

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA ĐIỆN TỬ - VIỄN THÔNG

NGUYỄN HUY HOÀNG

**BEAMFORMING TRONG HỆ THỐNG MASSIVE MIMO
SỬ DỤNG DEEP LEARNING**

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN
NGÀNH KỸ THUẬT ĐIỆN TỬ - VIỄN THÔNG
CHUYÊN NGÀNH VIỄN THÔNG - MẠNG

Thành phố Hồ Chí Minh - Năm 2025

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA ĐIỆN TỬ - VIỄN THÔNG

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN
NGÀNH KỸ THUẬT ĐIỆN TỬ - VIỄN THÔNG
CHUYÊN NGÀNH VIỄN THÔNG - MẠNG

BEAMFORMING TRONG HỆ THỐNG MASSIVE MIMO
SỬ DỤNG DEEP LEARNING

Họ và tên sinh viên: Nguyễn Huy Hoàng
Mã số sinh viên: 21200091

NGƯỜI HƯỚNG DẪN KHOA HỌC:
1. ThS. Nguyễn Thái Công Nghĩa
2. HVCH. Vũ Thảo Vi

Thành phố Hồ Chí Minh - Năm 2025

XÁC NHẬN HOÀN TẤT CHỈNH SỬA BÁO CÁO TỐT NGHIỆP

TÊN ĐỀ TÀI: BEAMFORMING TRONG HỆ THỐNG MASSIVE MIMO SỬ DỤNG DEEP LEARNING

Tên tiếng Anh: BEAMFORMING FOR MASSIVE MIMO SYSTEMS USING DEEP LEARNING

Họ và tên sinh viên: Nguyễn Huy Hoàng

Mã số sinh viên: 21200091

Hội đồng đánh giá họp ngày tháng năm

1. Chủ tịch: TS. Đặng Lê Khoa
2. Thư ký: ThS. Nguyễn Thị Xuân Uyên
3. Ủy viên: TS. Nguyễn Minh Trí
4. Ủy viên: ThS. Trương Tấn Quang
5. Ủy viên: ThS. Ngô Minh Nghĩa

Người hướng dẫn: ThS. Nguyễn Thái Công Nghĩa, HVCH. Vũ Thảo Vi

Xác nhận của người hướng dẫn:

LỜI CẢM ƠN

Em xin gửi lời cảm ơn chân thành nhất tới những người đã đồng hành và hỗ trợ em trong suốt quá trình tìm hiểu và hoàn thành khóa luận. Đặc biệt, em xin gửi lời cảm ơn chân thành nhất tới GV. ThS. Nguyễn Thái Công Nghĩa và HVCH. Vũ Thảo Vi, các thầy cô đã đồng hành và truyền đạt kiến thức quý báu cho em trong suốt thời gian vừa qua.

Từ khi thực hiện đề tài “Beamforming trong hệ thống massive MIMO sử dụng Deep Learning”, em đã nhận ra rằng sự hướng dẫn tận tâm của giảng viên là yếu tố vô cùng quan trọng để thành công trong quá trình thực hiện đề tài. Thầy, cô không chỉ là người hướng dẫn, mà còn là người chia sẻ những kinh nghiệm quý giá cho em.

Trong suốt quá trình, thầy, cô đã cung cấp cho em những ý kiến và định hướng quan trọng. Thầy, cô luôn dành thời gian nghe em trình bày ý tưởng và những khó khăn em đang gặp phải. Nhờ sự tận tâm và kiên nhẫn của thầy, cô, em đã có được sự phát triển và hoàn thiện trong đề tài nghiên cứu của mình.

Em cũng xin cảm ơn đến tất cả các quý thầy cô đã tận tình truyền đạt những tri thức quý báu cho em trong suốt bốn năm học vừa qua và những người bạn đã chia sẻ những niềm vui, nỗi buồn, và những khó khăn trong quá trình học tập.

Cuối cùng, em muốn bày tỏ lòng biết ơn chân thành đến tất cả những người đã đóng góp, hỗ trợ và đồng hành cùng em trong quá trình nghiên cứu và hoàn thành khóa luận này. Sự giúp đỡ của mọi người đã tạo điều kiện thuận lợi để em có thể tiến tới thành công. Em sẽ luôn mang trong lòng những kỷ niệm đáng quý và tôn trọng sự đóng góp của mọi người.

TP. Hồ Chí Minh, ngày tháng năm 2025

Nguyễn Huy Hoàng

LỜI CAM KẾT

Em xin cam đoan rằng các số liệu và kết quả được trình bày trong khoá luận là trung thực và những tài liệu tham khảo được sử dụng trong khoá luận đã được trích dẫn một cách đầy đủ, rõ ràng về nguồn gốc theo quy định.

Em xin chịu trách nhiệm hoàn toàn về tính chính xác và trung thực của các thông tin được trình bày trong khoá luận.

Sinh viên thực hiện

Nguyễn Huy Hoàng

TÓM TẮT

Trong các hệ thống Massive MIMO, việc thiết kế búp sóng hiệu quả đang gặp phải những thách thức về độ phức tạp tính toán, cũng như hiệu suất có thể suy giảm đáng kể khi hoạt động với thông tin trạng thái kênh không hoàn hảo. Khóa luận “Beamforming trong hệ thống Massive MIMO sử dụng Deep Learning” đề xuất một giải pháp dựa trên Deep Learning kết hợp với beamforming, trong đó mô hình mạng neuron tạo búp sóng được huấn luyện theo hướng end-to-end. Mô hình học cách ánh xạ trực tiếp từ thông tin trạng thái kênh không hoàn hảo đến các búp sóng tối ưu bằng cách tối đa hóa trực tiếp hiệu suất phổ. Kết quả mô phỏng trên các cấu hình antenna đa dạng (MISO 1×64 , MIMO 2×32 , MIMO 4×64 , MIMO 8×64) đã khẳng định tính hiệu quả của phương pháp, các kết quả cho thấy hiệu suất phổ được cải thiện đáng kể khi tăng số lượng antenna thu nhờ lợi ích phân tập. Đồng thời, đề tài cũng cho thấy việc huấn luyện trên bộ dữ liệu có chất lượng cao (PNR cao) là yếu tố then chốt để đảm bảo tính mạnh mẽ của mô hình và Deep Learning là một hướng tiềm năng để thiết kế các giải pháp tạo búp sóng có độ trễ thấp cũng như hiệu năng cao cho các hệ thống không dây thế hệ mới.

ABSTRACT

In Massive MIMO systems, designing efficient beamforming faces significant challenges, particularly regarding computational complexity and potential performance degradation when operating with imperfect CSI. The thesis titled "Beamforming in Massive MIMO Systems using Deep Learning" proposes a novel solution that integrates Deep Learning with beamforming. Specifically, a neural network model for beamforming is trained in an end-to-end manner. This model learns to directly map from imperfect CSI to optimal beamforming weights by maximizing spectral efficiency. Simulation results, conducted across various antenna configurations, have validated the effectiveness of the proposed approach. These results demonstrate a substantial improvement in spectral efficiency with an increased number of receive antennas, attributed to diversity gains. Furthermore, the study highlights that training on high-quality datasets (with high PNR) is crucial for ensuring the model's robustness. In summary, Deep Learning presents a promising direction for designing low-latency and high-performance beamforming solutions for next-generation wireless communication systems.

MỤC LỤC

TÓM TẮT	1
MỤC LỤC	2
DANH SÁCH CHỮ VIẾT TẮT.....	4
DANH SÁCH CÁC HÌNH	6
DANH SÁCH CÁC BẢNG.....	7
CHƯƠNG 1: TỔNG QUAN VỀ ỨNG DỤNG DEEP LEARNING TRONG KỸ THUẬT BEAMFORMING	8
1.1. Tổng quan về Beamforming trong hệ thống Massive MIMO	8
1.1.1. Massive MIMO và kỹ thuật Beamforming.....	8
1.1.2. Antenna mảng – Nền tảng vật lý của Beamforming.....	9
1.1.3. Các kiến trúc Beamforming và thách thức phần cứng.....	10
1.2. Hướng tiếp cận học sâu với Beamforming	11
1.3. Mục tiêu của đề tài.....	13
1.4. Phương pháp thực hiện đề tài	14
1.5. Bố cục báo cáo.....	15
1.6. Giới hạn của đề tài	16
CHƯƠNG 2: CƠ SỞ LÝ THUYẾT VỀ KỸ THUẬT BEAMFORMING VÀ DEEP LEARNING	17
2.1. Kỹ thuật Beamforming	17
2.1.1. Digital Beamforming	17
2.1.2. Analog Beamforming	18
2.1.3. Hybrid Beamforming.....	19
2.2. Cơ sở toán học và mô hình hệ thống	20
2.2.1. Mô hình kênh truyền sóng mmWave.....	20
2.2.2. Array Steering Vector	22
2.2.3. Ước tính kênh tuần tự trong môi trường đa đường.....	23
2.3. Bài toán tối ưu hoá Beamforming.....	24
2.3.1. Mục tiêu tối ưu hoá	24
2.3.2. Bài toán tối ưu hóa hiệu suất phổ trong hệ thống Beamforming.....	26
2.4. Deep Learning.....	27

2.4.1. Hàm sai số (Loss Function)	29
2.4.2. Hàm kích hoạt (Activation Function)	29
2.4.3. Mạng thần kinh học sâu (Deep Neural Network)	30
2.4.4. Thuật toán tối ưu	31
2.5. Bài toán tối ưu hoá Analog Beamforming với DNN (ABF DNN).....	32
CHƯƠNG 3: THIẾT KẾ BỘ TẠO BÚP SÓNG TƯƠNG TỰ BẰNG PHƯƠNG PHÁP DEEP LEARNING	33
3.1. Quy trình tạo dữ liệu huấn luyện	33
3.1.1. Chuẩn bị môi trường mô phỏng và tham số hệ thống	33
3.1.2. Tạo kênh truyền và ước lượng kênh	34
3.1.3. Tổ chức bộ dữ liệu huấn luyện	36
3.2. Kiến trúc và hoạt động của ABF DNN.....	36
3.3. Hàm sai số và mục tiêu tối ưu hoá.....	42
3.4. Huấn luyện mô hình.....	42
CHƯƠNG 4: KẾT QUẢ VÀ THẢO LUẬN	44
4.1. Kết quả quá trình huấn luyện.....	44
4.2. Kết quả quá trình đánh giá.....	48
4.2.1. Phân tích và so sánh hiệu năng giữa các cấu hình antenna.....	48
4.2.2. Khảo sát ảnh hưởng của chất lượng dữ liệu huấn luyện.....	49
CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN ĐỀ TÀI	56
5.1. Kết luận.....	56
5.2. Hướng phát triển đề tài	57
TÀI LIỆU THAM KHẢO	58

DANH SÁCH CHỮ VIẾT TẮT

Từ viết tắt	Từ đầy đủ	Ý nghĩa
ABF DNN	Analog Beamforming Deep Neural Network	Bộ tạo búp sóng tương tự sử dụng mạng thần kinh học sâu
AOA	Angle of Arrival	Góc tới của tín hiệu
AOD	Angle of Departure	Góc phát của tín hiệu
AWGN	Additive White Gaussian Noise	Nhiều Gauss trắng cộng
BB	Baseband	Băng tần gốc
BF	Beamforming	Kỹ thuật tạo búp sóng
BFNN	Beamforming Neural Network	Mạng thần kinh tạo búp sóng
BS	Base Station	Trạm gốc
CSI	Channel State Information	Thông tin trạng thái kênh
eMBB	Enhanced Mobile Broadband	Băng rộng tăng cường
MIMO	Multi-Input Multi-Output	Hệ thống đa antenna phát đa antenna thu
MMSE	Minimum Mean Square Error	Lỗi trung bình bình phương tối thiểu
mMTC	Massive Machine Type Communications	Kết nối lượng lớn thiết bị
NMSE	Normalized Mean Squared Error	Lỗi trung bình bình phương chuẩn hóa
OMP	Orthogonal Matching Pursuit	Thuật toán đối sánh trực giao
PNR	Pilot-to-Noise Ratio	Tỉ số công suất pilot trên nhiễu
RF	Radio Frequency	Tần số vô tuyến
SE	Spectral Efficiency	Hiệu suất phổ

SIC	Successive Interference Cancellation	Loại bỏ nhiễu tuần tự
SNR	Signal-to-Noise Ratio	Tỉ số tín hiệu trên nhiễu
UCA	Uniform Circular Array	Mảng tròn đồng nhất
UE	User Equipment	Thiết bị đầu cuối người dùng
ULA	Uniform Linear Array	Mảng tuyến tính đồng nhất
URA	Uniform Rectangular Array	Mảng chữ nhật đồng nhất
URLLC	Ultra-Reliable and Low-Latency Communications	Truyền thông siêu đáng tin cậy và độ trễ cực thấp

DANH SÁCH CÁC HÌNH

Hình 1.1. Minh hoạ kỹ thuật Beamforming.....	9
Hình 1.2. Minh hoạ các loại antenna mảng.	10
Hình 2.1. Kiến trúc Digital Beamforming trong Massive MIMO.....	18
Hình 2.2. Kiến trúc Analog Beamforming trong Massive MIMO.	18
Hình 2.3. Kiến trúc Hybrid Beamforming trong Massive MIMO.....	19
Hình 2.4. Mô hình hệ thống mmWave Massive MIMO với 1 chuỗi RF.	21
Hình 2.5. Mối quan hệ giữa trí tuệ nhân tạo, Machine Learning và Deep Learning.....	28
Hình 2.6. Kiến trúc mạng thần kinh học sâu.	31
Hình 3.1. Minh họa phương pháp thiết kế Analog Beamforming hai giai đoạn.	37
Hình 3.2. Quá trình huấn luyện mô hình với cấu hình 1×64	43
Hình 4.1. Biểu đồ Loss với $\text{PNR} = 10\text{dB}$ tại cấu hình 1×64	45
Hình 4.2. Biểu đồ Loss với $\text{PNR} = 10\text{dB}$ tại cấu hình 2×32	45
Hình 4.3. Biểu đồ Loss với $\text{PNR} = 10\text{dB}$ tại cấu hình 4×64	46
Hình 4.4. Biểu đồ Loss với $\text{PNR} = 10\text{dB}$ tại cấu hình 8×64	47
Hình 4.5. So sánh hiệu suất phổ của mô hình BFNN, $\text{PNR} = 10\text{dB}$ với các cấu hình antenna $1 \times 64, 2 \times 32, 4 \times 64, 8 \times 64$	48
Hình 4.6. So sánh hiệu suất phổ của mô hình BFNN cấu hình 1×64 với các mức PNR khác nhau.	50
Hình 4.7. So sánh hiệu suất phổ của mô hình BFNN cấu hình 2×32 với các mức PNR khác nhau.	51
Hình 4.8. So sánh hiệu suất phổ của mô hình BFNN cấu hình 4×64 với các mức PNR khác nhau.	52
Hình 4.9. So sánh hiệu suất phổ của mô hình BFNN cấu hình 8×64 với các mức PNR khác nhau.	53
Hình 4.10. So sánh hiệu suất phổ của mô hình BFNN cấu hình 1×64 , $\text{PNR} = 20\text{dB}$, với các tham số L_{est} khác nhau.....	55

DANH SÁCH CÁC BẢNG

Bảng 2.1. So sánh các kỹ thuật Beamforming.....	20
Bảng 2.2. Các hàm kích hoạt phổ biến.	30
Bảng 3.1. Các tham số cho quá trình tạo dữ liệu mô phỏng kênh truyền.....	33
Bảng 3.2. Kiến trúc mạng DNN cho Analog Beamforming.	40
Bảng 4.1. Các tham số cho quá trình huấn luyện	44

CHƯƠNG 1: TỔNG QUAN VỀ ỨNG DỤNG DEEP LEARNING TRONG KỸ THUẬT BEAMFORMING

1.1. Tổng quan về Beamforming trong hệ thống Massive MIMO

1.1.1. Massive MIMO và kỹ thuật Beamforming

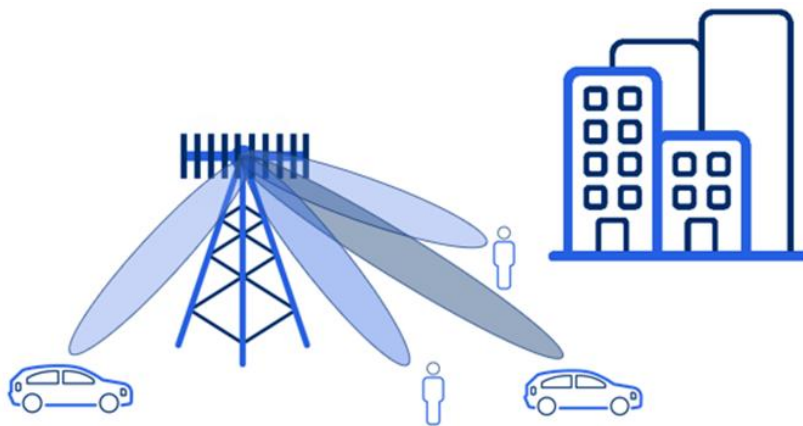
Các hệ thống truyền thông không dây thế hệ mới như 5G đang được phát triển để đáp ứng nhu cầu ngày càng tăng về tốc độ dữ liệu cao và lưu lượng truy cập lớn, phục vụ cho các dịch vụ đa dạng như băng rộng tăng cường (eMBB), truyền thông độ trễ cực thấp và đáng tin cậy (URLLC), truyền thông kết nối đa thiết bị (mMTC). Để thực hiện được điều này, ngành công nghiệp viễn thông đã và đang hướng đến việc khai thác các dải tần số mới và cao hơn. Trong khi phổ tần hiện có cho các dịch vụ di động đang trở nên quá “chật hẹp”, băng tần sóng mmWave, trong khoảng từ 30 đến 300 GHz, cung cấp một lượng lớn phổ tần để sử dụng cho truyền thông di động.

Tuy nhiên, việc sử dụng sóng mmWave đối mặt với một thách thức rất lớn là sự suy hao đường truyền trong không gian tự do rất cao, cùng với tác động hấp thụ của khí quyển và sự suy giảm tín hiệu do các điều kiện thời tiết như mưa hay sương mù gây ra. Chính những yếu tố này làm giảm đáng kể phạm vi và khả năng cung cấp dịch vụ của mmWave so với các hệ thống di động băng rộng hiện có. Do đó, để khai thác hiệu quả tiềm năng của sóng mmWave, việc tìm ra giải pháp công nghệ để bù đắp cho suy hao đường truyền trở thành một yêu cầu tiên quyết.

Để giải quyết thách thức về suy hao đường truyền của sóng và đảm bảo một liên kết truyền thông đáng tin cậy, hệ thống Massive MIMO đã được đề xuất và được xem là một công nghệ trụ cột cho các hệ thống 5G và xa hơn. Massive MIMO là một kỹ thuật truyền thông không dây có thể mở rộng quy mô của MIMO, trong đó các trạm gốc được trang bị một số lượng rất lớn các antenna nhằm cải thiện hiệu suất phổ và hiệu suất năng lượng. Kiến trúc này cho phép trạm gốc (BS) giao tiếp đồng thời với nhiều thiết bị người dùng (UE) trên cùng một tài nguyên tần số - thời gian.

Tuy nhiên, do đặc tính suy hao của mmWave, việc truyền tín hiệu đẳng hướng là không hiệu quả. Thay vào đó, các hệ thống 5G bắt buộc phải sử dụng các antenna định

hướng có độ lợi cao, với khả năng điều khiển hướng một cách linh hoạt. Điều này được thực hiện thông qua kỹ thuật tạo búp sóng. Beamforming là một kỹ thuật xử lý tín hiệu không gian được sử dụng để cải thiện tỉ số tín hiệu trên nhiễu (SNR), loại bỏ các nguồn nhiễu không mong muốn và tập trung tín hiệu truyền đến các vị trí cụ thể. Kỹ thuật này là trung tâm của các hệ thống truyền thông không dây MIMO như 5G, LTE và WLAN. Beamforming trong các ứng dụng không dây cũng có thể được sử dụng để tăng dung lượng luồng dữ liệu giữa trạm gốc và các thiết bị người dùng. Các kỹ thuật Beamforming dựa trên tối ưu hóa đang trở nên phổ biến hơn trong các hệ thống truyền thông không dây hiện đại [1].



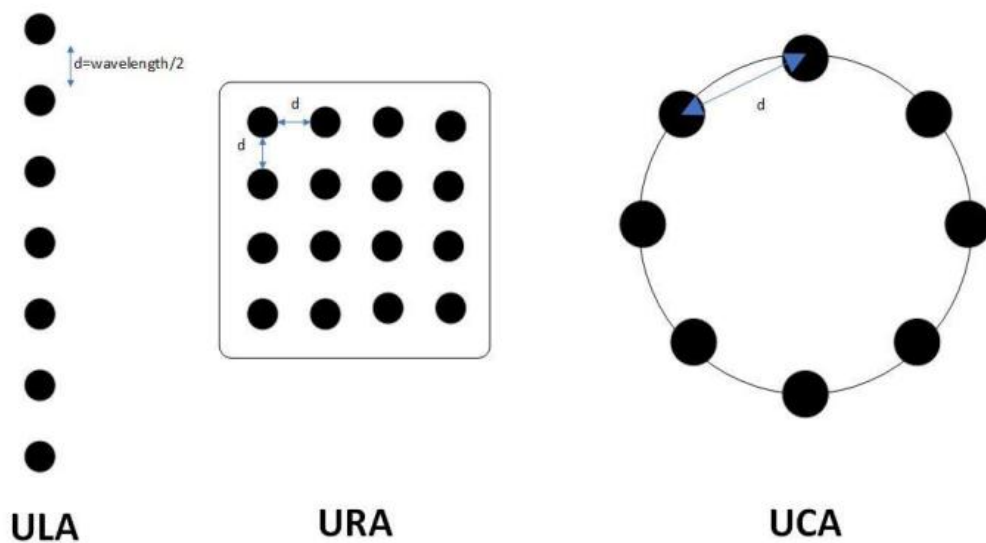
Hình 1.1. Minh họa kỹ thuật Beamforming

1.1.2. Antenna mảng – Nền tảng vật lý của Beamforming

Nền tảng vật lý để hiện thực hóa kỹ thuật beamforming chính là các antenna mảng (antenna arrays). Một antenna mảng bao gồm nhiều phần tử antenna riêng lẻ, hoạt động dựa trên nguyên tắc giao thoa sóng để tạo ra một đáp ứng bức xạ tổng hợp có định hướng và độ lợi cao hơn nhiều so với một antenna đơn lẻ. Búp sóng chính sẽ được quét tới các hướng mong muốn, các búp sóng này được quét thông qua các bộ dịch pha thường hoạt động ở các tần số RF. Antenna mảng thích nghi là antenna có khả năng chống nhiễu tốt, thu được tín hiệu chính xác và tự hiệu chỉnh trong hệ thống truyền thông, đặc tính bức xạ của những antenna này sẽ chuyển đổi thích nghi theo sự chuyển đổi của môi trường.

Antenna mảng có kích thước khác nhau được sử dụng để cung cấp độ lợi beamforming nhằm giảm thiểu suy hao đường truyền. Trong thiết kế antenna mảng, hình

học và khoảng cách giữa các phần tử là yếu tố quyết định. Các cấu trúc phổ biến bao gồm mảng thẳng (ULA), mảng chữ nhật (URA) và mảng tròn (UCA). Hình 1.2 mô tả hình học và kích thước mảng khác nhau. Mỗi phần tử antenna trong một mảng được tách biệt bằng một nửa bước sóng của tần số truyền để tránh sự ghép nối lẫn nhau giữa các phần tử antenna và các vấn đề về grating lobes (các búp sóng phụ không mong muốn, chúng có kích thước xấp xỉ búp sóng chính và xuất hiện ở các hướng khác với hướng mà chúng ta muốn tập trung năng lượng [2]).



Hình 1.2. Minh hoạ các loại Antenna mảng

1.1.3. Các kiến trúc Beamforming và thách thức phần cứng

Vấn đề tiêu thụ năng lượng cao của nhiều chuỗi tần số vô tuyến (RF) trở nên quan trọng trong truyền thông sóng mmWave và do đó, các mảng antenna tại phía phát (BS) yêu cầu Beamforming để truyền tín hiệu định hướng. Kiến trúc của các kỹ thuật Beamforming được phân loại thành ba dạng, cụ thể là Digital Beamforming, Analog Beamforming và Hybrid Beamforming.

Một chuỗi RF duy nhất được dành riêng cho mỗi antenna được gọi là định hình chùm tia kỹ thuật số (Digital Beamforming) và có thể triển khai trong các hệ thống MIMO thông thường nhưng không thực tế khi triển khai trong các hệ thống Massive MIMO, bởi vì hệ thống Massive MIMO chứa số lượng lớn antenna dẫn đến số lượng chuỗi RF cũng

lớn. Mặt khác, các chuỗi RF có vai trò quan trọng trong các hệ thống Massive MIMO mmWave nhưng lại có giá thành đắt hơn và mức tiêu thụ năng lượng cao.

Analog beamforming là một kỹ thuật tạo búp sóng được thực hiện hoàn toàn trong miền tần số vô tuyến (RF), sử dụng các bộ dịch pha (phase shifters) và bộ suy hao (attenuators) để điều chỉnh tín hiệu từ mỗi phần tử antenna trước khi tổng hợp. Kiến trúc này nổi bật nhờ độ phức tạp phần cứng và tiêu thụ năng lượng thấp, đặc biệt phù hợp cho các hệ thống Massive MIMO với số lượng antenna lớn. Lợi ích chính của nó là giảm số lượng bộ chuyển đổi ADC/DAC và đường truyền RF, từ đó giảm chi phí và điện năng tiêu thụ trong khi vẫn mang lại khả năng định hướng tín hiệu và tăng cường độ lợi búp sóng đáng kể, cải thiện hiệu suất truyền dẫn trong môi trường không dây.

Kiến trúc còn lại là chùm tia lai, việc thiết kế được chia thành hai giai đoạn, với giai đoạn một là thiết kế một bộ tạo búp sóng tương tự ở miền tần số vô tuyến (RF), được cấu thành từ mạng lưới các bộ dịch pha và giai đoạn hai là thiết kế một bộ tiền mã hóa số được xử lý ở miền băng tần gốc (baseband). Trong bối cảnh hệ thống Massive MIMO đang phát triển nhanh chóng, việc triển khai hiệu quả là một thách thức lớn. Mặc dù các kiến trúc Hybrid Beamforming đang thu hút nhiều sự chú ý nhờ khả năng cân bằng giữa hiệu suất và chi phí, nhưng để thực sự tối ưu hóa và hiểu rõ giới hạn của chúng, việc nghiên cứu chuyên sâu về Analog Beamforming là cần thiết.

1.2. Hướng tiếp cận học sâu với Beamforming

Việc thiết kế búp sóng trong các hệ thống MIMO, đặc biệt là trong kiến trúc tạo búp sóng tương tự (Analog Beamforming), về bản chất là một bài toán tối ưu hóa phức tạp. Mục tiêu là tìm ra các búp sóng tối ưu để tối đa hóa một chỉ số hiệu năng hệ thống, chẳng hạn như hiệu suất phổ, dưới các ràng buộc vật lý khắc nghiệt của phần cứng. Một trong những thách thức lớn nhất là ràng buộc về điều kiện biên độ không đổi (constant modulus constraint) trên các bộ dịch pha, khiến cho bài toán trở thành một vấn đề tối ưu hóa phi lồi (non-convex) và rất khó để tìm ra lời giải tối ưu toàn cục.

Để giải quyết bài toán này, các công trình nghiên cứu trước đây trong lĩnh vực viễn thông đã phát triển nhiều hướng tiếp cận dựa trên mô hình toán học và các thuật toán

tối ưu kinh điển. Một hướng đi phổ biến là xem việc thiết kế bộ tiền mã hóa (precoder) như một bài toán khôi phục tín hiệu thưa (sparse signal recovery). Lợi dụng đặc tính thưa về mặt không gian của kênh truyền sóng mmWave, các thuật toán như OMP đã được đề xuất để tìm ra một bộ búp sóng tiệm cận với bộ tiền mã hóa thuần số tối ưu [3]. Một cách tiếp cận khác là biến đổi kênh truyền trong không gian góc sang miền không gian chùm tia (beam space), nơi kênh trở nên thưa hơn. Dựa trên đặc tính này, các thuật toán như phát hiện tập hỗ trợ (Support Detection) được phát triển để ước lượng kênh một cách hiệu quả với chi phí thấp, tạo tiền đề cho việc lựa chọn chùm tia tối ưu [4]. Các kỹ thuật như tối thiểu hóa xen kẽ (Alternating Minimization) được sử dụng để giải quyết riêng biệt cho các biến khác nhau, thường với mục tiêu tối thiểu hóa MMSE [5].

Đối với các bài toán có cấu trúc phức tạp hơn, các kỹ thuật nâng cao như lập trình phân thức (Fractional Programming) và phân rã kép phạt (Penalty Dual Decomposition) cũng đã được áp dụng để tìm ra lời giải tiệm cận [6] [7]. Một phương pháp khác nhằm giảm độ phức tạp phần cứng là kích hoạt một tập con các antenna, các thuật toán lựa chọn dựa trên norm (norm-based) hoặc lựa chọn tuần tự (successive selection) được sử dụng để tìm ra tập con antenna tối ưu nhằm tối đa hóa dung lượng hoặc SNR [8]. Tuy nhiên, một hạn chế chung của các phương pháp truyền thống này là chúng thường đòi hỏi các vòng lặp phức tạp, tốn thời gian, dẫn đến độ trễ tính toán cao, khiến chúng không phù hợp cho các ứng dụng thời gian thực nơi trạng thái kênh thay đổi nhanh chóng. Hơn nữa, hiệu suất của chúng có thể suy giảm đáng kể khi hoạt động với CSI không hoàn hảo, vốn là điều không thể tránh khỏi trong thực tế.

Trước những thách thức đó, các phương pháp dựa trên dữ liệu, đặc biệt là Deep Learning, đang nổi lên như một hướng tiếp cận đầy tiềm năng cho việc xử lý tín hiệu trong các hệ thống truyền thông không dây phải đối mặt với sự méo dạng kênh và nhiễu phức tạp. Một trong những ý tưởng đột phá là xem toàn bộ hệ thống truyền thông như một bộ tự mã hóa (autoencoder), có khả năng tối ưu hóa đầu cuối (end-to-end) [9]. Ưu điểm cốt lõi của phương pháp này là khả năng chuyển toàn bộ gánh nặng tính toán phức tạp sang giai đoạn huấn luyện ngoại tuyến (offline training). Cụ thể, các mạng thần kinh học sâu (DNN) có thể được huấn luyện trên một lượng lớn dữ liệu mô phỏng, bao gồm nhiều trạng thái kênh và chuỗi thông tin khác nhau. Sau khi giai đoạn huấn luyện tón

nhiều tài nguyên này hoàn tất, mô hình có thể đưa ra dự đoán gần như tức thời ở giai đoạn triển khai trực tuyến (online deployment) chỉ bằng các phép tính toán lan truyền xuôi đơn giản, giúp giảm thiểu đáng kể độ phức tạp và độ trễ [10].

Phát huy những ưu điểm này, với khả năng học các mối quan hệ phức tạp và phi tuyến tính từ dữ liệu, các mạng thần kinh học sâu có thể được huấn luyện ngoại tuyến để học ánh xạ trực tiếp từ thông tin trạng thái kênh đến các búp sóng tối ưu [7]. Đây là phương pháp được lựa chọn để hiện thực hóa trong khuôn khổ của đề tài này, lấy cảm hứng từ công trình của Lin, Zhu [11]. Thay vì đi theo một giải pháp có sẵn, phương pháp này sử dụng mạng DNN như một công cụ tối ưu hóa trực tiếp. Mô hình nhận đầu vào là kênh ước tính, kênh hoàn hảo và SNR, sau đó trực tiếp dự đoán các tham số của búp sóng. Quá trình huấn luyện sẽ điều chỉnh các trọng số của mạng sao hiệu suất phổ (Spectral Efficiency - SE) của hệ thống được tối đa hóa. Hướng tiếp cận này có tiềm năng tìm ra các giải pháp mới mà không bị giới hạn bởi các thuật toán kinh điển, đồng thời tối ưu hóa trực tiếp trên mục tiêu mong muốn (chẳng hạn như SE) của một hệ thống truyền thông thực tế.

1.3. Mục tiêu của đề tài

Để giải quyết các câu hỏi nghiên cứu đã được đề ra, đề tài này sẽ thực hiện một chuỗi các mục tiêu cụ thể, được cấu trúc như sau:

Thứ nhất, tìm hiểu và phân tích cơ sở lý thuyết về hệ thống truyền thông sóng mmWave Massive MIMO, kiến trúc tạo búp sóng tương tự (Analog Beamforming - ABF) và các phương pháp Deep Learning.

Thứ hai, tạo một bộ dữ liệu kênh truyền bằng phương pháp mô phỏng. Dữ liệu này được tạo ra dựa trên mô hình kênh hình học và kiến trúc mảng tuyến tính đồng nhất (ULA), nhằm phản ánh các đặc tính của kênh truyền, làm cơ sở cho việc huấn luyện và đánh giá mô hình học sâu.

Thứ ba, hiện thực hóa và kiểm chứng một mô hình mạng thần kinh tạo búp sóng (Beamforming Neural Network - BFNN) cho kịch bản cơ sở là hệ thống MISO. Mô hình

này, vốn chỉ tối ưu hóa búp sóng ở phía phát, sẽ đóng vai trò là cơ sở để so sánh và đối chiếu hiệu năng.

Thứ tư, mở rộng và phát triển kiến trúc BFNN để giải quyết bài toán phức tạp hơn trong hệ thống MIMO. Đây là mục tiêu chính của đề tài, trong đó mô hình phải có khả năng tối ưu hóa đồng thời cả búp sóng phát (precoder) và búp sóng thu (combiner).

Cuối cùng, khảo sát và đánh giá toàn diện hiệu năng của các mô hình Massive MIMO mở rộng trên các cấu hình antenna đa dạng. Việc đánh giá dựa trên hiệu suất phổ, qua đó cho phép lượng hóa lợi ích của việc tối ưu hóa phía thu và rút ra các kết luận về tính hiệu quả cũng như khả năng mở rộng của phương pháp đề xuất.

1.4. Phương pháp thực hiện đề tài

Để đạt được các mục tiêu đề ra, quy trình sẽ được chia thành hai giai đoạn chính. Giai đoạn đầu tiên tập trung vào việc mô hình hoá hệ thống và tạo dựng bộ dữ liệu. Một môi trường mô phỏng được xây dựng bằng các công cụ tính toán ma trận để tạo ra một bộ dữ liệu lớn, bao gồm các mẫu kênh truyền trong kịch bản sóng mmWave. Cụ thể, kênh truyền sẽ được mô hình hoá theo mô hình kênh Saleh – Valenzuela, cả phía phát (BS) và phía thu (UE) đều được trang bị mảng tuyến tính đồng nhất (ULA) với khoảng cách giữa các phần tử bằng một nửa bước sóng. Mỗi mẫu dữ liệu được tạo ra bao gồm một ma trận thông tin kênh hoàn hảo và một phiên bản ước tính không hoàn hảo tương ứng, mô phỏng việc hệ thống hoạt động dưới điều kiện thông tin trạng thái kênh không hoàn hảo.

Giai đoạn thứ hai tập trung vào việc thiết kế và huấn luyện mô hình học sâu. Một mô hình mạng neuron tạo búp sóng được xây dựng và huấn luyện bằng ngôn ngữ lập trình *Python* cùng các thư viện học sâu mã nguồn mở tiêu chuẩn như *TensorFlow* và *Keras*. Mô hình nhận đầu vào là thông tin kênh ước tính đã được xử lý thành dạng vector thực. Thay vì học theo một nhãn có sẵn, mạng neuron được tối ưu hoá trực tiếp trên mục tiêu cuối cùng của hệ thống là tối ưu hoá hiệu suất phổ. Do đó, hàm sai số được định nghĩa là giá trị âm của hiệu suất phổ, thuật toán tối ưu *Adam* được sử dụng để cập nhật các trọng số của mạng nhằm đạt được mục tiêu này. Sau khi huấn luyện, hiệu năng của

mô hình sẽ được đánh giá trên một tập dữ liệu kiểm thử riêng biệt để đảm bảo tính khách quan của kết quả.

1.5. Bố cục báo cáo

Nội dung chính của đề tài được trình bày và diễn giải qua các chương sau:

Chương 1. Tổng quan về ứng dụng Deep Learning trong kỹ thuật Beamforming: Chương này sẽ trình bày về tầm quan trọng của kỹ thuật Beamforming trong hệ thống Massive MIMO trong việc khắc phục nhược điểm của dải tần mmWave. Với sự phát triển của Deep Learning, việc áp dụng vào quá trình tạo búp sóng mở ra một hướng tiếp cận đầy hứa hẹn để giải quyết bài toán tối ưu hoá phức tạp này. Đồng thời, chương cũng giới thiệu tổng quan các công trình nghiên cứu tiêu biểu đã áp dụng mạng neuron vào bài toán thiết kế búp sóng. Cuối cùng, chương này cũng trình bày mục tiêu, phương pháp, phạm vi nghiên cứu và bố cục của đề tài.

Chương 2. Cơ sở lý thuyết về kỹ thuật Beamforming và Deep Learning: Nội dung chương này trình bày tập trung vào các kiến thức nền tảng của hệ thống truyền thông Massive MIMO và các kiến trúc tạo búp sóng, bao gồm Analog Beamforming, Digital Beamforming và Hybrid Beamforming, từ đó đưa ra mô hình toán học của kênh truyền và bài toán tối ưu hóa Beamforming. Đồng thời, chương này cũng trình bày nền tảng về Deep Learning để ứng dụng trong bài toán về Beamforming.

Chương 3. Thiết kế bộ tạo búp sóng tương tự với Deep Learning: Chương này tập trung vào việc mô tả chi tiết các phương pháp được sử dụng trong đề tài. Nội dung bao gồm việc trình bày cụ thể mô hình hệ thống được mô phỏng, quy trình tạo bộ dữ liệu kênh truyền bằng công cụ mô phỏng. Cùng với đó, chương này sẽ trình bày chi tiết kiến trúc của mạng neuron tạo búp sóng (BFNN) được đề xuất, bao gồm các lớp mạng, hàm kích hoạt và hàm sai số là âm của hiệu suất phổ.

Chương 4. Kết quả mô phỏng và thảo luận: Chương này trình bày và phân tích các kết quả từ quá trình mô phỏng và huấn luyện. Các kết quả về quá trình hội tụ của mô hình, hiệu suất phổ đạt được ở các mức SNR khác nhau và so sánh hiệu năng giữa các

cấu hình antenna sẽ được trình bày dưới dạng bảng biểu và đồ thị cùng với các phần thảo luận, diễn giải ý nghĩa của các kết quả đó.

Chương 5. Kết luận và hướng phát triển đề tài: Chương cuối cùng sẽ tóm tắt lại những kết quả chính mà đề tài đã đạt được, đối chiếu với các mục tiêu đã đề ra và đưa ra kết luận cuối cùng. Đồng thời, chương cũng đề xuất một số hướng nghiên cứu và phát triển cho đề tài trong tương lai.

1.6. Giới hạn của đề tài

Để tập trung vào các mục tiêu chính, đề tài này được thực hiện trong một số giới hạn và giả định nhất định. Việc hiểu rõ các giới hạn này là cần thiết để đánh giá đúng phạm vi và kết quả của công trình:

- **Kiến trúc Beamforming:** Đề tài chỉ tập trung và kiến trúc Analog Beamforming (ABF), với phía phát và phía thu chỉ sử dụng một chuỗi RF ($N_{RF} = 1$). Các kiến trúc Hybrid Beamforming hay Digital Beamforming, có thể mang lại hiệu suất cao hơn nhưng có độ phức tạp cao, không thuộc phạm vi của đề tài.
- **Kịch bản người dùng:** Mô hình hệ thống được xem xét là kịch bản đơn người dùng. Các yếu tố phức tạp trong môi trường đa người dùng như nhiễu liên người dùng và các thuật toán lập lịch không được xem xét.
- **Mô hình kênh truyền:** Kênh truyền được mô phỏng theo mô hình dải tần băng hẹp (narrowband), có độ phân tán hạn chế và là đặc trưng của sóng mmWave. Các hiệu ứng của kênh băng rộng như fading lựa chọn tần số và mô hình OFDM không được tích hợp. Kênh cũng được giả định là không thay đổi theo thời gian và bỏ qua hiệu ứng Doppler do sự di chuyển của người dùng.
- **Giả định về phần cứng:** Các thành phần phần cứng trong hệ thống được giả định là lý tưởng. Cụ thể hơn, các bộ dịch pha (phase shifters) có thể tạo ra các giá trị pha liên tục mà không bị ảnh hưởng bởi lỗi lượng tử hoá. Các hiệu ứng phi tuyến của bộ khuếch đại công suất cũng được bỏ qua.

CHƯƠNG 2: CƠ SỞ LÝ THUYẾT VỀ KỸ THUẬT BEAMFORMING VÀ DEEP LEARNING

Như đã được thảo luận ở chương trước, việc sử dụng công nghệ Massive MIMO kết hợp với kỹ thuật tạo búp sóng Beamforming là giải pháp then chốt để khắc phục suy hao đường truyền cao của dải tần sóng mmWave. Để có thể phân tích, đánh giá, và đề xuất các phương pháp thiết kế búp sóng hiệu quả, việc xây dựng một mô hình toán học chính xác và nắm vững các nguyên lý hoạt động là yêu cầu cơ bản.

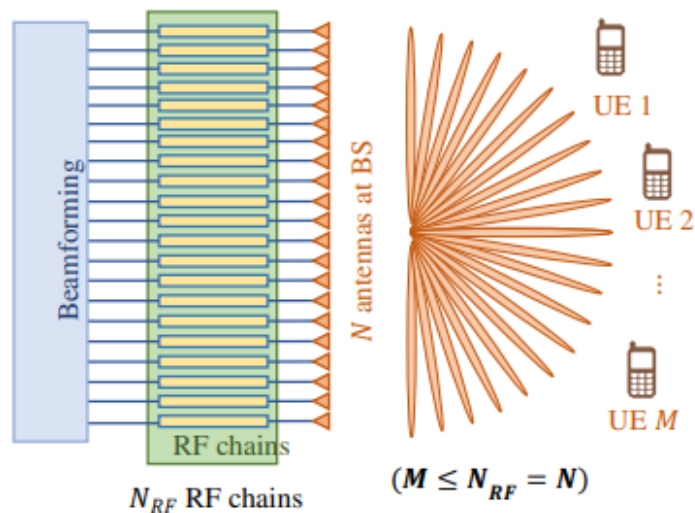
Chương này sẽ tập trung vào việc trình bày các cơ sở lý thuyết và công cụ toán học cần thiết để mô tả kỹ thuật Beamforming, từ việc mô hình hóa các thành phần vật lý như antenna, kênh truyền, cho đến việc phân tích các thuật toán và công nghệ cốt lõi. Sau đó, chương này cũng trình bày lý thuyết của mô hình DL, bao gồm hàm sai số, hàm kích hoạt, cấu trúc của mạng DNN và các thuật toán tối ưu. Cuối cùng, chương sẽ trình bày phương pháp thiết kế búp sóng dựa trên DL và xây dựng bài toán tối ưu hoá Analog Beamforming với DNN.

2.1. Kỹ thuật Beamforming

Như đã mô tả trong Chương 1, Beamforming được phân thành ba loại dựa trên kiến trúc kỹ thuật, đó là Digital, Analog và Hybrid Beamforming. Dưới đây sẽ đi sâu vào chi tiết của từng loại Beamforming.

2.1.1. Digital Beamforming

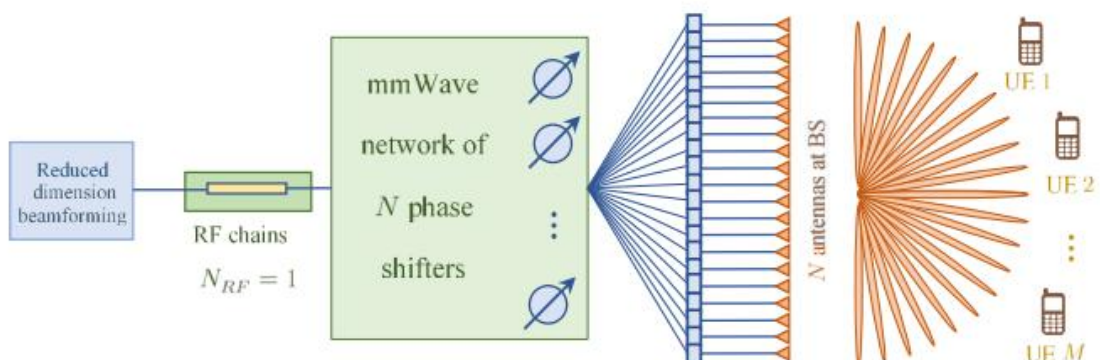
Trong kỹ thuật định hình chùm tia kỹ thuật số, mỗi phần tử antenna có một chuỗi RF chuyên dụng riêng (Hình 2.1). Điều này có nghĩa là số chuỗi RF N_{RF} phải bằng với số phần tử antenna N , phục vụ cho M người dùng ($M \leq N_{RF} = N$). Kiến trúc này cung cấp bậc tự do cao, vì nó cho phép điều khiển pha và biên độ của tín hiệu trên mỗi antenna một cách độc lập. Theo lý thuyết, Beamforming kỹ thuật số là kiến trúc tốt nhất trong các kỹ thuật định hình chùm tia, nhưng hạn chế của kỹ thuật này là mỗi antenna phải có một chuỗi RF chuyên dụng. Đây là một ràng buộc lớn về phần cứng, đặc biệt là trong truyền thông sóng mmWave. Do đó, việc triển khai toàn bộ búp sóng số trong hệ thống mmWave Massive MIMO rất tốn kém và phức tạp.



Hình 2.1. Kiến trúc Digital Beamforming trong Massive MIMO

2.1.2. Analog Beamforming

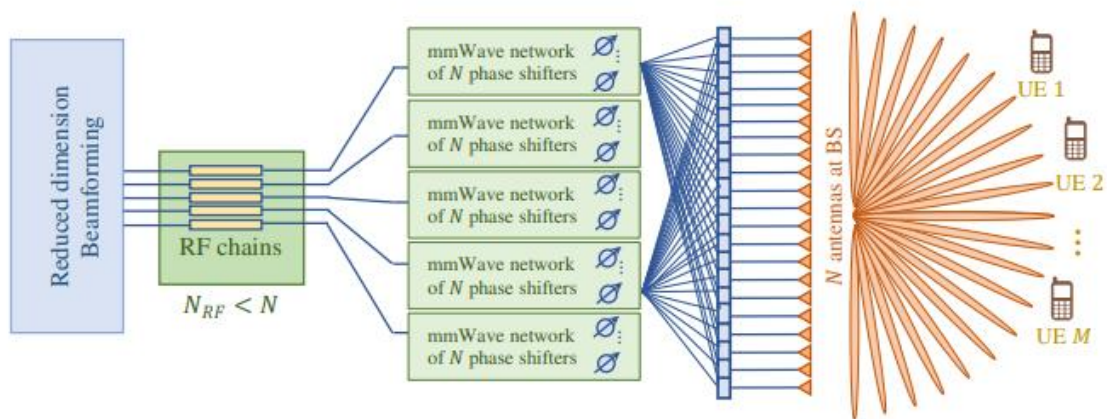
Khác với kiến trúc Beamforming kỹ thuật số, Beamforming tương tự là một cách tiếp cận đơn giản và ít tốn kém hơn, được sử dụng trong việc triển khai các hệ thống mmWave Massive MIMO. Kỹ thuật này được thực hiện bằng cách kết nối các antenna thông qua một mạng lưới các bộ dịch pha được điều khiển kỹ thuật số với một chuỗi RF duy nhất ($N_{RF} = 1$). Các bộ dịch pha ít tiêu thụ điện năng và đỡ tốn kém hơn so với các chuỗi RF. Nhược điểm chính của kỹ thuật này là nó cung cấp ít bậc tự do hơn do chỉ có một bộ tạo búp sóng duy nhất và chỉ có thể điều khiển pha của tín hiệu. Việc tinh chỉnh các búp sóng và tạo các điểm triệt tiêu nhiều để điều khiển các mảng antenna lớn là những thách thức trong kiến trúc này.



Hình 2.2. Kiến trúc Analog Beamforming trong Massive MIMO

2.1.3. Hybrid Beamforming

Để tận dụng lợi ích của cả hai phương pháp, kỹ thuật định hình chùm tia lai (Hybrid Beamforming) được triển khai. Đây được coi là giải pháp cốt lõi cho các hệ thống Massive MIMO trong việc giảm độ phức tạp của phần cứng và hiệu suất đạt được gần như khi triển khai Digital. Trong kiến trúc này, số lượng chuỗi RF được giảm xuống tùy thuộc vào số lượng người dùng ($N_{RF} < N$). Việc tạo búp sóng được thực hiện ở cả hai miền, bao gồm một bộ tiền mã hóa số ở băng tần gốc và một bộ tạo búp sóng tương tự ở tần số RF, hai miền này được kết nối với nhau qua mạng lưới các bộ dịch pha.



Hình 2.3. Kiến trúc Hybrid Beamforming trong Massive MIMO

Để có cái nhìn rõ hơn về các kỹ thuật định hình chùm tia, Bảng 2.1 thể hiện sự so sánh giữa các kỹ thuật. Digital Beamforming được thiết lập với $N_{RF} = N$ có kiến trúc phức tạp cũng như khả năng tiêu thụ năng lượng và giá thành cao. Trong khi với Analog Beamforming với $N_{RF} = 1$, là một giải pháp phù hợp vì có kiến trúc đơn giản hơn. Tuy nhiên, đối với trường hợp hệ thống mmWave Massive MIMO với điều kiện $N_{RF} < N$, Hybrid Beamforming mang lại lựa chọn hứa hẹn vì đóng vai trò như một sự dung hoà giữa hiệu năng và độ phức tạp. Mỗi kiến trúc sẽ có ưu nhược điểm khác nhau, khi thiết kế hệ thống Beamforming cần lựa chọn kỹ thuật phù hợp với bài toán tương ứng.

Đề tài sẽ tập trung phân tích sâu vào kiến trúc **Analog Beamforming** như một bài toán nền tảng, do đây là kiến trúc có độ phức tạp phần cứng đơn giản và các nguyên lý của nó là cơ sở để phát triển các kiến trúc lai phức tạp hơn.

Bảng 2.1. So sánh các kỹ thuật Beamforming

	Digital Beamforming	Analog Beamforming	Hybrid Beamforming
Số lượng người dùng	Đa người dùng	Đơn người dùng	Đa người dùng
Khả năng điều khiển tín hiệu	Điều khiển pha và biên độ	Chỉ điều khiển pha	Điều khiển pha và biên độ
Độ phức tạp phần cứng	Phức tạp nhất; Chuỗi RF bằng với số lượng antenna tại phía phát (BS) $N_{RF} = N$	Đơn giản; Một chuỗi RF duy nhất $N_{RF} = 1$	Độ phức tạp trung bình; Số lượng chuỗi RF thấp hơn so với số lượng antenna tại phía phát (BS) $N_{RF} < N$
Tiêu thụ năng lượng	Cao	Thấp	Trung bình
Giá thành	Cao	Thấp	Trung bình

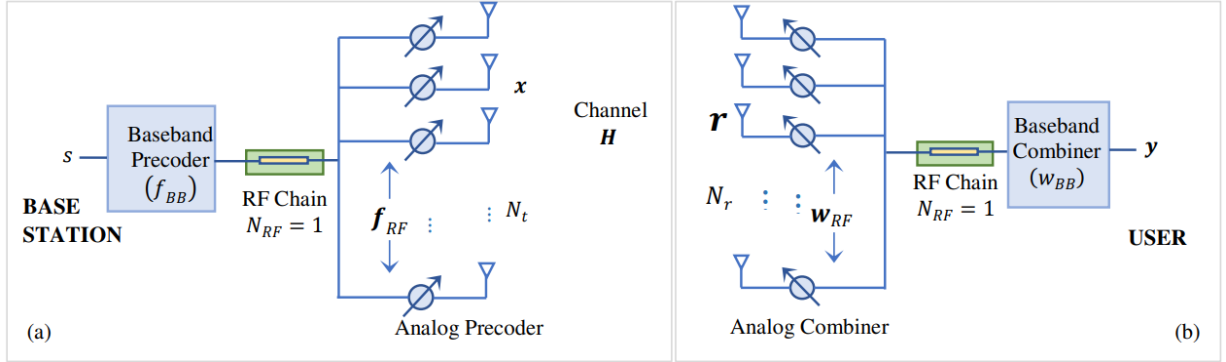
2.2. Cơ sở toán học và mô hình hệ thống

Để có thể phân tích và thiết kế các thuật toán beamforming một cách có hệ thống, việc xây dựng một mô hình toán học chính xác cho toàn bộ hệ thống truyền thông là bước đi đầu tiên và quan trọng nhất. Mô hình này không chỉ mô tả các thành phần vật lý như antenna và kênh truyền mà còn định nghĩa các công cụ toán học cần thiết để biểu diễn tín hiệu và các quá trình xử lý không gian.

2.2.1. Mô hình kênh truyền sóng mmWave

Xét một hệ thống massive MIMO sóng mmWave đường xuống, đơn người dùng, hoạt động với kiến trúc Analog Beamforming (ABF). Trong hệ thống này, một trạm gốc (BS) được trang bị một mảng antenna thẳng đồng nhất (ULA) gồm N_t antenna và một chuỗi RF duy nhất, thực hiện truyền một luồng dữ liệu đến một người dùng. Phía người

dùng (UE) cũng được trang bị một mảng ULA gồm N_r antenna và một chuỗi RF duy nhất, như được minh họa trong Hình 2.4.



Hình 2.4. Mô hình hệ thống mmWave Massive MIMO với 1 chuỗi RF

Tại phía phát, tín hiệu được xử lý bởi một bộ tiền mã hóa số ở băng tần gốc f_{BB} theo sau là một bộ tiền mã hóa tương tự $\mathbf{f}_{RF} \in \mathcal{F}(N_t \times 1)$. Trong khi đó, tại phía thu, một bộ tổ hợp tương tự $\mathbf{w}_{RF} \in \mathcal{W}(N_r \times 1)$ và một bộ tổ hợp số ở băng tần gốc w_{BB} được áp dụng. Các thuộc tính của các tập hợp nghiệm \mathcal{F} và \mathcal{W} được quyết định bởi cấu trúc phần cứng tương tự cụ thể được sử dụng trong các bộ dịch pha. Tín hiệu đã được tiền mã hóa và phát đi là một vector, ký hiệu là \mathbf{x} ($N_t \times 1$), với:

$$\mathbf{x} = \mathbf{f}_{RF} f_{BB} S. \quad (2.1)$$

Kênh truyền sóng mmWave có những đặc điểm riêng biệt so với các băng tần thấp hơn. Do suy hao trong không gian tự do lớn và tính nhạy cảm với các vật cản, tín hiệu mmWave thường lan truyền qua một số lượng hữu hạn các đường truyền, tạo ra một kênh truyền có tính thưa về mặt không gian (spatially sparse). Mô hình kênh hình học Saleh-Valenzuela được sử dụng rộng rãi vì nó phản ánh chính xác đặc tính này. Ma trận kênh không gian \mathbf{H} có kích thước $N_r \times N_t$, mô tả mối quan hệ truyền sóng giữa N_t antenna tại trạm gốc BS và N_r antenna tại thiết bị người dùng (UE), được biểu diễn như sau:

$$\mathbf{H} = \sqrt{\frac{N_t N_r}{L}} \left(\sum_{l=1}^L \alpha^{(l)} \mathbf{a}_{UE}(\theta^{(l)}) \mathbf{a}_{BS}^H(\psi^{(l)}) \right), \quad (2.2)$$

trong đó, L là tổng số đường truyền. Đường truyền thứ nhất ($l = 1$) được coi là đường truyền thẳng (Line-of-Sight - LoS), và $L - 1$ đường còn lại là các đường truyền không

thẳng (Non-Line-of-Sight - NLoS) do phản xạ, tán xạ. Ngoài ra, $\alpha^{(l)}$ là độ lợi phức của đường truyền thứ l . Trong môi trường mmWave, biên độ của thành phần LoS là $|\alpha^{(1)}|$, thường lớn hơn các thành phần NLoS là $\{|\alpha^l|\}_{l=2}^L$. Kế tiếp, $a_{UE}(\theta^{(l)})$ và $a_{BS}(\psi^{(l)})$ lần lượt là các vector lái mảng tại phía thu và phía phát. Chúng phụ thuộc vào góc tới (AoA) $\theta^{(l)}$ và góc phát (AoD) $\psi^{(l)}$ của đường truyền thứ l . Ký hiệu $(\cdot)^H$ biểu thị cho phép chuyển vị liên hợp.

2.2.2. Array Steering Vector

Vector lái mảng là một công cụ toán học cơ bản nhưng cực kỳ quan trọng, nó mô tả một cách chính xác đáp ứng pha của một mảng antenna đối với một sóng phẳng đến từ một hướng cụ thể. Đối với một mảng tuyến tính đồng nhất (ULA), vector lái mảng tại phía thu với $a_{UE}(\theta^{(1)})$, áp dụng cho một mảng antenna ULA gồm N_r antenna, là một vector có kích thước $(N_r \times 1)$ và được cho bởi công thức

$$\mathbf{a}_{UE}(\theta^{(1)}) = \frac{1}{\sqrt{N_r}} \begin{bmatrix} 1 \\ e^{-j2\pi\theta^{(1)}} \\ \vdots \\ e^{-j2\pi(N_r-1)\theta^{(1)}} \end{bmatrix} = \frac{1}{\sqrt{N_r}} \begin{bmatrix} 1 \\ e^{-\frac{j2\pi}{\lambda}d\sin\phi^{(1)}} \\ \vdots \\ e^{-\frac{j2\pi}{\lambda}(N_r-1)d\sin\phi^{(1)}} \end{bmatrix}, \quad (2.3)$$

trong biểu thức trên, $\theta^{(1)}$ là góc không gian chuẩn hoá, có mối liên hệ với góc phương vị vật lý $\phi^{(1)}$ của đường truyền thứ l qua biểu thức $\theta^{(1)} = \left(\frac{d}{\lambda}\right) \sin(\phi^{(1)})$. Góc $\phi^{(1)}$ biểu thị hướng vật lý bao phủ một nửa mặt phẳng không gian, với miền giá trị từ $-\frac{\pi}{2} \leq \phi^{(1)} \leq \frac{\pi}{2}$. Các tham số vật lý bao gồm d là khoảng cách giữa hai phần tử antenna khác nhau, và λ là bước sóng của tín hiệu.

Theo quy ước thiết kế chuẩn, khoảng cách d thường được chọn bằng một nửa bước sóng, tức là $d = \frac{\lambda}{2}$. Lựa chọn này có ý nghĩa kỹ thuật quan trọng nhằm tránh hiện tượng búp sóng lưới (grating lobes) không mong muốn giữa các phần tử, qua đó đảm bảo hiệu năng bức xạ của toàn bộ mảng antenna [2]. Tương tự, vector lái mảng tại phía phát, $a_{BS}(\psi^{(1)})$, cũng được xây dựng dựa trên cấu trúc và các nguyên tắc trên.

2.2.3. Ước lượng kênh tuần tự trong môi trường đa đường

Để thu được thông tin trạng thái kênh (CSI) trong môi trường mmWave đa đường và thưa, một giải thuật ước tính kênh tuần tự và thích ứng được triển khai, dựa trên phương pháp tìm kiếm phân cấp của Alkhateeb [12]. Giải thuật này hoạt động theo cơ chế "bóc tách" kênh truyền để xác định các tham số của L_{est} đường truyền trội nhất.

Quá trình này bao gồm một vòng lặp lớn, mỗi vòng lặp xác định một đường truyền thông qua một quy trình tìm kiếm phân cấp đa giai đoạn. Ở mỗi giai đoạn, các búp sóng thăm dò được sử dụng để xác định vùng góc có năng lượng cao nhất, sau đó "zoom-in" bằng các búp sóng hẹp hơn ở giai đoạn sau để tăng độ chính xác.

Để xử lý sự hiện diện của nhiều đường truyền, một phương pháp loại bỏ đóng góp tuần tự được áp dụng. Sau khi một đường truyền được xác định, ảnh hưởng của nó sẽ được ước tính và trừ đi khỏi tín hiệu nhận được trong các lần tìm kiếm tiếp theo. Cơ chế này, tuy không hoàn toàn giống với phép chiếu trực giao trong giải thuật Loại bỏ Nhiều Tuần tự (SIC) kinh điển, nhưng vẫn tuân theo ý tưởng cốt lõi là làm lộ diện các đường truyền yếu hơn để có thể phát hiện chúng ở các vòng lặp sau.

Sau khi đã xác định được toàn bộ các cặp góc đến và góc đi $(\widehat{\theta}_l, \widehat{\phi}_l)$ cho L_{est} đường truyền, độ lợi phức của chúng được ước tính đồng thời. Một ma trận cảm biến $\Phi \in \mathbb{C}^{M \times L_{est}}$ được xây dựng, trong đó mỗi cột đại diện cho "dấu hiệu" (signature) của một đường truyền đã tìm được, dựa trên các búp sóng đã sử dụng trong một phép đo tham chiếu. Mối quan hệ giữa tín hiệu đo được $\mathbf{y}_{measure} \in \mathbb{C}^{M \times 1}$ và vector độ lợi phức chưa biết $\alpha \in \mathbb{C}^{L_{est} \times 1}$ được mô tả bởi hệ phương trình sau:

$$\mathbf{y}_{measure} = \Phi \alpha + \mathbf{n}, \quad (2.4)$$

trong đó \mathbf{n} là vector nhiễu. Độ lợi phức ước tính $\hat{\alpha}$ được tìm ra bằng cách giải bài toán bình phương tối thiểu có điều chuẩn hóa (regularized least squares):

$$\hat{\alpha} = (\Phi^H \Phi + \lambda \mathbf{I})^{-1} \Phi^H \mathbf{y}_{measure}, \quad (2.5)$$

với λ là một tham số điều chuẩn hóa nhỏ để đảm bảo tính ổn định về mặt số học. Cuối cùng, ma trận kênh \mathbf{H}_{est} được tái tạo lại bằng cách tổng hợp các đường truyền đã được ước tính.

Sau khi đã có vector độ lợi phức ước tính ($\hat{\alpha}$) và các cặp góc đến/góc đi, ma trận kênh \mathbf{H}_{est} được tái tạo lại. Quá trình này thực hiện bằng cách nhân độ lợi phức của mỗi đường truyền với ma trận kênh con tương ứng (được tạo ra từ các vector lái mảng của đường truyền đó) rồi tổng hợp tất cả các đóng góp lại. Cụ thể, ma trận kênh ước tính được xây dựng theo công thức:

$$\mathbf{H}_{est} = \sum_{l=1}^{L_{est}} \hat{\alpha}_l \cdot \mathbf{a}_{UE}(\hat{\theta}_l) \mathbf{a}_{BS}^H(\hat{\phi}_l). \quad (2.6)$$

Trong khuôn khổ của đề tài, để đánh giá hiệu năng cốt lõi của thuật toán một cách công bằng, một kịch bản lý tưởng trong đó thuật toán được cung cấp thông tin tiên nghiệm chính xác về độ phức tạp của kênh được thiết lập. Do đó, số lượng đường truyền ước tính được đặt bằng với số lượng đường truyền vật lý thực tế, tức là $L_{est} = L$. Việc khảo sát ảnh hưởng của việc chọn các giá trị L_{est} thấp hơn (ví dụ: $L_{est} = 1$ hoặc $L_{est} = 2$) sẽ được trình bày ở phần kết quả (Chương 4).

2.3. Bài toán tối ưu hoá Beamforming

2.3.1. Mục tiêu tối ưu hoá

Trong các hệ thống truyền thông hiện đại, một trong những mục tiêu quan trọng nhất là truyền được càng nhiều dữ liệu càng tốt trên một lượng băng thông giới hạn. Chỉ số đo lường hiệu quả này chính là hiệu suất phổ (SE), có đơn vị là bits/s/Hz. Do đó, bài toán thiết kế beamforming thường nhắm đến việc tối đa hóa chỉ số này. Tại UE, tín hiệu nhận được phía người dùng, ký hiệu là \mathbf{r} được biểu diễn là:

$$\mathbf{r} = \mathbf{H} \mathbf{f}_{RF} \mathbf{f}_{BB} s + \mathbf{n}, \quad (2.7)$$

trong đó, \mathbf{H} là ma trận kênh truyền, \mathbf{f}_{RF} và \mathbf{f}_{BB} lần lượt là các bộ tiền mã hóa (precoder) ở miền tần số vô tuyến (RF) và băng tần gốc (BB) tại phía phát. Giá trị s là tín hiệu được

truyền đi từ phía phát, \mathbf{n} là vector AWGN có kích thước là $N_r \times 1$, tuân theo phân bố $\mathcal{N}_c(0, \sigma^2 \mathbf{I}_{N_r})$ với σ^2 là phương sai nhiễu và \mathbf{I}_{N_r} là ma trận đơn vị có kích thước $N_r \times N_r$.

Để trích xuất thông tin từ tín hiệu thu \mathbf{r} , tín hiệu này sẽ được xử lý qua hai giai đoạn tổ hợp (combiner) tại phía thu, gồm một bộ tổ hợp tương tự \mathbf{w}_{RF} và một bộ tổ hợp số \mathbf{w}_{BB} . Tín hiệu đầu ra sau các bộ tổ hợp này, ký hiệu là \mathbf{y} , được xác định như sau:

$$\mathbf{y} = \mathbf{w}_{BB}^H \mathbf{w}_{RF}^H \mathbf{r}. \quad (2.8)$$

Thay thế biểu thức của \mathbf{r} vào phương trình trên, tín hiệu \mathbf{y} được viết lại dưới dạng:

$$\mathbf{y} = \mathbf{w}_{BB}^H \mathbf{w}_{RF}^H \mathbf{H} \mathbf{f}_{RF} f_{BB} s + \mathbf{w}_{BB}^H \mathbf{w}_{RF}^H \mathbf{n}. \quad (2.9)$$

Phương trình (2.9) cho phép nhận diện rõ công suất tín hiệu và công suất nhiễu tại đầu ra của bộ tổ hợp. Thành phần $\mathbf{w}_{BB}^H \mathbf{w}_{RF}^H \mathbf{H} \mathbf{f}_{RF} f_{BB} s$, đại diện cho tín hiệu mong muốn sau khi đã qua toàn bộ chuỗi truyền dẫn và xử lý tại phía thu. Thành phần $\mathbf{w}_{BB}^H \mathbf{w}_{RF}^H \mathbf{n}$ là nhiễu đã được lọc bởi các bộ tổ hợp. Với $(.)^H$ là biểu thị phép chuyển vị liên hợp.

Dựa trên tín hiệu đầu ra đã được tổ hợp \mathbf{y} , hiệu suất phổ (SE) của hệ thống, ký hiệu là \mathbf{R} , có thể được tính toán. Hiệu suất phổ là một chỉ số quan trọng đo lường khả năng truyền tải thông tin của hệ thống trên một đơn vị băng thông (thường đo bằng bit/s/Hz). Theo định lý Shannon-Hartley cho kênh AWGN, \mathbf{R} được xác định bởi tỷ số tín hiệu trên nhiễu (SNR) tại đầu ra của bộ tổ hợp:

$$\mathbf{R} = \log_2 \left(1 + \frac{\|\mathbf{w}_{BB}^H \mathbf{w}_{RF}^H \mathbf{H} \mathbf{f}_{RF} f_{BB}\|^2}{\|\mathbf{w}_{BB}^H \mathbf{w}_{RF}^H\|^2 \sigma^2} \right). \quad (2.10)$$

Trong phương trình (2.10), tử số $\|\mathbf{w}_{BB}^H \mathbf{w}_{RF}^H \mathbf{H} \mathbf{f}_{RF} f_{BB}\|^2$ biểu thị công suất của thành phần tín hiệu mong muốn sau khi đã được định hình bởi các bộ tiền mã hóa và tổ hợp, cũng như trải qua ảnh hưởng của kênh truyền. Ký hiệu $\|\cdot\|^2$ biểu thị bình phương chuẩn Euclid của một đại lượng phức, tương ứng với công suất tín hiệu. Mẫu số $\|\mathbf{w}_{BB}^H \mathbf{w}_{RF}^H\|^2 \sigma^2$ biểu thị công suất của nhiễu tại đầu ra của bộ tổ hợp và cho thấy nhiễu gốc σ^2 đã được

biến đổi như thế nào bởi các bộ tổ hợp \mathbf{w}_{RF} và \mathbf{w}_{BB} . Kích thước của các bộ tổ hợp ảnh hưởng trực tiếp đến công suất nhiễu.

Phương trình (2.10) đóng vai trò là một công cụ phân tích hiệu quả nhằm định lượng hiệu suất thông lượng tiềm năng của hệ thống dưới các điều kiện kênh và cấu hình phần cứng xác định. Để tối đa hóa hiệu suất phổ của hệ thống, việc tối ưu hóa đồng thời các bộ tiền mã hóa ($\mathbf{f}_{RF}, \mathbf{f}_{BB}$) và các bộ tổ hợp ($\mathbf{w}_{RF}, \mathbf{w}_{BB}$) là một yêu cầu then chốt. Quá trình tối ưu hóa này là nền tảng để đạt được hiệu suất phổ cực đại, qua đó cải thiện đáng kể hiệu quả sử dụng tài nguyên phổ và dung lượng truyền dẫn tổng thể của hệ thống.

2.3.2. Bài toán tối ưu hóa hiệu suất phổ trong hệ thống Beamforming

Trong kiến trúc Analog Beamforming, do một chuỗi RF duy nhất được kết nối tới tất cả các phần tử antenna thông qua các bộ dịch pha, các thành phần của bộ tạo búp sóng tương tự phải thỏa mãn ràng buộc biên độ không đổi.

Trong một hệ thống thực tế, các thành phần như bộ dịch pha (phase shifters) chỉ có thể thay đổi pha của tín hiệu, chúng không thể khuếch đại hay làm suy giảm biên độ (độ lớn) của tín hiệu. Cụ thể, biên độ của mỗi phần tử trong vector bộ tiền mã hóa RF tại phía phát \mathbf{f}_{RF} phải bằng 1 (C_1), tức là $|\mathbf{f}_{RF}|_i = 1$, với mọi $i = 1, 2, \dots, N_t$. Tương tự, đối với vector bộ tổ hợp RF tại phía thu \mathbf{w}_{RF} , biên độ của mỗi phần tử phải bằng 1 (C_2), tức là $|\mathbf{w}_{RF}|_j = 1$, với mọi $j = 1, 2, \dots, N_r$.

Bên cạnh các ràng buộc về biên độ không đổi, hệ thống còn xem xét các ràng buộc về chuẩn hóa công suất phát và thu. Cụ thể, công suất phát truyền đi được chuẩn hóa sao cho $\|\mathbf{f}_{BB}\mathbf{f}_{RF}\|^2 \leq 1$, và công suất thu được chuẩn hóa sao cho $\|\mathbf{w}_{BB}\mathbf{w}_{RF}\|^2 \leq 1$. Với các ràng buộc này, để tối đa hóa hiệu suất phổ \mathbf{R} ở công thức (2.10), các giá trị tối ưu của \mathbf{f}_{BB} và \mathbf{w}_{BB} lần lượt là $\frac{1}{\sqrt{N_t}}$ và $\frac{1}{\sqrt{N_r}}$.

Khi áp dụng đồng thời các ràng buộc về công suất phát/thu và ràng buộc biên độ không đổi phụ thuộc vào bộ dịch pha, bài toán tối ưu hóa bộ tạo búp sóng tương tự có thể được phát biểu như sau, với $\gamma = \frac{1}{\sigma^2}$ biểu thị tỷ số tín hiệu trên nhiễu (SNR) tại phía thu:

$$\max_{\mathbf{f}_{RF}, \mathbf{w}_{RF}} \log_2 \left(1 + \frac{\gamma \|\mathbf{w}_{RF}^H \mathbf{H} \mathbf{f}_{RF}\|^2}{N_t \|\mathbf{w}_{RF}^H\|^2} \right) \quad (2.11)$$

$$\begin{aligned} \text{subject to } C_1: & |[f_{RF}]_i| = 1, \text{ với mọi } i = 1, 2, \dots, N_t, \\ C_2: & |[\mathbf{w}_{RF}]_j| = 1, \text{ với mọi } j = 1, 2, \dots, N_r. \end{aligned}$$

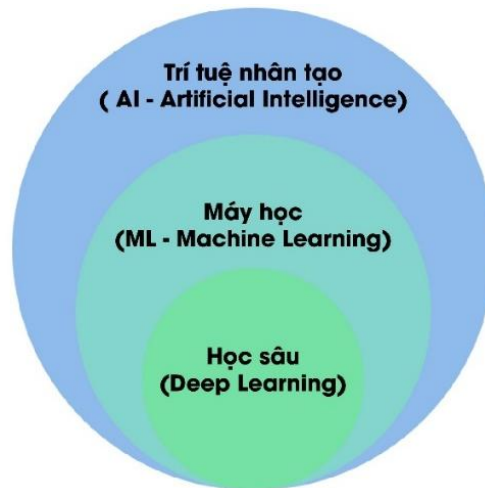
Mô hình trong đề tài được thiết kế và áp dụng cho đường xuống (downlink). Tuy nhiên, tính đối ngẫu (duality) trong truyền thông cho phép mở rộng mô hình này sang kênh đường lên (uplink). Điều này được thực hiện thông qua việc thay thế ma trận kênh \mathbf{H} bằng ma trận chuyển vị liên hợp của nó (\mathbf{H}^H), đồng thời hoán đổi vai trò của các bộ tiền mã hoá ($\mathbf{f}_{BB}, \mathbf{f}_{RF}$) và các bộ tổ hợp ($\mathbf{w}_{BB}, \mathbf{w}_{RF}$). Bài toán tối đa hóa hiệu suất phổ R như đã trình bày trong công thức (2.10) yêu cầu tối ưu hóa đồng thời bộ tạo búp sóng phát \mathbf{f}_{RF} và bộ tạo búp sóng thu \mathbf{w}_{RF} . Đây là một vấn đề phi lồi, gây ra thách thức đáng kể do sự hiện diện của ràng buộc biên độ không đổi và số lượng lớn các biến cần xem xét. Trong bối cảnh này, các mạng thần kinh học sâu DNN nổi lên như một công cụ tiềm năng, có khả năng giải quyết hiệu quả loại bài toán tối ưu phức tạp này.

2.4. Deep Learning

Học sâu (Deep Learning - DL), một phân nhánh của học máy (Machine Learning - ML), tập trung vào việc ứng dụng các thuật toán máy tính để tự học và cải thiện hiệu suất. DL sử dụng các mạng neuron nhân tạo (Artificial Neural Networks - ANN) nhằm mô phỏng cơ chế tư duy và học hỏi của con người. Những tiến bộ vượt bậc trong năng lực xử lý dữ liệu đã tăng cường đáng kể sức mạnh của DL, cho phép máy tính quan sát, học hỏi và phản ứng nhanh hơn con người trong các tình huống phức tạp.

Trong lĩnh vực viễn thông, học sâu được triển khai thành công trong nhiều kỹ thuật xử lý tín hiệu và các ứng dụng không dây khác nhau. Cụ thể, các ứng dụng bao gồm phát hiện tín hiệu, ước lượng kênh, giảm thiểu nhiễu, điều chế tín hiệu,... Ưu điểm nổi bật của các mô hình học sâu nằm ở khả năng xử lý và phân tích các tập dữ liệu lớn, phi cấu trúc và phức tạp. Bằng cách tự động điều chỉnh các tham số dựa trên các mẫu dữ liệu quan sát được, mô hình học sâu có thể liên tục cải thiện hiệu suất và hiệu quả hoạt động.

Bản chất thích ứng này khiến học sâu trở nên đặc biệt phù hợp để giải quyết các bài toán tối ưu hóa đầy thách thức trong truyền thông, chẳng hạn như tối ưu hóa bộ tạo búp sóng trong các hệ thống Analog Beamforming, đồng thời có khả năng xử lý khối lượng dữ liệu ngày càng gia tăng trong các hệ thống không dây hiện đại.



Hình 2.5. Mối quan hệ giữa trí tuệ nhân tạo, Machine Learning và Deep Learning

Các thuật toán học sâu (DL) được kế thừa từ học máy (ML) và được phân loại thành các nhóm chính như học có giám sát (Supervised Learning - SL), học không giám sát (Unsupervised Learning - UL) và học tăng cường (Reinforcement Learning - RL). Trong đó, có hai thuật toán chính để biểu diễn mối quan hệ giữa các lớp liên kề, đó là hồi quy tuyến tính (Linear Regression) và hồi quy phi tuyến (Non-linear Regression).

Thuật toán hồi quy tuyến tính là một dạng cơ bản của học máy có giám sát, nhằm mục đích xác định một đường thẳng tối ưu để mô tả xu hướng chính trong tập dữ liệu huấn luyện. Ở dạng cơ bản nhất, với đầu vào x và đầu ra y , mô hình của thuật toán này tuân theo công thức toán học cơ bản của một đường thẳng và được huấn luyện để tìm ra trọng số w và độ lệch (bias) b :

$$y = wx + b. \quad (2.12)$$

Các thuật toán hồi quy phi tuyến là các bộ phân loại mà đầu ra của chúng được biểu diễn dưới dạng xác suất. Thuật toán này được đặc trưng bởi một hàm xác suất của phương trình đường thẳng, đây chính là đầu ra của thuật toán hồi quy tuyến tính. Hàm

xác suất này được gọi là hàm kích hoạt, ký hiệu là $\sigma(\cdot)$. Phương trình của thuật toán hồi quy phi tuyến được trình bày như sau:

$$y = \sigma(wx + b). \quad (2.13)$$

2.4.1. Hàm sai số (Loss Function)

Hàm sai số trong học sâu được sử dụng để so sánh sự khác biệt giữa kết quả dự đoán của mô hình và giá trị thực tế của dữ liệu. Mục tiêu là tìm một tập hợp tối ưu các trọng số và tham số nhằm tối thiểu hóa giá trị của hàm lỗi, giúp kết quả dự đoán của mô hình tiệm cận với giá trị thực tế.

Một hàm lỗi tiêu chuẩn trong DL là sai số bình phương trung bình (Mean Squared Error - MSE). Hàm này tính toán sai số trung bình của bình phương giữa giá trị dự đoán và giá trị thực tế. MSE đo lường mức độ sai lệch và cũng phản ánh phương hướng của sự phân tán. Tuy nhiên, do phép bình phương, các lỗi có giá trị lớn hơn sẽ ảnh hưởng đáng kể hơn đến giá trị của hàm lỗi. Công thức MSE cho bài toán hồi quy tuyến tính với N quan sát được biểu diễn như sau:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (2.14)$$

trong đó y_i là giá trị thực tế, \hat{y}_i là giá trị dự đoán, và N là số lượng dữ liệu quan sát.

2.4.2. Hàm kích hoạt (Activation Function)

Hàm kích hoạt là một thành phần thiết yếu trong cấu trúc mạng neuron, được áp dụng cho đầu ra của mỗi neuron (hoặc lớp neuron) trong mạng, ngoại trừ lớp đầu vào. Mục đích chính của hàm kích hoạt là đưa yếu tố **phi tuyến tính** vào mô hình, cho phép mạng học và biểu diễn các mối quan hệ phức tạp, phi tuyến trong dữ liệu. Nếu không có các hàm kích hoạt phi tuyến, một mạng thần kinh học sâu sẽ chỉ hoạt động như một chuỗi các phép biến đổi tuyến tính, do đó hạn chế khả năng học các mô hình dữ liệu phức tạp. Hàm kích hoạt giúp mạng có thể ánh xạ các đầu vào thành đầu ra một cách phức tạp, làm tăng khả năng học hỏi và biểu diễn của mô hình. Bảng 2.2 dưới đây liệt kê một số loại hàm kích hoạt phổ biến được sử dụng trong Deep Learning.

Bảng 2.2. Các hàm kích hoạt phổ biến

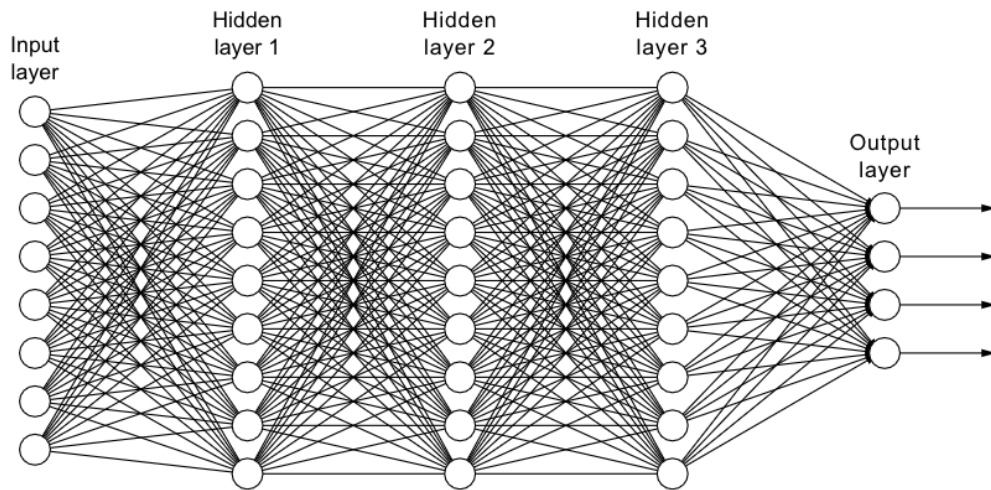
Hàm kích hoạt	Công thức	Kết quả
Sigmoid	$\sigma(x) = \frac{1}{1 + e^{-x}}$	[0,1]
Tanh	$\sigma(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	[-1,1]
ReLU	$\sigma(x) = \max(0, x)$	[0,∞]

2.4.3. Mạng thần kinh học sâu (Deep Neural Network)

Mạng thần kinh học sâu (DNNs) là một phần mở rộng trực tiếp của mạng neuron truyền thẳng (Feedforward Neural Networks) một lớp ẩn, được xây dựng bằng cách xếp chồng các lớp phi tuyến lên nhau.

Nguyên lý hoạt động của một DNN dựa trên quá trình lan truyền tín hiệu qua các lớp. Trong quá trình lan truyền xuôi (forward propagation), dữ liệu đầu vào sẽ được đưa vào lớp đầu tiên. Đầu ra của mỗi lớp sẽ trở thành đầu vào cho lớp kế tiếp. Tại mỗi neuron của một lớp ẩn, một phép biến đổi tuyến tính (nhân với ma trận trọng số và cộng với một vector thiên vị) được thực hiện, theo sau là một hàm kích hoạt phi tuyến tính (non-linear activation function) như *Sigmoid*, *Tanh* hoặc *ReLU*. Chính sự kết hợp của nhiều lớp biến đổi phi tuyến tính này cho phép DNN học được cách biểu diễn dữ liệu phức tạp. Các lớp đầu tiên gần với đầu vào có xu hướng học các đặc trưng ở mức độ thấp và cơ bản (ví dụ như các cạnh, góc trong một hình ảnh). Các lớp sâu hơn sau đó sẽ kết hợp các đặc trưng cấp thấp này lại để hình thành các đặc trưng ngày càng phức tạp và trừu tượng hơn (ví dụ như các bộ phận hoặc toàn bộ vật thể).

DNNs được xem là phương pháp cho các bài toán nhận dạng hệ thống do tính chất xấp xỉ phổ quát (universal approximation property) của mạng neuron. Tính chất này khẳng định rằng mạng neuron có khả năng xấp xỉ bất kỳ hàm liên tục nào trên các tập con nhỏ gọn trong không gian Euclid n chiều (\mathbb{R}^n) với độ chính xác tùy ý [13].



Hình 2.6. Kiến trúc mạng thần kinh học sâu

Mặc dù DNNs sở hữu tính chất xấp xỉ phổ quát, khả năng học hỏi hiệu quả từ dữ liệu vẫn là một vấn đề nghiên cứu mở. Do dung lượng mô hình cao, DNNs có xu hướng dễ bị quá khớp (overfitting). Hiện tượng này xảy ra khi mô hình "học thuộc lòng" các chi tiết và nhiễu trong tập huấn luyện thay vì học các quy luật tổng quát, dẫn đến hiệu năng kém trên dữ liệu mới.

Để giải quyết vấn đề này, các mô hình phải được ràng buộc thông qua các kỹ thuật điều chuẩn hóa (regularization) hoặc thiên kiến quy nạp (inductive bias) phù hợp. Một kỹ thuật điều chuẩn hóa phổ biến là *Dropout*, kỹ thuật này sẽ tạm thời "tắt" một tỷ lệ các neuron một cách ngẫu nhiên trong quá trình huấn luyện để mạng học các đặc trưng một cách mạnh mẽ hơn. Một kỹ thuật quan trọng khác là *Batch Normalization*, giúp chuẩn hoá đầu vào của mỗi lớp để ổn định và tăng tốc quá trình huấn luyện, đồng thời cũng có tác dụng điều chuẩn hoá nhẹ.

2.4.4. Thuật toán tối ưu

Trong quá trình huấn luyện mạng thần kinh học sâu (DNN), thuật toán tối ưu đóng vai trò điều chỉnh trọng số của các neuron nhằm mục đích tối thiểu hóa hàm sai số (Loss function). Các thuật toán tối ưu phổ biến bao gồm *Stochastic Gradient Descent (SGD)*, *Adam (Adaptive Moment Estimation)* và *RMSprop*.

Trong đề tài này, thuật toán *Adam* được lựa chọn cho mô hình DNN. Lựa chọn này dựa trên ưu điểm về tốc độ hội tụ nhanh vượt trội so với các thuật toán tối ưu hóa

khác, cùng với khả năng ổn định hơn trong quá trình học do cơ chế tự điều chỉnh tốc độ học. Đặc biệt, *Adam* thể hiện hiệu quả cao trong các bài toán hồi quy, giúp giảm thiểu lỗi ước lượng và cải thiện độ chính xác của các vector kênh ước lượng. Theo [14], cơ chế hoạt động của *Adam* là duy trì hai biến động lượng (momentum) riêng biệt cho từng tham số mô hình, với m là động lượng bậc một, đại diện cho trung bình động của các gradient và v là động lượng bậc hai, đại diện cho trung bình động của bình phương gradient.

2.5. Bài toán tối ưu hoá Analog Beamforming với DNN (ABF DNN)

Mục tiêu của việc thiết kế búp sóng là xây dựng một bài toán tối ưu hóa phức tạp nhằm tìm ra cặp búp sóng phát và thu tối ưu $(\mathbf{f}_{RF}, \mathbf{w}_{RF})$ để tối đa hóa hiệu suất phổ R (2.11), dưới các ràng buộc vật lý của phần cứng. Ngược lại, mục tiêu của quá trình huấn luyện mạng neuron là tìm ra một bộ trọng số tối ưu (θ) để tối thiểu hóa một hàm sai số λ được định nghĩa trước. Để ứng dụng học sâu vào bài toán này, cần phải thiết lập một sự liên kết toán học chặt chẽ và có ý nghĩa giữa hai bài toán tối ưu trên.

Phương pháp được lựa chọn trong đề tài này là học End-to-End, một hướng tiếp cận mạnh mẽ cho phép tối ưu hóa trực tiếp hiệu năng hệ thống. Phương pháp này định nghĩa một hàm sai số λ chính là giá trị **âm của hiệu suất phổ R** . Từ đó, bài toán tối đa hóa hiệu suất phổ được chuyển đổi một cách trực tiếp thành một bài toán tối thiểu hóa hàm sai số. Hàm sai số trong mô hình được định nghĩa như sau:

$$\lambda(\theta) = -R(\mathbf{f}_{RF}(\theta), \mathbf{w}_{RF}(\theta)) = -\log_2 \left(1 + \frac{\gamma \|\mathbf{w}_{RF}(\theta)^H \mathbf{H} \mathbf{f}_{RF}(\theta)\|^2}{N_t \|\mathbf{w}_{RF}(\theta)^H\|^2} \right), \quad (2.15)$$

trong đó, các búp sóng $\mathbf{f}_{RF}(\theta)$ và $\mathbf{w}_{RF}(\theta)$ được xem như là các hàm số theo bộ trọng số θ của mạng neuron. Bằng cách định nghĩa hàm sai số theo cách này, bài toán huấn luyện mạng neuron trở thành bài toán tối ưu hóa trực tiếp hiệu năng của hệ thống Beamforming. Mạng neuron không còn là một công cụ làm theo một thuật toán hay giải pháp có sẵn, mà đã trở thành một công cụ tối ưu hóa mạnh mẽ, có khả năng tự khám phá không gian giải pháp để tìm ra các búp sóng hiệu quả nhằm giải quyết bài toán kỹ thuật ban đầu.

CHƯƠNG 3: THIẾT KẾ BỘ TẠO BÚP SÓNG TƯƠNG TỰ BẰNG PHƯƠNG PHÁP DEEP LEARNING

3.1. Quy trình tạo dữ liệu huấn luyện

Việc huấn luyện hiệu quả mô hình Analog Beamforming Deep Neural Network (ABF DNN) đòi hỏi một tập dữ liệu lớn và đa dạng về trạng thái kênh. Quy trình tạo dữ liệu này bao gồm việc mô phỏng các ma trận kênh và tín hiệu thu trong hệ thống mmWave Massive MIMO.

3.1.1. Chuẩn bị môi trường mô phỏng và tham số hệ thống

Trước khi bắt đầu quá trình tạo dữ liệu, môi trường mô phỏng cần được cấu hình và các tham số hệ thống phải được định nghĩa một cách cụ thể. Các tham số này được thể hiện trong Bảng 3.1.

Bảng 3.1. Tham số cho quá trình tạo dữ liệu mô phỏng kênh truyền

Tham số hệ thống	Giá trị
Số antenna ở phía phát (BS) (N_t)	64/32
Số antenna ở phía thu (UE) (N_r)	1/2/4/8
Số chuỗi RF ở phía phát	1
Số chuỗi RF ở phía thu	1
Tỉ số công suất pilot trên nhiễu (PNR)	20/10/0 (dB)
Số đường truyền thực tế (L)	3 (1 LoS, 2 NLoS)
Số lượng đường truyền ước tính (L_{est})	1/2/3
Số mẫu tạo cho quá trình huấn luyện	100000

Kích thước của hệ thống MIMO được xác định bằng số lượng antenna tại BS (N_t) và số lượng antenna tại UE (N_r). Số lượng chuỗi RF tại cả phía phát và thu là 01 cho trường hợp Analog Beamforming theo mục tiêu của đề tài. Tỉ số tín hiệu trên nhiễu (SNR) đầu vào của hệ thống được thiết lập để đánh giá hiệu suất trong các điều kiện nhiễu khác

nhau. Số lượng đường truyền trong kênh L (ví dụ trong trường hợp này là 3 đường, gồm 1 đường truyền thẳng và 2 đường truyền không thẳng), phản ánh sự phức tạp của môi trường truyền. Số lượng đường truyền ước tính L_{est} quyết định cấu trúc đa đường của kênh truyền ước lượng. Các thông số vật lý khác của kênh mmWave như tần số sóng mang, bước sóng, và các yếu tố liên quan đến suy hao đường truyền trung bình.

Tổng số lượng mẫu ma trận kênh được tạo ra là 100,000 mẫu cho tập huấn luyện và 10,000 mẫu cho tập kiểm tra. Tập kiểm định được tiếp tục tách ra từ tập huấn luyện và với 10,000 mẫu (ứng với 10% tập huấn luyện; 90,000 mẫu còn lại dùng để huấn luyện).

3.1.2. Tạo kênh truyền và ước lượng kênh

Quá trình này tập trung vào việc tạo ra các cặp dữ liệu kênh bao gồm kênh hoàn hảo (Perfect Channel State Information - PCSI) và kênh ước lượng (Estimated Channel State Information - ECSI), đồng thời mô phỏng tín hiệu nhận được để phản ánh điều kiện thực tế của hệ thống.

Đối với mỗi mẫu dữ liệu, một ma trận kênh hoàn hảo (\mathbf{H}) được sinh ra. Ma trận này mô tả chính xác các đặc tính của môi trường truyền dẫn. Các góc tới (AoA) và góc phát (AoD) của mỗi đường truyền được khởi tạo ngẫu nhiên trong một phạm vi xác định (ví dụ, $[-\frac{\pi}{2}, \frac{\pi}{2}]$). Độ lợi phức của mỗi đường truyền cũng được gán ngẫu nhiên từ phân bố Gauss phức, trong đó độ lợi của đường LoS thường được chuẩn hóa và có biên độ cao hơn đáng kể so với các đường NLoS, làm nổi bật tính chất thừa của kênh mmWave. Các vector lái mảng (array steering vector) cho phía phát và phía thu được xây dựng dựa trên các góc AoA/AoD và cấu hình mảng tuyến tính đồng nhất (ULA) của antenna. Ma trận kênh hoàn hảo \mathbf{H} được tổng hợp từ đóng góp của tất cả các đường truyền đã được mô hình hóa từ công thức (2.2).

Để mô phỏng điều kiện thực tế của CSI không hoàn hảo, một quá trình ước lượng kênh được thực hiện. Quá trình này tạo ra một ma trận kênh ước lượng \mathbf{H}_e , được mô phỏng từ thuật toán ước lượng kênh trong công trình của Alkhateeb [12]. Các bộ tạo búp sóng tương tự (\mathbf{f}_{RF} , \mathbf{w}_{RF}) cho phía phát và thu được tạo ra. Các bộ tạo búp sóng này được thiết kế để tuân thủ các ràng buộc vật lý, đặc biệt là ràng buộc biên độ không đổi do sử

dụng bộ dịch pha. Tín hiệu thu được tại người dùng (\mathbf{r}) được mô phỏng bằng cách truyền tín hiệu qua kênh hoàn hảo \mathbf{H} với các búp sóng phát, sau đó tín hiệu nhiễu AWGN được cộng vào. Tín hiệu đầu ra sau khi qua bộ tổ hợp (\mathbf{y}) được trích xuất từ \mathbf{r} bằng cách áp dụng các búp sóng thu theo công thức (2.9)

Chất lượng của kênh ước tính \mathbf{H}_e được kiểm soát trực tiếp bởi tham số PNR trong quá trình mô phỏng. PNR đại diện cho tỉ số giữa công suất của các tín hiệu pilot được dùng để "dò" kênh và công suất của nhiễu nền. Độ chính xác của \mathbf{H}_e so với \mathbf{H} được định lượng bằng NMSE:

$$NMSE = 10 \log_{10} \left(E \left[\frac{||\mathbf{H} - \mathbf{H}_e||_F^2}{||\mathbf{H}||_F^2} \right] \right). \quad (3.1)$$

Mối quan hệ này có ý nghĩa quan trọng, với mức PNR cao tương ứng với một kịch bản thuận lợi (ví dụ như người dùng ở gần trạm phát). Tín hiệu pilot sẽ mạnh và rõ ràng, cho phép thuật toán ước lượng kênh hoạt động chính xác hơn, dẫn đến một kênh ước tính \mathbf{H}_e có NMSE thấp.

Với mức PNR thấp tương ứng với một kịch bản khó khăn (ví dụ như người dùng ở xa trạm phát hoặc bị che khuất). Tín hiệu pilot yếu và bị nhiễu che lấp, khiến cho thuật toán ước lượng dễ mắc sai lầm, dẫn đến một kênh ước lượng có NMSE cao. Ở các mức PNR rất thấp, NMSE có thể lớn hơn 1 đáng kể, cho thấy công suất của nhiễu đã gấp nhiều lần công suất của tín hiệu.

Cần nhấn mạnh rằng, việc khảo sát NMSE theo PNR ở đây không nhằm mục đích tối thiểu hóa NMSE. Thay vào đó, nó đóng vai trò như một công cụ để xác định chất lượng của các bộ dữ liệu đầu vào. Bằng cách kiểm soát PNR, các tập dữ liệu \mathbf{H}_e sẽ có mức độ sai số khác nhau, từ đó tạo ra một môi trường huấn luyện đa dạng và thực tế để kiểm chứng tính mạnh mẽ của mô hình được đề xuất.

Đối với mỗi mẫu dữ liệu được sinh ra, cặp ma trận kênh hoàn hảo \mathbf{H} và ma trận kênh ước lượng \mathbf{H}_e sẽ được lưu trữ. Quá trình này được lặp lại cho đến khi đạt được tổng số lượng mẫu kênh mong muốn cho tập dữ liệu huấn luyện.

3.1.3. Tổ chức bộ dữ liệu huấn luyện

Sau khi hoàn thành toàn bộ quá trình tạo dữ liệu, bộ dữ liệu huấn luyện sẽ được tổ chức thành hai tập, sẵn sàng cho việc huấn luyện mô hình DNN:

Perfect CSI (H): Đây là tập hợp các ma trận kênh hoàn hảo được sinh ra từ mô hình kênh Saleh-Valenzuela. Tập dữ liệu này đóng vai trò là "thông tin kênh hoàn hảo" cho đầu ra của mô hình, được sử dụng trong việc tính toán hàm sai số trong quá trình huấn luyện DNN.

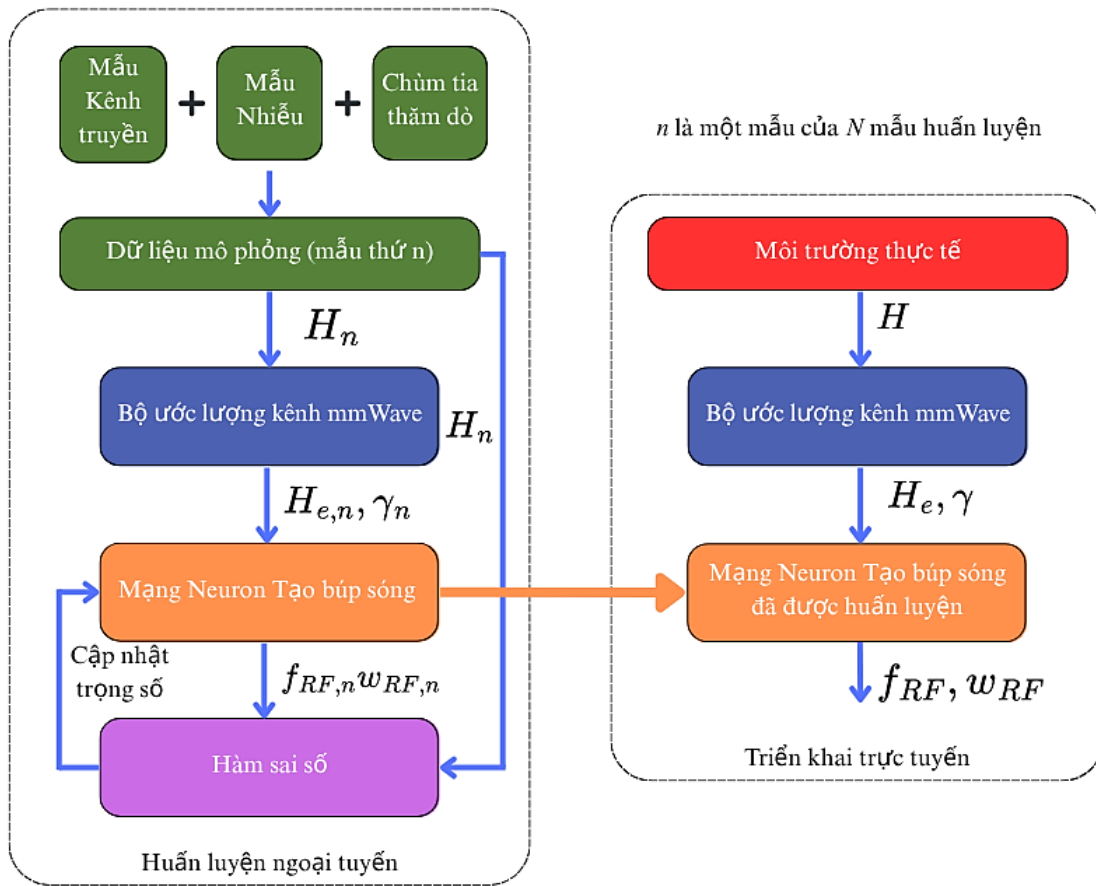
Estimated CSI (H_e): Đây là tập hợp các ma trận kênh ước lượng được tạo ra thông qua quá trình mô phỏng ước lượng kênh thích nghi [12]. Tập dữ liệu này đại diện cho "thông tin kênh không hoàn hảo" và sẽ được sử dụng làm đầu vào cho mô hình DNN trong quá trình học.

Mỗi tập hợp dữ liệu này sẽ có kích thước tương ứng với tổng số mẫu kênh được tạo, số lượng antenna thu và số lượng antenna phát (ví dụ: $100000 \times N_r \times N_t$). Việc chuẩn bị bộ dữ liệu này là nền tảng quan trọng để đảm bảo mô hình DNN có thể học một cách hiệu quả các mối quan hệ phức tạp giữa thông tin kênh ước lượng và các bộ tạo bức sóng tương tự tối ưu.

3.2. Kiến trúc và hoạt động của ABF DNN

Để giải quyết bài toán tối ưu phi lồi đã nêu, đề tài này áp dụng một phương pháp dựa trên học sâu theo hướng end-to-end. Kiến trúc mạng neuron tạo bức sóng tương tự (ABF DNN) được thiết kế để trực tiếp học một ánh xạ phức tạp từ không gian của kênh ước tính đến không gian của các bộ tạo bức sóng tương tự tối ưu. Thay vì giải quyết bài toán bằng các thuật toán lặp, mạng neuron sẽ hoạt động như một công cụ xấp xỉ hàm, có khả năng đưa ra giải pháp gần tối ưu chỉ qua một lượt tính toán duy nhất.

Quy trình thiết kế và sử dụng mô hình này được chia thành hai giai đoạn chính, đó là huấn luyện ngoại tuyến và triển khai trực tuyến, được minh họa trong Hình 3.1.



Trong giai đoạn huấn luyện ngoại tuyến, mạng DNN được huấn luyện để học mối quan hệ giữa thông tin kênh và các bộ tạo búp sóng tối ưu nhằm đạt được SE cực đại.

Dữ liệu đầu vào là dữ liệu huấn luyện được tạo ra từ các mẫu kênh mô phỏng, bao gồm thông số kênh thực tế (H_n), ước lượng kênh ($H_{e,n}$) và SNR (γ_n). Ước lượng kênh (H_e) được tìm ra bằng phương pháp từ [12] sử dụng các tham số kênh mô phỏng tạo ra H . Cụ thể, bộ ước lượng này sử dụng các chùm tia thăm dò (sounding beams) được thiết kế theo một sổ mã phân cấp để thu thập thông tin từ kênh truyền có nhiễu. Do đó, H_e là một phiên bản không hoàn hảo của kênh thật H , chứa đựng cả sai số do bản chất lượng tử hóa của thuật toán và ảnh hưởng của nhiễu nền. Dữ liệu này được tiền xử lý bằng cách tách phần thực và ảo, sau đó làm phẳng thành một vector đặc trưng trước khi đưa vào mạng DNN.

ĐẦU VÀO: H_e (kênh ước lượng), H (kênh hoàn hảo) và γ (SNR).

ĐẦU RA: Bộ tiền mã hoá tương tự (f_{RF}) và bộ tổ hợp (w_{RF}) đã được tối ưu hoá.

HÀM MỤC TIÊU: Quá trình huấn luyện sẽ tối thiểu hóa hàm sai số λ .

1. **Khởi tạo:** Bộ dữ liệu huấn luyện bao gồm $H_e \in \mathbb{C}^{N_r \times N_t}$ và $H \in \mathbb{C}^{N_r \times N_t}$ cho các giá trị kết hợp khác nhau của N_r và N_t từ thuật toán ước lượng kênh [12] được khởi tạo. Với $N_r \in 2^{\mathbb{N}}$ và $N_t \in 2^{\mathbb{N}}$.
2. γ được khởi tạo thành một phần tử duy nhất.
3. H_e được tiền xử lý bằng cách tách phần thực và ảo để tạo thành một vector số thực đầu vào, ký hiệu là H_{est} . Vì vậy, $H_{est} \in \mathbb{R}^{2 \times N_r \times N_t}$.
4. **Kiến trúc mô hình:**

Đầu vào = $[H_{est}, H, \gamma]$, Đầu ra = $[f_{RF}, w_{RF}]$, optimizer = 'Adam'

- a. *BatchNormalization* (H_{est})
- b. H_{est} được làm phẳng thành một vector có kích thước $(1 \times (2N_r N_t))$
- c. Các lớp ẩn:
 - i. *Dense*(i , activation = 'ReLU')
 - ii. *BatchNormalization* (i)
 - iii. *Dropout*()
- d. Các lớp đầu ra:

Nhánh phát: $\theta = \text{Dense}(i, \text{activation} = 'sigmoid')$,

$$f_{RF} = \cos(\theta) + j\sin(\theta).$$

Nhánh thu: $\phi = \text{Dense}(N_r, \text{activation} = 'sigmoid')$,

$$w_{RF} = \cos(\phi) + j\sin(\phi).$$

f. $\lambda = -\frac{1}{N} \sum_{n=1}^N \log_2 \left(1 + \frac{\gamma_n \|w_{RF,n}^H H_n f_{RF,n}\|^2}{N_t \|w_{RF,n}^H\|^2} \right).$

5. Lưu trọng số mô hình để sử dụng trong giai đoạn kiểm thử.
-

Kiến trúc mạng thần kinh học sâu (DNN) được đề xuất trong nghiên cứu này sở hữu một cấu trúc đa lớp tinh vi, bao gồm một lớp đầu vào (Input layer), nhiều lớp Dense (fully-connected layers), các lớp chuẩn hóa hàng loạt (Batch Normalization layers) và một lớp Lambda tùy chỉnh.

Ban đầu, để khởi động quá trình trích xuất đặc trưng, đầu ra của bộ ước lượng kênh mmWave, ký hiệu là \mathbf{H}_e , cùng với kênh hoàn hảo là \mathbf{H} và tỉ số tín hiệu trên nhiễu (SNR) γ , được cung cấp cho lớp đầu vào của DNN. Do bản chất của DNN được thiết kế để xử lý các giá trị thực, vector đầu vào phải được chuyển đổi cho phù hợp. Cụ thể, phần thực $Re(\mathbf{H}_e)$ và phần ảo $Im(\mathbf{H}_e)$ của ước lượng kênh \mathbf{H}_e (vốn là một vector phức) được nối tiếp với nhau để tạo thành một vector giá trị thực \mathbf{H}_{est} có kích thước $2 \times N_t N_r$. Vector giá trị thực này sau đó được đưa vào lớp *Dense* sơ khởi của mạng.

Kiến trúc mạng bao gồm từ ba lớp *Dense* trở lên, được triển khai trong các lớp ẩn (*Hidden layers*) với số lượng neuron giảm dần một cách có hệ thống theo chiều sâu của mạng. Số lượng neuron trong mỗi lớp này được điều chỉnh dựa trên chiều của chuỗi huấn luyện, đảm bảo quá trình tạo ra các neuron đầu ra diễn ra hiệu quả thông qua việc sử dụng các hàm kích hoạt. Hàm *ReLU* (trong Bảng 2.2), được lựa chọn làm hàm kích hoạt chính cho các lớp này, cung cấp khả năng phi tuyến cần thiết cho mô hình và giúp giảm thiểu vấn đề vanishing gradient.

Đặc biệt, một lớp *Batch Normalization* luôn được đặt ngay sau mỗi lớp *Dense*. Vai trò của lớp *Batch Normalization* là tiền xử lý đầu vào cho lớp *Dense* kế tiếp, giúp chuẩn hóa dữ liệu về trung bình 0 và phương sai 1, qua đó cải thiện đáng kể sự ổn định, tốc độ hội tụ và hiệu suất tổng thể của quá trình huấn luyện.

Ngoài ra, các lớp *Dropout* cũng được tích hợp vào kiến trúc mạng để cải thiện mô hình bằng cách giảm thiểu overfitting. *Dropout* hoạt động bằng cách ngẫu nhiên loại bỏ một lượng các neuron trong một lớp trong quá trình huấn luyện, buộc mạng phải học các biểu diễn mạnh mẽ hơn và không quá phụ thuộc vào bất kỳ neuron cụ thể nào.

Lớp cuối cùng của mạng DNN thực hiện nhiệm vụ quyết định: tạo ra các bộ tạo búp sóng phức ($\mathbf{f}_{RF}, \mathbf{w}_{RF}$) đồng thời đảm bảo chúng tuân thủ ràng buộc biên độ không

đổi của phần cứng. Để thực hiện điều này, các lớp *Lambda* tùy chỉnh được sử dụng. Đầu vào cho các lớp *Lambda* là các vector pha θ và ϕ do các lớp ẩn của mạng tạo ra. Lớp *Lambda* sau đó biến đổi các góc pha (giá trị thực) này thành các búp sóng (giá trị phức):

$$\mathbf{f}_{RF} = \exp(j\theta) = \cos(\theta) + j.\sin(\theta), \quad (3.2)$$

$$\mathbf{w}_{RF} = \exp(j\phi) = \cos(\phi) + j.\sin(\phi), \quad (3.3)$$

trong đó, θ và ϕ lần lượt biểu thị pha của các hệ số bộ tạo búp sóng tương tự trong \mathbf{f}_{RF} và \mathbf{w}_{RF} . Các vector phức \mathbf{f}_{RF} và \mathbf{w}_{RF} sau đó được sử dụng làm đầu vào cho quá trình tính toán hàm mục tiêu của mô hình (λ).

Bảng 3.2. Triển khai mạng DNN cho Analog Beamforming

Cấu hình $N_r \times N_t$	Lớp	Kích thước đầu ra	Hàm kích hoạt	Số lượng tham số
1×64	Input Layer	128×1		0
	Dense Layer 1	256×1	ReLU	33024
	Dense Layer 2	128×1	ReLU	32896
	Dense Layer 3	64×1		8256
	Lambda Layer	64×1		0
2×32	Input Layer	128×1		0
	Dense Layer 1	256×1	ReLU	33024
	Dense Layer 2	128×1	ReLU	32896
	Dense Layer 3	32×1		4128
	Dense Layer 4	2×1		258
	Lambda Layer 1	32×1		0
	Lambda Layer 2	2×1		0

Cấu hình $N_r \times N_t$	Lớp	Kích thước đầu ra	Hàm kích hoạt	Số lượng tham số
4×64	Input Layer	512×1		0
	Dense Layer 1	512×1	ReLU	262656
	Dense Layer 2	256×1	ReLU	131328
	Dense Layer 3	128×1	ReLU	32896
	Dense Layer 4	64×1		8256
	Dense Layer 5	4×1		516
	Lambda Layer 1	64×1		0
	Lambda Layer 2	4×1		0
8×64	Input Layer	1024×1		0
	Dense Layer 1	1024×1	ReLU	1049600
	Dense Layer 2	512×1	ReLU	524800
	Dense Layer 3	256×1	ReLU	131328
	Dense Layer 4	128×1	ReLU	32896
	Dense Layer 5	64×1		8256
	Dense Layer 6	8×1		1032
	Lambda Layer 1	64×1		0
	Lambda Layer 2	8×1		0

Sau khi quá trình huấn luyện ngoại tuyến hoàn tất, mô hình DNN cùng với bộ trọng số tối ưu được triển khai trên hệ thống phần cứng thực tế. Đây là giai đoạn triển khai trực tuyến của mô hình, nơi ưu điểm về tốc độ cao và độ phức tạp thấp của phương pháp học sâu được thể hiện rõ rệt, phù hợp cho các ứng dụng yêu cầu độ trễ thấp của mạng 5G.

Quy trình hoạt động trong thời gian thực diễn ra như sau: Đầu tiên, hệ thống sẽ thực hiện một phép đo kênh nhanh để thu được thông tin trạng thái kênh ước tính (\mathbf{H}_e) tại thời điểm hiện tại. Ma trận \mathbf{H}_e này sau đó được đưa vào làm đầu vào cho mô hình DNN đã được tích hợp sẵn trên bộ xử lý tín hiệu của trạm phát (BS). Mạng neuron sẽ thực hiện một lượt lan truyền xuôi duy nhất để đưa ra các vector pha tối ưu cho búp sóng phát và thu. Cuối cùng, các giá trị pha này được dùng để điều khiển trực tiếp các bộ dịch pha vật lý trong mảng antenna, qua đó tạo ra búp sóng và thiết lập một liên kết truyền thông hiệu năng cao gần như tức thời.

3.3. Hàm sai số và mục tiêu tối ưu hoá

Quá trình huấn luyện một mạng thần kinh học sâu (DNN) dựa trên phương pháp Gradient Descent và thuật toán lan truyền ngược (backpropagation). Mục tiêu của quá trình này là điều chỉnh các trọng số của mạng một cách lặp đi lặp lại để tối thiểu hóa một hàm sai số (Loss Function) đã được định nghĩa trước.

Để liên kết hai bài toán tối ưu này, như đã trình bày ở mục 2.5 của Chương 2, phương pháp được đưa ra là sử dụng một hàm sai số λ tự định nghĩa, như công thức (2.15) dựa trên chính hàm mục tiêu của hệ thống. Cụ thể, hàm sai số được định nghĩa là giá trị âm trung bình của Hiệu suất phổ trên một batch các mẫu huấn luyện:

$$\lambda = -\frac{1}{N} \sum_{n=1}^N \log_2 \left(1 + \frac{\gamma_n \|\mathbf{w}_{RF,n}^H \mathbf{H}_n \mathbf{f}_{RF,n}\|^2}{N_t \|\mathbf{w}_{RF,n}^H\|^2} \right). \quad (3.4)$$

Trong quá trình huấn luyện, các trọng số được cập nhật sau mỗi batch chứ không phải sau mỗi mẫu đơn lẻ. Việc chia cho N là để tính toán giá trị trung bình của loss trên toàn bộ các mẫu trong batch đó. Điều này giúp cho giá trị của hàm sai số không phụ thuộc vào kích thước của batch (batch size), đảm bảo quá trình cập nhật gradient là ổn định và có ý nghĩa thống kê, phản ánh đúng hiệu năng trung bình của mô hình.

3.4. Huấn luyện mô hình

Quá trình huấn luyện các mô hình mạng neuron tạo búp sóng (BFNN) được thực hiện trên nền tảng Google Colab, sử dụng GPU để tăng tốc độ tính toán. Ngôn ngữ lập

trình chính là Python, kết hợp với thư viện học sâu mã nguồn mở *TensorFlow* và *Keras* để xây dựng và tối ưu hóa các mô hình.

Mỗi mô hình, tương ứng với một cấu hình antenna cụ thể được mô tả ở Bảng 3.2, được huấn luyện trên một bộ dữ liệu đã được tạo ra ở các mức PNR khác nhau (ví dụ: 0dB, 10dB và 20dB), mô phỏng các điều kiện kênh truyền đa dạng. Các lớp *Dropout* cũng được sử dụng để tự động điều chỉnh tốc độ học và tránh hiện tượng *Overfitting*.

```
Epoch 1/500
352/352 [=====] - 7s 14ms/step - loss: -1.9876 - val_loss: -2.1930 - lr: 0.0010
Epoch 2/500
352/352 [=====] - 3s 9ms/step - loss: -2.3588 - val_loss: -2.3619 - lr: 0.0010
Epoch 3/500
352/352 [=====] - 3s 9ms/step - loss: -2.4891 - val_loss: -2.4310 - lr: 0.0010
Epoch 4/500
352/352 [=====] - 3s 9ms/step - loss: -2.5638 - val_loss: -2.4902 - lr: 0.0010
Epoch 5/500
352/352 [=====] - 5s 13ms/step - loss: -2.6141 - val_loss: -2.5195 - lr: 0.0010
Epoch 6/500
352/352 [=====] - 3s 8ms/step - loss: -2.6511 - val_loss: -2.5510 - lr: 0.0010
Epoch 7/500
352/352 [=====] - 3s 9ms/step - loss: -2.6812 - val_loss: -2.5794 - lr: 0.0010
Epoch 8/500
352/352 [=====] - 3s 9ms/step - loss: -2.7047 - val_loss: -2.5895 - lr: 0.0010
...
Epoch 492/500
352/352 [=====] - 4s 13ms/step - loss: -3.0913 - val_loss: -2.7831 - lr: 0.0010
Epoch 493/500
352/352 [=====] - 4s 12ms/step - loss: -3.0885 - val_loss: -2.7825 - lr: 0.0010
Epoch 494/500
352/352 [=====] - 3s 10ms/step - loss: -3.0895 - val_loss: -2.7913 - lr: 0.0010
Epoch 495/500
352/352 [=====] - 3s 10ms/step - loss: -3.0898 - val_loss: -2.7931 - lr: 0.0010
Epoch 496/500
352/352 [=====] - 5s 14ms/step - loss: -3.0878 - val_loss: -2.7810 - lr: 0.0010
Epoch 497/500
352/352 [=====] - 4s 10ms/step - loss: -3.0901 - val_loss: -2.7807 - lr: 0.0010
Epoch 498/500
352/352 [=====] - 3s 10ms/step - loss: -3.0889 - val_loss: -2.7934 - lr: 0.0010
Epoch 499/500
352/352 [=====] - 3s 10ms/step - loss: -3.0917 - val_loss: -2.7720 - lr: 0.0010
Epoch 500/500
352/352 [=====] - 5s 15ms/step - loss: -3.0886 - val_loss: -2.7847 - lr: 0.0010
```

Hình 3.2. Quá trình huấn luyện mô hình với cấu hình 1×64

Sự hội tụ của mô hình được theo dõi và đánh giá thông qua hàm sai số vốn được định nghĩa là giá trị âm của hiệu suất phổ. Hình 3.2 minh họa một ví dụ điển hình về quá trình huấn luyện cho cấu hình 1×64 . Giá trị sai số đã giảm đáng kể từ giá trị khởi điểm khoảng (-1.9876) ở epoch đầu tiên xuống còn (-3.0886) ở epoch thứ 500. Xu hướng giảm đều đặn và sự hội tụ của đường cong sai số là minh chứng rõ ràng cho thấy mô hình học hiệu quả và tối ưu thành công các trọng số để có thể tối đa hàm mục tiêu đã đề ra.

CHƯƠNG 4: KẾT QUẢ VÀ THẢO LUẬN

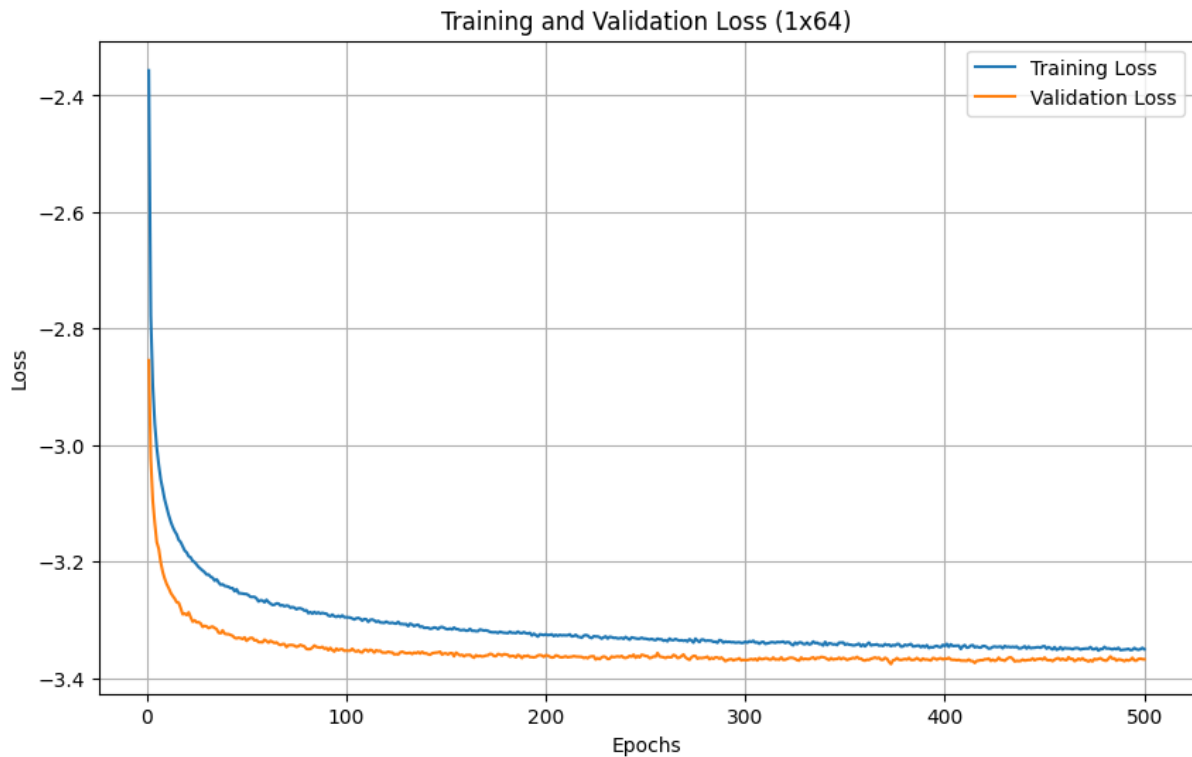
4.1. Kết quả quá trình huấn luyện

Dựa trên cơ sở lý thuyết đã trình bày ở Chương 2 và phương pháp thực hiện ở Chương 3, Chương 4 sẽ trình bày các kết quả thu được trong quá trình huấn luyện mô hình với các tham số được liệt kê trong Bảng 4.1. Kết quả thu được từ các mô hình BFNN với các cấu hình antenna khác nhau đã trình bày ở Bảng 3.2 sẽ được đánh giá và đối sánh để thấy được hiệu quả của mô hình đề xuất. Ngoài ra, ảnh hưởng của việc lựa chọn tham số cũng được khảo sát trong quá trình huấn luyện mô hình.

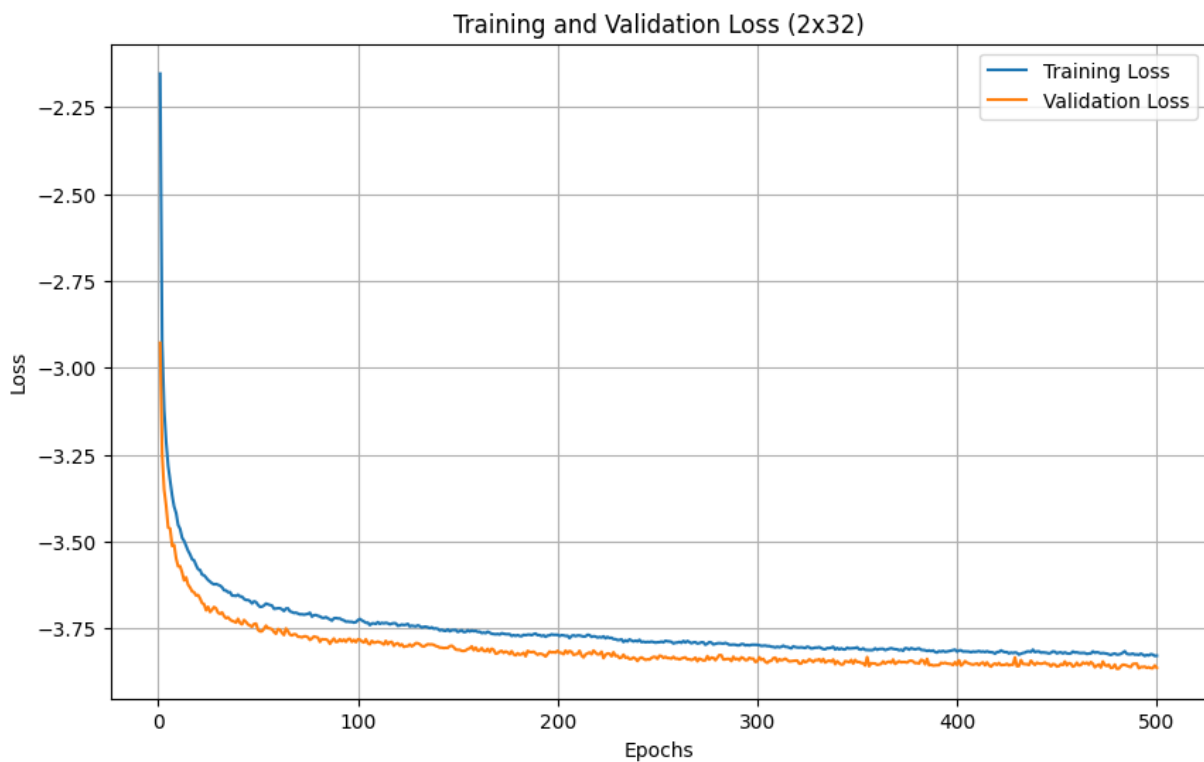
Bảng 4.1. Tham số cho quá trình huấn luyện

Tham số hệ thống	Giá trị
Tỷ số công suất pilot trên nhiễu (PNR)	0/10/20 (dB)
Số Epoch	500
Learning rate (Tốc độ học tập)	0.001
Thuật toán tối ưu	Adam
Batch size	256
Hàm kích hoạt	ReLU
Dữ liệu huấn luyện	90,000
Dữ liệu kiểm định	10,000
Dữ liệu kiểm tra	10,000

Trong quá trình huấn luyện, tập dữ liệu gồm 100,000 mẫu được phân chia thành 90,000 mẫu cho tập huấn luyện và 10,000 mẫu cho tập kiểm định. Mô hình được huấn luyện trong 500 epoch bằng thuật toán tối ưu Adam. Các siêu tham số chính bao gồm tốc độ học được thiết lập ở mức 0.001 và kích thước batch là 256. Hàm kích hoạt Rectified Linear Unit (ReLU) được sử dụng cho các lớp ẩn. Cấu hình huấn luyện này được tham khảo dựa trên công trình của Lin và Zhu [11].



Hình 4.1. Biểu đồ Loss với PNR = 10dB tại cấu hình 1×64

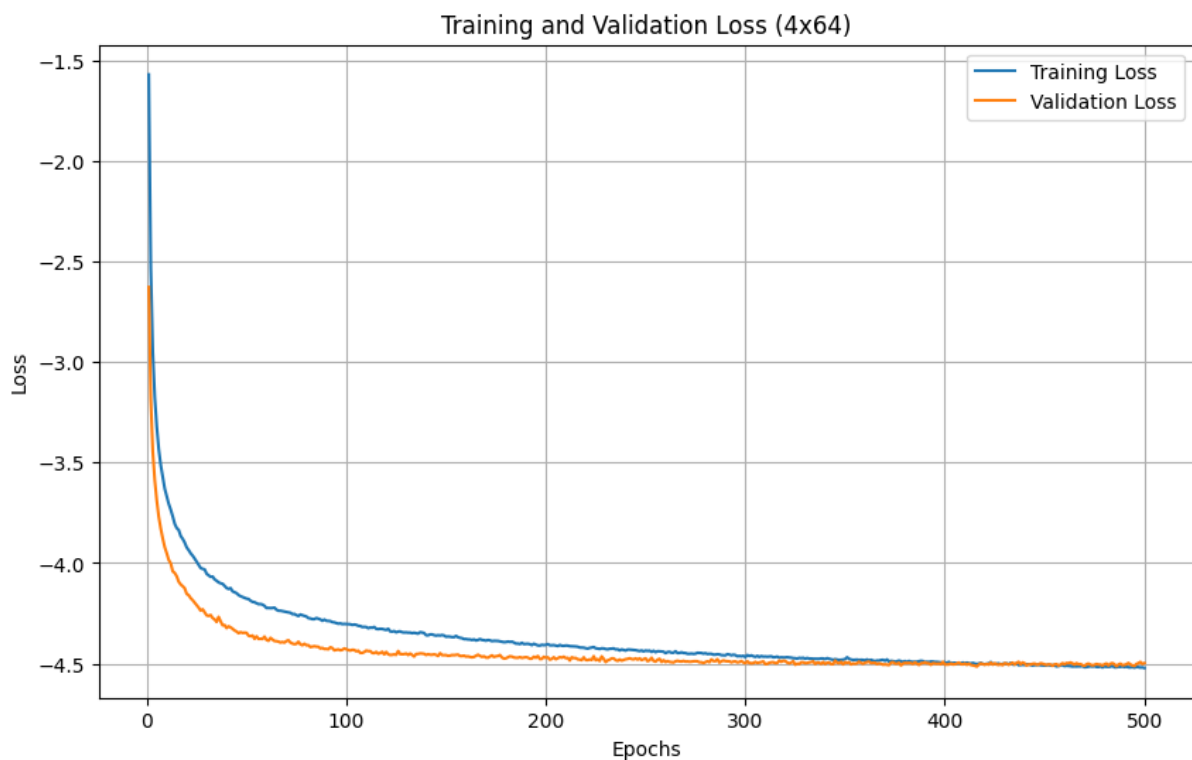


Hình 4.2. Biểu đồ Loss với PNR = 10dB tại cấu hình 2×32

Quá trình huấn luyện mô hình ở cấu hình 1×64 được thể hiện trong Hình 4.1 diễn ra trong 500 epoch. Ban đầu, *training loss* giảm mạnh từ giá trị khởi điểm khoảng

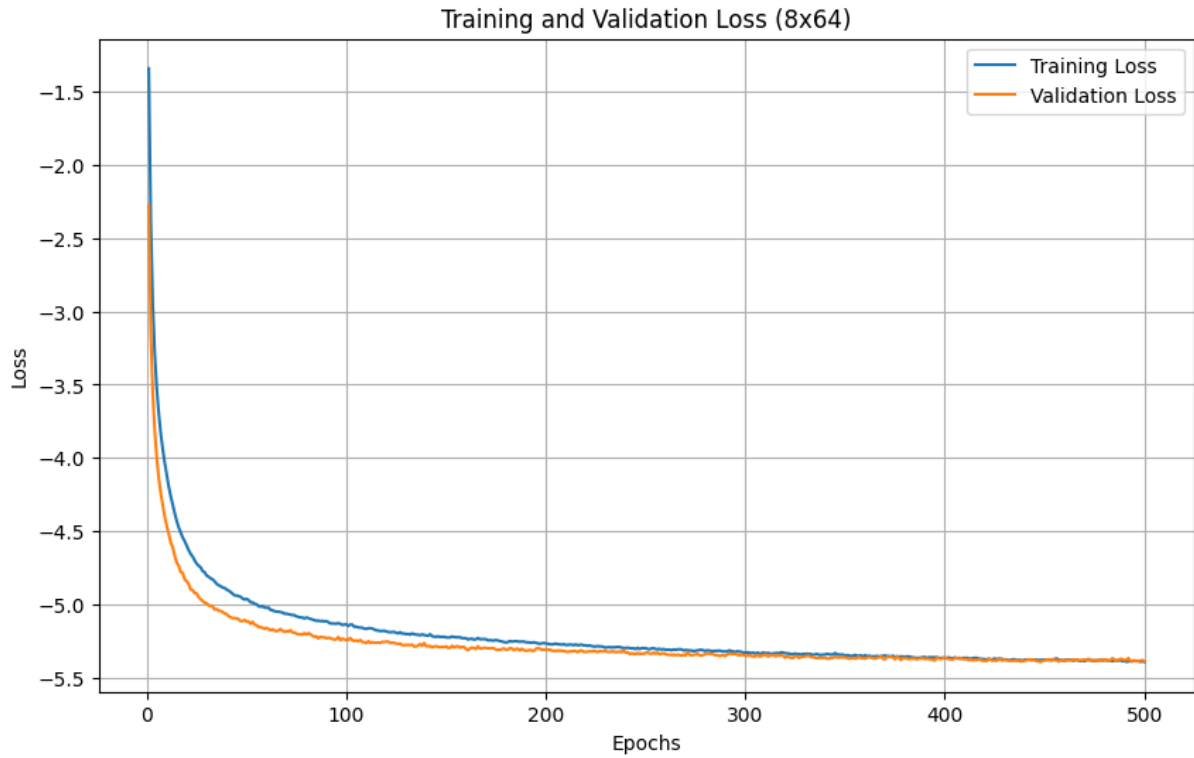
(-2.2), trong khi *validation loss* bắt đầu từ khoảng (-2.85). Cả hai nhanh chóng giảm sâu trong 100 epochs đầu tiên rồi dần hội tụ. Sau 500 epoch, mô hình ổn định khi *training loss* đạt xấp xỉ (-3.35) và *validation loss* đạt kết quả còn tốt hơn ở mức (-3.38).

Quá trình học của mô hình có cấu hình $N_r \times N_t$ là 2×32 bắt đầu với *training loss* khoảng (-2.2) và *validation loss* ở mức (-2.95). Ta có thể quan sát ở Hình 4.2, mô hình học hỏi cực kỳ nhanh chóng trong khoảng 100 epochs đầu tiên, được thể hiện qua sự sụt giảm mạnh mẽ của cả hai đường cong sai số. Sau giai đoạn khởi đầu, quá trình học chậm lại và đạt đến trạng thái hội tụ một cách ổn định. Tại epoch 500, *training loss* đạt giá trị xấp xỉ (-3.80), trong khi *validation loss* đạt mức tốt hơn là khoảng (-3.83).



Hình 4.3. Biểu đồ Loss với $\text{PNR} = 10\text{dB}$ tại cấu hình 4×64

Quá trình học của cấu hình 4×64 bắt đầu với *training loss* ở mức (-1.5) và *validation loss* khoảng (-2.6). Mô hình cho thấy khả năng học hỏi nhanh chóng trong khoảng 100 epoch đầu tiên trước khi dần đi vào trạng thái hội tụ ổn định. Tại điểm cuối cùng ở epoch 500, *training loss* đạt giá trị xấp xỉ (-4.50), trong khi *validation loss* đạt kết quả ấn tượng hơn ở mức (-4.55).



Hình 4.4. Biểu đồ Loss với $PNR = 10\text{dB}$ tại cấu hình 8×64

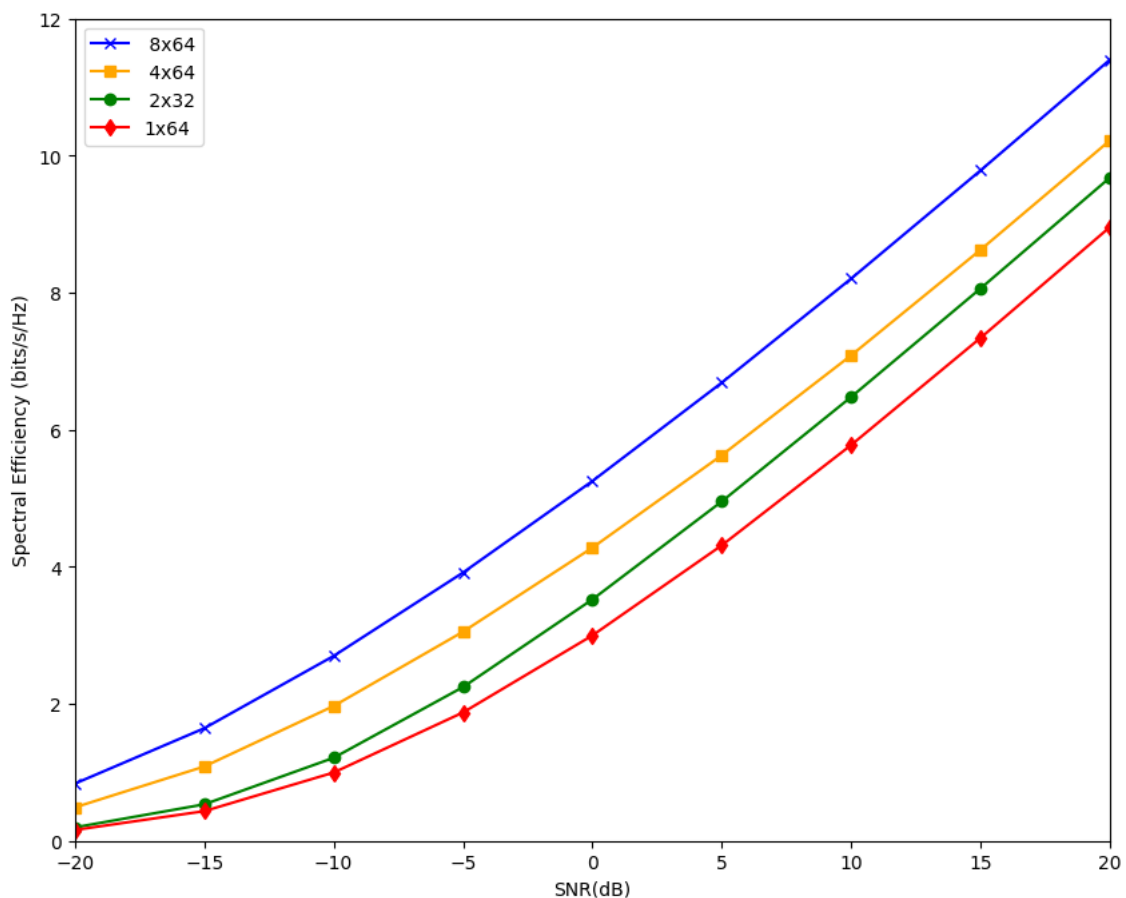
Hình 4.4 thể hiện quá trình học của mô hình với cấu hình 8×64 bắt đầu với *training loss* khoảng (-1.4) và *validation loss* khoảng (-2.2) . Mô hình thể hiện khả năng học hỏi cực kỳ nhanh chóng trong khoảng 100 epochs đầu tiên, thể hiện qua sự sụt giảm mạnh mẽ của cả hai đường loss giống với các mô hình trên. Sau giai đoạn học ban đầu, quá trình tinh chỉnh chậm lại và đạt đến trạng thái hội tụ. Tại điểm cuối cùng ở epoch 500, *training loss* ổn định tại mức (-5.4) , trong khi *validation loss* đạt xấp xỉ (-5.3) .

Khi huấn luyện ở cùng một mức $PNR = 10\text{dB}$, kết quả phân tích so sánh bốn cấu hình antenna $N_r \times N_t$ (1×64 , 2×32 , 4×64 , và 8×64) cho thấy tất cả các mô hình đều có quá trình huấn luyện ổn định và khả năng tổng quát hóa tốt, các đường cong *training loss* và *validation loss* đều giảm mạnh trong khoảng 100 epoch đầu tiên và sau đó dần ổn định cho thấy mô hình đã học hiệu quả và không có dấu hiệu overfitting nhờ việc tinh chỉnh các lớp *Dropout* và sự hiệu quả của các kỹ thuật điều chuẩn hóa. Một điểm đáng chú ý là *validation loss* ở các cấu hình thường đạt giá trị tốt hơn cả *training loss*. Điều này có thể được giải thích bởi hiệu ứng điều chuẩn hóa (regularization) của các lớp *Dropout*, vốn chỉ được kích hoạt trong quá trình huấn luyện.

4.2. Kết quả quá trình đánh giá

Sau khi các mô hình BFNN cho từng cấu hình antenna đã được huấn luyện thành công, giai đoạn tiếp theo tập trung vào việc đánh giá và so sánh hiệu năng cuối cùng của chúng. Hiệu suất phổ (Spectral Efficiency) được sử dụng làm chỉ số đo lường chính để định lượng hiệu quả của các búp sóng do mạng neuron dự đoán. Quá trình đánh giá được thực hiện trên một tập dữ liệu kiểm thử riêng biệt, với tỷ số tín hiệu trên nhiễu (SNR) được quét qua một dải rộng từ -20dB đến 20dB.

4.2.1. Phân tích và so sánh hiệu năng giữa các cấu hình antenna



Hình 4.5. So sánh hiệu suất phổ của mô hình BFNN, PNR = 10dB với các cấu hình antenna 1×64 , 2×32 , 4×64 , 8×64

Kết quả so sánh hiệu suất phổ giữa các mô hình BFNN, tương ứng với bốn cấu hình antenna khác nhau, được trình bày trong Hình 4.5. Để thực hiện một phép so sánh công bằng và khách quan về ảnh hưởng của cấu hình antenna đến hiệu năng của mô hình BFNN, một điều kiện huấn luyện nhất quán đã được thiết lập cho tất cả các trường hợp.

Cụ thể, tất cả các mô hình cho các cấu hình 1×64 , 2×32 , 4×64 và 8×64 đều được huấn luyện trên các bộ dữ liệu được tạo ra ở cùng một mức PNR là 10dB và trên cùng một dải SNR. Từ kết quả thực nghiệm, có thể rút ra hai nhận xét quan trọng.

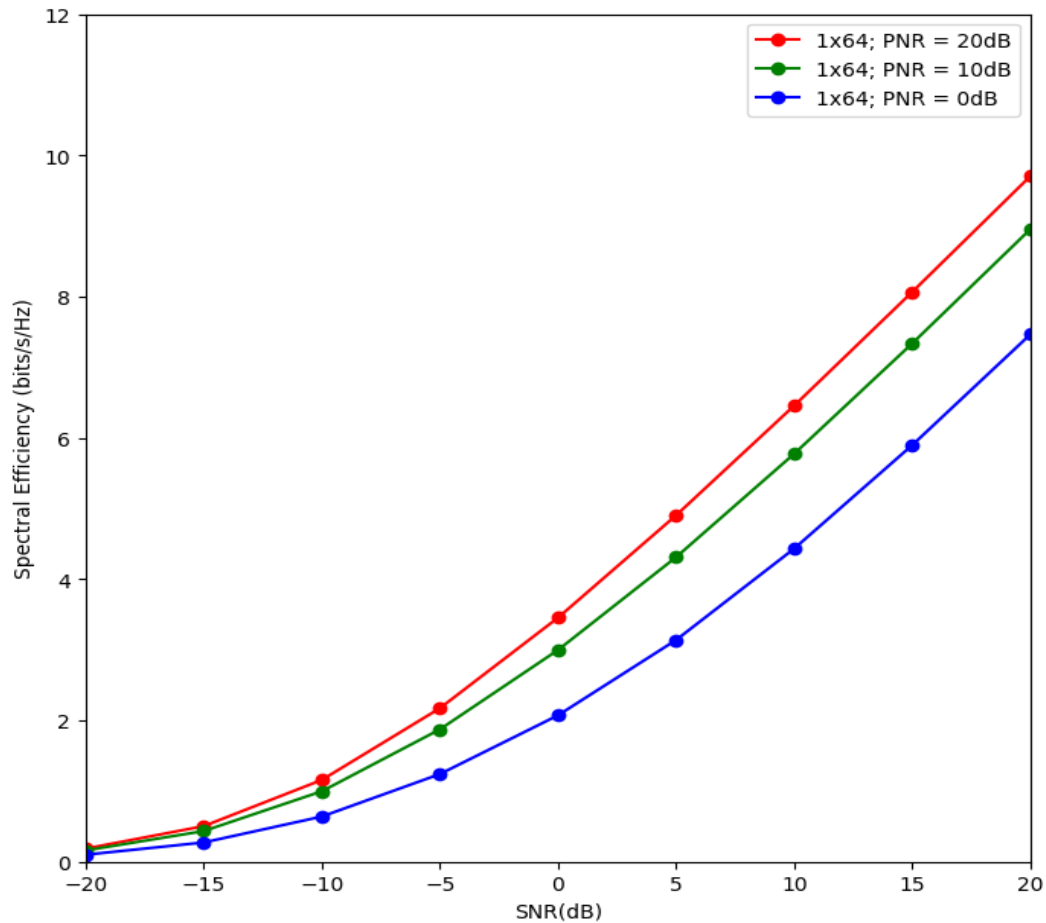
Trước hết, có thể nhận thấy một sự tương quan rõ rệt giữa SNR và hiệu suất phổ trên tất cả các cấu hình. Khi SNR tăng, chất lượng của liên kết truyền thông được cải thiện, cho phép hệ thống truyền tải được nhiều bit dữ liệu hơn trên mỗi đơn vị băng thông. Điều này là phù hợp với quy luật vật lý cơ bản của hệ thống truyền thông và cơ sở lý thuyết đã nêu ở Chương 2.

Quan trọng hơn, kết quả thực nghiệm đã cho thấy được vai trò quyết định của số lượng antenna thu (N_r) đối với hiệu năng của hệ thống. Để lượng hóa lợi ích của việc tối ưu hóa đồng thời cả phía thu, chúng ta sử dụng kết quả của mô hình MISO 1×64 làm đường cơ sở. Sự vượt trội của các cấu hình MIMO được thể hiện rõ rệt, đặc biệt là ở các vùng có SNR cao. Tại mức $SNR = 10dB$, cấu hình 2×32 đã cho thấy sự cải thiện hiệu suất phổ khoảng 12% so với cấu hình 1×64 . Khi tăng số lượng antenna thu lên 4 và 8, sự cải thiện này càng trở nên đáng kể, lần lượt đạt mức khoảng 24.1% cho cấu hình 4×64 và lên đến khoảng 41.4% cho cấu hình 8×64 .

Kết quả này cho thấy rằng mô hình BFNN đề xuất có khả năng mở rộng tốt và học được cách khai thác hiệu quả lợi thế từ các cấu hình phân cứng mạnh mẽ hơn. Việc tăng số lượng antenna thu không chỉ giúp chống lại các hiệu ứng bất lợi của kênh truyền mà còn trực tiếp nâng cao giới hạn hiệu năng của toàn hệ thống. Do đó, có thể nhận thấy rằng, ngay cả trong kiến trúc Analog Beamforming, việc áp dụng kiến trúc MIMO với nhiều antenna thu là một chiến lược cực kỳ hiệu quả để tối ưu hóa hiệu suất phổ.

4.2.2. Khảo sát ảnh hưởng của chất lượng dữ liệu huấn luyện

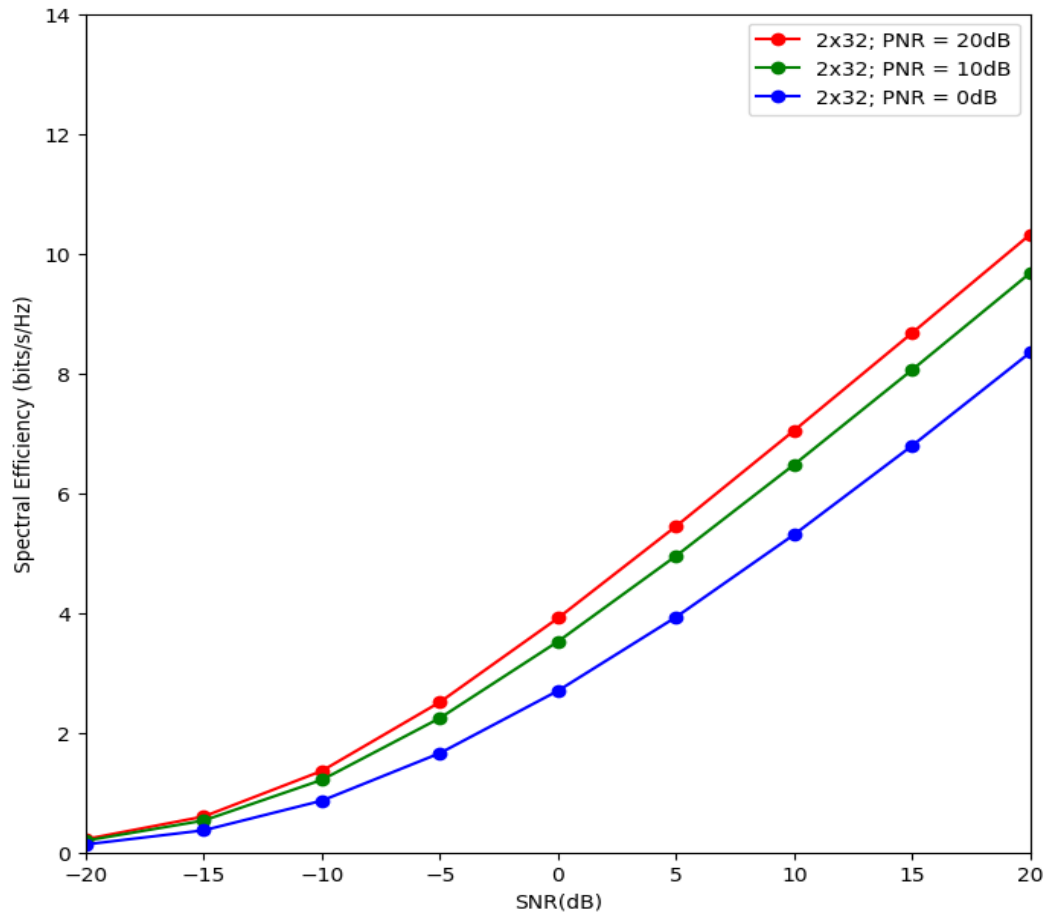
Để khảo sát tác động của chất lượng thông tin trạng thái kênh (CSI) đầu vào đến hiệu năng của mô hình học sâu, một thí nghiệm so sánh đã được thực hiện. Kết quả phân tích cho thấy chất lượng của dữ liệu huấn luyện, được đặc trưng bởi PNR, có tác động quyết định đến hiệu suất phổ của mô hình trên toàn dải SNR khảo sát.



Hình 4.6. So sánh hiệu suất phổ của mô hình BFNN cấu hình 1×64 với các mức PNR khác nhau

Đối với cấu hình cơ bản nhất là 1×64 (MISO), Hình 4.6 cho thấy hiệu năng của mô hình có sự phụ thuộc lớn vào chất lượng của dữ liệu huấn luyện. Ở vùng có tỷ số tín hiệu trên nhiễu (SNR) thấp, chẳng hạn tại $SNR = -10dB$, hiệu suất phổ tăng mạnh khoảng 43% khi PNR huấn luyện được nâng từ 0dB lên 10dB, và tiếp tục tăng thêm 30% ở bước cải thiện tiếp theo lên 20dB. Tại vùng có SNR cao, ví dụ tại $SNR = 10dB$, các mức tăng tương đối này trở nên khiêm tốn hơn, lần lượt là khoảng 21% và 12%.

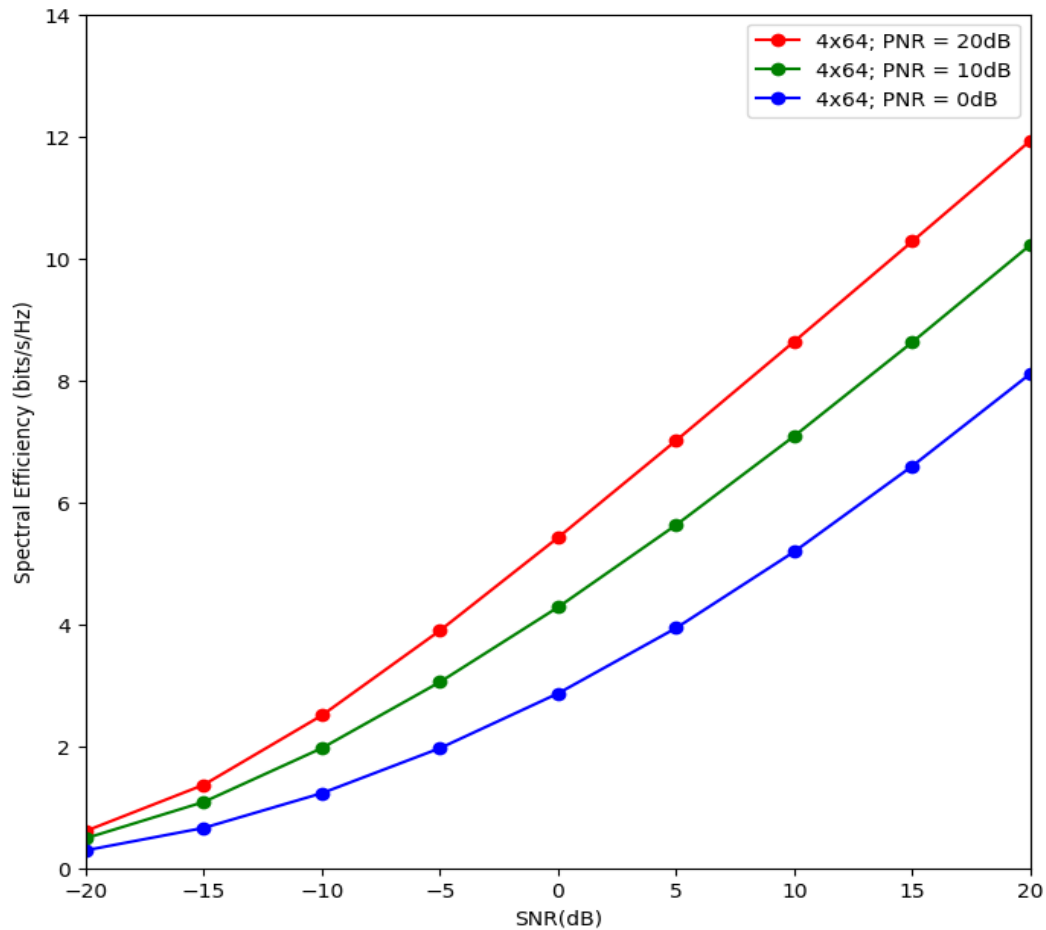
Các quan sát này chỉ ra rằng, ngay cả với cấu hình antenna đơn giản nhất, việc đảm bảo chất lượng dữ liệu huấn luyện là rất quan trọng. Lợi ích tương đối lớn nhất đạt được quan sát được khi thay đổi từ $PNR = 0dB$ sang $PNR = 10dB$, đặc biệt là trong các kịch bản nhiễu nặng, giúp mô hình hoạt động hiệu quả hơn.



Hình 4.7. So sánh hiệu suất phổ của mô hình BFNN cấu hình 2×32 với các mức PNR khác nhau

Đối với cấu hình 2×32 tại Hình 4.7, phân tích ở các mức SNR khác nhau tiếp tục khẳng định tầm quan trọng của chất lượng của kênh ước tính đầu vào \mathbf{H}_{est} . Trong vùng có tỷ số tín hiệu trên nhiễu (SNR) thấp, chẳng hạn tại $SNR = -10dB$, hiệu suất phổ tăng mạnh 50% khi PNR huấn luyện được nâng từ 0dB lên 10dB, nhưng chỉ tăng thêm khoảng 17% ở bước cải thiện tiếp theo lên 20dB.

Tương tự, tại vùng có SNR cao ($SNR = 10dB$), các mức tăng tương đối này trở nên ôn hòa hơn, lần lượt là khoảng 18% và 8%. Các kết quả này cho thấy, đối với cấu hình 2×32 , lợi ích lớn nhất đến từ việc đảm bảo một chất lượng dữ liệu huấn luyện cơ bản (chuyển từ $PNR = 0dB$ lên $PNR = 10dB$). Việc tiếp tục tăng mức PNR lên 20dB vẫn mang lại hiệu quả nhưng với mức tăng thêm không còn đột phá như bước đầu, đặc biệt là trong các kịch bản nhiễu nặng.

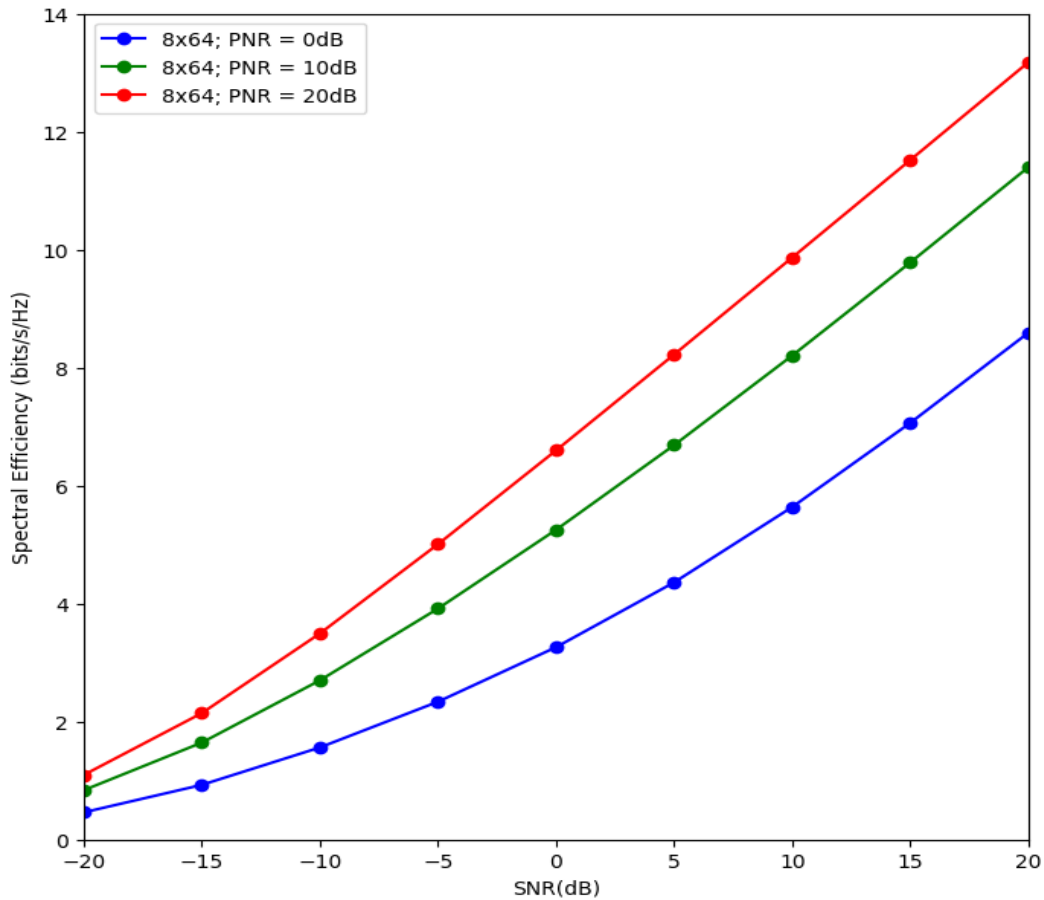


Hình 4.8. So sánh hiệu suất phổ của mô hình BFNN cấu hình 4×64 với các mức PNR khác nhau

Quá trình khảo sát trên cấu hình MIMO 4×64 tại Hình 4.8 một lần nữa khẳng định tác động quyết định của chất lượng dữ liệu huấn luyện đến hiệu năng cuối cùng của mô hình BFNN. Sự cải thiện này được thể hiện rõ rệt nhất trong vùng có tỷ số tín hiệu trên nhiễu (SNR) thấp. Cụ thể, tại mức $SNR = -10dB$, hiệu suất phổ tăng khoảng 58% khi PNR huấn luyện được cải thiện từ 0dB lên 10dB và tiếp tục tăng thêm 32% khi PNR tại mức 20dB.

Trong khi đó, ở vùng có SNR cao, chẳng hạn tại $SNR = 10dB$, các mức tăng tương đối này trở nên ôn hòa hơn, lần lượt là khoảng 39% và 19%. Những kết quả này cho thấy rằng, trong khi việc cải thiện chất lượng dữ liệu huấn luyện luôn mang lại lợi ích, thì lợi ích tương đối lớn nhất đạt được khi chuyển từ một bộ dữ liệu có chất lượng rất thấp ($PNR = 0dB$) sang mức vừa phải ($PNR = 10dB$) là yếu tố tối quan trọng để

đảm bảo sự mạnh mẽ và hiệu quả, với lợi ích về mặt tỉ lệ được thể hiện rõ nhất trong các kịch bản hoạt động có SNR thấp.



Hình 4.9. So sánh hiệu suất phổ của mô hình BFNN cấu hình 8×64 với các mức PNR khác nhau

Phân tích chi tiết trên cấu hình 8×64 , vốn là kiến trúc phức tạp nhất, cho thấy hiệu suất phổ cuối cùng của hệ thống có sự phụ thuộc rất lớn vào chất lượng của dữ liệu được sử dụng trong quá trình huấn luyện. Kết quả thực nghiệm ở Hình 4.9 đã định lượng được một sự cải thiện hiệu suất tuần tự và nhất quán khi chất lượng kênh ước tính đầu vào, đặc trưng bởi PNR, được nâng cao. Cụ thể, tại mức $SNR = -10dB$, một sự cải thiện đột phá về hiệu suất đã được ghi nhận: mô hình được huấn luyện ở $PNR = 10dB$ cho hiệu suất phổ cao hơn tới 80% so với mô hình được huấn luyện ở $PNR = 0dB$. Khi chất lượng dữ liệu tiếp tục được cải thiện lên $PNR = 20dB$, hiệu suất vẫn tăng thêm gần 30%. Trong khi đó, ở vùng có SNR cao, chẳng hạn tại $SNR = 10dB$, mức tăng tương đối ở

mỗi bước ôn hòa hơn, lần lượt là khoảng 26% và 21%, mặc dù mức tăng tuyệt đối về hiệu suất ở vùng này vẫn rất đáng kể.

Một nhận xét chung và quan trọng có thể rút ra từ kết quả thực nghiệm là tính nhất quán trong hành vi của phương pháp thiết kế búp sóng dựa trên học sâu (BFNN) khi đối mặt với các điều kiện kênh và cấu hình phần cứng khác nhau. Các kết quả này không chỉ định lượng hiệu suất mà còn cho thấy những năng lực nội tại của mô hình được đề xuất.

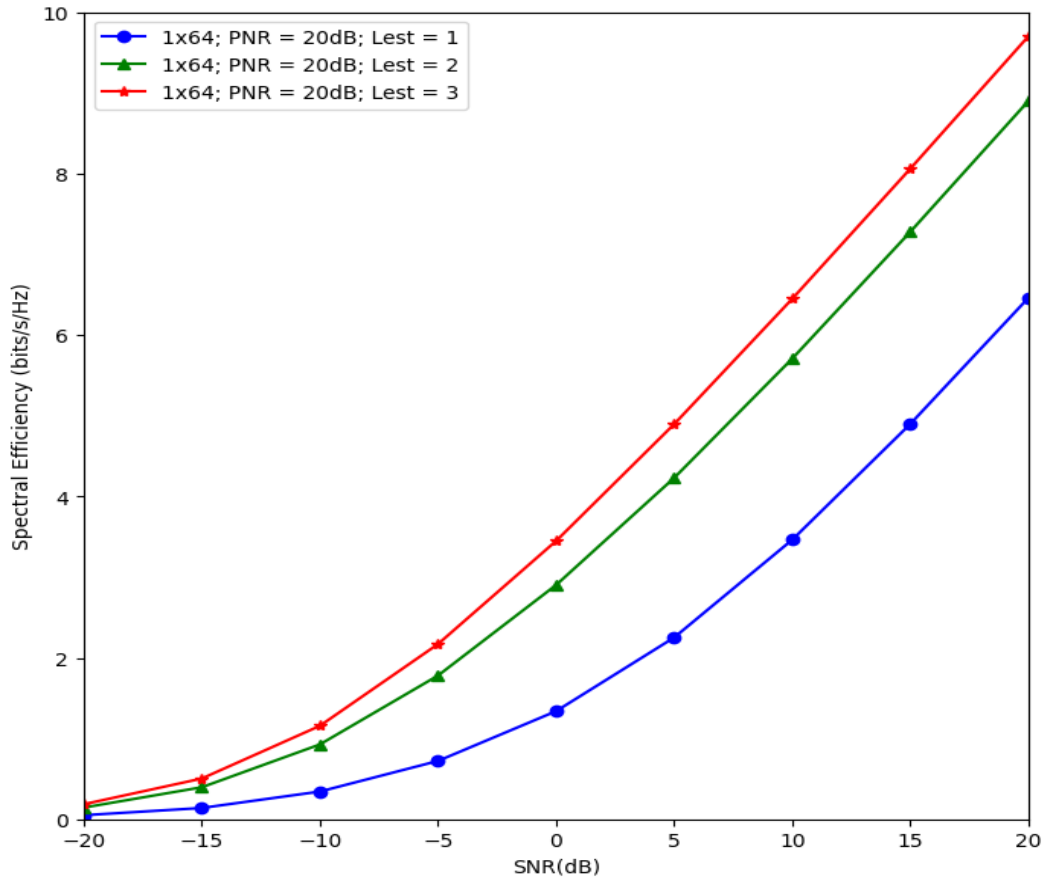
Thứ nhất, các kết quả đã chứng minh được khả năng của phương pháp học sâu trong việc tạo ra các mô hình chuyên biệt, thích ứng với chất lượng của thông tin trạng thái kênh (CSI) có sẵn trong quá trình huấn luyện. Cụ thể, trên tất cả các cấu hình antenna, hiệu suất của các mô hình được phân cấp một cách rõ ràng theo chất lượng của bộ dữ liệu huấn luyện (đặc trưng bởi PNR). Mô hình được huấn luyện trên dữ liệu có chất lượng cao ($PNR = 20dB$) đã khai thác tối đa thông tin chính xác để tạo ra các búp sóng gần như tối ưu. Ngược lại, mô hình được huấn luyện trên dữ liệu nhiễu nặng ($PNR = 0dB$) đã học được tính mạnh mẽ (robustness), thể hiện qua khả năng trích xuất các đặc trưng cơ bản của kênh để duy trì hoạt động hiệu quả trong các điều kiện khó khăn.

Thứ hai, một phát hiện quan trọng là lợi ích tương đối của việc huấn luyện trên dữ liệu tốt được thể hiện rõ rệt nhất trong các kịch bản SNR thấp. Điều này cho thấy việc huấn luyện trên một bộ dữ liệu có chất lượng đủ tốt là yếu tố then chốt để mô hình có thể phát huy hiệu quả, đặc biệt là trong việc duy trì sự ổn định và hiệu quả hoạt động ở các vùng nhiễu nặng, vốn là các kịch bản vận hành đầy thách thức trong thực tế.

Cuối cùng, sự nhất quán của các xu hướng này trên cả bốn cấu hình antenna, từ MISO 1×64 đến MIMO 8×64 , cho thấy rằng đây là một đặc tính nội tại và mạnh mẽ của phương pháp học sâu được đề xuất. Nó cho thấy phương pháp này không chỉ hoạt động trong một trường hợp riêng lẻ mà còn có khả năng mở rộng và áp dụng hiệu quả cho các hệ thống có cấu trúc phần cứng đa dạng hơn.

Để khảo sát tác động của độ chính xác trong quá trình ước tính kênh, một thí nghiệm so sánh đã được thực hiện trên cấu hình MISO 1×64 . Hình 4.10 trình bày hiệu suất phổ của mô hình khi kênh ước tính đầu vào được tạo ra bằng cách sử dụng các số

lượng đường truyền ước tính (L_{est}) khác nhau, cụ thể là $L_{est} = 1$, $L_{est} = 2$ và $L_{est} = 3$, trong khi số lượng đường truyền vật lý thực tế là $L = 3$. Cùng với đó, mức PNR được đặt ở 20dB cho cả hai trường hợp để đảm bảo tính công bằng khi so sánh.



Hình 4.10. So sánh hiệu suất phổ của mô hình BFNN cấu hình 1×64 , $\text{PNR} = 20\text{dB}$, với các tham số L_{est} khác nhau

Kết quả so sánh cho thấy một xu hướng rất rõ rệt: hiệu suất phổ được cải thiện một cách nhất quán khi tăng số lượng đường truyền ước tính (L_{est}) từ 1 lên 3. Sự khác biệt này bắt nguồn trực tiếp từ chất lượng của kênh ước tính được cung cấp cho mô hình BFNN. Để định lượng sự cải thiện này, tại một điểm $\text{SNR} = 10\text{dB}$, việc tăng L_{est} từ 1 lên 2 đã cải thiện hiệu suất phổ thêm khoảng 57%. Tiếp tục tăng L_{est} từ 2 lên 3 mang lại sự cải thiện thêm khoảng 29%. Kết quả này khẳng định rằng độ chính xác của thuật toán ước tính kênh, cụ thể là khả năng nắm bắt được đặc tính đa đường của kênh truyền, là một yếu tố cực kỳ quan trọng, có ảnh hưởng trực tiếp và mạnh mẽ đến hiệu suất tổng thể của hệ thống.

CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN ĐỀ TÀI

5.1. Kết luận

Khóa luận này đã trình bày việc thiết kế và đánh giá một phương pháp tạo búp sóng tương tự (Analog Beamforming) cho hệ thống Massive MIMO dựa trên học sâu. Phương pháp đề xuất sử dụng một mạng thần kinh (ABF DNN) được huấn luyện theo hướng end-to-end để tối ưu hóa trực tiếp hiệu suất phổ, cung cấp một giải pháp hiệu quả cho bài toán tối ưu hóa phi lồi vốn có độ phức tạp tính toán cao và nhạy cảm với thông tin trạng thái kênh (CSI) không hoàn hảo. Các kết quả mô phỏng đã khẳng định tính khả thi và hiệu quả của phương pháp, đồng thời cung cấp các kết luận quan trọng.

Thứ nhất, đề tài cho thấy phương pháp tối ưu hóa hiệu suất phổ trực tiếp bằng Deep Learning là một hướng tiếp cận hiệu quả. Bằng cách định nghĩa hàm sai số là giá trị âm của hiệu suất phổ, mô hình ABF DNN đã học được cách ánh xạ từ CSI không hoàn hảo đến các búp sóng tối ưu. Sự hội tụ ổn định của quá trình huấn luyện trên nhiều cấu hình cho thấy khả năng của mạng trong việc tự khám phá không gian giải pháp của bài toán tối ưu hóa phi lồi này.

Thứ hai, cấu trúc phần cứng antenna có vai trò quyết định đến hiệu năng hệ thống. Kết quả cho thấy hiệu suất phổ tăng một cách nhất quán khi tăng số lượng antenna thu, bắt nguồn trực tiếp từ lợi ích phân tập của công nghệ massive MIMO. Điều này khẳng định khả năng mở rộng và khai thác hiệu quả mô hình đề xuất.

Thứ ba, chất lượng của dữ liệu huấn luyện là yếu tố then chốt. Các thí nghiệm cho thấy hiệu suất mô hình phụ thuộc sâu sắc vào chất lượng của CSI đầu vào, đặc trưng bởi tỉ số công suất pilot trên nhiễu (PNR). Lợi ích của việc huấn luyện trên dữ liệu chất lượng cao (PNR cao) thể hiện rõ rệt nhất ở các kịch bản có tỉ số tín hiệu trên nhiễu (SNR) thấp, qua đó khẳng định tầm quan trọng của dữ liệu trong việc đảm bảo tính mạnh mẽ của hệ thống trong môi trường nhiễu nặng.

Cuối cùng, hiệu năng của mô hình phụ thuộc lớn vào độ chính xác của thuật toán ước tính kênh. Kết quả thực nghiệm chỉ ra rằng việc cung cấp một mô hình kênh chính xác hơn, thông qua việc ước tính đủ số lượng đường truyền (L_{est}), có thể cải thiện

hiệu suất phổ lên tới 57%. Điều này cũng cho thấy rằng chất lượng thông tin đầu vào là một trong những yếu tố quyết định đến hiệu quả của phương pháp học sâu.

Tóm lại, khóa luận đã hoàn thành các mục tiêu đề ra, cung cấp những phân tích rõ ràng về việc ứng dụng học sâu cho bài toán Analog Beamforming và cho thấy đây là một hướng tiếp cận tiềm năng cho các hệ thống truyền thông không dây thế hệ mới.

5.2. Hướng phát triển đề tài

Dựa trên các kết quả đã đạt được và những giới hạn đã được xác định, một số hướng nghiên cứu và phát triển tiềm năng trong tương lai có thể được đề xuất như sau:

Mở rộng qua kiến trúc Hybrid Beamforming: Đề tài hiện tại chỉ tập trung vào kiến trúc Analog Beamforming thuần túy với một chuỗi RF. Một hướng phát triển tự nhiên và có tính thực tiễn cao là mở rộng phương pháp này cho kiến trúc Hybrid Beamforming. Bài toán sẽ trở nên phức tạp hơn khi mô hình DNN cần phải tối ưu hóa đồng thời cả bộ tạo búp sóng tương tự ở miền RF và bộ tiền mã hóa số ở miền băng gốc.

Nghiên cứu kịch bản đa người dùng (Multi-User MIMO): Mô hình hệ thống trong đề tài được giới hạn ở kịch bản đơn người dùng. Việc mở rộng sang môi trường đa người dùng sẽ là một bước tiến quan trọng. Thách thức chính trong kịch bản này là việc quản lý và khử nhiễu liên người dùng. Hướng nghiên cứu trong tương lai có thể tập trung vào việc thiết kế một hàm sai số mới, không chỉ tối đa tổng hiệu suất phổ (sum-rate) mà còn nhằm tối thiểu nhiễu liên người dùng.

Khảo sát các kiến trúc mạng tiên tiến hơn: Đề tài đã sử dụng thành công mô hình DNN truyền thẳng. Một hướng cải tiến là khảo sát các kiến trúc mô hình khác như mạng tích chập (CNN) hay các mô hình Transformer, vốn có khả năng học sâu và khai thác các đặc trưng không gian trong ma trận kênh một cách hiệu quả hơn.

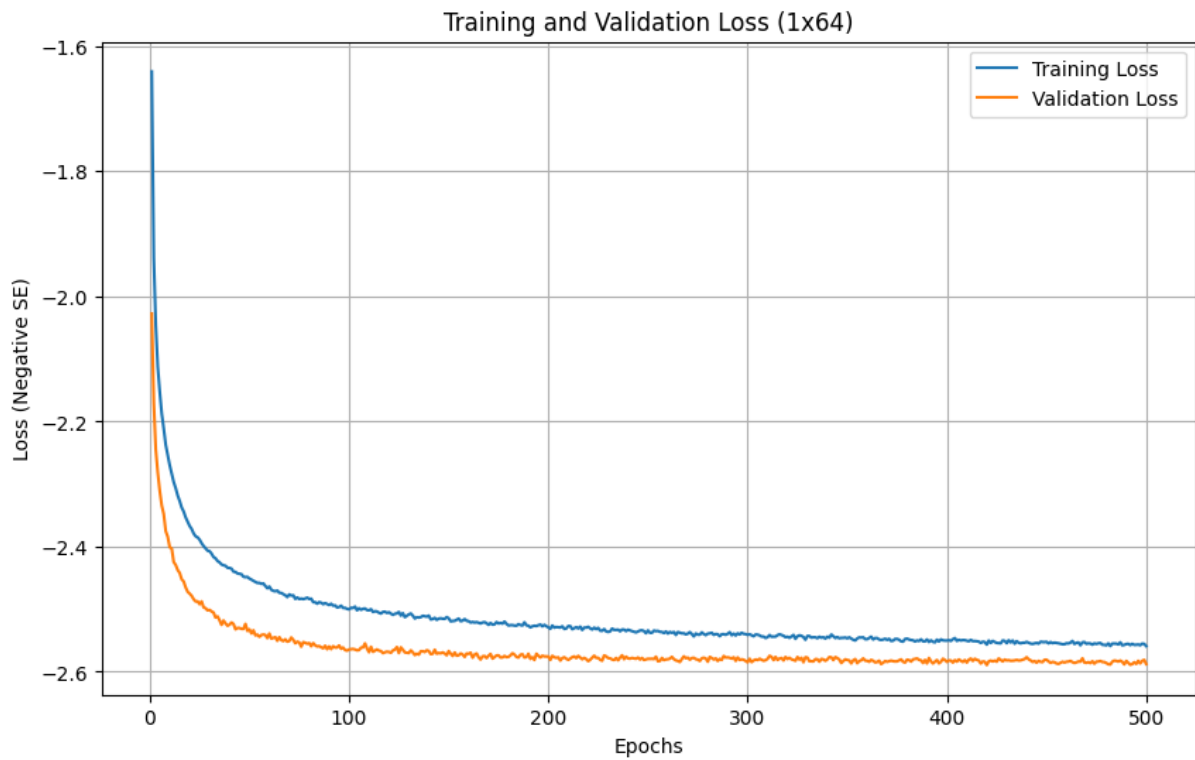
Những hướng phát triển này sẽ góp phần hoàn thiện phương pháp tiếp cận dựa trên học sâu, từng bước thu hẹp khoảng cách giữa các kết quả nghiên cứu trong môi trường mô phỏng và các yêu cầu phức tạp của các hệ thống truyền thông 5G và xa hơn.

TÀI LIỆU THAM KHẢO

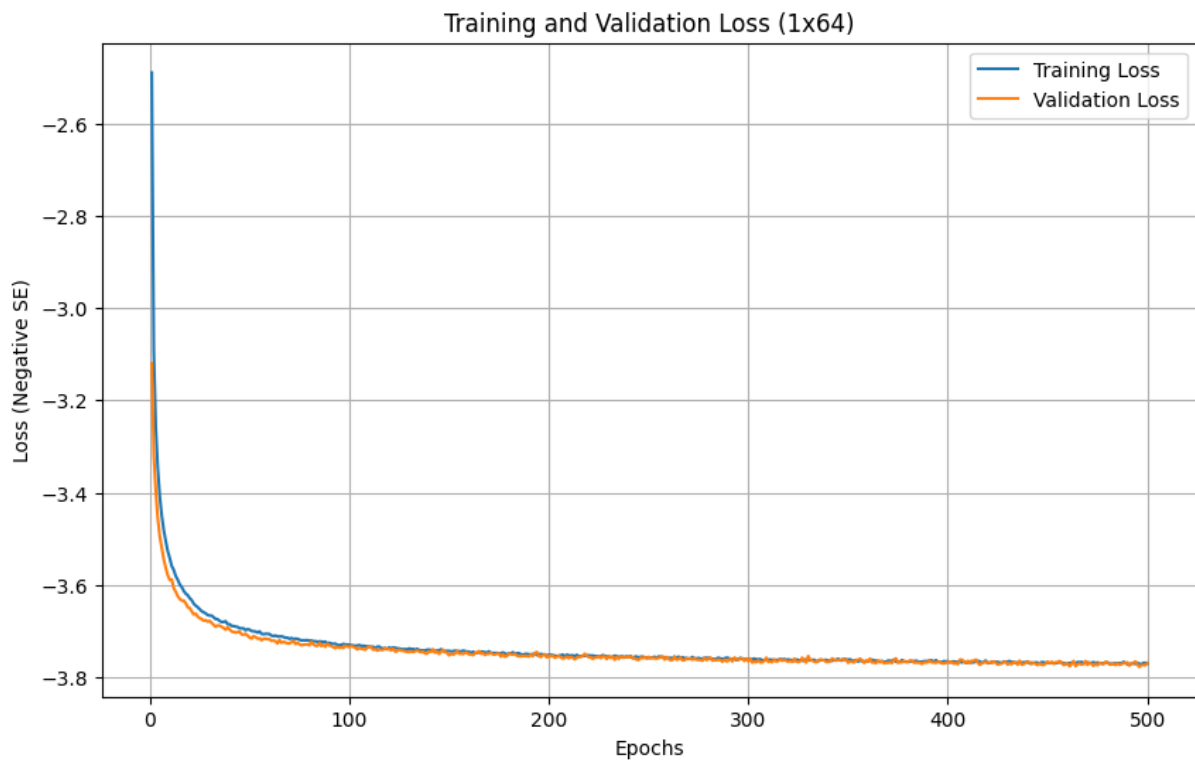
- [1] “What Is Beamforming?” Accessed: July 05, 2025. [Online]. Available: <https://www.mathworks.com/discovery/beamforming.html>
- [2] S. Ahmed, “BEAMFORMING MANAGEMENT AND BEAM TRAINING IN 5G SYSTEM”.
- [3] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, “Spatially Sparse Precoding in Millimeter Wave MIMO Systems,” *IEEE Trans. Wirel. Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014, doi: 10.1109/TWC.2014.011714.130846.
- [4] X. Gao, L. Dai, S. Han, C.-L. I, and X. Wang, “Reliable Beamspace Channel Estimation for Millimeter-Wave Massive MIMO Systems with Lens Antenna Array,” *IEEE Trans. Wirel. Commun.*, vol. 16, no. 9, pp. 6010–6021, Sept. 2017, doi: 10.1109/TWC.2017.2718502.
- [5] T. Lin, J. Cong, Y. Zhu, J. Zhang, and K. B. Letaief, “Hybrid Beamforming for Millimeter Wave Systems Using the MMSE Criterion,” *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3693–3708, May 2019, doi: 10.1109/TCOMM.2019.2893632.
- [6] Z. Cheng, N. Li, R. Long, J. Zhu, C. Ouyang, and P. Chen, “Secure Antenna Selection and Beamforming in MIMO Systems,” Nov. 04, 2024, *arXiv*: arXiv:2302.11127. doi: 10.48550/arXiv.2302.11127.
- [7] W. Xia, G. Zheng, K.-K. Wong, and H. Zhu, “Model-Driven Beamforming Neural Networks,” *IEEE Wirel. Commun.*, vol. 27, no. 1, pp. 68–75, Feb. 2020, doi: 10.1109/MWC.001.1900239.
- [8] S. Sanayei and A. Nosratinia, “Antenna selection in MIMO systems,” *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 68–73, Oct. 2004, doi: 10.1109/MCOM.2004.1341263.

- [9] S. Dörner, S. Cammerer, J. Hoydis, and S. ten Brink, “Deep Learning-Based Communication Over the Air,” *IEEE J. Sel. Top. Signal Process.*, vol. 12, no. 1, pp. 132–143, Feb. 2018, doi: 10.1109/JSTSP.2017.2784180.
- [10] H. Ye, G. Y. Li, and B.-H. F. Juang, “Power of Deep Learning for Channel Estimation and Signal Detection in OFDM Systems,” Aug. 28, 2017, *arXiv*: arXiv:1708.08514. doi: 10.48550/arXiv.1708.08514.
- [11] T. Lin and Y. Zhu, “Beamforming Design for Large-Scale Antenna Arrays Using Deep Learning,” *IEEE Wirel. Commun. Lett.*, vol. 9, no. 1, pp. 103–107, Jan. 2020, doi: 10.1109/LWC.2019.2943466.
- [12] A. Alkhateeb, O. E. Ayach, G. Leus, and R. W. H. Jr, “Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems,” *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014, doi: 10.1109/JSTSP.2014.2334278.
- [13] L. Zancato and A. Chiuso, “A novel Deep Neural Network architecture for non-linear system identification,” June 06, 2021, *arXiv*: arXiv:2106.03078. doi: 10.48550/arXiv.2106.03078.
- [14] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” Jan. 30, 2017, *arXiv*: arXiv:1412.6980. doi: 10.48550/arXiv.1412.6980.

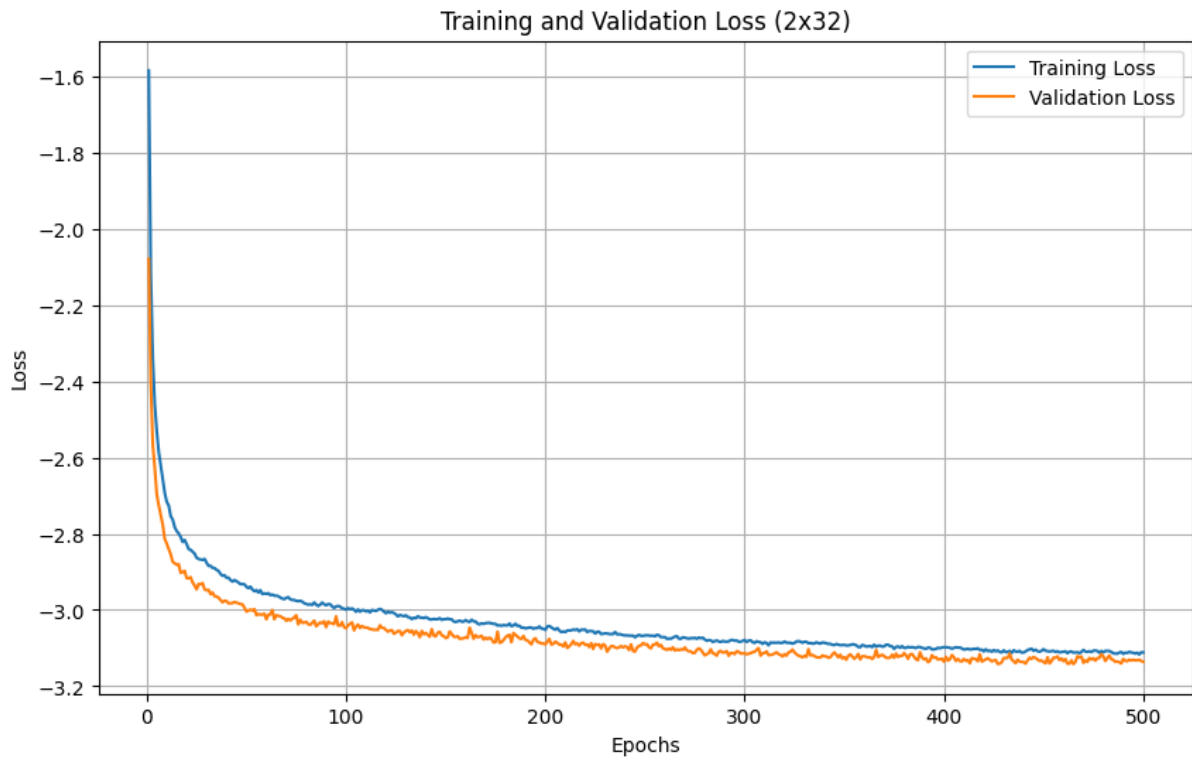
PHỤ LỤC: BIỂU ĐỒ LOSS CỦA CÁC CẤU HÌNH ANTENNA



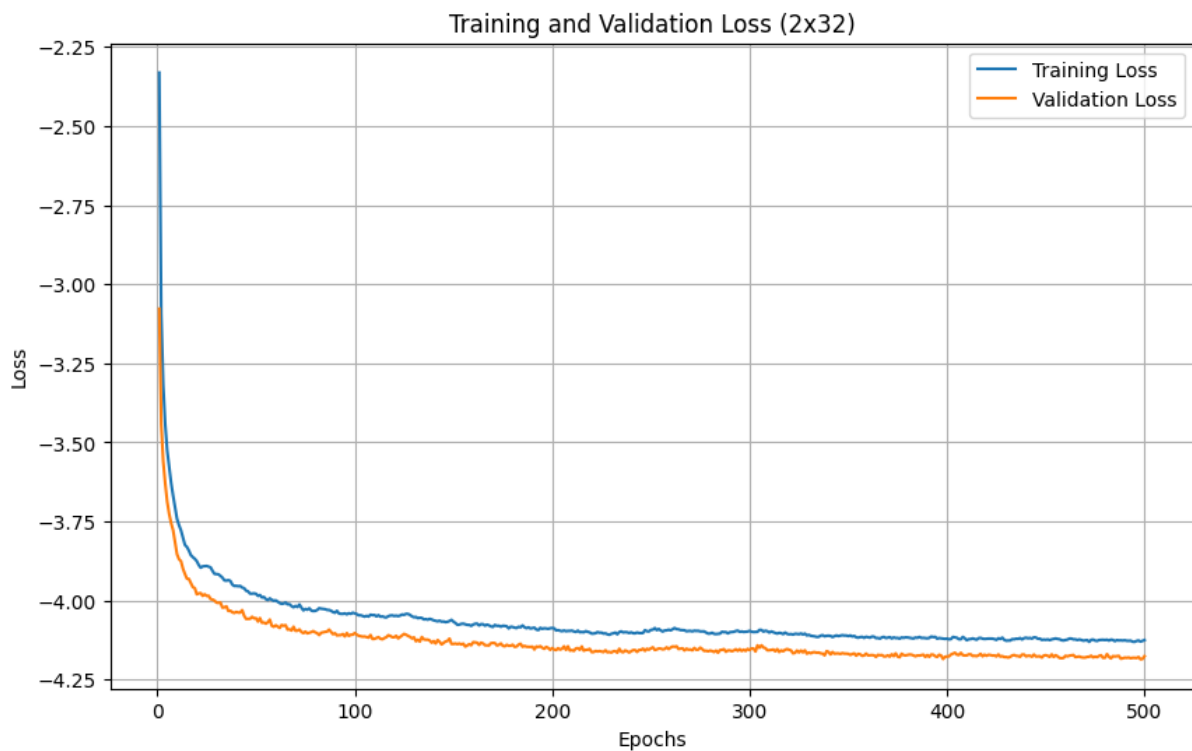
Hình A.1. Biểu đồ Loss của mô hình BFNN tại PNR = 0dB với cấu hình 1×64



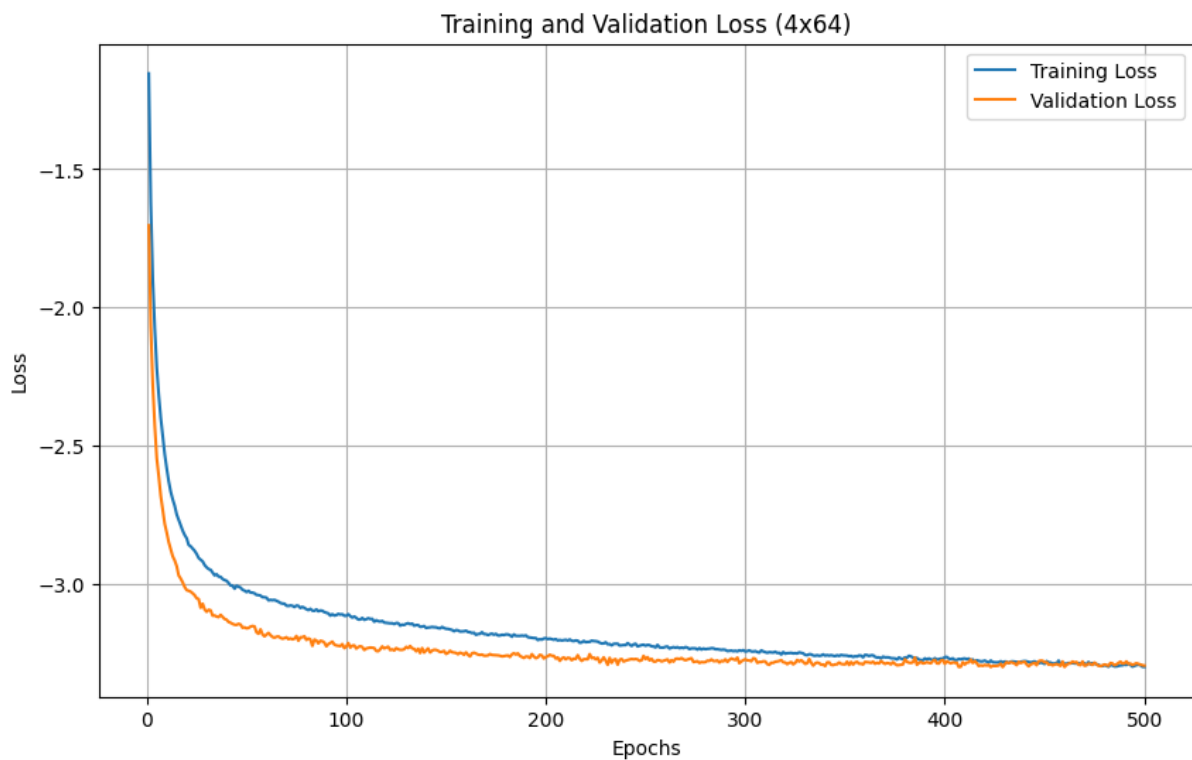
Hình A.2. Biểu đồ Loss của mô hình BFNN tại PNR = 20dB với cấu hình 1×64



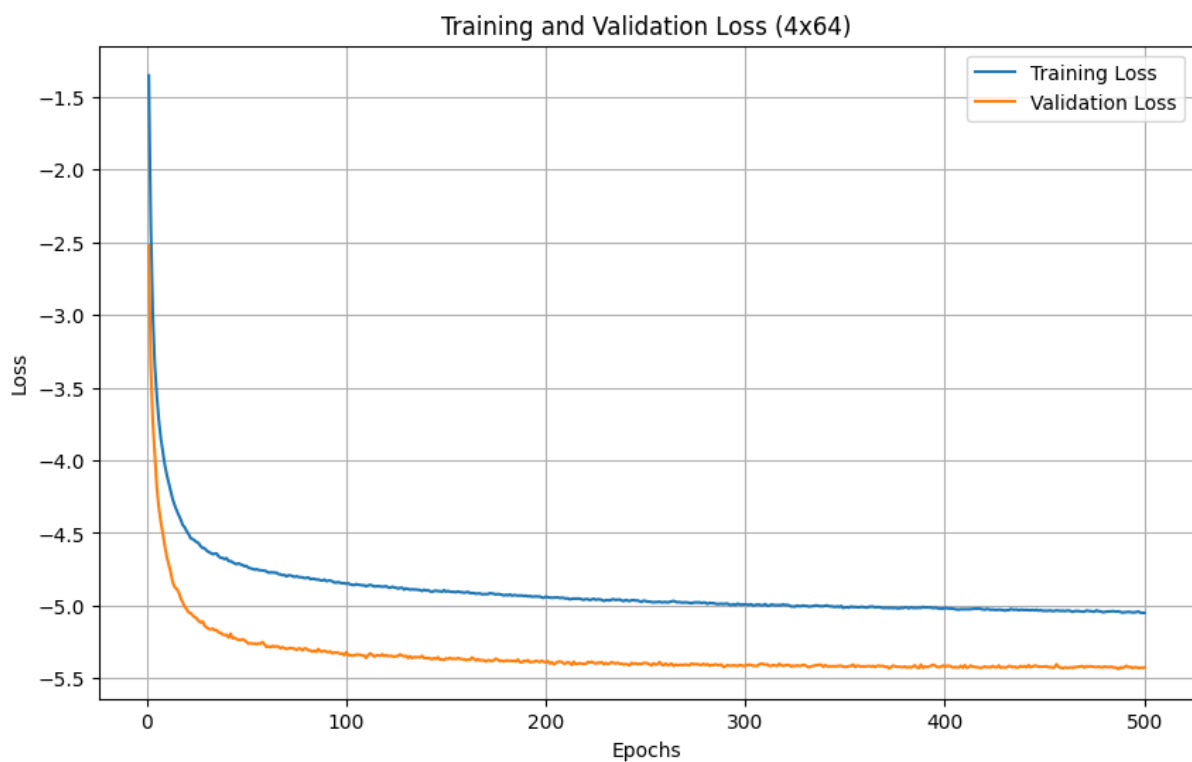
Hình A.3. Biểu đồ Loss của mô hình BFNN tại $\text{PNR} = 0\text{dB}$ với cấu hình 2×32



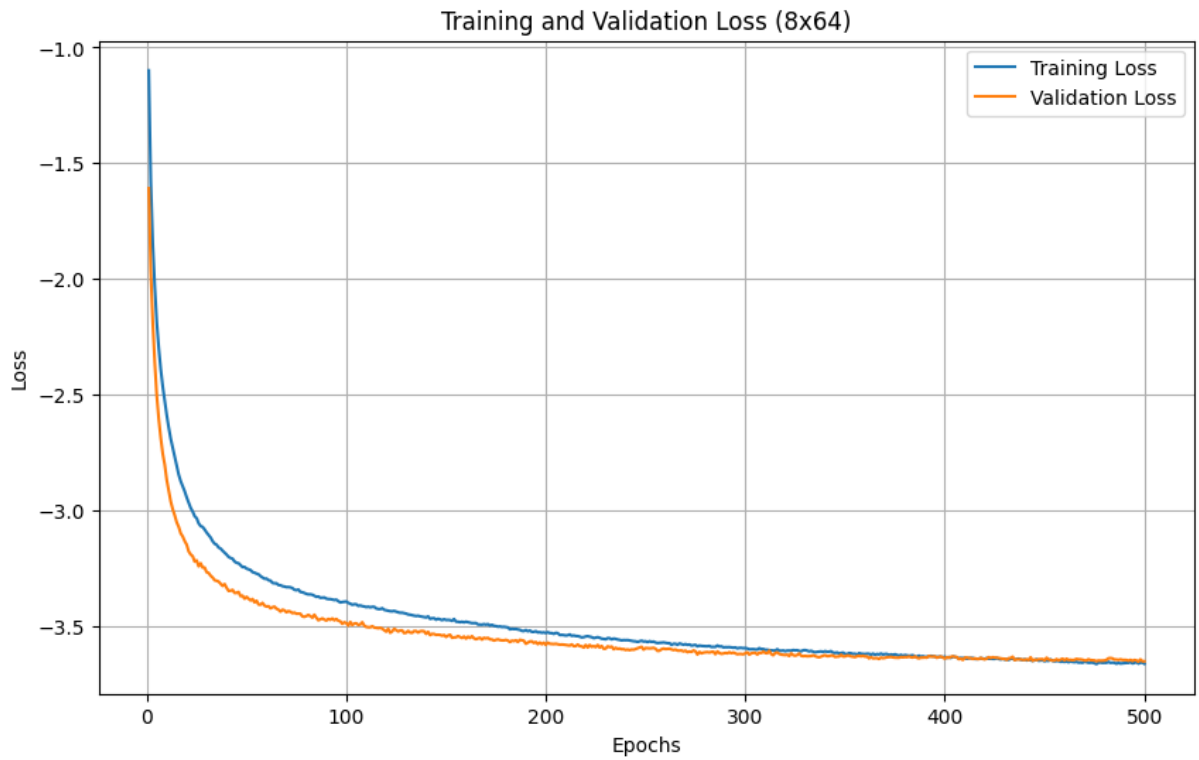
Hình A.4. Biểu đồ Loss của mô hình BFNN tại $\text{PNR} = 20\text{dB}$ với cấu hình 2×32



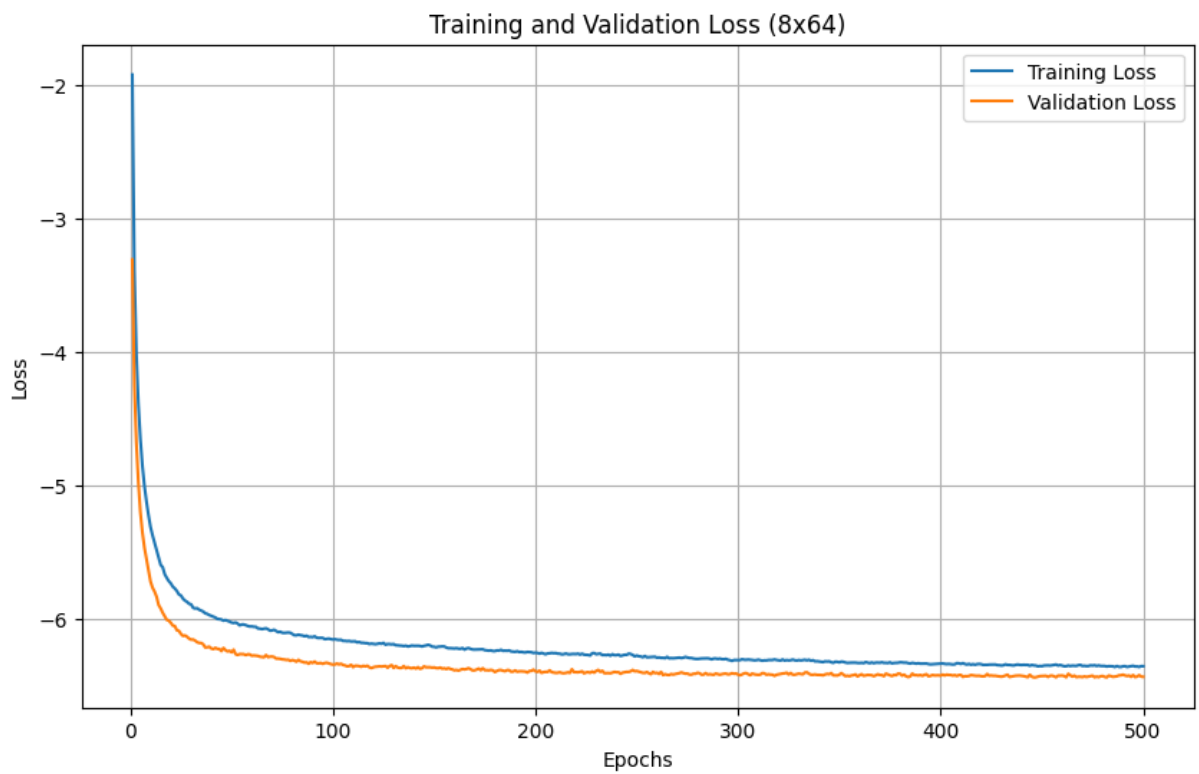
Hình A.5. Biểu đồ Loss của mô hình BFNN tại PNR = 0dB với cấu hình 4×64



Hình A.6. Biểu đồ Loss của mô hình BFNN tại PNR = 20dB với cấu hình 4×64



Hình A.7. Biểu đồ Loss của mô hình BFNN tại PNR = 0dB với cấu hình 8×64



Hình A.8. Biểu đồ Loss của mô hình BFNN tại PNR = 20dB với cấu hình 8×64