into the input $\hat{\mathbf{H}}_{AE}^{est}$ of the AE module for the MSE operation. In the analog precoding matrix, we pass the output channel signal from AAE to AE to extract the real and imaginary components of the channel signal consisting of the phases of the elements in \mathbf{F}_{RF} . Then we can get the optimal digital precoding matrix \mathbf{F}_{RF} .



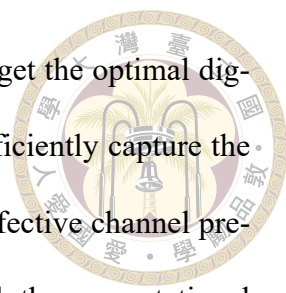
4.4 Digital precoding

The previously studied digital precoding uses the autoencoder model[16] and alternating optimization (AO) Optimization[60] that is more difficult to capture features in the frequency and spatial domains. As CNNs have the advantages of local neuron connectivity and weight sharing, they can avoid overfitting and do not learn the same features under different signal processing, and at the same time improve the computational efficiency and accuracy. In terms of channel precoding, CNN has better prediction and recognition ability than other general neural networks. Therefore, we use CNN to convert the input channel signal data into 2D Fourier transform, and utilize the characteristics of CNN with local neuron connections, translation invariant, and weight sharing to capture the frequency and spatial domain characteristics of the channel precoding, not only the front and back direction of the neighboring channel data, but also the up and down, left and right direction of the neighboring channel data, so as to enable our digital precoding module to have a better prediction and recognition ability.

The two-stage hybrid structure combines the benefits of analog phase shifters and flexible digital processing. We put the output $\tilde{\mathbf{H}}_{AE}^{est}$ of the AE module into the input $\hat{\mathbf{H}}_{CNN}^{est}$ of the CNN module, and we put the output $\tilde{\mathbf{H}}_{CNN}^{est}$ of the CNN module into the input of the CNN module for the MSE operation. Then, the digital precoding matrix can be written as

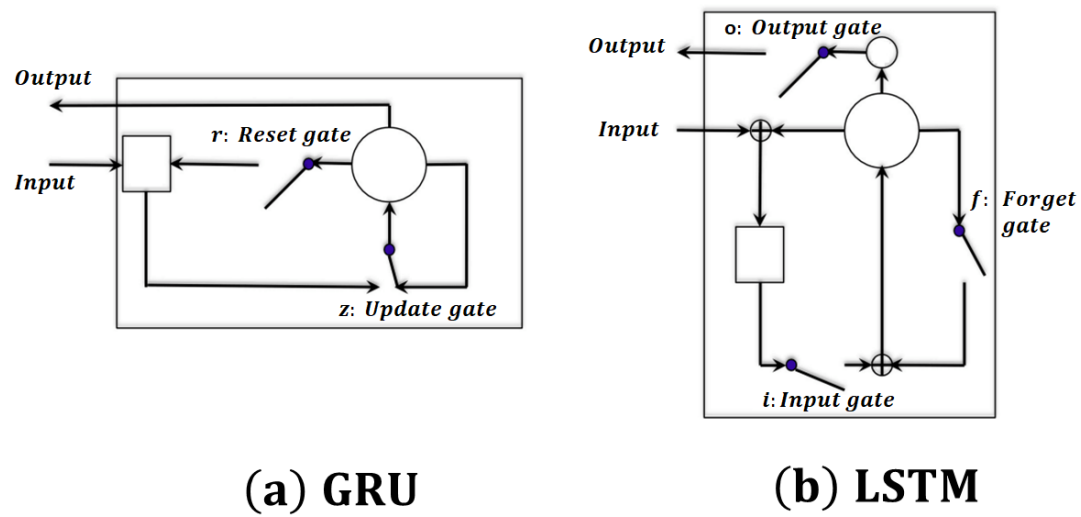
$$\mathbf{F}_{BB}[k] = \sqrt{P_t} \frac{\tilde{\mathbf{F}}_{BB}[k]}{\left\| \mathbf{F}_{RF} \tilde{\mathbf{F}}_{BB}[k] \right\|_F}. \quad (4.16)$$

In the digital precoding matrix, we first pass the output channel signals from the AE to the CNN model to obtain the real and imaginary components of the channel signals



composed of the phases of the elements in the $\mathbf{F}_{\text{BB}}[k]$. Then we can get the optimal digital precoding matrix $\mathbf{F}_{\text{BB}}[k]$. These attributes enable 2D-CNN to efficiently capture the intricate frequency and spatial domain characteristics essential for effective channel precoding. Moreover, they help in reducing interference, enhancing both the computational efficiency and the accuracy of the precoding process.

4.5 Analog beamforming



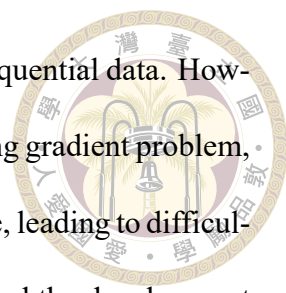
(a) GRU

(b) LSTM

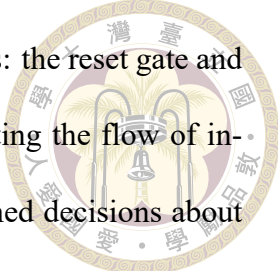
Figure 4.3: The architecture of GRU and LSTM.

The analog beamforming in our previous study used the autoencoder model[16] and alternating optimization (AO) Optimization[60] that is more difficult to capture whether the features reconstruct the original signals. Therefore, we use autoencoder to reconstruct the latent space back to the input channel signal data as much as possible, and extract from the bottleneck features, so that our analog beamforming module can improve the accuracy of capturing the features.

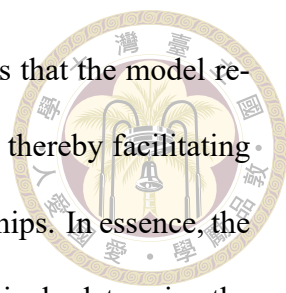
In the realm of deep learning, recurrent neural networks (RNNs) have emerged as



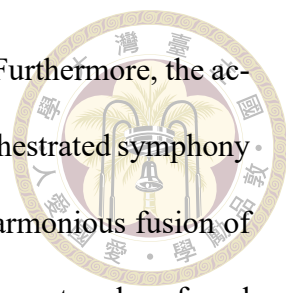
powerful tools for capturing temporal dependencies and patterns in sequential data. However, traditional RNNs often grapple with the challenge of the vanishing gradient problem, where the gradients propagated backward through time tend to dwindle, leading to difficulties in learning long-range dependencies. This limitation has motivated the development of more sophisticated RNN variants, such as Fig. 4.3 (a) Gated Recurrent Unit (GRU) is a simplified variation of the Fig. 4.3 (b) LSTM (Long Short-Term Memory) model. Fig. 4.3 (a) highlights the GRU, a more streamlined variation of the LSTM. The GRU was specifically devised to tackle the vanishing gradient problem while maintaining the efficiency of the model. And, the GRU strikes a balance between computational simplicity and effective information retention, making it a popular choice for modeling sequential data. The introduction of GRU has significantly enriched the landscape of RNN architectures, offering a practical and efficient solution for addressing long-range dependencies. Its simplified design, coupled with its capacity to effectively capture temporal patterns, has solidified its position as a crucial component in various applications. As such, the GRU continues to play a vital role in advancing the capabilities of RNNs, paving the way for more sophisticated and intelligent sequence modeling techniques. Fig. 4.3 (b) illustrates the LSTM model, renowned for its ability to effectively handle long-range dependencies in sequences. This mechanism allows the LSTM to retain essential information over extended time periods, enabling it to capture and utilize valuable long-term dependencies in the data. GRU offer several advantages, which have contributed to their popularity in various sequence modeling tasks. GRU are easier to train compared to standard RNNs, such as vanilla RNNs. This is because they use gating mechanisms that allow them to capture long-range dependencies without suffering from the vanishing gradient problem to the same extent. GRU have gained prominence in the realm of deep learning for their

The logo of National Taiwan University (NTU) is located in the upper right quadrant of the page. It is a circular emblem with a gold border. Inside the circle, there is a central figure of a person holding a torch, surrounded by the university's name in Chinese characters: "國立台灣大學" (National Taiwan University) and "1946" at the bottom.

unique architecture, characterized by two essential gating mechanisms: the reset gate and the update gate. These gating elements play a crucial role in regulating the flow of information within the GRU cell, allowing the network to make informed decisions about the handling of temporal dependencies. The reset gate, a fundamental component of the GRU, serves as a mechanism for controlling the retention or dismissal of information from the previous time step. By selectively resetting certain elements of the hidden state, the reset gate empowers the network to disregard obsolete or irrelevant information, thus enabling it to focus on the most relevant and significant temporal patterns within the sequence. This ability to effectively reset and filter out outdated information resembles the initial processing of the first symbol in an input sequence, facilitating the initialization of the network's memory for subsequent inputs. Notably, units characterized by short-term dependencies are likely to exhibit active reset gates, enabling them to focus on recent information, whereas units characterized by long-term dependencies are more likely to display active update gates, enabling them to retain crucial historical context and patterns. The interplay between the reset and update gates in the GRU architecture serves as a mechanism for the network to dynamically adjust its memory and information processing, ensuring the effective modeling of temporal sequences with varying degrees of complexity and length. This dynamic adaptability and sophisticated control over information flow have solidified the GRU's position as a versatile and powerful tool for sequence modeling tasks across diverse domains. This gating helps in modeling complex temporal dependencies effectively. The GRU architecture serves as a pivotal component that grants the model the flexibility to determine the optimal proportion that should be updated with new information. This fine-grained control mechanism enables the GRU to adapt its memory utilization based on the specific characteristics of the input sequence and the inherent temporal dependen-

The logo of National Taiwan University (NTU) is positioned in the upper right quadrant of the page. It is a circular emblem with a gold border. Inside the circle, there are two palm trees flanking a central bell. The Chinese characters '國立台灣大學' (National Taiwan University) are written around the perimeter of the circle.

cies present within it. This nuanced decision-making process ensures that the model remains responsive to both immediate changes and longer-term trends, thereby facilitating the accurate capture and representation of complex temporal relationships. In essence, the update gate functions as a dynamic filter, allowing the GRU to selectively determine the relevance and significance. The gate's adaptive nature empowers the network to adjust its memory retention and information integration based on the specific requirements of the task at hand, facilitating robust and effective sequence modeling across a diverse array of applications and domains. The sophisticated control provided by the update gate within the GRU architecture not only mitigates the challenges associated with vanishing gradients but also ensures that the model can effectively capture intricate temporal dependencies. This helps mitigate the vanishing gradient problem by preventing gradients from becoming too small during backpropagation. GRU are computationally more efficient than some other RNN variants, such as LSTM, while still providing competitive performance. This efficiency is particularly valuable in applications where computational resources are limited. GRU often perform well on tasks involving short to medium-length sequences. They can capture dependencies within these sequences without excessive computational overhead. The gating mechanisms in GRU can help reduce overfitting in certain scenarios by allowing the model to control the flow of information and avoid memorizing noise in the training data. GRU have a simpler architecture compared to LSTMs, which makes them easier to understand, implement, and tune. In its elegant and ingenious design, the GRU seamlessly amalgamates the traditionally distinct forget and input gates into a singular and unified "update gate," while simultaneously introducing the novel and impactful concept of the "reset gate." This unique amalgamation of key gating mechanisms imbues the GRU with a level of sophistication and flexibility that is not only unparalleled but also



increasingly sought after in contemporary research and applications. Furthermore, the activation h_t^j of the GRU at a given time step t unfolds as a delicately orchestrated symphony of intricate and nuanced computations, embodying a seamless and harmonious fusion of the previous activation h_{t-1}^j and the candidate activation \tilde{h}_t^j . This elegant and profound fusion can be conceptualized as a linear interpolation, a delicate dance between the temporal continuity and dynamic evolution of the underlying data, seamlessly blending the past and present into a coherent and unified representation that captures the essence of the data's underlying temporal dynamics. It is within this delicate interplay of unified gating mechanisms and seamless temporal interpolations that the GRU distinguishes itself, transcending the conventional boundaries of standard LSTM models and emerging as a beacon of simplicity, efficiency, and remarkable performance. In the ever-evolving landscape of recurrent neural networks, the GRU's simplicity belies its remarkable power, standing as a testament to the relentless pursuit of more effective, streamlined, and versatile models that push the boundaries of what is possible in the realm of sequential data analysis and processing.

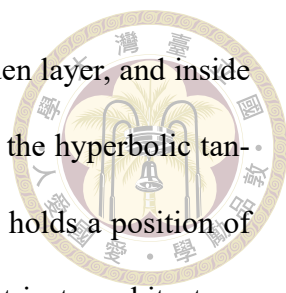
$$h_t^j = (1 - z_t^j)h_{t-1}^j + z_t^j\tilde{h}_t^j \quad (4.17)$$

The update gate z_t^j have a prominent place in regulating the extent to which the unit updates its activation.

$$z_t^j = \sigma(W_z \hat{\mathbf{H}}_{GRU}^{est} + U_z h_{t-1})^j. \quad (4.18)$$

The candidate activation \tilde{h}_t^j is computed in a manner similar to the update gate:

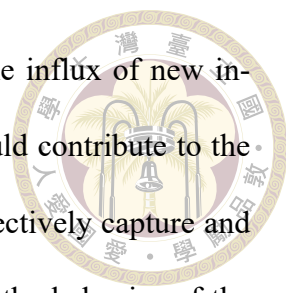
$$\tilde{h}_t^j = \tanh(W \hat{\mathbf{H}}_{GRU}^{est} + U(r_t h_{t-1}))^j, \quad (4.19)$$



U, W represent weight matrices from the input layer to the hidden layer, and inside the hidden layer. σ represents the sigmoid function, \tanh represents the hyperbolic tangent function. In the realm of GRU, the concept of the reset gate r_t^j holds a position of critical importance. Serving as a central component of the GRU's intricate architecture, the reset gate's role is multifaceted, and its activation state profoundly influences the information processing dynamics within the unit. When the reset gate is in an inactive state, it effectively enables the unit to selectively discard or ignore information from the past, thus facilitating a mechanism akin to the unit's initial processing of the very first symbol within an input sequence. This intricate interplay of gate activation and information retention lies at the heart of the GRU's ability to adaptively process sequential data, enabling it to effectively capture the temporal dynamics and dependencies present within the data. The computation of the reset gate's activation involves a carefully orchestrated series of calculations that intricately weigh the relevance and significance of past information. Through a meticulously designed computational formula, the reset gate is endowed with the capability to selectively control the flow of information, allowing the GRU to dynamically adjust its processing based on the current context and the specific temporal dynamics inherent in the data. This delicate dance of gate activation and information processing forms the foundation of the GRU's adaptability, enabling it to effectively capture and leverage the intricate temporal dependencies present within sequential data. The reset gate can be written as:

$$r_t^j = \sigma(W_r \hat{\mathbf{H}}_{GRU}^{est} + U_r h_{t-1})^j \quad (4.20)$$

The update gate z within the framework of GRU assumes a pivotal role in regulating the balance between the influence of past information and the integration of new information within the current state. As a dynamic control mechanism, the update gate



orchestrates the intricate interplay between the existing state and the influx of new information, thereby dictating the degree to which historical data should contribute to the present state. This mechanism is crucial in enabling the GRU to effectively capture and model the temporal dynamics inherent in sequential data. Notably, the behavior of the update gate reveals essential insights into the nature of temporal dependencies within the data. In instances where short-term dependencies hold sway, the GRU tends to activate the reset gates r , thereby signaling the unit's focus on leveraging recent information to inform its current state. Conversely, when the unit encounters long-term dependencies that necessitate a more comprehensive understanding of the data's historical context, the update gate z assumes a more prominent role, guiding the unit in integrating pertinent past information to shape its current state effectively. This intricate orchestration of gate dynamics underscores the GRU's capacity to flexibly adapt to diverse data patterns, allowing it to strike an optimal balance between capturing short-term nuances and comprehending broader, long-term trends. By modulating the flow of information through the interplay of update and reset gates, the GRU emerges as a powerful tool for modeling complex temporal dependencies, offering an invaluable asset across a wide spectrum of applications, including natural language processing, speech recognition, and time series analysis, among others. Then, the analog beamforming matrix can be written as

$$\mathbf{B}_{\text{RF}} = \frac{1}{\sqrt{N_t}} e^{j\angle \left(\left[\mathbf{H}_{\text{eff}}(:, 1:N_{\text{RF}}) \right] [k] \right)}, \quad (4.21)$$

4.6 Digital beamforming

The previously studied digital beamforming uses the autoencoder model[16] and alternating optimization (AO) Optimization[60] that is more difficult to capture features in

the frequency and spatial domains. 1D-CNNs excel at capturing local patterns and dependencies within sequential data. They can automatically learn features such as time series analysis, so as to enable our digital beamforming module to have better anti-interference ability.



We put the output $\tilde{\mathbf{H}}_{GRU}^{est}$ of the GRU module into the input $\hat{\mathbf{H}}_{1D-CNN}^{est}$ of the CNN module, and we put the output $\tilde{\mathbf{H}}_{1D-CNN}^{est}$ of the 1D-CNN module into the input of the 1D-CNN module for the MSE operation. Then, the digital beamforming matrix can be written as

$$\mathbf{B}_{BB}[k] = \sqrt{P_t} \frac{\tilde{\mathbf{B}}_{BB}[k]}{\|\mathbf{B}_{RF} \tilde{\mathbf{B}}_{BB}[k]\|_F}. \quad (4.22)$$

Therefore, the overall loss function can be written as

$$Loss = \left\| \hat{\mathbf{H}}_{AAE}^{est} - \tilde{\mathbf{H}}_{AAE}^{est} \right\|_{AAE}^2 + \left\| \hat{\mathbf{H}}_{AE}^{est} - \tilde{\mathbf{H}}_{AE}^{est} \right\|_{AE}^2 + \left\| \hat{\mathbf{H}}_{2D-CNN}^{est} - \tilde{\mathbf{H}}_{2D-CNN}^{est} \right\|_{2D-CNN}^2 + \left\| \hat{\mathbf{H}}_{GRU}^{est} - \tilde{\mathbf{H}}_{GRU}^{est} \right\|_{GRU}^2 + \left\| \hat{\mathbf{H}}_{1D-CNN}^{est} - \tilde{\mathbf{H}}_{1D-CNN}^{est} \right\|_{1D-CNN}^2. \quad (4.23)$$

In the digital beamforming matrix, we first pass the output channel signals from the GRU to the 1D-CNN model to obtain the real and imaginary components of the channel signals composed of the phases of the elements in the $\mathbf{B}_{BB}[k]$. Then we can get the optimal digital beamforming matrix $\mathbf{B}_{BB}[k]$.

4.7 Computational Complexity Analysis



Table 4.1 provides a comprehensive comparison of the computational complexities associated with different models. Notably, our proposed VAE-CATT-Beamforming model exhibits significantly lower complexity when contrasted with the LSTM-DAE-Beamforming model. Its computational demands are on par with the DAE-Beamforming model. In terms of overall system performance, VAE-CATT-Beamforming outperforms the compared models concerning Mean Squared Error (MSE). It does possess a slightly higher complexity compared to DAE-Beamforming, DAE-CNN-Beamforming, and DAE-CNN-ATT-Beamforming. However, it remains considerably more efficient when contrasted with the complexities of DAE-CNN-Beamforming, DAE-CNN-ATT-Beamforming, and DAE-Beamforming approaches. In essence, our proposed model strikes a well-balanced equilibrium between robust performance and reasonable computational complexity.

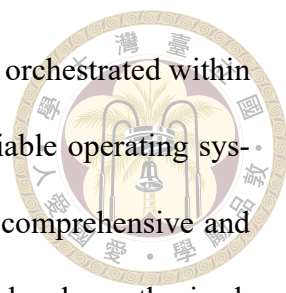
Furthermore, we conducted inference time measurements for the considered models, resulting in the following times: 0.39 seconds for VAE-CATT-Beamforming, 0.26 seconds for DAE-Beamforming, 0.43 seconds for LSTM-DAE-Beamforming, 0.34 seconds for DAE-CNN-Beamforming, and 0.37 seconds for DAE-CNN-ATT-Beamforming. Importantly, the variability in the inference times is notably low. Additionally, it's important to highlight that the parallelization capabilities of GPUs significantly accelerate the execution time and processing speed of deep learning models.





Chapter 5 Results

In the scope of this thesis, our investigation delves into the intricate dynamics of the MIMO system, investigating the influence of different system parameters on the overall performance and effectiveness of the entire configuration. To provide a comprehensive analysis, we meticulously set up the experimental environment with a carefully chosen array of system parameters, ensuring a robust and reliable framework for our evaluations. The system parameters were meticulously tailored to capture a diverse range of scenarios, including the numbers of transmitting antennas N_t , which were meticulously varied across 64 and 32, alongside the numbers of receiving antennas N_r , selected as 32 and 16. Additionally, the propagation paths L were systematically manipulated, spanning 1, 2, 3, and 4, in order to capture the multifaceted nature of real-world propagation scenarios. The robustness and consistency of our findings were ensured through the use of a well-calibrated batch size of 128, coupled with a carefully selected learning rate of 0.0001. Furthermore, a meticulous training strategy encompassing 200 epochs was employed to meticulously explore and dissect the various intricacies of the MIMO channel. Our hardware configuration, comprising the high-performance Intel Core i9-11900K CPU processor operating at an impressive 5.30 GHz, supported by a substantial 64 GB of RAM, and the state-of-the-art NVIDIA® GeForce RTX™ 3080Ti GPU equipped with 12 GB GDDR6X memory, provided the essential computational power necessary for conducting the rigorous simu-



lations and analyses. The entire experimental setup was meticulously orchestrated within the Linux Mint version 21.2 environment, ensuring a stable and reliable operating system conducive to robust research endeavors. In order to generate a comprehensive and diverse dataset, a total of 50,000 MIMO channel samples were meticulously synthesized, guaranteeing a rich and varied pool of data for in-depth analysis. Python 3.9 served as the primary programming framework for implementing all the models, with TensorFlow 2.8.0 serving as the foundational tool for training and analysis. The Adam optimizer, known for its robust and efficient performance in various machine learning tasks, was carefully selected as the primary optimization algorithm, further enhancing the reliability and effectiveness of our training procedures. The system parameters are specified in Table 5.1. We show comparison results of the proposed AAE-AATT-Beamforming, DAE-Beamforming, LSTM-DAE-Beamforming, DAE-CNN-Beamforming, and DAE-CNN-ATT-Beamforming. These are expressed as follows:

1) Proposed AAE-AATT-Beamforming: uses the DAE similar to [16], and uses the additive attention mechanism to enhance the capability of the beamforming design.

2) DAE-Beamforming: uses the DAE similar to [16] enhance the capability of the beamforming design.

3) DAE-CNN-Beamforming: uses CNN similar to [23] with the DAE to enhance the capability of the beamforming design.

4) DAE-CNN-ATT-Beamforming: uses a CNN with the DAE, and utilizes the additive attention method similar to [37] to enhance the performance of the beamforming design.

5) LSTM-DAE-Beamforming: uses a LSTM model similar to [54] to retrieve pre-

Table 5.1: SIMULATION PARAMETERS

Parameters	Value
Carrier frequency (f_c)	28 GHz
Paths ($L_t = L_g = L$)	4
Learning rate α	0.0001
Transmitter antennas (N_b)	32, 64
Angle (θ_l)	$\mathcal{U}(-\pi/2, \pi/2)$
Receiver antennas (N_u)	16, 32
Subcarriers	128
IRS reflecting elements (M)	32, 64
Bandwidth (f_s)	2 GHz

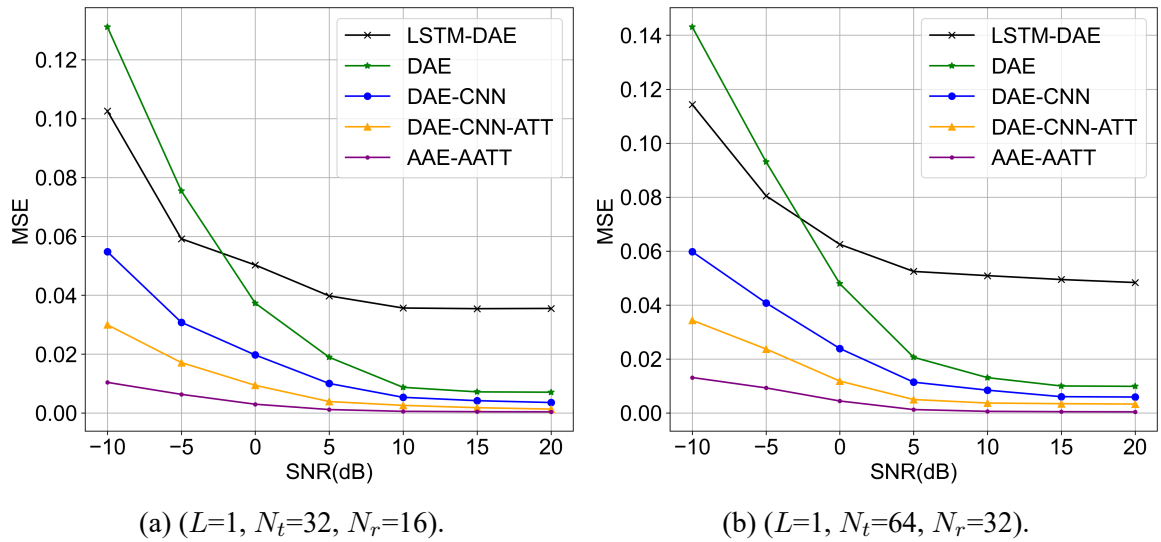
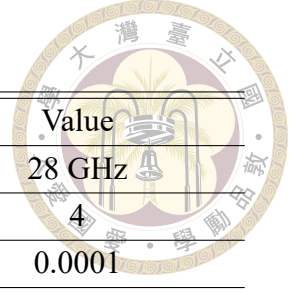
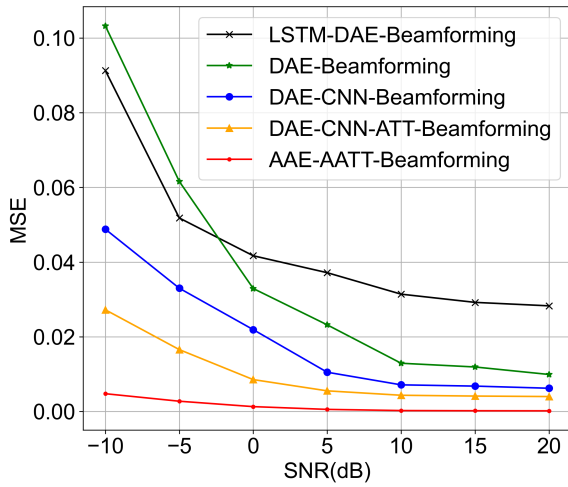


Figure 5.1: Comparison of MSE versus SNR between different methods without beamforming.

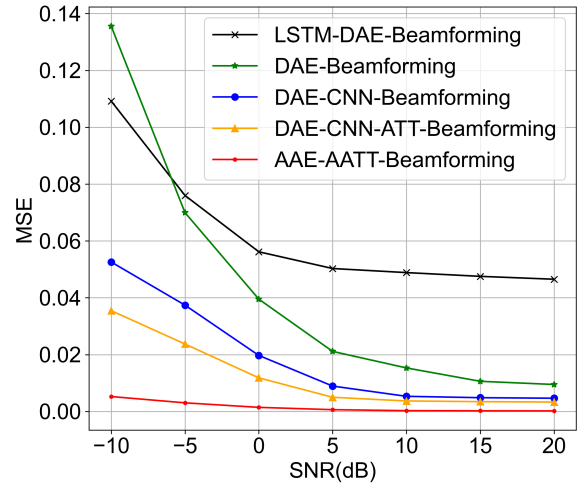
vious channel information sequences and then utilize DAE to improve the efficiency of beamforming designs.

5.1 The Investigation of System Parameters

In the evaluation of the IRS-assisted beamforming model within a mmWave MIMO-OFDM system under different SNR conditions, several noteworthy observations can be made in Fig.5.1a and Fig.5.1b: DAE-CNN-ATT and DAE-CNN models exhibit subopti-

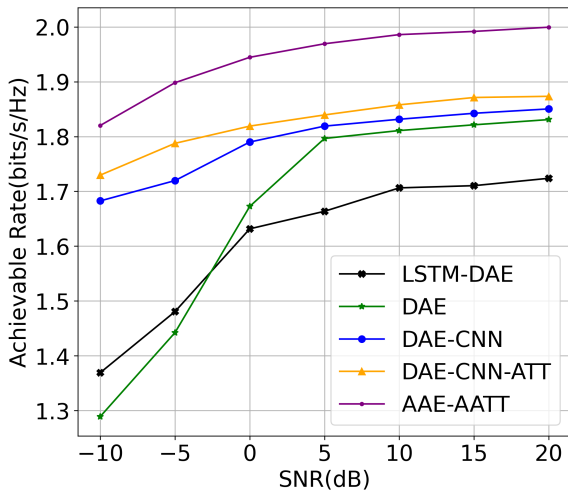


(a) ($L=1, N_t=32, N_r=16$).

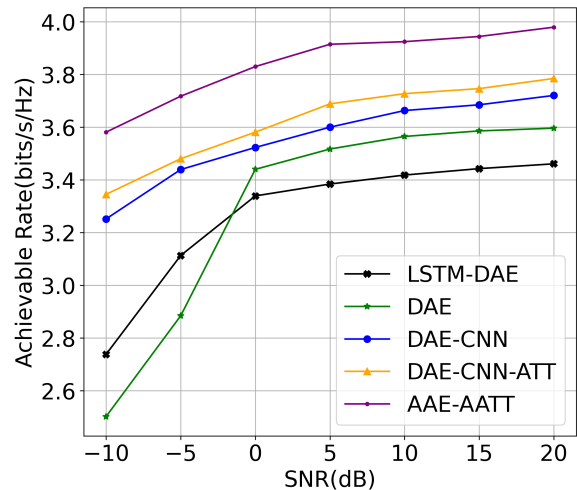


(b) ($L=1, N_t=64, N_r=32$).

Figure 5.2: Comparison of MSE versus SNR between different methods with beamforming.

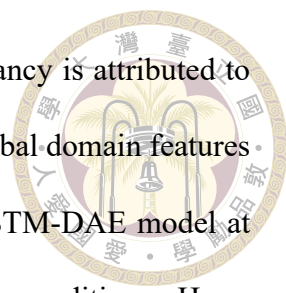


(a) ($L=1, N_t=32, N_r=16$).

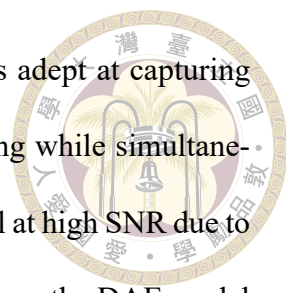


(b) ($L=1, N_t=64, N_r=32$).

Figure 5.3: Comparison of achievable rate versus SNR between different methods without beamforming.



mal performance compared to the AAE-AATT model. This discrepancy is attributed to their smaller receptive fields, which hinder their ability to capture global domain features effectively under severe noise conditions. DAE outperforms the LSTM-DAE model at high SNR due to its superior noise reduction capabilities under such conditions. However, LSTM-DAE outperforms the DAE model at low SNR due to LSTM's capacity to capture dependencies between channels. As SNR increases, the proposed AAE-AATT model consistently achieves the lowest MSE among all the models. This is attributed to the additive attention mechanism incorporated in the AAE-AATT model, which adeptly captures inter-channel data, including frequency, spatial, and temporal aspects. Additionally, AAEs have the unique ability to generate realistic data samples based on this learned latent space. This capability proves valuable in various applications such as signal processing and the generation of new data points resembling the training data. Importantly, the learned latent space plays a pivotal role in capturing essential features for intra-channel beamforming in both frequency and spatial domains. Overall, the AAE-AATT model provides the best MSE performance at both low and high SNR. A slight degradation in the MSE performance of all models is observed as the number of antennas increases. This decline can be attributed to variations in the parameters of the estimated number of antennas. With precoding, the MSE performance of all models shows improvement. This enhancement is attributed to the AAE-AATT-Beamforming method, which utilizes the AAE model and the additive attention mechanism to mitigate noise effects and optimize beamforming designs. In Fig.5.2a and Fig.5.2b, as the number of transmitting and receiving antennas increases, a slight decrease in the MSE of all models is observed. This is due to the higher number of antenna parameters involved. AAE-AATT demonstrates superior performance compared to DAE-CNN-ATT and DAE-CNN models. This advantage



is attributed to the translation invariance of the CNN model. This is adept at capturing crucial frequency and spatial features necessary for channel precoding while simultaneously reducing interference. DAE outperforms the LSTM-DAE model at high SNR due to its noise-robust reconstructed signal. Conversely, LSTM-DAE surpasses the DAE model at low SNR by leveraging its long-term memory function to model channel signal sequences. As SNR increases in Fig.5.3a, the proposed AAE-AATT model consistently achieves the highest achievable rate performance among all models. This is facilitated by the additive attention mechanism within the AAE-AATT model, which captures inter-channel data, including sequences, spatial, and temporal information. Additionally, AAE is employed to learn a structured latent space, resulting in superior achievable rate results. Fig.5.3b illustrates that the achievable rate comparison results for all the models experience a slight degradation. This degradation is attributed to the challenges posed by estimating antenna parameters, especially as the number of antennas increases. In Fig.5.4b, we present achievable rate comparison results for various models, including the proposed AAE-AATT-Beamforming, DAE-Beamforming, DAE-CNN-Beamforming, DAE-CNN-ATT-Beamforming, and LSTM-DAE-Beamforming models with beamforming. These results indicate that all models exhibit enhanced achievable rate performance. The proposed AAE-AATT-Beamforming method leverages the additive attention mechanism to allow the models to analyze input sequences from diverse perspectives simultaneously. This mechanism captures inter-channel data, encompassing frequency, spatial, and temporal aspects, enabling the models to comprehend different subcarrier dependencies and extract vital channel information. The additive attention mechanism improves channel beamforming performance by reevaluating the significance of frequency, spatial, and temporal domains. Additionally, AAEs have the unique ability to generate realistic data samples

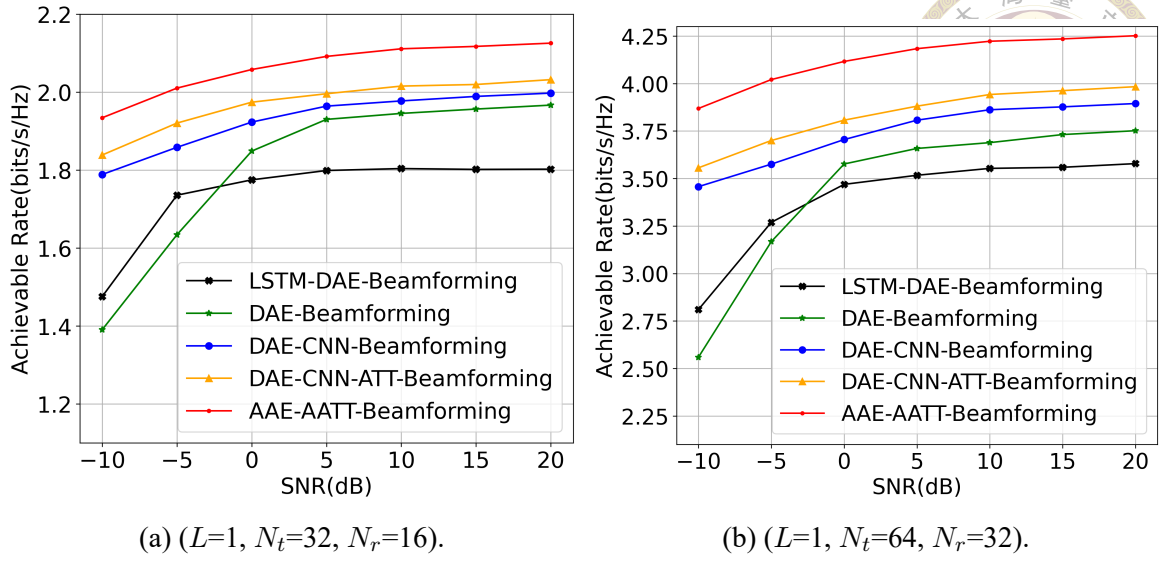
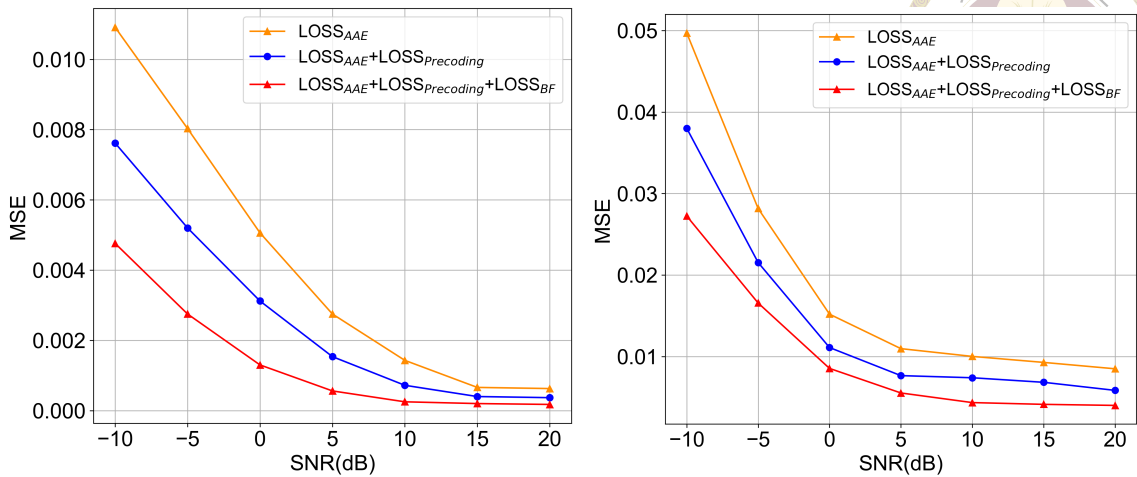


Figure 5.4: Comparison of achievable rate versus SNR between different methods with beamforming.

based on this learned latent space. This capability proves valuable in various applications such as signal processing and the generation of new data points resembling the training data. Importantly, the learned latent space plays a pivotal role in capturing essential features for intra-channel beamforming in both frequency and spatial domains. In Fig.5.4b, we observe that MSE comparison results for all the models exhibit a slight degradation, mainly due to the challenges associated with estimating antenna parameters as the number of antennas increases.

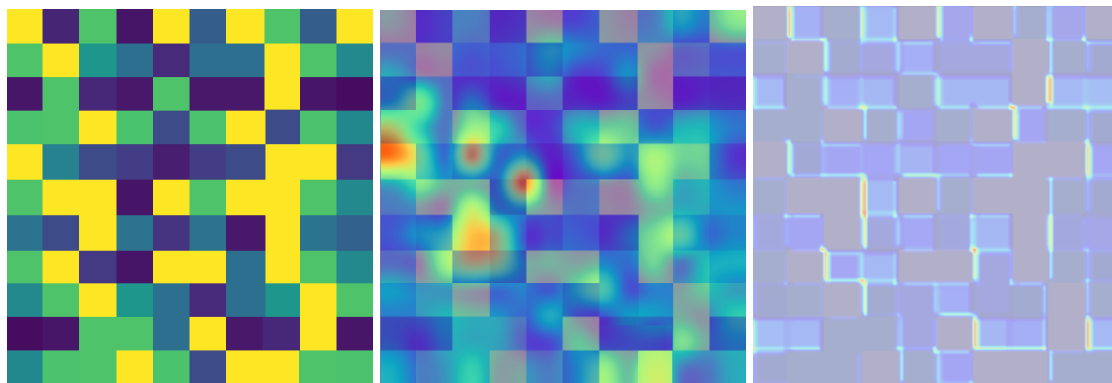
5.2 The Investigation of Loss Functions

Fig.5.5a presents MSE comparison results for the proposed AAE-AATT-Beamforming model with different loss functions. The MSE performance difference between $LOSS_{AAE}$ and $LOSS_{AAE} + LOSS_{Precoding}$ is 21.63%, and between $LOSS_{AAE} + LOSS_{Precoding}$ and $LOSS_{AAE} + LOSS_{Precoding} + LOSS_{BF}$ is 31.836%. Fig.5.5b displays MSE comparison results for the DAE-CNN-ATT-Beamforming model with different loss functions.



(a) AAE-AATT-Beamforming ($L=1$, $N_t=32$), (b) DAE-CNN-ATT-Beamforming ($L=1$, $N_r=16$).

Figure 5.5: Comparison of loss functions versus SNR between different methods.

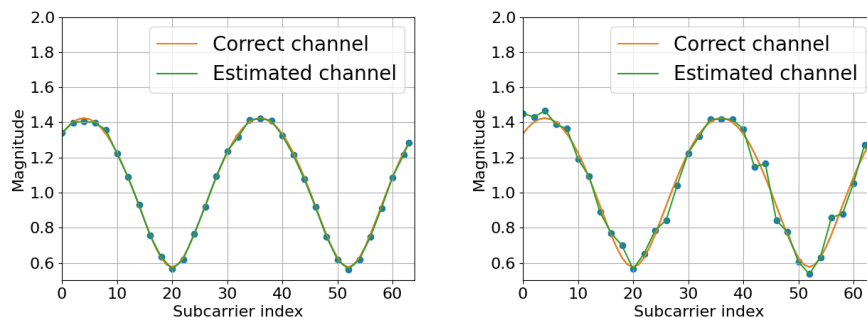


(a) Original 2-D Fourier spectrum.

(b) AAE model.

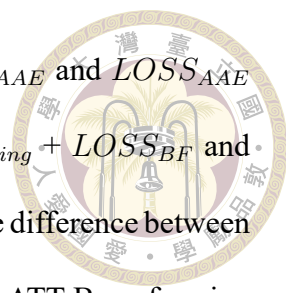
(c) CNN model.

Figure 5.6: Comparison of attention map between different methods.



(a) AAE model with additive attention. (b) AAE model without additive attention.

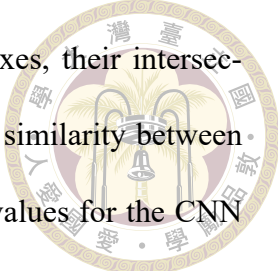
Figure 5.7: Comparison of subcarrier channel prediction for different methods.



Here, we observe that MSE performance difference between $LOSS_{AAE}$ and $LOSS_{AAE} + LOSS_{Precoding}$ is 26.48%, and between $LOSS_{AAE} + LOSS_{Precoding} + LOSS_{BF}$ and $LOSS_{AAE} + LOSS_{Precoding}$ is 37.57%. The overall MSE performance difference between the proposed AAE-AATT-Beamforming model and the DAE-CNN-ATT-Beamforming model is 71.83%. This divergence in MSE performance can be attributed to the AAE's optimization approach, AAEs have the unique ability to generate realistic data samples based on this learned latent space. This capability proves valuable in various applications such as signal processing and the generation of new data points resembling the training data. Importantly, the learned latent space plays a pivotal role in capturing essential features for intra-channel beamforming in both frequency and spatial domains, facilitating comprehensive contextual information acquisition from channel signals. Additionally, the analog beamforming module benefits from feature capture enhancements via an autoencoder and GRU, while the digital beamforming module effectively reduces interference through the utilization of a CNN.

5.3 The Investigation of Spectral Analysis

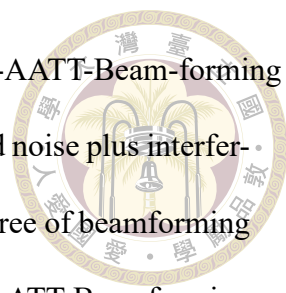
The objective of spectral analysis is to break down a signal into its various frequency components, a technique employed to scrutinize the frequency aspects of channel signals. This transformation from time and spatial domain functions to frequency domain functions is achieved through the Fourier transform, as depicted in Fig.5.6a. The X-axis and Y-axis represent the 2D signal spectrum in the time and spatial domains, respectively. The color spectrum in the Fourier transform's frequency domain intensity ranges from yellow (strongest) to black (weakest), following the order: Yellow > Lime > Teal > Blue > Black. To gauge the relative overlap of two bounding boxes, we employ Intersection



over Union (IoU) measurements. When there are two bounding boxes, their intersection and union areas can be calculated, allowing us to determine the similarity between the two bounding boxes. In our observations, we note that the IoU values for the CNN model are Yellow=0.48, Lime=0.33, Teal=0.42, Blue=0.46, Black=0.41, whereas for the VAE model, these values are significantly higher: Yellow=0.93, Lime=0.86, Teal=0.84, Blue=0.79, Black=0.88. Additionally, in our experiments, we found that the attention map's area in the frequency domain is larger for the AAE model compared to the CNN model, as depicted in Fig.5.6b and Fig.5.6c. This phenomenon is attributed to the AAE's incorporation of an encoder that capability proves valuable in various applications such as signal processing and the generation of new data points resembling the training data. Importantly, the learned latent space plays a pivotal role in capturing essential features for intra-channel beamforming in both frequency and spatial domains. Furthermore, we conducted experiments comparing the AAE model with and without additive attention. The results demonstrated that the AAE model with additive attention outperforms the one without it in the time domain subcarrier channel prediction, as illustrated in Fig.5.7a and Fig.5.7b. The rationale behind this improvement lies in the additive attention mechanism within our AAE-AATT-Beamforming approach. This mechanism adeptly captures inter-channel global data, encompassing time domain information, thereby enabling the model to effectively comprehend intricate relationships, dependencies, and interactions among various sequences of channel signal subcarriers.

5.4 The Investigation of Interference Analysis

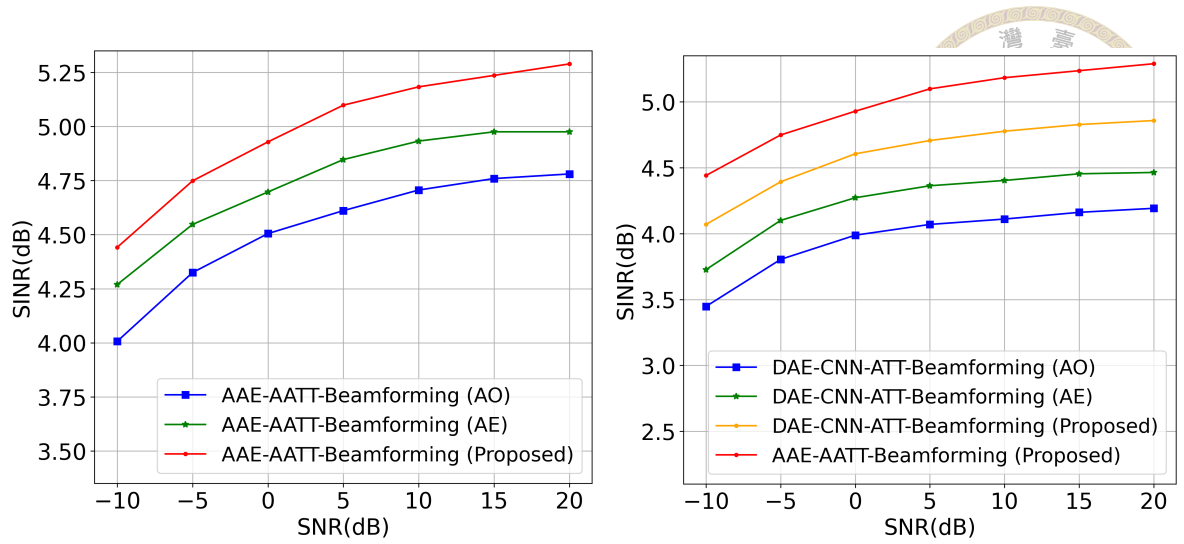
In Fig.5.8a, we conduct a comparison between Signal-to-Interference-plus-Noise Ratio (SINR) and Signal-to-Noise Ratio (SNR) for three different combinations: AAE-AATT-



Beamforming (Proposed), AAE-AATT-Beamforming (AE), and AAE-AATT-Beamforming (AO). SINR represents the ratio of received signal strength to received noise plus interference signal strength. Using Shannon's Theorem, we calculate the degree of beamforming interference with Equation(4.1). Our observations reveal that AAE-AATT-Beamforming (Proposed) exhibits the best performance, boasting a 4.35% improvement in system performance over DAE-CNN-ATT-Beamforming. When transitioning to AAE-AATT-Beamforming(AE) and AAE-AATT-Beamforming (AO), the system performance decreases by 6.23% and 10.61%, respectively. In Fig.5.8b, the DAE-CNN-ATT-Beam-forming section illustrates that replacing DAE-CNN-ATT-Beamforming (AE) and DAE-CNN-ATT-Beamforming (AO) leads to a degradation of system performance by 9.56% and 18.34%, respectively. (Proposed) outperforms (AE) because in digital precoding, CNN leverages advantages like local neuron connections and weight sharing. This ensures that the same features are not redundantly learned under different signal processing, avoiding over-computation and enhancing efficiency and accuracy. (AE) surpasses (AO) because the autoencoder effectively restores the latent space to the original input channel signal data. This enables our analog precoding module to capture global domain features more effectively. In contrast, the order in which variables are updated in AO can significantly affect its convergence rate and final solution. Weak choices of variable ordering can lead to slow convergence or getting stuck in suboptimal solutions.

5.5 The Investigation of Generalization

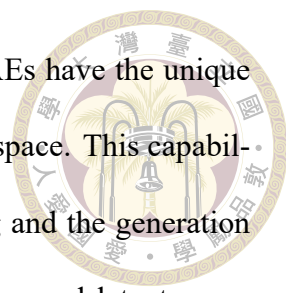
Generalization refers to the ability of deep neural networks to perform effectively when exposed to new, unseen real-world data. A robust model maintains high performance on test datasets that were not part of its training data. To assess generalizability,



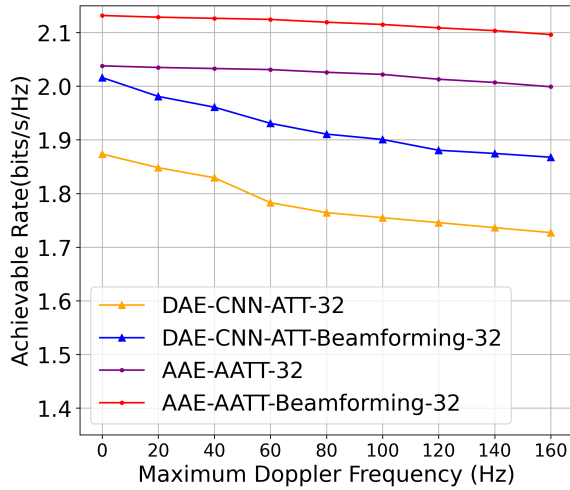
(a) AAE-AATT-Beamforming ($L=1$, $N_t=32$, $N_r=16$), (b) DAE-CNN-ATT-Beamforming ($L=1$, $N_t=32$, $N_r=16$).

Figure 5.8: Comparison of SNR versus SINR between different methods.

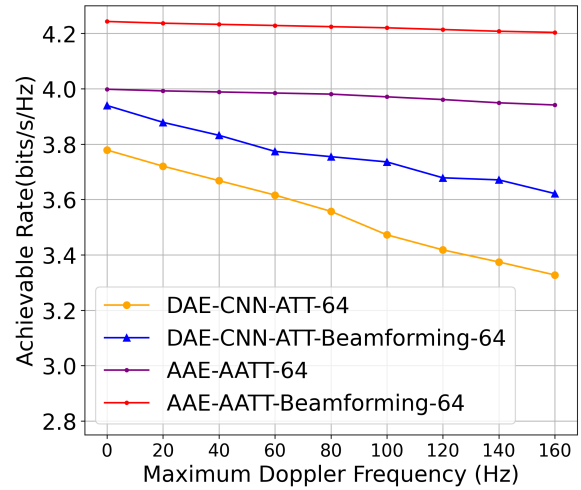
we introduce modifications to different Doppler frequencies. In Fig. 5.9a, we present achievable rate comparison results for the AAE-AATT-Beamforming and DAE-CNN-ATT-Beamforming models with and without beamforming for 32 antennas. The achievable rate comparison between DAE-CNN-ATT-32 and DAE-CNN-ATT-Beamforming-32 reveals a 8.56% difference. Achievable rate comparison results for DAE-CNN-ATT-Beamforming-32 are 8.34%, while those for AAE-AATT-32 are 1.72%. The performance difference in achievable rate is 2.13% for AAE-AATT-Beamforming-32 and 7.82% for DAE-CNN-ATT-Beamforming-32. In Fig. 5.9b, we showcase achievable rate comparison results for the AAE-AATT-Beamforming and DAE-CNN-ATT-Beamforming models with and without beamforming for 64 antennas. Here, we observe that achievable rate comparison results for DAE-CNN-ATT-64 are 7.16%, while those for DAE-CNN-ATT-Beamforming-64 are 4.86%. The achievable rate performance difference for AAE-AATT-Beamforming-64 is 2.18%. Notably, the achievable rate performance difference between AAE-AATT-Beamforming-64 and DAE-CNN-ATT-Beamforming-64 is 4.63%. This behavior can be attributed to the AAE model, channel data is input and processed through

The logo of National Taiwan University (NTU) is located in the upper right quadrant of the page. It is a circular emblem with a gold border. Inside the circle, there are two stylized figures holding a banner, and the university's name is written in Chinese characters around the perimeter.

an encoder, which maps the input data into the bottleneck layer. AAEs have the unique ability to generate realistic data samples based on this learned latent space. This capability proves valuable in various applications such as signal processing and the generation of new data points resembling the training data. Importantly, the learned latent space plays a pivotal role in capturing essential features for intra-channel beamforming in both frequency and spatial domains. And, additive attention mechanism effectively captures global inter-channel data, including crucial time domain information. This capability allows the model to gain a deep understanding of intricate relationships, dependencies, and interactions among different sequences of channel signal subcarriers. In the analog precoding module, an autoencoder is employed to optimize the latent space and faithfully reconstruct it to match the original input channel signal data that greatly enhancing the accuracy of capturing global domain features. The digital precoding module utilizes a 2D-CNN with characteristics like local neuron connectivity, translation invariance, and weight sharing. This configuration excels at capturing crucial frequency and spatial features necessary for channel precoding while effectively reducing interference. Within the analog beamforming module, a GRU is utilized, featuring two gating mechanisms: the reset gate and the update gate. These gates are essential in controlling the flow of information within the cell. They enable the network to make informed decisions about retaining information from the previous time step and incorporating new information, effectively modeling complex temporal dependencies. Importantly, the update gate addresses the vanishing gradient problem. This mechanism significantly enhances the accuracy of feature capture within the analog beamforming module. The digital beamforming module employs a 1D-CNN, which excels at capturing local patterns and dependencies within sequential data. This configuration is particularly well-suited for tasks like time series



(a) ($L=1, N_t=32, N_r=16$).



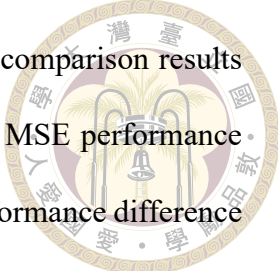
(b) ($L=1, N_t=64, N_r=32$).

Figure 5.9: The generalization evaluation of different maximum Doppler frequencies between different methods.

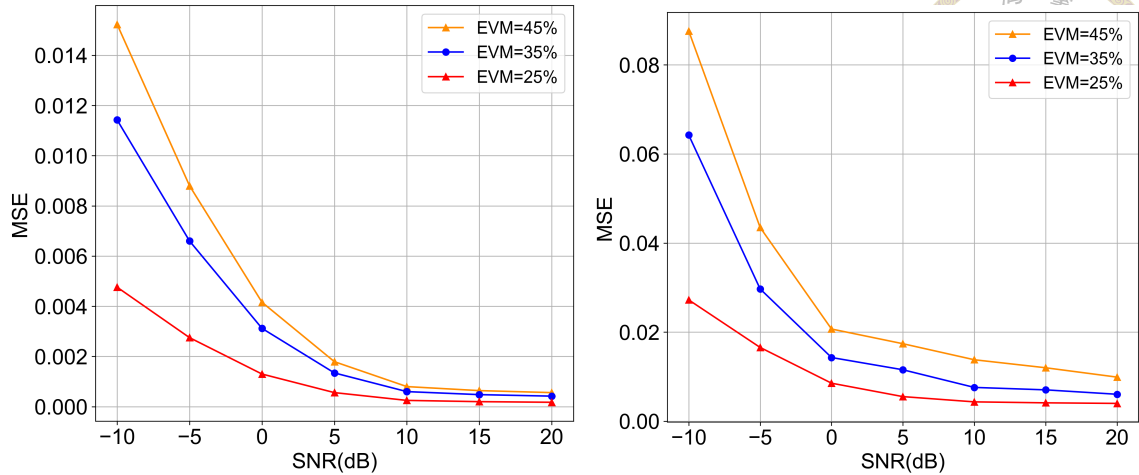
analysis. It effectively captures the necessary features for channel beamforming while minimizing interference.

5.6 The Investigation of Robustness

Robustness in the context of deep neural networks pertains to their ability to adapt to perturbations in source signals. It measures the network's capacity to maintain reliability in the face of real-world imperfections such as noisy inputs, distributional shifts, and adversarial attacks. Robust models excel at anomaly detection and consistently produce accurate predictions even in the presence of minor disturbances. To evaluate robustness, We evaluate the robustness by modifying different EVM (Error Vector Magnitude). In Fig. 5.10a, Fig. 5.10a compares the MSE performance of the proposed AAE-AATT-Beamforming model with different EVM, we can observe that the MSE performance difference between AAE-EVM=45% and AAE-EVM=35% is 17.48%, and the MSE performance difference between AAE-EVM= 35% and AAE-EVM=25% the MSE performance difference is 24.53%. Fig. 5.10b shows MSE comparison results of DAE-CNN-ATT-



Beamforming model with different EVM, we can observe that MSE comparison results between CNN-EVM=45% and CNN-EVM=35% is 26.43%, and the MSE performance difference between CNN-EVM=35% and CNN-EVM=25% MSE performance difference is 34.81%, and the MSE performance difference between AAE-EVM=25% and CNN-EVM=25% is 71.32%. The reason is that AAE have the unique ability to generate realistic data samples based on this learned latent space. This capability proves valuable in various applications such as signal processing and the generation of new data points resembling the training data. Importantly, the learned latent space plays a pivotal role in capturing essential features for intra-channel beamforming in both frequency and spatial domains. At the same time, additive attention is a powerful mechanism for modeling global relationships among elements within an input sequence. It operates without imposing rigid assumptions regarding the nature of these relationships, making it highly adaptable for capturing diverse dependencies. In the context of channel data analysis, this layer assumes a crucial role by effectively capturing inter-channel features in the time domain characteristics, with a specific emphasis on the intricate and complex interdependencies present in the channel data. In the analog precoding module, an autoencoder is used to optimize the latent space and faithfully reconstruct it to match the original input channel signal data that resulting in a significant enhancement in the accuracy of capturing global domain features. In the digital precoding module, a 2D-CNN is employed, characterized by features like local neuron connectivity, translation invariance, and weight sharing. This configuration excels at capturing crucial frequency and spatial features essential for channel precoding while concurrently reducing interference. Within the analog beamforming module, a GRU is utilized that make decisions regarding information retention from the previous time step and the incorporation of new information. The GRU effectively models complex



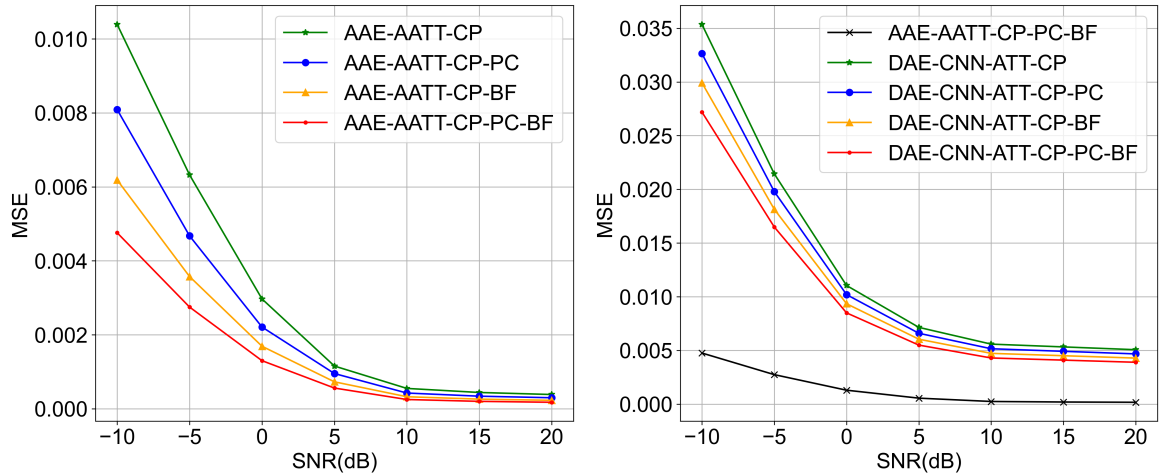
(a) AAE-AATT-Beamforming ($L=1$, $N_t=32$), (b) DAE-CNN-ATT-Beamforming ($L=1$, $N_r=16$).

Figure 5.10: Comparison of MSE performance against the impact of different EVM between different methods.

temporal dependencies, with the update gate addressing the vanishing gradient problem, thereby significantly enhancing feature capture accuracy within the analog beamforming module. Lastly, the digital beamforming module utilizes a 1D-CNN, which excels at capturing local patterns and dependencies within sequential data. This configuration is highly effective in capturing the necessary features for channel beamforming while minimizing interference.

5.7 The Investigation of System Architecture Ablation

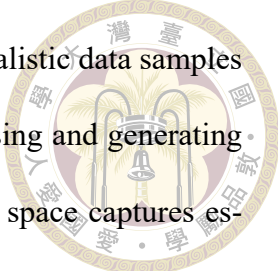
The Fig.5.11a shows MSE comparison results of the AAE-AATT-Beamforming model with different system architecture ablation, and we can observe that the MSE performance difference between AAE-AATT-Channel prediction-Precoding-Beamforming (AAE-AATT-CP-PC-BF) and (AAE-AATT-CP-BF) is 11.43%. The MSE performance difference between (AAE-AATT-CP-BF) and (AAE-AATT-CP-PC) is 12.36%. The MSE performance difference between (AAE-AATT-CP-PC) and (AAE-AATT-CP) is 13.68%. Fig.5.11b



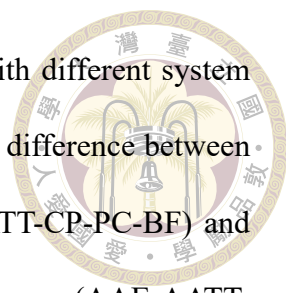
(a) AAE-AATT-Beamforming ($L=1$, $N_t=32$, $N_r=16$), (b) DAE-CNN-ATT-Beamforming ($L=1$, $N_t=32$, $N_r=16$).

Figure 5.11: Comparison of MSE performance against the impact of different system architecture ablation.

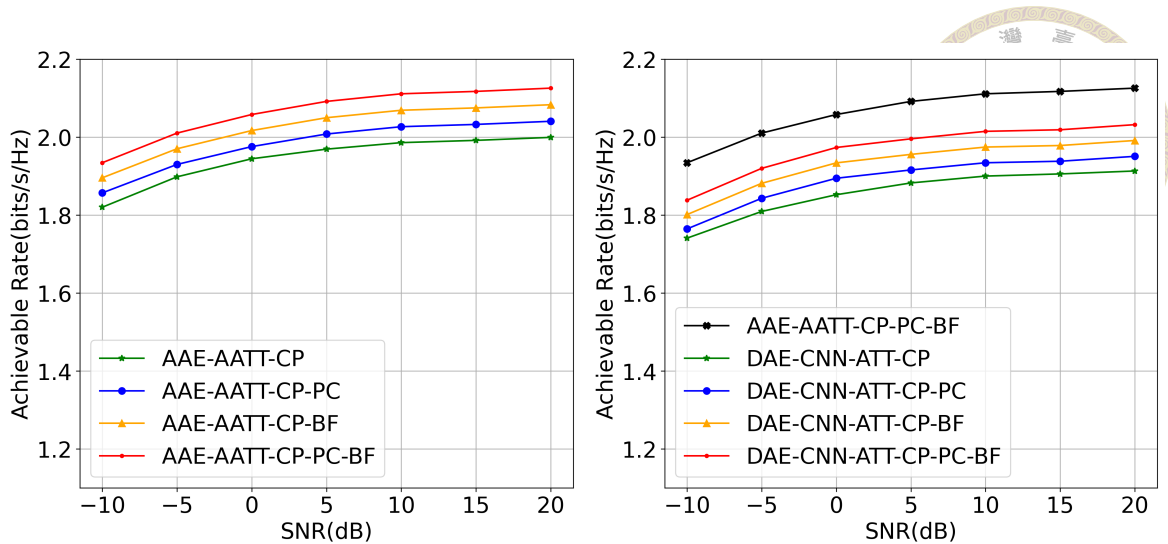
shows MSE comparison results of DAE-CNN-ATT-Beamforming model with different system architecture ablation, and we can observe that the MSE performance difference between DAE-CNN-ATT-Channel prediction-Precoding-Beamforming (DAE-CNN-ATT-CP-PC-BF) and (DAE-CNN-ATT-CP-BF) is 23.57%. The MSE performance difference between (DAE-CNN-ATT-CP-BF) and (DAE-CNN-ATT-CP-PC) is 31.61%. The MSE performance difference between (DAE-CNN-ATT-CP-PC) and (DAE-CNN-ATT-CP) is 36.74%. The reason is that AAE-AATT-Beamforming addresses various aspects of interference and channel optimization in the following ways: To tackle interference that varies with time and frequency across different channels, we leverage additive attention, a versatile mechanism that captures global relationships within input sequences. It excels at modeling diverse dependencies without imposing rigid assumptions. In the context of channel data analysis, it plays a vital role by effectively capturing inter-channel features, especially in the time domain, and addressing complex interdependencies within the channel data. Our AAE-AATT-Beamforming model also addresses interference specific to sub-carriers. It does so by processing channel data through an autoencoder within the AAE



model. This autoencoder-based approach enables the generation of realistic data samples from a learned latent space, which proves valuable for signal processing and generating data resembling the training dataset. Importantly, this learned latent space captures essential features for intra-channel beamforming, both in terms of frequency and spatial domains. In the analog precoding module, our model employs an autoencoder to optimize the latent space and faithfully reconstruct it to match the original input channel signal data that significantly improving the accuracy of capturing global domain features. In contrast, the digital precoding module utilizes a 2D-CNN with characteristics such as local neuron connectivity, translation invariance, and weight sharing. These features enable it to capture crucial frequency and spatial features essential for channel precoding while minimizing interference. Within the analog beamforming module, we utilize the GRU featuring reset and update gates that enabling decisions about retaining information from the previous time step and incorporating new data. This mechanism effectively models complex temporal dependencies and mitigates the vanishing gradient problem. Consequently, it enhances the accuracy of feature capture within the analog beamforming module. For digital beamforming, our model employs a 1D-CNN known for its ability to capture local patterns and dependencies within sequential data. This configuration is particularly suited for tasks such as time series analysis. It proves advantageous in capturing the necessary features for channel beamforming while minimizing interference. And, our AAE-AATT-Beamforming model employs a combination of techniques, including additive attention, autoencoders, and specialized neural network architectures, to address various types of interference and optimize channel performance across time, frequency, and spatial domains. This multifaceted approach enhances the accuracy and efficiency of channel data analysis and beamforming processes. Fig.5.12a shows achievable rate com-



parison results of the proposed AAE-AATT-Beamforming model with different system architecture ablation, and we can observe that the MSE performance difference between AAE-AATT-Channel prediction-Precoding-Beamforming (AAE-AATT-CP-PC-BF) and (AAE-AATT-CP-BF) is 9.65%. The MSE performance difference between (AAE-AATT-CP-BF) and (AAE-AATT-CP-PC) is 13.52%. The MSE performance difference between (AAE-AATT-CP-PC) and (AAE-AATT-CP) is 14.73%. Fig.5.12b shows achievable rate comparison results of DAE-CNN-ATT-Beamforming model with different system architecture ablation, and we can observe that the MSE performance difference between (DAE-CNN-ATT-CP-PC-BF) and (DAE-CNN-ATT-CP-BF) is 25.46%. The MSE performance difference between (DAE-CNN-ATT-CP-BF) and (DAE-CNN-ATT-CP-PC) is 34.21%. The MSE performance difference between (DAE-CNN-ATT-CP-PC) and (DAE-CNN-ATT-CP) is 37.12%. The reason is that AAE-AATT-Beamforming framework addresses interference that varies over time and frequency across different channels. It leverages additive attention, a potent mechanism capable of modeling intricate relationships among elements within an input sequence without imposing rigid assumptions. This adaptability is particularly valuable for capturing diverse dependencies present in the complex inter-channel relationships. In the context of channel data analysis, additive attention plays a pivotal role by effectively capturing interchannel features in the time domain, emphasizing the intricate and complex dependencies in the channel data. Furthermore, our AAE-AATT-Beamforming system mitigates subcarrier channel interference. In the AAE model, channel data is input and processed through an encoder, which maps the data into a latent space representation. AAEs possess a unique ability to generate realistic data samples based on this learned latent space, proving valuable in applications like signal processing and generating data points resembling the training data. Crucially, the learned latent space



(a) AAE-AATT-Beamforming ($L=1$, $N_t=32$, $N_r=16$), (b) DAE-CNN-ATT-Beamforming ($L=1$, $N_t=32$, $N_r=16$).

Figure 5.12: Comparison of achievable rate performance against the impact of different system architecture ablation.

is instrumental in capturing essential features for intra-channel beamforming, spanning both frequency and spatial domains. Within the analog precoding module, an autoencoder optimizes the latent space and reconstructs it as faithfully as possible to the original input channel signal data that resulting in significantly improved accuracy in capturing global domain features. The digital precoding module employs a 2D-CNN with features like local neuron connectivity, translation invariance, and weight sharing. This configuration excels at capturing crucial frequency and spatial features essential for channel precoding while simultaneously reducing interference. In the analog beamforming module, the GRU is utilized that allowing the network to decide whether to retain information from the previous time step or incorporate new data. The GRU effectively models complex temporal dependencies and addresses the vanishing gradient problem through the update gate, significantly enhancing feature capture accuracy. Finally, the digital beamforming module employs a 1D-CNN, which excels at capturing local patterns in sequential data. This configuration is suitable for time series analysis and proves advantageous in capturing the necessary features for channel beamforming while minimizing interference.



Chapter 6 Conclusions

6.1 Conclusions

In the conclusions, we proposed the AAE-AATT-Beamforming in IRS-assisted mm-wave MIMO OFDM system. Additive attention is a potent mechanism for modeling global relationships within input sequences in the time domain. The AAE-learned latent space plays a central role in capturing crucial features for intra-channel beamforming in both frequency and spatial domains. In the analog precoding module, an autoencoder is used to optimize the latent space and faithfully reconstruct it to match the original input channel signal data, significantly enhancing the accuracy of capturing global domain features. The digital precoding module employs a 2D-CNN with features like translation invariance, enabling it to capture vital frequency and spatial features for channel precoding while minimizing interference. Within the analog beamforming module, GRU is utilized, featuring reset and update gates that control information flow within the cell. This enhances feature capture accuracy. The digital beamforming module employs a 1D-CNN, proficient in capturing local patterns and dependencies in sequential data, making it well-suited for tasks like time series analysis. This module effectively captures essential channel beamforming features while reducing interference. Numerical results showcase significant improvements in MSE, achievable rate, generalizability, and robustness com-

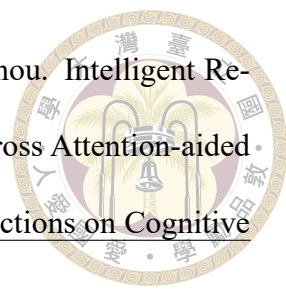
pared to prior research[18, 25, 49, 54].



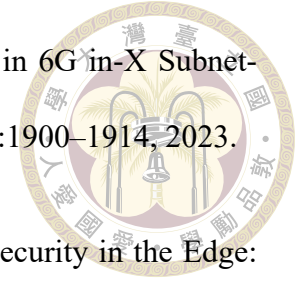


References

- [1] N. A. Abbasi, J. L. Gomez, R. Kondaveti, S. M. Shaikbepari, S. Rao, S. Abu-Surra, G. Xu, J. Zhang, and A. F. Molisch. THz Band Channel Measurements and Statistical Modeling for Urban D2D Environments. IEEE Transactions on Wireless Communications, 22(3):1466–1479, 2023.
- [2] L. Bai, Z. Huang, and X. Cheng. A Non-Stationary Model With Time-Space Consistency for 6G Massive MIMO mmWave UAV Channels. IEEE Transactions on Wireless Communications, 22(3):2048–2064, 2023.
- [3] A. Bazzi and M. Chafii. On Outage-Based Beamforming Design for Dual-Functional Radar-Communication 6G Systems. IEEE Transactions on Wireless Communications, 22(8):5598–5612, 2023.
- [4] G. Chen and Q. Wu. Fundamental Limits of Intelligent Reflecting Surface Aided Multiuser Broadcast Channel. IEEE Transactions on Communications, 71(10):5904–5919, 2023.
- [5] H.-Y. Chen, M.-H. Wu, T.-W. Yang, C.-W. Huang, and C.-F. Chou. Attention-aided Autoencoder-based Channel Prediction for Intelligent Reflecting Surface-Assisted Millimeter Wave Communications. IEEE Transactions on Green Communications and Networking, pages 1–1, 2023.

- 
- [6] H.-Y. Chen, M.-H. Wu, T.-W. Yang, C.-W. Huang, and C.-F. Chou. Intelligent Reflecting Surface-Assisted Millimeter Wave Communications: Cross Attention-aided Variational Autoencoder-based Precoding Design. IEEE Transactions on Cognitive Communications and Networking, pages 1–1, 2024.
- [7] M. Chu, A. Liu, V. K. N. Lau, C. Jiang, and T. Yang. Deep Reinforcement Learning Based End-to-End Multiuser Channel Prediction and Beamforming. IEEE Transactions on Wireless Communications, 21(12):10271–10285, 2022.
- [8] M. Cui, L. Dai, Z. Wang, S. Zhou, and N. Ge. Near-Field Rainbow: Wideband Beam Training for XL-MIMO. IEEE Transactions on Wireless Communications, 22(6):3899–3912, 2023.
- [9] F. de Oliveira Torres, V. A. de Santiago Júnior, D. B. da Costa, D. L. Cardoso, and R. C. L. de Oliveira. Throughput Maximization for a Multicarrier Cell-Less NOMA Network: A Framework Based on Ensemble Metaheuristics. IEEE Transactions on Wireless Communications, 22(1):348–361, 2023.
- [10] F. Dong, F. Liu, Y. Cui, W. Wang, K. Han, and Z. Wang. Sensing as a Service in 6G Perceptive Networks: A Unified Framework for ISAC Resource Allocation. IEEE Transactions on Wireless Communications, 22(5):3522–3536, 2023.
- [11] W. Du, Z. Chu, G. Chen, P. Xiao, Z. Lin, C. Huang, and W. Hao. Weighted Sum-Rate and Energy Efficiency Maximization for Joint ITS and IRS Assisted Multiuser MIMO Networks. IEEE Transactions on Communications, 70(11):7351–7364, 2022.
- [12] X. Du, T. Wang, Q. Feng, C. Ye, T. Tao, L. Wang, Y. Shi, and M. Chen. Multi-Agent

Reinforcement Learning for Dynamic Resource Management in 6G in-X Subnetworks. IEEE Transactions on Wireless Communications, 22(3):1900–1914, 2023.



[13] Z. M. Fadlullah, B. Mao, and N. Kato. Balancing QoS and Security in the Edge: Existing Practices, Challenges, and 6G Opportunities With Machine Learning. IEEE Communications Surveys And Tutorials, 24(4):2419–2448, 2022.


[14] X. Fang, W. Feng, Y. Wang, Y. Chen, N. Ge, Z. Ding, and H. Zhu. NOMA-Based Hybrid Satellite-UAV-Terrestrial Networks for 6G Maritime Coverage. IEEE Transactions on Wireless Communications, 22(1):138–152, 2023.

[15] Z. Gao, M. Wu, C. Hu, F. Gao, G. Wen, D. Zheng, and J. Zhang. Data-Driven Deep Learning Based Hybrid Beamforming for Aerial Massive MIMO-OFDM Systems With Implicit CSI. IEEE Journal on Selected Areas in Communications, 40(10):2894–2913, 2022.

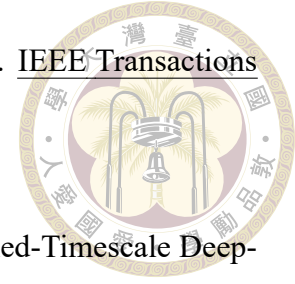
[16] J. Guo, C.-K. Wen, and S. Jin. Deep Learning-Based CSI Feedback for Beamforming in Single- and Multi-Cell Massive MIMO Systems. IEEE Journal on Selected Areas in Communications, 39(7):1872–1884, 2021.

[17] M. Gupta, I. P. Roberts, and J. G. Andrews. System-Level Analysis of Full-Duplex Self-Backhauled Millimeter Wave Networks. IEEE Transactions on Wireless Communications, 22(2):1130–1144, 2023.

[18] Y. Hei, C. Liu, W. Li, L. Ma, and M. Lan. CNN Based Hybrid Precoding for MmWave MIMO Systems With Adaptive Switching Module and Phase Modulation Array. IEEE Transactions on Wireless Communications, 21(12):10489–10501, 2022.

- 
- [19] J. Hu, T. Shui, L. Xiang, and K. Yang. Multi-Domain Resource Scheduling for Simultaneous Wireless Computing and Power Transfer in Fog Radio Access Network. IEEE Transactions on Communications, 71(1):269–281, 2023.
- [20] Q. Hu, S. Shi, Y. Cai, and G. Yu. DDPG-Driven Deep-Unfolding With Adaptive Depth for Channel Estimation With Sparse Bayesian Learning. IEEE Transactions on Signal Processing, 70:4665–4680, 2022.
- [21] W. Huang, W. Ding, C. Kai, Y. Yi, and Y. Huang. Joint Placement and Beamforming Design for IRS-Enhanced Multiuser MISO Systems. IEEE Transactions on Communications, 70(10):6678–6692, 2022.
- [22] Y. Huang, L. Zhu, and R. Zhang. Integrating Intelligent Reflecting Surface Into Base Station: Architecture, Channel Model, and Passive Reflection Design. IEEE Transactions on Communications, 71(8):5005–5020, 2023.
- [23] J. M. J. Huttunen, D. Korpi, and M. Honkala. DeepTx: Deep Learning Beamforming With Channel Prediction. IEEE Transactions on Wireless Communications, 22(3):1855–1867, 2023.
- [24] T. Ji, M. Hua, C. Li, Y. Huang, and L. Yang. Robust Max-Min Fairness Transmission Design for IRS-Aided Wireless Network Considering User Location Uncertainty. IEEE Transactions on Communications, 71(8):4678–4693, 2023.
- [25] H. Jiang, Y. Lu, X. Li, B. Wang, Y. Zhou, and L. Dai. Attention-Based Hybrid Precoding for mmWave MIMO Systems. In 2021 IEEE Information Theory Workshop (ITW), pages 1–6, 2021.
- [26] J. Johnston and X. Wang. Model-Based Deep Learning for Joint Activity Detection

and Channel Estimation in Massive and Sporadic Connectivity. IEEE Transactions on Wireless Communications, 21(11):9806–9817, 2022.



[27] K. Kang, Q. Hu, Y. Cai, G. Yu, J. Hoydis, and Y. C. Eldar. Mixed-Timescale Deep-Unfolding for Joint Channel Estimation and Hybrid Beamforming. IEEE Journal on Selected Areas in Communications, 40(9):2510–2528, 2022.

[28] V. Kumar, R. Zhang, M. D. Renzo, and L.-N. Tran. A Novel SCA-Based Method for Beamforming Optimization in IRS/RIS-Assisted MU-MISO Downlink. IEEE Wireless Communications Letters, 12(2):297–301, 2023.

[29] C. Liao, F. Wang, and V. K. N. Lau. Optimized Design for IRS-Assisted Integrated Sensing and Communication Systems in Clutter Environments. IEEE Transactions on Communications, 71(8):4721–4734, 2023.

[30] C. Liu, Z. Wei, D. W. K. Ng, J. Yuan, and Y.-C. Liang. Deep Transfer Learning for Signal Detection in Ambient Backscatter Communications. IEEE Transactions on Wireless Communications, 20(3):1624–1638, 2021.

[31] C. Liu, W. Yuan, S. Li, X. Liu, H. Li, D. W. K. Ng, and Y. Li. Learning-Based Predictive Beamforming for Integrated Sensing and Communication in Vehicular Networks. IEEE Journal on Selected Areas in Communications, 40(8):2317–2334, 2022.

[32] H. Liu, Y. Zhang, X. Zhang, M. El-Hajjar, and L.-L. Yang. Deep Learning Assisted Adaptive Index Modulation for mmWave Communications With Channel Estimation. IEEE Transactions on Vehicular Technology, 71(9):9186–9201, 2022.

[33] P. Liu, Y. Li, W. Cheng, X. Dong, and L. Dong. Active Intelligent Reflecting Surface

Aided RSMA for Millimeter-Wave Hybrid Antenna Array. IEEE Transactions on Communications, 71(9):5287–5302, 2023.



[34] X. Liu, Y. Deng, and T. Mahmoodi. Wireless Distributed Learning: A New Hybrid Split and Federated Learning Approach. IEEE Transactions on Wireless Communications, 22(4):2650–2665, 2023.

[35] Y. Lu and L. Dai. Near-Field Channel Estimation in Mixed LoS/NLoS Environments for Extremely Large-Scale MIMO Systems. IEEE Transactions on Communications, 71(6):3694–3707, 2023.

[36] Y. Ma, M. Li, Y. Liu, Q. Wu, and Q. Liu. Optimization for Reflection and Transmission Dual-Functional Active RIS-Assisted Systems. IEEE Transactions on Communications, 71(9):5534–5548, 2023.

[37] S. Mourya, S. Amuru, and K. K. Kuchi. A Spatially Separable Attention Mechanism for Massive MIMO CSI Feedback. IEEE Wireless Communications Letters, 12(1):40–44, 2023.

[38] X. Ou, X. Xie, H. Lu, and H. Yang. Resource Allocation in MU-MISO Rate-Splitting Multiple Access With SIC Errors for URLLC Services. IEEE Transactions on Communications, 71(1):229–243, 2023.

[39] C. Pan, G. Zhou, K. Zhi, S. Hong, T. Wu, Y. Pan, H. Ren, M. D. Renzo, A. Lee Swindlehurst, R. Zhang, and A. Y. Zhang. An Overview of Signal Processing Techniques for RIS/IRS-Aided Wireless Systems. IEEE Journal of Selected Topics in Signal Processing, 16(5):883–917, 2022.

[40] H. Pan, Y. Liu, G. Sun, J. Fan, S. Liang, and C. Yuen. Joint Power and 3D Trajectory

Optimization for UAV-Enabled Wireless Powered Communication Networks With Obstacles. IEEE Transactions on Communications, 71(4):2364–2380, 2023.



[41] R. H. Y. Perdana, T.-V. Nguyen, and B. An. Adaptive User Pairing in Multi-IRS-Aided Massive MIMO-NOMA Networks: Spectral Efficiency Maximization and Deep Learning Design. IEEE Transactions on Communications, 71(7):4377–4390, 2023.


[42] Q. Qi, X. Chen, A. Khalili, C. Zhong, Z. Zhang, and D. W. K. Ng. Integrating Sensing, Computing, and Communication in 6G Wireless Networks: Design and Optimization. IEEE Transactions on Communications, 70(9):6212–6227, 2022.


[43] X. Qin, T. Ma, Z. Tang, X. Zhang, H. Zhou, and L. Zhao. Service-Aware Resource Orchestration in Ultra-Dense LEO Satellite-Terrestrial Integrated 6G: A Service Function Chain Approach. IEEE Transactions on Wireless Communications, 22(9):6003–6017, 2023.


[44] J. Qiu, J. Yu, A. Dong, and K. Yu. Joint Beamforming for IRS-Aided Multi-Cell MISO System: Sum Rate Maximization and SINR Balancing. IEEE Transactions on Wireless Communications, 21(9):7536–7549, 2022.

[45] I. P. Roberts, S. Vishwanath, and J. G. Andrews. LoneSTAR: Analog Beamforming Codebooks for Full-Duplex Millimeter Wave Systems. IEEE Transactions on Wireless Communications, 22(9):5754–5769, 2023.

[46] S. H. A. Shah and S. Rangan. LSTM-Aided Selective Beam Tracking in Multi-Cell Scenario for mmWave Wireless Systems. IEEE Transactions on Wireless Communications, pages 1–1, 2023.

- 
- [47] L. Sun, Y. Wang, A. L. Swindlehurst, and X. Tang. Generative-Adversarial-Network Enabled Signal Detection for Communication Systems With Unknown Channel Models. IEEE Journal on Selected Areas in Communications, 39(1):47–60, 2021.
- [48] J. Tang, Z. Peng, D. K. C. So, X. Zhang, K.-K. Wong, and J. A. Chambers. Energy Efficiency Optimization for a Multiuser IRS-Aided MISO System With SWIPT. IEEE Transactions on Communications, 71(10):5950–5962, 2023.
- [49] J. Tao, J. Chen, J. Xing, S. Fu, and J. Xie. Autoencoder Neural Network Based Intelligent Hybrid Beamforming Design for mmWave Massive MIMO Systems. IEEE Transactions on Cognitive Communications and Networking, 6(3):1019–1030, 2020.
- [50] H. Wang, J. Fang, H. Duan, and H. Li. Spatial Channel Covariance Estimation and Two-Timescale Beamforming for IRS-Assisted Millimeter Wave Systems. IEEE Transactions on Wireless Communications, 22(9):6048–6060, 2023.
- [51] J. Wang, L. Dai, L. Yang, and B. Bai. Clustered Cell-Free Networking: A Graph Partitioning Approach. IEEE Transactions on Wireless Communications, 22(8):5349–5364, 2023.
- [52] J. Wang, W. Zhang, Y. Chen, Z. Liu, J. Sun, and C.-X. Wang. Time-Varying Channel Estimation Scheme for Uplink MU-MIMO in 6G Systems. IEEE Transactions on Vehicular Technology, 71(11):11820–11831, 2022.
- [53] K. Wang, H. Li, Z. Ding, and P. Xiao. Reinforcement Learning Based Latency Minimization in Secure NOMA-MEC Systems With Hybrid SIC. IEEE Transactions on Wireless Communications, 22(1):408–422, 2023.

- 
- [54] T. Wang, C.-K. Wen, S. Jin, and G. Y. Li. Deep Learning-Based CSI Feedback Approach for Time-Varying Massive MIMO Channels. IEEE Wireless Communications Letters, 8(2):416–419, 2019.
- [55] Y. Wang, G. Wang, R. He, B. Ai, and C. Tellambura. Doppler Shift and Channel Estimation for Intelligent Transparent Surface Assisted Communication Systems on High-Speed Railways. IEEE Transactions on Communications, 71(7):4204–4215, 2023.
- [56] J. Wu, S. Kim, and B. Shim. Parametric Sparse Channel Estimation for RIS-Assisted Terahertz Systems. IEEE Transactions on Communications, 71(9):5503–5518, 2023.
- [57] Z. Xie, W. Chen, and H. V. Poor. A Unified Framework for Pushing in Two-Tier Heterogeneous Networks With mmWave Hotspots. IEEE Transactions on Wireless Communications, 22(1):19–31, 2023.
- [58] S. Xu, Y. Cao, C. Li, D. Wang, and L. Yang. Spanning Tree Method for Over-the-Air Channel Calibration in 6G Cell-Free Massive MIMO. IEEE Transactions on Wireless Communications, 22(8):5567–5582, 2023.
- [59] W. Yan, W. Hao, C. Huang, G. Sun, O. Muta, H. Gacanin, and C. Yuen. Beamforming Analysis and Design for Wideband THz Reconfigurable Intelligent Surface Communications. IEEE Journal on Selected Areas in Communications, 41(8):2306–2320, 2023.
- [60] J. Yaswanth, S. K. Singh, K. Singh, and M. F. Flanagan. Energy-Efficient Beamforming Design for RIS-Aided MIMO Downlink Communication With SWIPT. IEEE Transactions on Green Communications and Networking, 7(3):1164–1180, 2023.

- 
- [61] S. Yu, W. Chen, and H. V. Poor. Real-Time Monitoring With Timing Side Information. IEEE Transactions on Communications, 71(4):1953–1969, 2023.
- [62] Q. Yuan, H. Liu, M. Xu, Y. Wu, L. Xiao, and T. Jiang. Deep Learning-Based Hybrid Precoding for Terahertz Massive MIMO Communication With Beam Squint. IEEE Communications Letters, 27(1):175–179, 2023.
- [63] Y. Zhang, J. Sun, J. Xue, G. Y. Li, and Z. Xu. Deep Expectation-Maximization for Joint MIMO Channel Estimation and Signal Detection. IEEE Transactions on Signal Processing, 70:4483–4497, 2022.
- [64] Z. Zhang, L. Dai, X. Chen, C. Liu, F. Yang, R. Schober, and H. V. Poor. Active RIS vs. Passive RIS: Which Will Prevail in 6G? IEEE Transactions on Communications, 71(3):1707–1725, 2023.
- [65] J. Zheng, H. Zhang, J. Kang, L. Gao, J. Ren, and D. Niyato. Covert Federated Learning via Intelligent Reflecting Surfaces. IEEE Transactions on Communications, 71(8):4591–4604, 2023.
- [66] F. Zhou, C. Wang, G. Wu, Y. Wu, Q. Wu, and N. Al-Dhahir. Accurate Spectrum Map Construction for Spectrum Management Through Intelligent Frequency-Spatial Reasoning. IEEE Transactions on Communications, 71(7):3932–3945, 2023.