

**ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**



ĐỖ HẢI SƠN

**NGHIÊN CỨU NHẬN DẠNG HỆ THỐNG
VỚI TRI THỨC MỚI CHO HỆ THỐNG TRUYỀN THÔNG
MIMO KÍCH THƯỚC LỚN**

**LUẬN VĂN THẠC SĨ
NGÀNH CÔNG NGHỆ KỸ THUẬT ĐIỆN TỬ, TRUYỀN THÔNG**

Hà Nội - 2023

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

ĐỒ HẢI SƠN

NGHIÊN CỨU NHẬN DẠNG HỆ THỐNG
VỚI TRI THỨC MỚI CHO HỆ THỐNG TRUYỀN THÔNG
MIMO KÍCH THƯỚC LỚN

Ngành: Công nghệ kỹ thuật điện tử, truyền thông

Chuyên ngành: Kỹ thuật viễn thông

Mã số: 8510302.02

LUẬN VĂN THẠC SĨ
NGÀNH CÔNG NGHỆ KỸ THUẬT ĐIỆN TỬ, TRUYỀN THÔNG

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. Trần Thị Thúy Quỳnh

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn thạc sĩ **Nghiên cứu nhận dạng hệ thống với tri thức mới cho hệ thống truyền thông MIMO kích thước lớn** là công trình nghiên cứu thực sự của tôi, được thực hiện dựa trên cơ sở lý thuyết, kiến thức chuyên ngành dưới sự hướng dẫn khoa học của TS. Trần Thị Thúy Quỳnh.

Tôi xin cam đoan những công việc trong luận văn thực hiện chưa từng được các tác giả khác đề xuất. Với sự hiểu biết của mình, tôi chắc chắn các số liệu, kết quả trong luận văn là trung thực và chưa được công bố ở đâu và trong bất cứ công trình nào trừ công trình của tác giả và tài liệu tham khảo.

Nếu có gì sai trái, tôi xin hoàn toàn chịu trách nhiệm.

Hà Nội, ngày ... tháng ... năm 2023

Học viên

Đỗ Hải Sơn

LỜI CẢM ƠN

Tôi xin chân thành cảm ơn TS. Trần Thị Thúy Quỳnh, người đã dẫn dắt, định hướng cho nghiên cứu nhận dạng hệ thống với tri thức mới cho hệ thống truyền thông MIMO kích thước lớn. Tôi cũng xin cảm ơn các thầy, cô Khoa Điện tử - Viễn thông, các thành viên của Viện Tiên tiến về Kỹ thuật và Công nghệ, cùng các thầy, cô phản biện đã cho nhiều góp ý quan trọng, giúp tôi khắc phục những điểm còn hạn chế trong luận văn.

Nghiên cứu này được tài trợ bởi Quỹ Phát triển khoa học và công nghệ Quốc gia (NAFOSTED) trong đề tài mã số: 01/2019/TN.

Xin gửi lời cảm ơn đến Giáo sư Karim Abed-Meraim và các thành viên thuộc phòng thí nghiệm PRISME, Đại học Orleans, Cộng hoà Pháp đã tận tình giúp đỡ, hướng dẫn tôi trong quá trình tôi là thực tập sinh tại phòng thí nghiệm PRISME từ tháng 09/2022 đến 12/2022.

Cuối cùng, tôi xin được gửi lời cảm ơn đến gia đình, bạn bè đã luôn bên cạnh động viên, khích lệ, và giúp đỡ tôi trong suốt thời gian qua.

Trong quá trình thực hiện luận văn, do thời gian, kiến thức, và điều kiện nghiên cứu còn một số hạn chế nên không tránh khỏi thiếu sót, rất mong nhận được ý kiến đóng góp từ người đọc để tôi có thể tiếp tục phát triển hướng nghiên cứu này.

Hà Nội, ngày ... tháng ... năm 2023

Học viên

Đỗ Hải Sơn

TÓM TẮT

Tóm tắt: Các thế hệ mạng di động như 5G hay WiFi 802.11ax hiện nay đang phải sử dụng một phần đáng kể băng thông và tài nguyên tính toán cho việc nhận dạng kênh truyền vô tuyến. Luận văn tập trung vào giảm thiểu chi phí và độ phức tạp của việc ước lượng kênh truyền trong các hệ thống MIMO kích thước lớn bằng việc sử dụng các “tri thức mới”, ví dụ phương pháp bán mù sử dụng thêm các thông tin bên lề về cấu trúc mảng ăng-ten hay học máy, học sâu sử dụng thông tin khác với thông tin kênh. Trước hết, đường bao Cramér Rao được sử dụng để xem xét sai số bất tối thiểu của việc nhận dạng kênh truyền khi thay đổi cấu hình của các mảng ăng-ten MIMO kích thước lớn cũng như khi sử dụng một giải thuật bán mù nhằm giảm thiểu chi phí. Kết quả mô phỏng chỉ ra rằng các cấu hình mảng ăng-ten 3D (ví dụ: UCyA) không những giúp tiết kiệm diện tích triển khai mà còn giảm sai số trong quá trình ước lượng kênh truyền. Bên cạnh đó, sử dụng một phần thông tin thống kê từ dữ liệu chưa biết trước theo hướng tiếp cận bán mù cũng có thể làm tăng tính chính xác của việc ước lượng. Sau đó, một mạng nơ-ron sâu có tên ISDNN được đề xuất cho cả hai mô hình kênh truyền có cấu trúc và phi cấu trúc để nhận dạng các hệ thống viễn thông MIMO kích thước lớn. Mạng nơ-ron được đề xuất cho sai số và độ phức tạp đều thấp hơn các giải thuật ước lượng tuyến tính như ZF hay MMSE với độ lợi có thể lên đến 10^3 về sai số bất và giảm $\mathcal{O}(L)$ về độ phức tạp. Ngoài ra, khi so sánh với một mạng nơ-ron sâu khác là DetNet, chỉ với 24 tham số học, độ chính xác của ISDNN cũng vượt trội tiệm cận tỷ lệ sai bất 10^{-4} . Các nghiên cứu trong luận văn chỉ ra tiềm năng của việc ứng dụng các “tri thức mới” trong việc ước lượng kênh truyền, khi có thể giải quyết nhiều bài toán như chi phí, độ phức tạp, và hiệu năng.

Từ khóa: Nhận dạng hệ thống, mMIMO, CRB, tri thức mới, học sâu.

MỤC LỤC

Lời cam đoan	
Lời cảm ơn	
Tóm tắt	
Mục lục	i
Danh mục ký hiệu và chữ viết tắt	iii
Danh mục hình vẽ	viii
Danh mục bảng biểu	x
MỞ ĐẦU	1
CHƯƠNG 1. MÔ HÌNH KÊNH VÀ CÁC PHƯƠNG PHÁP NHẬN DẠNG HỆ THỐNG TRONG MIMO KÍCH THƯỚC LỚN	7
1.1. Mô hình kênh trong hệ thống MIMO kích thước lớn	7
1.1.1. Mô hình kênh truyền xác định	7
1.1.2. Mô hình kênh truyền ngẫu nhiên	10
1.2. Các phương pháp nhận dạng hệ thống MIMO kích thước lớn	15
1.2.1. Nhận dạng kênh không mù	15
1.2.2. Nhận dạng kênh mù	18
1.2.3. Nhận dạng kênh bán mù	19
1.2.4. Nhận dạng kênh sử dụng học máy	19
CHƯƠNG 2. ĐÁNH GIÁ ẢNH HƯỞNG CỦA CẤU HÌNH MẢNG ĂNG-TEN TRONG NHẬN DẠNG HỆ THỐNG MIMO KÍCH THƯỚC LỚN	21
2.1. Mô hình kênh truyền có cấu trúc với các cấu hình mảng ăng-ten khác nhau	21
2.2. Cramér Rao Bound cho giải thuật nhận dạng hệ thống không mù và bán mù	23
2.2.1. CRB trong trường hợp chỉ sử dụng pilot	23
2.2.2. CRB trong trường hợp bán mù	25
2.3. Mô phỏng và đánh giá	26
2.4. Kết luận chương	30
CHƯƠNG 3. NHẬN DẠNG HỆ THỐNG SỬ DỤNG MẠNG HỌC SÂU	31
3.1. Giới thiệu về mạng nơ-ron sâu và mở rộng sâu	31
3.2. Mạng nơ-ron sâu DetNet	32
3.3. Đề xuất mạng nơ-ron sâu ISDNN cho nhận dạng kênh truyền	35

3.3.1. Bộ nhận dạng ISD cho hệ thống mMIMO	35
3.3.2. Đề xuất mạng nơ-ron sâu ISDNN cho mô hình kênh truyền phi cấu trúc	37
3.3.3. Đề xuất mạng nơ-ron sâu ISDNN cho mô hình kênh truyền có cấu trúc	41
3.4. Mô phỏng và đánh giá	42
3.4.1. Tạo bộ dữ liệu	42
3.4.2. Đào tạo và đánh giá kiến trúc mạng nơ-ron sâu đề xuất	45
3.5. Kết luận chương	53
KẾT LUẬN	55
DANH MỤC CÔNG TRÌNH KHOA HỌC CỦA TÁC GIẢ LIÊN QUAN ĐẾN LUẬN VĂN	57
TÀI LIỆU THAM KHẢO	58
PHỤ LỤC	64
A. Thuật toán tối ưu Adam	64

DANH MỤC KÝ HIỆU VÀ CHỮ VIẾT TẮT

Danh mục ký hiệu	
Ký hiệu	Giải thích
in thường	Vô hướng
in thường, đậm	Vector
in hoa, đậm	Ma trận
$(.)^\top$	Chuyển vị
$(.)^{-1}$	Nghịch đảo
$(.)^H$	Chuyển vị liên hợp phức
α	Tham số học của mạng ISDNN
δ	Tốc độ học / độ dài bước
\Im	Phần ảo
λ	Bước sóng
$\mathbb{E}(\cdot)$	Công suất trung bình
\mathbf{e}	Véc-tơ phần dư / lỗi
\mathcal{L}	Hàm mất mát
\mathcal{O}	Độ phức tạp
\mathcal{U}	Phân bố đều
Ψ	Toán tử phi tuyến tính
Ψ_t	Hàm phi tuyến tính phân đoạn
\Re	Phần thực
σ	Độ lệch chuẩn
θ	Góc ngẩng / Bộ tham số của việc học
\mathcal{F}	Ma trận Fourier rời rạc
$\mathbf{0}_T$	Véc-tơ cột gồm các phần tử 0 có kích thước $T \times 1$
arg	Vị trí
b	Bias của bộ biến đổi tuyến tính

\mathcal{CN}	Tạp âm AWGN phức
con	Phép nối
∂	Đạo hàm riêng
dB	Đơn vị decibel
diag	Thành phần đường chéo
\mathbf{G}	Ma trận của bộ nhận dạng tuyến tính
\mathbf{G}_H	Ma trận Gram của \mathbf{H}
\mathbf{H}	Ma trận kênh truyền
\mathbf{I}_K	Ma trận đơn vị kích thước K
K	Số các ký hiệu chứa dữ liệu hoặc pilot trong một ký hiệu OFDM
L	Số ăng-ten thu
M	Số lượng đường truyền giữa một cặp ăng-ten thu phát
min	Giá trị nhỏ nhất
n	Thời điểm
K_p	Số lượng ký hiệu pilot
K_d	Số lượng ký hiệu dữ liệu
ρ	Toán tử tuyến tính
\mathbf{s}	Các ký hiệu được gửi đi từ bộ phát
T	Số ăng-ten phát / người dùng bên phát
$\text{vec}(\mathbf{X})$	Véc-tơ hoá ma trận \mathbf{X}
\mathbf{w}	Tạp âm AWGN
w	Trọng số của bộ biến đổi tuyến tính
\mathbf{x}	Các ký hiệu bên thu nhận được
Θ	Véc-tơ các tham số không biết trước cần được ước lượng
β	Hệ số khuếch đại phức
ϕ	Góc phương vị

Danh mục chữ viết tắt		
Chữ viết tắt	Giải thích tiếng Anh	Giải thích tiếng Việt
AI	Artificial intelligence	Trí tuệ nhân tạo
AWGN	Additive white Gaussian noise	Tạp âm trắng cộng sinh
B	Blind	Kỹ thuật nhận dạng mù
BCCH	Broadcast control channel	Kênh điều khiển phát sóng
BER	Bit error rate	Tỷ lệ lỗi bit
CBSM	Correlation-based stochastic mode	Mô hình ngẫu nhiên dựa trên tương quan
CCCH	Common control channel	Kênh điều khiển chung
CMA	Constant modulus algorithm	Thuật toán mô-đun không đổi
CP	Cyclic prefix	Tiền tố vòng
CRB	Cramér Rao bound	Đường bao Cramér Rao Bound
CSI	Channel state information	Thông tin về trạng thái kênh truyền
CSI-RS	Channel state information - reference signal	Tín hiệu tham chiếu thông tin trạng thái kênh truyền
Data-driven	Data-driven	Hướng dữ liệu
DCCH	Dedicated control channel	Kênh điều khiển chuyên dụng
DDCE	Decision-directed channel estimation	Phương pháp hướng quyết định
DetNet	Detection Network	Mạng nơ-ron học sâu phát hiện
DL	Deep learning	Học sâu
DL-SCH	Downlink shared channel	Kênh chia sẻ đường xuống
DMRS	Demodulation reference signal	Tín hiệu tham chiếu giải điều chế
DNN	Deep-neural network	Mạng nơ-ron sâu
DoA	Direction of arrival	Hướng sóng đến
DoD	Direction of departure	Hướng phát sóng
DTCH	Dedicated traffic channel	Kênh lưu lượng chuyên dụng
FFT	Fast Fourier transform	Biến đổi Fourier nhanh
FIM	Fisher information matrix	Ma trận thông tin Fisher
GBSM	Geometry-based stochastic model	Mô hình ngẫu nhiên dựa trên hình học
GPR	Gaussian process regression	Thuật toán hồi quy Gaussian
HOS	Higer-order statistics	Đặc tính thống kê bậc cao

i.i.d	Independent and identically distributed	Biến độc lập và phân phối đồng nhất
ICA	Independent component analysis	Phân tích thành phần độc lập
ISD	Iterative sequential detection	Mạng học sâu phát hiện tuần tự lặp lại
ISDNN	Iterative sequential Deep-neural network	Mạng nơ-ron sâu tuần tự lặp lại
LMMSE	Linear Minimum mean square error	Kỹ thuật ước lượng lỗi bình phương trung bình nhỏ nhất tuyến tính
LMS	Least mean squares	Kỹ thuật trung bình bình phương nhỏ nhất
LS	Least square	Kỹ thuật bình phương nhỏ nhất
LSTM	Long/short-term memory	Mạng trí nhớ dài hạn/ngắn hạn
MIMO	Multi-input multi-output	Hệ thống đa đầu vào đa đầu ra
ML	Machine learning	Học máy
MLE	Maximum likelihood estimator	Bộ ước lượng hợp lệ cực đại
mMIMO	Massive Multi-input multi-output	Hệ thống đa đầu vào đa đầu ra kích thước lớn
MMSE	Minimum mean square error	Kỹ thuật lỗi trung bình phương sai tối thiểu
mmWave	Millimeter wave	Bước sóng mi-li-mét
Model-driven	Model-driven	Hướng mô hình
MRE	Mutually referenced equalizers	Bộ cân bằng kênh tham chiếu
NB	Non-blind	Kỹ thuật nhận dạng không mù
NN	Neural network	Mạng nơ-ron
OFDM	Orthogonal frequency-division multiplexing	Ghép kênh phân chia theo tần số trực giao
OP	Only pilot	Chỉ sử dụng các ký hiệu hoa tiêu
PC	Pilot contamination	Ô nhiễm pilot
PCA	Principal components analysis	Phân tích thành phần chính
PCCH	Paging control channel	Kênh điều khiển paging
pdf	Probability density function	Hàm mật độ xác suất
Pilot	Pilot	Ký hiệu hoa tiêu
Pilot-assisted	Pilot-assisted	Phương pháp dựa trên tín hiệu hoa tiêu

PSK	Phase-shift keying	Điều chế pha
PSS	Primary synchronization signal	Tín hiệu đồng bộ sơ cấp
PTRS	Phase tracking reference signal	Tín hiệu tham chiếu theo dõi pha
QAM	Quadrature amplitude modulation	Điều chế biên độ cầu phương
RACH	Random access channel	Kênh truy cập ngẫu nhiên
RL	Reinforcement learning	Học tăng cường
RLS	Recursive least squares	Kỹ thuật bình phương đệ quy nhỏ nhất
SB	Semi-blind	Kỹ thuật nhận dạng bán mù
SIMO	Single-input multi-output	Hệ thống đơn đầu vào đa đầu ra
SISO	Single-input single-output	Hệ thống đơn đầu đơn vào đầu ra
SNR	Signal noise ratio	Tỷ lệ tín hiệu tạp âm
SOS	Second-order statistics	Đặc tính thống kê bậc hai
SRS	Sounding reference signal	Tín hiệu tham chiếu âm thanh
SSS	Secondary synchronization signal	Tín hiệu đồng bộ thứ cấp
SC	Sub-cluster	Cụm con trong các TC
TC	Time cluster	Cụm phân chia theo thời gian đến
Training-based	Training-based	Phương pháp dựa trên việc đào tạo
RT	Ray-tracing	Phân tích theo tia
UCA	Uniform circle array	Mảng tròn cách đều
UCyA	Uniform cylindrical array	Mảng trụ đồng nhất
ULA	Uniform linear array	Mảng thẳng cách đều
UL-SCH	Uplink shared channel	Kênh chia sẻ đường lên
ZF	Zero forcing	Kỹ thuật ép không

DANH MỤC HÌNH VẼ

Hình 1	Các tín hiệu tham chiếu (bo bằng đường nét đứt màu xanh) tại kênh vật lí khác nhau trong mạng 5G.	3
Hình 1.1	Phân loại mô hình kênh truyền xác định và ngẫu nhiên trong MIMO kích thước lớn.	7
Hình 1.2	Mô hình minh hoạ hệ thống truyền thông mMIMO.	8
Hình 1.3	Các bước tạo mô hình kênh truyền trong tiêu chuẩn 3GPP phiên bản 16 [1].	12
Hình 1.4	Phân loại các phương pháp ước lượng kênh truyền viễn thông.	16
Hình 2.1	Minh hoạ hai cấu hình 1D và 3D của các mảng ăng-ten.	22
Hình 2.2	CRB của hai cấu hình ULA và UCyA ứng với mô hình kênh truyền có cấu trúc (structured) và phi cấu trúc (unstructured). Cấu hình của mảng ăng-ten như sau $N_{ULA} = 96$, $N_{UCA} = 24$, và $N_{3D} = 4$	27
Hình 2.3	CRB của hai cấu hình ULA và UCyA khi thay đổi N_{3D} . Các thông số mô phỏng như sau $N_{UCA} = 24$, $N_{ULA} = 24 * N_{3D}$, và $SNR = 5$ dB.	28
Hình 2.4	CRB của hai cấu hình ULA và UCyA khi thay đổi N_{UCA} . Các thông số mô phỏng như sau $N_{3D} = 4$, $N_{ULA} = 4 * N_{UCA}$, và $SNR = 5$ dB.	29
Hình 3.1	Minh hoạ sử dụng DNN để nhận dạng hệ thống viễn thông.	31
Hình 3.2	Kiến trúc của một lớp trong kiến trúc mạng DetNet [50].	35
Hình 3.3	Hàm phi tuyến tính phân đoạn $\psi_{t_k}(s)$ được sử dụng trong DetNet.	36
Hình 3.4	Kiến trúc của một lớp trong mạng nơ-ron sâu ISDNN đề xuất cho mô hình kênh truyền phi cấu trúc.	40
Hình 3.5	Minh hoạ một số hàm kích hoạt được dùng trong kiến trúc đề xuất.	41
Hình 3.6	Kiến trúc của một lớp trong mạng nơ-ron sâu ISDNN đề xuất cho mô hình kênh truyền có cấu trúc. Giả sử biết thông tin của DoA và cấu hình mảng ăng-ten tại bên thu $\varphi(\theta, \phi)$	43
Hình 3.7	Ánh xạ các nhóm 4 bit thành các ký hiệu sử dụng điều chế 16-QAM.	45
Hình 3.8	Phân bố khi gieo ngẫu nhiên của các hệ số phần thực của $h_{l,t}$ và $\hat{\beta}_{l,t}$ trong hai ma trận \mathbf{H} , $\hat{\beta}$	46
Hình 3.9	Sự hội tụ của quá trình đào tạo mạng ISDNN và DetNet.	48

Hình 3.10 Độ chính xác của mạng ISDNN so sánh với DetNet và các bộ nhận dạng tuyến tính.	49
Hình 3.11 Sự hội tụ của quá trình đào tạo các mạng ISDNN với các sai số kênh truyền đầu vào khác nhau.	50
Hình 3.12 Độ chính xác của mạng ISDNN với các sai số kênh truyền đầu vào khác nhau.	51
Hình 3.13 Sự hội tụ của quá trình đào tạo các mạng ISDNN có cấu trúc với các sai số kênh truyền và thông tin bên lề đầu vào khác nhau.	52
Hình 3.14 Độ chính xác của mạng ISDNN có cấu trúc với các sai số kênh truyền và thông tin bên lề đầu vào khác nhau.	53

DANH MỤC BẢNG BIỂU

Bảng 1.1	Các mô đặc trưng kênh truyền mMIMO dựa trên đo đạc [7].	9
Bảng 1.2	Số lượng các TC và sub-path trong mỗi TC theo chuẩn 3GPP phiên bản 16 [1].	13
Bảng 1.3	Độ lệch pha của các sub-path trong một TC.	14
Bảng 1.4	Thông tin về công suất và độ trễ của các sub-path trong mỗi TC. . .	14
Bảng 2.1	Các tham số mô phỏng hệ thống truyền thông không dây để ước lượng CRB.	27
Bảng 3.1	Các tham số mô phỏng hệ thống truyền thông không dây của mạng nơ-ron sâu ISDNN được đề xuất.	44
Bảng 3.2	Ánh xạ các nhóm 4 bit thành các ký hiệu sử dụng điều chế 16-QAM. . .	44
Bảng 3.3	So sánh độ phức tạp của các thuật toán nhận dạng kênh truyền. . . .	47

MỞ ĐẦU

Lý do chọn đề tài

Theo [4], trong năm 2021, chỉ riêng các thiết bị di động đã đòi hỏi tổng lưu lượng truy cập internet đạt 48,27 Petabytes/tháng và tăng trưởng 46% hàng năm. Tốc độ truy cập của các thế hệ mạng di động phổ biến ở Việt Nam như 4G là trung bình trên 28,2 Megabits/giây và tăng lên từ 29 đến 47% hàng năm. Với sự gia tăng nhanh chóng của số lượng các thiết bị di động và đòi hỏi về chất lượng nội dung của người dùng, yêu cầu về truyền tải nhanh và hiệu quả trong các hệ thống truyền thông không dây luôn là chủ đề nghiên cứu được quan tâm.

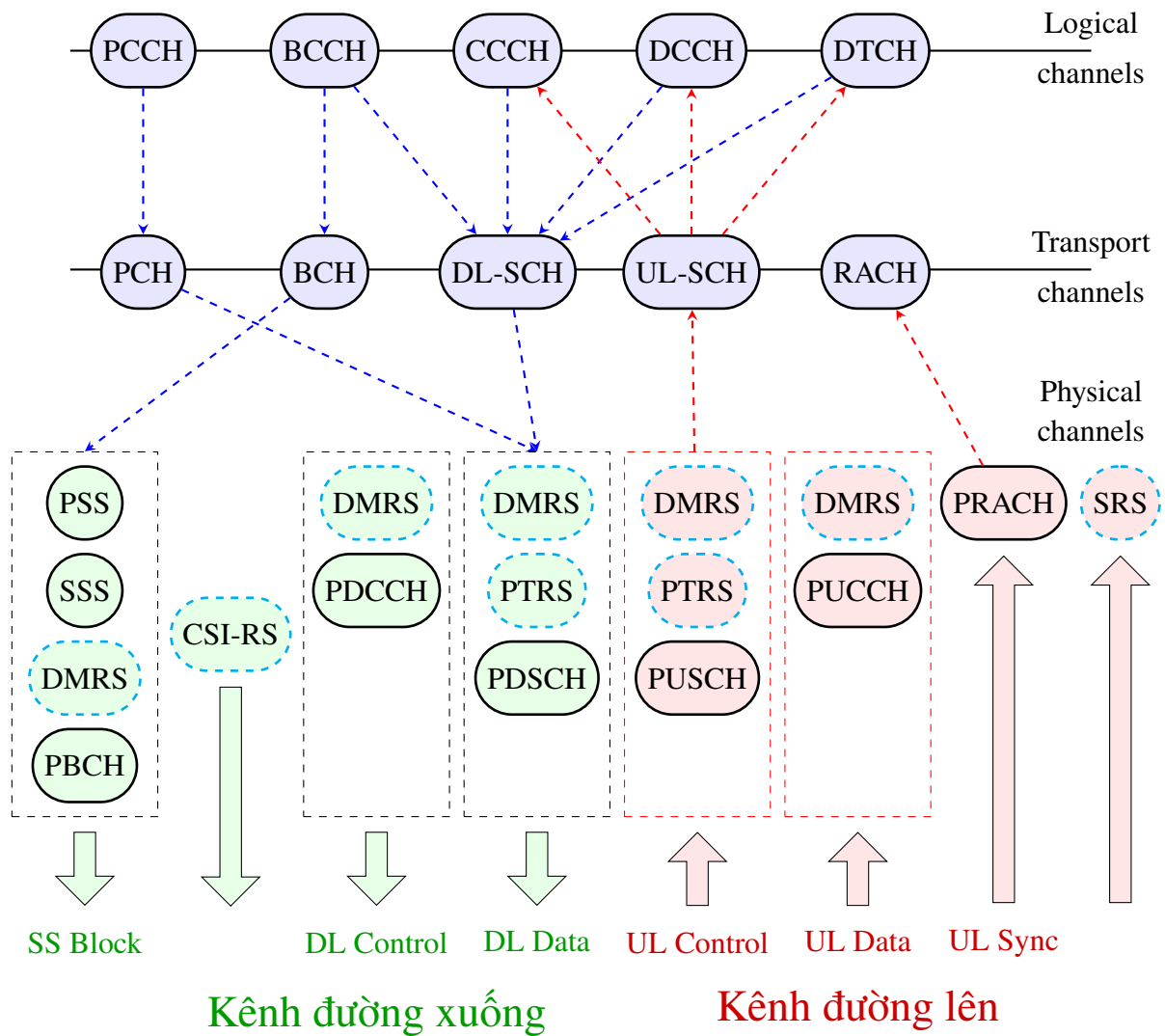
Trong thế hệ mạng di động tiên tiến như 5G và cao hơn, để tăng hiệu quả truyền tải, các hệ truyền thông đơn ăng-ten (antenna) hoặc đa đầu vào đa đầu ra (MIMO - Multi-input multi-output) kích thước nhỏ (số lượng ăng-ten của trạm cơ sở nhỏ hơn 10) đã được thay thế bằng các hệ thống **MIMO kích thước lớn** (mMIMO - Massive MIMO) với số lượng ăng-ten trạm cơ sở từ 100 đến hàng nghìn phần tử và phục vụ cùng lúc hàng chục đến hàng trăm người dùng [25]. Kênh truyền của các hệ thống mMIMO có hai tính chất quan trọng bao gồm: (i) Làm cứng kênh (hardening) [64], khi số lượng phần tử ăng-ten trong mMIMO tăng lên rất lớn, các kênh truyền vô tuyến có thể trở thành dạng large-scale fading và thay đổi chậm hơn; (ii) Truyền thuận lợi (favorable propagation) [39], trong hệ thống mMIMO, tín hiệu truyền tải trên kênh đường lên từ nhiều người dùng khác nhau là trực giao với nhau (không gây nhiễu cho nhau).

Để đạt được hiệu quả truyền tải, bất kỳ thế hệ mạng di động nào như 5G hay các thế hệ mạng WiFi như 802.11ac, 802.11ax đều cần thực hiện ước lượng kênh truyền vô tuyến nhằm khôi phục lại tín hiệu gốc được gửi đi. Trong luận văn này, “**Nhận dạng hệ thống truyền thông**” được hiểu là “ước lượng kênh truyền”. Cụ thể, mạng 5G theo chuẩn 3GPP TS 38.211 phiên bản 16 [1], có đến bốn loại tín hiệu tham chiếu được sử dụng cho việc đồng bộ và ước lượng kênh truyền, bao gồm: tín hiệu tham chiếu giải điều chế (DMRS - Demodulation reference signal), tín hiệu tham chiếu theo dõi pha (PTRS - Phase tracking reference signal), tín hiệu tham chiếu phát thanh (SRS - Sounding reference signal), và tín hiệu tham chiếu thông tin trạng thái kênh truyền (CSI-RS - Channel state information - reference signal). Trên hình 1 mô tả liên kết giữa các tín hiệu tham chiếu ở kênh vật lý với các lớp phía trên của mạng lõi 5G. Có thể nhận thấy, cả kênh đường lên và đường

xuống của lớp dữ liệu và điều khiển đều có thể có sự xuất hiện của các tín hiệu tham chiếu này. Ví dụ, tín hiệu tham chiếu DMRS có thể chiếm đến 50% số sóng mang con (sub-carrier) trong một slot và từ 1 đến 2 slots trong một khung con (sub-frame) của 5G. Tùy thuộc vào điều kiện kênh truyền, trạm cơ sở sẽ tăng/giảm số lượng các tín hiệu tham chiếu này, có thể có các trường hợp lý tưởng khi DMRS chỉ cần ở kênh đường lên, nhờ tính đối xứng và cứng của kênh truyền mMIMO. Tuy nhiên, vẫn có thể nhận xét rằng, trong 5G, khi số lượng tín hiệu tham chiếu lớn, thì chi phí trong việc truyền tải và độ phức tạp trong việc ước lượng kênh truyền đều tăng lên. Điều tương tự với các thế hệ mạng WiFi, trong chuẩn WiFi phổ biến hiện nay như 802.11ac [2], các tín hiệu tham chiếu gọi là hoa tiêu (pilot) được sử dụng. Để tăng tốc độ truyền tải, 802.11ac hoạt động ở nhiều chế độ băng thông khác nhau, từ nhỏ nhất là 20MHz yêu cầu 4 pilots trên tổng 64 sub-carriers, đến lớn nhất là 160MHz yêu cầu 16 pilots trên tổng 512 sub-carriers. Trong luận văn này, thuật ngữ pilot sẽ được sử dụng để chỉ các tín hiệu tham chiếu như trong 5G hay pilot trong WiFi.

Nhiều phương pháp nghiên cứu đã được đề xuất để giảm thiểu số lượng pilot (chi phí) và độ phức tạp tính toán cần thiết cho việc ước lượng kênh truyền. Mà tiêu biểu trong số đó là ba hướng tiếp cận:

1. Các thuật toán nhận dạng “mù” (B - Blind) ở đây được hiểu là khi xử lý (nhận dạng), bộ xử lý “không nhìn thấy” (không có thông tin) của đầu vào. Các thuật toán xử lý tín hiệu mù phát triển mạnh trong thập kỷ 90 [3], tuy nhiên các phương pháp xử lý mù thường yêu cầu các thông số thống kê của tín hiệu mà thông thường không biết trước trong các hệ thống truyền thông thực, hơn nữa, độ chính xác mà các thuật toán này đưa ra cũng thấp hơn đáng kể khi so sánh với các phương pháp sử dụng pilot truyền thống. Do vậy, các thuật toán mù cũng ít được quan tâm trong những thế hệ mạng viễn thông di động trước 5G.
2. Các thuật toán nhận dạng “bán mù” (SB - Semi-blind) là phương pháp cải tiến của B và được quan tâm trong các năm gần đây [23, 24, 52]. Đây là kỹ thuật kết hợp các thông tin từ hướng tiếp cận mù truyền thống và các dạng thông tin khác, ví dụ [46]: số lượng nhỏ pilot, hướng sóng đến (DoA - Direction of arrival), tọa độ người dùng, ... Điều này giúp giảm đi số lượng pilot cần thiết cho việc nhận dạng hệ thống nhằm tăng hiệu quả phổ nhưng vẫn giữ được độ chính xác cần thiết, giảm độ phức tạp, và cho khả năng ứng dụng rộng rãi hơn.
3. Các thuật toán nhận dạng sử dụng học máy, học sâu (ML - Machine learning; DL - Deep learning) cũng là lĩnh vực nghiên cứu dành được nhiều sự quan tâm [67]



Hình 1. Các tín hiệu tham chiếu (bo bằng đường nét đứt màu xanh) tại kênh vật lý khác nhau trong mạng 5G.

trong các năm gần đây. Ưu điểm của các kỹ thuật sử dụng ML, DL là tính đa dạng, khi hướng tiếp cận ML, DL sử dụng cho mục đích xử lý các loại tín hiệu như hình ảnh, âm thanh đã đạt được các bước tiến rõ rệt. Đầu vào của các mạng DL được sử dụng để nhận dạng hệ thống rất linh hoạt, có thể tương ứng với cả ba hướng tiếp cận: pilot, mù, và bán mù kể trên. Sau quá trình huấn luyện, các mô hình (model) học máy có thể hoạt động độc lập như một bộ cân bằng mù/bán mù, khi chỉ cần đưa các tín hiệu thu được đi qua model, và các tín hiệu được khôi phục sẽ được trả về mà không cần đến các chuỗi pilot, hay thông tin về trạng thái kênh truyền (CSI - Channel state information).

Trong luận văn này, các phương pháp thuộc hai hướng tiếp cận SB khi sử dụng các thông tin bên lề về cấu trúc ăng-ten và sử dụng DL để tách sóng được gọi là “**tri thức mới**” [40] do sử dụng hoặc thông tin khác với thông tin kênh. Vì vậy, SB và DL là các

phương pháp tiên tiến và hứa hẹn trong việc ước lượng kênh truyền trong các hệ thống truyền thông hiện đại.

Ngoài việc thay đổi các thuật toán ước lượng kênh truyền, khi số lượng ăng-ten của mMIMO rất lớn thì cấu hình (kiến trúc) của các mảng ăng-ten này cũng cần được xem xét. Trong các thể hệ mạng di động cũ, các mảng ăng-ten thường ở dạng mảng thẳng cách đều (ULA - Uniform linear array). Tuy nhiên, nếu số lượng phần tử ăng-ten lên đến hàng trăm, trước hết dễ nhận thấy rằng kích thước của các mảng ULA này sẽ trở nên quá lớn. Ngoài ra, việc thay đổi kiến trúc mảng hoàn toàn có thể ảnh hưởng đến hiệu suất của các phép ước lượng, như trong [45] đã chỉ ra, các kiến trúc mảng 3D cho độ chính xác vượt trội khi ước lượng DoA khi so sánh với ULA truyền thống. Mà DoA lại có thể là một thông số hữu ích bổ sung cho các phương pháp ước lượng SB. Do vậy, các kiến trúc hình học 3D nên được xem xét trong mMIMO để giảm thiểu kích thước và tăng cường độ chính xác của các hệ thống thu phát. Bên cạnh thay đổi cấu trúc vật lý của mảng ăng-ten, các mô hình kênh truyền cũng có thể được xem xét nhằm khai thác thêm các thông tin bên lề nhằm giảm thiểu sai số cho việc ước lượng.

Từ thực tế và những phân tích nêu trên, luận văn tập trung nghiên cứu hai mục tiêu bao gồm

1. Xem xét sự ảnh hưởng của các kiến trúc mảng ăng-ten, mô hình kênh truyền đến hiệu suất của các thuật toán ước lượng NB và SB trong các hệ thống mMIMO.
2. Phát triển một phương pháp nhận dạng sử dụng DL cho các hệ thống mMIMO.

Phương pháp nghiên cứu

Trong luận văn, để đạt được mục tiêu nghiên cứu học viên đã tìm hiểu các tài liệu, bài báo, tạp chí quốc tế, ... có uy tín, thực hiện việc tính toán mô hình dữ liệu, phân tích số học để đưa ra các hướng giải quyết hợp lý, và sau đó kiểm nghiệm lại kết quả bằng hình thức mô phỏng trên Matlab, Python.

Cụ thể để phát triển các thuật toán mới như mục tiêu đề ra các bước nghiên cứu sau đã được thực hiện trong luận văn:

1. Tìm hiểu về các mô hình kênh truyền và thuật toán nhận dạng hệ thống trong truyền thông không dây với các hướng tiếp cận khác nhau.
2. Xác định bài toán cụ thể và mục tiêu của nghiên cứu.

3. Xây dựng mô hình toán học của kênh truyền vô tuyến và lựa chọn phương pháp đánh giá hiệu suất của các kiến trúc mảng ăng-ten/giải thuật ước lượng kênh truyền NB, SB.
4. Lựa chọn một phương pháp ước lượng kênh truyền có thể phát triển thành mạng học sâu.
5. Xây dựng mô hình toán học, huấn luyện mô hình, và tối ưu hóa các thuật toán đề xuất.
6. Đánh giá và phân tích kết quả ở nhiều kịch bản khác nhau, so sánh với các nghiên cứu đi trước.

Nội dung nghiên cứu

- Tìm hiểu về các mô hình kênh truyền mMIMO, các thuật toán ước lượng kênh truyền trong những năm gần đây.
- Xây dựng mô hình toán học của kênh truyền có cấu trúc (structured) cho việc truyền thông, từ đó khảo sát ảnh hưởng của các kiến trúc mảng ăng-ten, thuật toán ước lượng đến hiệu năng chung của việc ước lượng kênh truyền thông qua đường bao Cramér Rao (CRB).
- Phát triển một mạng nơ-ron sâu dựa trên giải thuật phát hiện tuần tự lặp lại (ISD - Iterative sequential detection) để ước lượng kênh truyền viễn thông.
- Đánh giá, kiểm tra các thuật toán được đề xuất trên các công cụ mô phỏng Matlab, Python.

Đóng góp của đề tài

Luận văn có hai đóng góp chính sau đây:

- Khảo sát sự ảnh hưởng của kiến trúc mảng ăng-ten 3D và mô hình kênh truyền có cấu trúc đến tính chính xác của việc ước lượng kênh truyền dựa trên CRB trong các hệ thống mMIMO.
- Đề xuất một mạng học sâu ISDNN cho việc ước lượng kênh truyền của hệ thống mMIMO cho cả hai mô hình kênh có cấu trúc và phi cấu trúc.

Kết quả nghiên cứu của chương 2 đã được công bố:

Do Hai Son and Tran Thi Thuy Quynh (2023), “Impact Analysis of Antenna Array Geometry on Performance of Semi-blind Structured Channel Estimation for massive MIMO-OFDM systems,” in *IEEE Statistical Signal Processing Workshop (SSP)*, Hanoi, Vietnam, July. [accepted]

Bố cục của luận văn

Nội dung tiếp theo của luận văn được trình bày như sau:

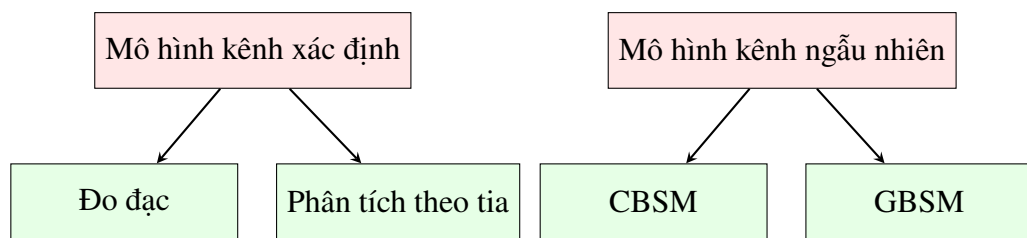
- Chương 1: Trình bày các mô hình kênh truyền trong hệ thống viễn thông mMIMO và tổng quan về các phương pháp nhận dạng hệ thống trong truyền thông không dây.
- Chương 2: Trình bày mô hình kênh truyền có cấu trúc cho các hệ thống mMIMO với các cấu hình mảng ăng-ten khác nhau. Sử dụng đường bao Cramér Rao để xem xét ảnh hưởng của kiến trúc mảng ăng-ten và giải thuật SB đến hiệu suất của việc ước lượng kênh truyền.
- Chương 3: Trình bày sơ lược về mạng DetNet (Detection network) phục vụ cho ước lượng kênh truyền viễn thông. Đề xuất một mô hình mạng ISDNN, cho cả mô hình kênh truyền có cấu trúc và phi cấu trúc, nhằm giảm độ phức tạp cho mô hình đầu ra. Tạo bộ dữ liệu, đào tạo, và đánh giá kết quả thông qua mô phỏng Python.
- Kết luận: Đưa ra kết luận về hai thuật toán được đề xuất và đề xuất các giải pháp để cải thiện hệ thống.

CHƯƠNG 1

MÔ HÌNH KÊNH VÀ CÁC PHƯƠNG PHÁP NHẬN DẠNG HỆ THỐNG TRONG MIMO KÍCH THƯỚC LỚN

Việc nhận dạng hệ thống trong truyền thông không dây đã luôn được phát triển ngay từ những thế hệ mạng không dây như 2G [59]. Bước đầu tiên trong bài toán nhận dạng là đo đạc và mô hình hoá kênh truyền vô tuyến dưới các dạng biểu diễn toán học khác nhau. Trong phần đầu tiên của chương này, tác giả sẽ trình bày khái quát về một số phương pháp mô hình kênh truyền xác định và ngẫu nhiên khác nhau để chọn ra phương pháp mô hình hoá kênh truyền phù hợp sử dụng trong luận văn. Tiếp đó, các phương pháp nhận dạng hệ thống MIMO kích thước lớn được phân loại thành 4 hướng tiếp cận và khảo sát một cách khái quát, để chỉ ra các phương pháp “tri thức mới” mà luận văn quan tâm.

1.1. Mô hình kênh trong hệ thống MIMO kích thước lớn

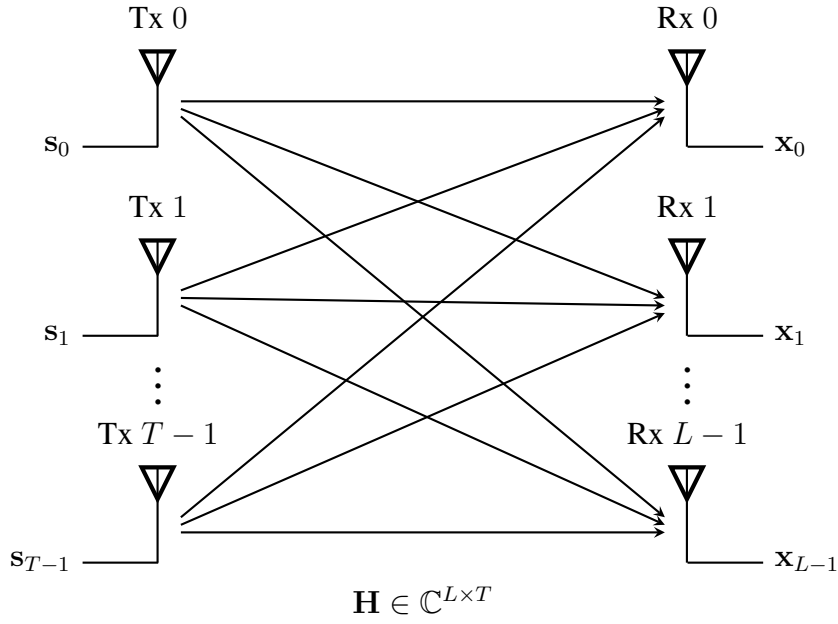


Hình 1.1. Phân loại mô hình kênh truyền xác định và ngẫu nhiên trong MIMO kích thước lớn.

Trong các hệ thống mMIMO, công bố [7] đã khảo sát các mô hình kênh truyền tiên tiến và phân loại chúng theo các hướng tiếp cận khác nhau. Từ đó, trong luận văn này, các mô hình kênh truyền trong hệ mMIMO được xem xét chia thành hai kiểu chính là mô hình kênh truyền xác định và mô hình kênh truyền ngẫu nhiên. Các mô hình kênh truyền xác định có thể chia thành hai loại nhỏ hơn là phương pháp đo đạc và kỹ thuật phân tích theo tia. Với mô hình kênh truyền ngẫu nhiên, hai kiểu chính đó là mô hình ngẫu nhiên dựa trên tương quan (CBSM - Correlation-based stochastic model) và mô hình ngẫu nhiên dựa trên hình học (GBSM - Geometry-based stochastic model).

1.1.1. Mô hình kênh truyền xác định

Để đi vào chi tiết hơn vào các mô hình kênh kể trên, trước hết, xem xét một mô hình hệ thống thu phát mMIMO đơn giản với T ăng-ten phát và L ăng-ten thu như trên



Hình 1.2. Mô hình minh hoạ hệ thống truyền thông mMIMO.

hình 1.2. Có thể minh hoạ hệ thống thu phát dưới dạng toán học, tại thời điểm n

$$\mathbf{x}(n) = \mathbf{H}(n)\mathbf{s}(n) + \mathbf{w} \quad (1.1)$$

trong đó $\mathbf{s}(n) \in \mathbb{C}^{T \times 1}$ là các ký hiệu được gửi đi từ T bộ phát. Kênh truyền được biểu diễn là ma trận $\mathbf{H} \in \mathbb{C}^{L \times T}$. Tiếp đến, $\mathbf{x}(n) \in \mathbb{C}^{L \times 1}$ là véc-tơ biểu diễn tín hiệu thu được từ L ăng-ten. Cuối cùng, $\mathbf{w} \in \mathbb{C}^{L \times 1}$ đại diện cho tạp âm trắng cộng sinh (AWGN - Additive white Gaussian noise).

1.1.1.1. Mô hình kênh truyền xác định dựa trên máy đo

Mô hình kênh truyền dựa trên đo đạc là phương pháp mô hình hoá kênh truyền tiệm cận với kênh truyền vô tuyến thực. Phương pháp mô hình kênh này mang tính thực nghiệm, trong các điều kiện cụ thể, như môi trường trong nhà [28, 64], khu ngoại ô [15], sân vận động [32], Việc đo đạc của các công bố kể trên thường dựa trên các thiết lập cụ thể, sử dụng các thiết bị chuyên dụng như máy phân tích mạng và các bộ phát sóng tiêu chuẩn. Trên bảng 1.1 là danh sách của 5 nghiên cứu kênh truyền thực tế dựa trên đo đạc. Có thể nhận thấy rằng, các mô hình này có những thông số rất riêng biệt như tần số hoạt động, băng thông, hay trong một môi trường cố định biết trước. Ngoài ra, trong mỗi công bố, các thông số đo được từ máy phân tích chuẩn chỉ phản ánh một số tính chất (DoA, DoD, số cụm, phân cực chéo, ...) của kênh truyền vô tuyến. Từ các thông số này, sử dụng giải thuật SEGA [8] để khôi phục lại các hệ số của kênh truyền một cách chính xác.

Giải thuật SAGE, viết tắt của Space Alternating Generalized Expectation Maxi-

Bảng 1.1. Các mô đặc trưng kênh truyền mMIMO dựa trên đo đạc [7].

Công bố	Tần số	Băng thông	Môi trường truyền	Cấu hình mảng ăng-ten ($T \times L$)	Thông số đo đạc
[28]	11 GHz	200 MHz	Rạp chiếu phim	1×256	Phân cực chéo, Góc phát (DoD)
[64]	2,6 GHz	40 MHz	Trong nhà, Ngoài trời	1×128	Hệ số khuếch đại
[15]	3,5 GHz	160 MHz	Khu ngoại ô	32×32	PDP - thống kê công suất, theo độ trễ, Hệ số K , Trễ trễ
[62]	3,5 GHz	200 MHz	Khu dân cư kích cỡ macro	16×256	DoA, DoD, Số lượng các cụm (TC), Vùng hiển thị (VR)
[32]	1,4725 GHz 4,45 GHz	91 MHz, 100 MHz	Sân vận động	1×128	Hệ số kênh truyền, Hệ số K của Ricean, Trễ trễ

mization, là một giải thuật lặp được sử dụng phổ biến để ước lượng kênh trong hệ thống truyền thông không dây. Phương pháp này lặp lại các bước ước lượng tham số kênh truyền bằng bộ ước lượng hợp lẽ cực đại (MLE - Maximum likelihood estimator) của tín hiệu nhận được dựa trên các tham số ước lượng và thông số từ đo đạc. Một vài bước tổng quan trong giải thuật SAGE như sau

1. Khởi tạo ước lượng tham số kênh ban đầu ($\hat{\mathbf{H}}$). Điều này có thể được thực hiện ngẫu nhiên hoặc dựa trên kiến thức trước đó, tùy thuộc vào thông tin có sẵn (Θ).
2. Với các ước lượng tham số kênh hiện tại, tính kỳ vọng có điều kiện của tín hiệu truyền tải dựa trên tín hiệu nhận được.

$$\hat{\mathbf{s}}_{\hat{\mathbf{H}}}(\Theta) = \mathbb{E}(\mathbf{s}|\mathbf{x}, \Theta)$$

3. Cập nhật ước lượng tham số kênh bằng cách tối đa hóa khả năng của tín hiệu nhận được dựa trên kỳ vọng có điều kiện tính toán ở bước trước đó.

$$\hat{\mathbf{H}} = \min_{\hat{\mathbf{H}}} \hat{\mathbf{s}}_{\hat{\mathbf{H}}}(\Theta)$$

4. Lặp lại các bước 2 và 3 cho đến khi các ước lượng tham số kênh hội tụ. Sự hội tụ thường được xác định bằng cách theo dõi một tiêu chí hội tụ, chẳng hạn như sự thay đổi trong ước lượng tham số giữa các lần lặp (hàm mất mát).
5. Sau khi đạt được sự hội tụ, thu được các ước lượng cuối cùng của các tham số kênh.

1.1.1.2. Kỹ thuật phân tích theo tia

Những năm gần đây, một loại mô hình xác định mới được đề xuất cho ra kết quả không thua kém quá nhiều so với việc sử dụng các máy đo chuyên dụng như đã trình bày ở trên, đó là phương pháp phân tích theo tia (RT - Ray-tracing) [31]. Dựa trên lợi thế của các mô hình 3D và kỹ thuật RT trong các lĩnh vực khác như trò chơi trực tuyến, nhiều đề xuất sử dụng RT cho mô hình kênh mMIMO với việc coi các đường truyền vô tuyến giống như các tia sáng lan truyền trong không gian mô phỏng. Ưu điểm của phương pháp này đó là không cần phải đo đạc quá nhiều thông số như các mô hình đo truyền thống. Thay vào đó, các thông số được quan tâm chỉ còn vị trí và hướng của bên thu, phát, cùng mô hình 3D chi tiết của môi trường phát. Đây chính là điểm đặc biệt của phương pháp này, khi mô hình 3D chi tiết của khu dân cư, văn phòng, cần được xây dựng chính xác và chi tiết để làm đầu vào cho RT. Quá trình ray-tracing trong mô hình kênh RT bao gồm các bước sau

1. Xác định vị trí và hướng di chuyển của các ăng-ten truyền và nhận trong mMIMO.
2. Xác định môi trường không gian, bao gồm các đối tượng như tòa nhà, cây cối, và địa hình.
3. Tạo ra các bức xạ ray từ anten truyền và theo dõi chúng khi chúng tương tác với các đối tượng trong môi trường.
4. Phân tích sự tương tác của các bức xạ ray với môi trường để tính toán sự suy giảm, phản xạ, giao thoa và nhiễu.
5. Tính toán các thông số truyền thông như tốc độ bit, tỷ lệ lỗi bit và công suất tín hiệu đến ở các ăng-ten nhận.

Một trong các ứng dụng có thể giúp người dùng thử nghiệm mô hình kênh truyền dựa trên phương pháp RT hiện nay là MatLab R2023a [35], tuy nhiên, các mô hình 3D có sẵn hiện tại chỉ là mô hình đường phố từ các nguồn dữ liệu công khai.

1.1.2. Mô hình kênh truyền ngẫu nhiên

Các mô hình kênh truyền ngẫu nhiên dựa trên các đặc trưng mang tính thống kê của kênh vô tuyến trong mMIMO để đưa ra các giả thiết cho ma trận kênh truyền H . Trong đó, CBSM là mô hình kênh được sử dụng phổ biến nhất trong các nghiên cứu nhận dạng kênh truyền lý thuyết do tính đơn giản của nó. Ngược lại, GBSM phân tích kênh truyền theo hướng vật lý, chia nhỏ việc truyền sóng vô tuyến thành các cụm, đường và xem xét ảnh hưởng của từng thành phần riêng lẻ đến ma trận kênh truyền.

1.1.2.1. CBSM

Xét một biểu diễn ma trận kênh truyền ở dạng đơn giản nhất như sau

$$\mathbf{H}_{total} = \alpha \mathbf{H} \quad (1.2)$$

trong đó, α đại diện cho các ảnh hưởng của kênh truyền ở kích thước lớn (large-scale fading) gồm suy hao (path loss), shadowing như được biểu diễn trên phương trình (1.7). Ma trận \mathbf{H} đại diện cho ảnh hưởng của kênh truyền ở kích thước nhỏ (small-scale fading), thông thường được biểu diễn dưới dạng ma trận

$$\mathbf{H} = \begin{bmatrix} h_{0,0} & \dots & h_{0,T-1} \\ \vdots & \ddots & \vdots \\ h_{L-1,0} & \dots & h_{L-1,T-1} \end{bmatrix} \quad (1.3)$$

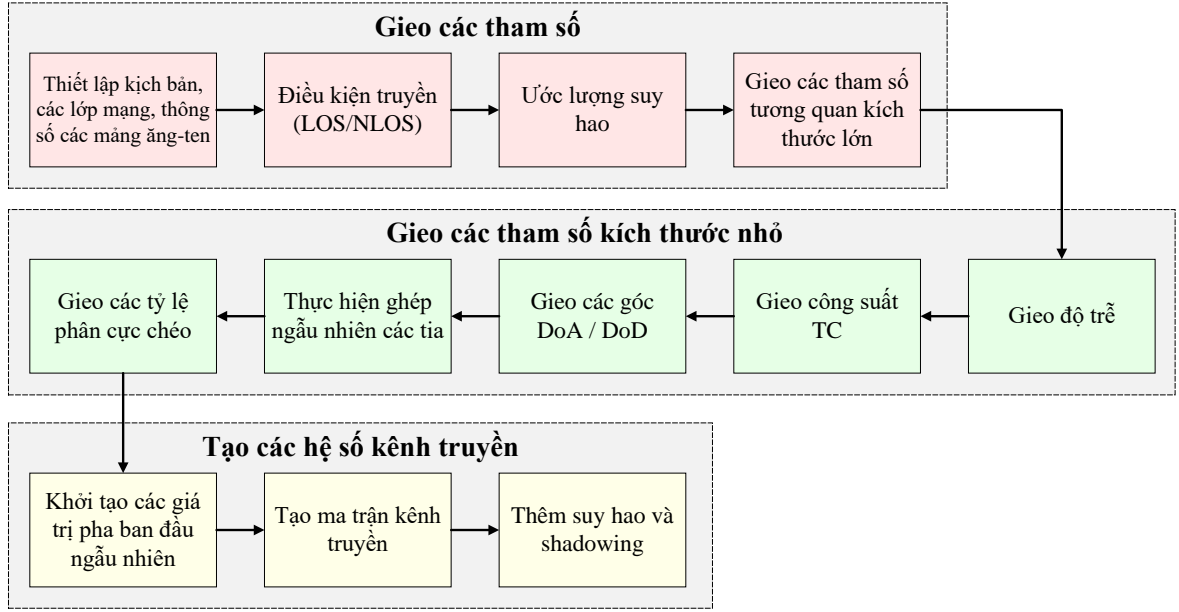
với mỗi phần tử thứ (l, t) tương ứng với đáp ứng xung của kênh giữa ăng-ten phát thứ t và ăng-ten nhận thứ l . Điều tương tự với các mô hình băng rộng gồm các độ trễ kênh khác nhau [48]. Nếu các hệ số trong ma trận kênh truyền được giả sử là có phân bố Gaussian phức, ma trận \mathbf{H} có thể được xác định chỉ từ ma trận hiệp phương sai như sau

$$\text{vec}(\mathbf{H}) = \mathbf{R}_{\text{mMIMO}}^{1/2} \text{vec}(\mathbf{G}) \quad (1.4)$$

trong đó, vec là phép véc-tơ hoá một ma trận, từ \mathbf{H} kích thước $L \times T$ thành $LT \times 1$. $\text{vec}(\mathbf{G})$ là véc-tơ chứa các biến ngẫu nhiên i.i.d Gaussian và phương sai σ . Ma trận $\mathbf{R}_{\text{mMIMO}}^{1/2}$ mô tả cấu trúc không gian của các mảng mMIMO, trong đó xem xét sự tương quan (correlation) lẫn nhau giữa các phần tử trong ma trận kênh truyền. Ma trận này được ước lượng như sau

$$\mathbf{R}_{\text{mMIMO}} = \mathbb{E}(\mathbf{H}\mathbf{H}^H) \quad (1.5)$$

với $(.)^H$ là ma trận Hermitian. Trong luận văn này, một trường hoặc đặc biệt của CBSM được xem xét đó là mô hình i.i.d Rayleigh. Mô hình này giả sử không có sự tương quan/ảnh hưởng lẫn nhau giữa các ăng-ten truyền/nhận. Điều này chỉ có được khi số lượng thành phần đa đường nhiều và phân bố đều trong miền không gian. Với giả thiết này, các giá trị trong ma trận tương quan của mMIMO ($\mathbf{R}_{\text{mMIMO}}$) chỉ tương ứng với một giá trị thực duy nhất. Do đó, theo phương trình (1.4), các hệ số trong ma trận \mathbf{H} sẽ là các biến ngẫu nhiên i.i.d Gaussian. Để đơn giản hoá hơn nữa, trong phương trình (1.2), các ảnh hưởng của kênh truyền ở kích thước lớn (α) được gộp vào ma trận \mathbf{H} thành các biến i.i.d Gaussian duy nhất. Trong phần sau của luận văn, mô hình kênh truyền này được gọi



Hình 1.3. Các bước tạo mô hình kênh truyền trong tiêu chuẩn 3GPP phiên bản 16 [1].

là “**phi cấu trúc**” (unstructured) do các ảnh hưởng của kênh đường minh hoạ qua các hệ số kênh duy nhất, không xét đến sự lan truyền của sóng vô tuyến như trong mô hình kênh “có cấu trúc” sẽ được trình bày ở mục 1.1.2.2.

1.1.2.2. GBSM

GBSM là mô hình kênh có độ phức tạp cao, khi xem xét sự lan truyền của sóng vô tuyến trong không gian dưới dạng các cụm, tia, đường truyền riêng biệt. Đây cũng là phương pháp được sử dụng phổ biến trong các chuẩn truyền thông di động, ví dụ như: 3GPP phiên bản 16 [1] cho 5G hay WINNER [18]. Phương pháp mô hình hoá kênh trong các tiêu chuẩn này được chia vào nhánh bán GBSM (semi-GBSM), do vừa dựa trên các thông số ngẫu nhiên và các thông số cố định thu được đo lường. Như nêu trên, tuy có tính chính xác cao, nhưng do hạn chế về độ phức tạp, nên một phương pháp mô hình kênh GBSM khác chỉ gồm các thành phần ngẫu nhiên có tên “có cấu trúc” hay tham số (parametric) cũng sẽ được trình bày ở phần sau.

Trên hình 1.3 là 12 bước trong quá trình mô hình hoá kênh truyền mMIMO được đề ra trong tiêu chuẩn 3GPP phiên bản 16 dành cho các hệ thống 5G. Phần tiếp theo sẽ trình bày sơ lược về các nét chính trong phương pháp mô hình hoá này. Dạng biểu diễn điển hình của phương pháp mô hình kênh truyền đo đạc như sau [20]

$$h_{l,t} = \sum_{n=1}^N \sum_{m=1}^{M_n} \alpha_{n,m,t} \cdot e^{i\varphi_{n,m,t}} \cdot \text{sinc}(l - \tau_{n,m,t}) \cdot e^{-ik_d s_t(\theta_{n,m,t}^d, \phi_{n,m,t}^d)} e^{-ik_a s_t(\theta_{n,m,t}^a, \phi_{n,m,t}^a)} \quad (1.6)$$

Bảng 1.2. Số lượng các TC và sub-path trong mỗi TC theo chuẩn 3GPP phiên bản 16 [1].

Thông số	UMi - Khu dân cư, kích cỡ micro			UMa - Khu dân cư, kích cỡ macro		
	LOS	NLOS	O2I	LOS	NLOS	O2I
Số lượng TC	12	19	12	12	20	12
Số lượng sub-path mỗi TC	20	20	20	20	20	20
Thông số	RMa - Khu nông thôn, kích cỡ macro			Văn phòng, trong nhà		
	LOS	NLOS	O2I	LOS	NLOS	
Số lượng TC	11	10	10	15	19	
Số lượng sub-path mỗi TC	20	20	20	20	20	

với $l = 0, 1, \dots, L - 1$. N, M_n lần lượt là số lượng các cụm phân chia theo thời gian (TC - Time cluster) đến bên thu và số lượng các đường nhỏ hơn (sub-path) trong mỗi cụm kể trên. Với sub-path thứ m trong cụm thứ n , α, φ, τ đại diện cho biên độ, pha của sub-path, và thời gian trễ tuyệt đối. Góc ngẩng (zenith) và góc phương vị (azimuth) của hướng sóng đi (DoD - Direction of departure) và hướng sóng đến (DoA) lần lượt là $\theta^d, \phi^d, \theta^a, \phi^a$. Ký hiệu (\cdot) tương ứng là phép nhân vô hướng. Các thành phần biên độ $\alpha_{n,m,t}$ của $\mathbf{h}_{l,t}$ lại phụ thuộc vào suy hao, như một hàm của tần số sóng mang (f_c), khoảng cách (d), cho bởi

$$\alpha_{n,m,t}(f_c, d)[\text{dB}] = \text{FSPL}(f_c, d_0) + 10n \log_{10} \left(\frac{d}{d_0} \right) + \chi_\sigma, \quad (1.7)$$

với $d \geq d_0$, khi $d_0 = 1\text{m}$

trong đó, $\text{FSPL}(f_c, d_0) = 20 \log_{10}(4\pi d_0 c / f_c)$, với c là vận tốc ánh sáng, n đại diện cho hệ số mất mát hàm mũ (PLE - Path loss exponent), và χ_σ là thành phần shadow fading được mô hình hoá dưới dạng một biến ngẫu nhiên lognormal với trung bình không và độ lệch chuẩn σ . Các thông số trong các phương trình (1.6) và (1.7) được quy định và đo đạc xác định cho từng môi trường khác nhau. Đầu tiên, cần xác định số lượng các TC trong một môi trường cụ thể, trên bảng 1.2 chỉ ra các thông số được đo đạc này cho 4 loại môi trường cụ thể, bao gồm: UMi - khu dân cư kích cỡ micro, UMa - khu dân cư kích cỡ macro, RMa - khu nông thôn kích cỡ macro, và điều kiện văn phòng trong nhà. Mỗi môi trường trên, các kịch bản đo được thiết lập ở 2 hoặc 3 tình huống gồm: tầm nhìn thẳng (LOS - Line of sight), không tầm nhìn thẳng (NLOS - non-LOS), và từ bên ngoài vào trong nhà (O2I - Outdoor to indoor).

Tiếp đến, độ dịch pha ban đầu của các sub-path trong mỗi TC cũng được 3GPP đưa ra như trên bảng 1.3.

Thông tin về tỷ lệ công suất của các sub-path trong mỗi TC và độ trễ của chúng so với sub-path đến sớm nhất được trình bày trên bảng 1.4. Cụ thể, mỗi TC sẽ được chia

Bảng 1.3. Độ lệch pha của các sub-path trong một TC.

sub-path #	$\varphi_{n,m,t}(^{\circ})$	sub-path #	$\varphi_{n,m,t}(^{\circ})$
1, 2	$\pm 0,0447$	11, 12	$\pm 0,6797$
3, 4	$\pm 0,1413$	13, 14	$\pm 0,8844$
5, 6	$\pm 0,2492$	15, 16	$\pm 1,1481$
7, 8	$\pm 0,3715$	17, 18	$\pm 1,5195$
9, 10	$\pm 0,5129$	19, 20	$\pm 2,1551$

Bảng 1.4. Thông tin về công suất và độ trễ của các sub-path trong mỗi TC.

SC #	sub-path #	Tỷ lệ công suất ($\alpha_{n,m,t} / \sum_{m=1}^M \alpha_{n,m,t}$)	Trễ tuyệt đối ($\tau_{n,m,t} - \tau_{n,1,t}$)
1	1, 2, 3, 4, 5, 6, 7, 8, 19, 20	10/20	0 ns
2	9, 10, 11, 12, 17, 18	6/20	5 ns
3	13, 14, 15, 16	4/20	10 ns

nhỏ hơn nữa thành 3 cụm con (SC - sub-cluster). Các sub-path sẽ được chia vào một trong ba SC trên dựa trên thời gian trễ của chúng so với tia đến sớm nhất. Bốn thông số còn lại bao gồm θ và ϕ của bên thu và phát lần lượt được biểu diễn dưới dạng các giá trị ngẫu nhiên thuộc các phân bố khác nhau tùy thuộc vào điều kiện môi trường, chi tiết xem trong các bảng từ 7.5-6 đến 7.5-10 tại [1].

Một phương pháp mô hình hoá kênh truyền GBSM khác, với độ phức tạp nhỏ hơn đó là parametric hay specular [26, 56]. Mô hình này vẫn kế thừa đặc trưng không gian của GBSM khi xem xét việc truyền sóng theo các đường khác nhau trong không gian 3D (φ). Tuy nhiên, thay vì chia thành các cụm, sub-path với các góc, độ trễ khác nhau, specular gộp các cụm (TC) thành các đường truyền lớn do xem xét đặc trưng là tính thừa trong các hệ truyền thông sử dụng sóng mi-li-mét. Các thành phần suy hao, dịch pha, và độ trễ được gộp chung lại thành một hệ số khuếch đại phức (β).

$$h_{l,t} = \sum_{n=0}^{N-1} \beta_{n,t} = \sum_{n=0}^{N-1} \beta_{n,t} \cdot e^{j\varphi_{l,n,t}} \quad (1.8)$$

các giá trị $\beta, \varphi(\theta, \phi)$ trong mô hình kể trên thường sẽ được gieo ngẫu nhiên theo các phân bố biết trước. Chi tiết hơn về mô hình kênh parametric có tại chương 2. Trong phần sau của luận văn, mô hình kênh truyền này được gọi là “**có cấu trúc**” (structured) để phân biệt với phi cấu trúc trong mục CBSM ở trên.

1.1.2.3. Đánh giá các phương pháp mô hình kênh cho mMIMO

Dựa trên [7], có ba thông số để đánh giá một mô hình kênh truyền bao gồm: tính chính xác (sự tương quan của mô hình và kênh thực), độ phức tạp (lượng tính toán cần để tạo ra một mô hình kênh), và tính tổng quát (mô hình kênh có phù hợp với nhiều kịch

bản kênh truyền khác nhau hay không?).

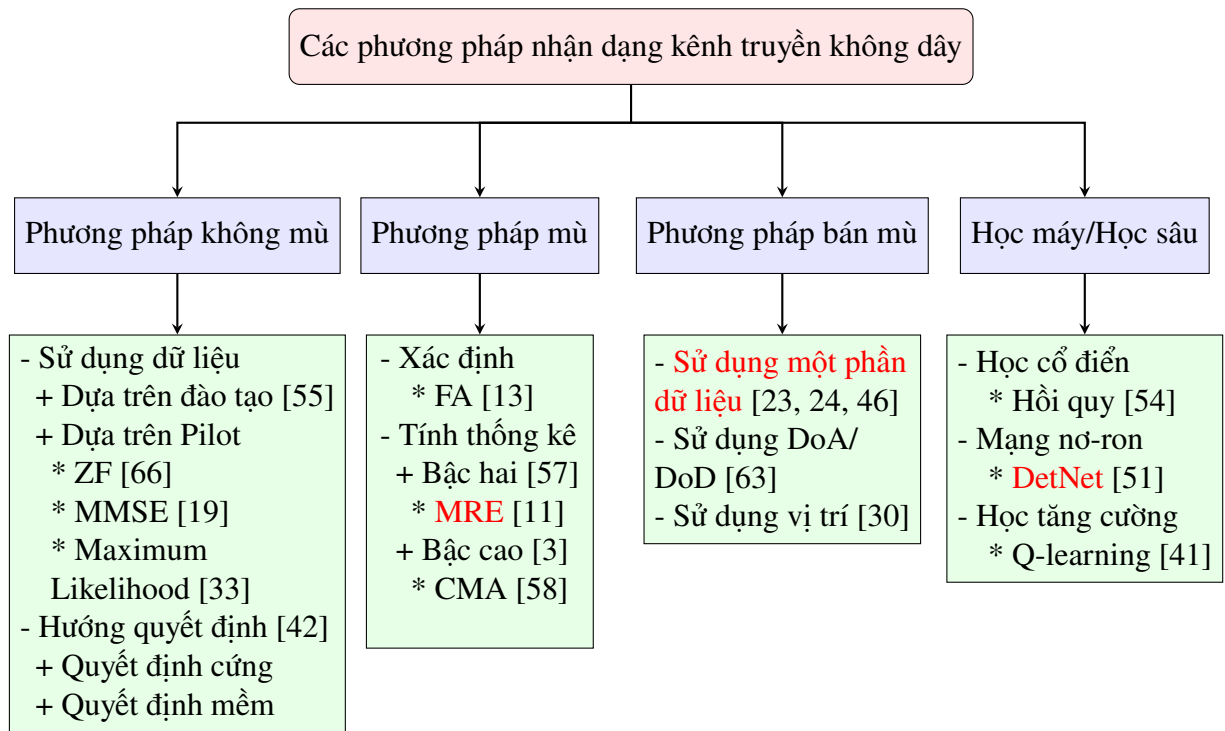
Cũng trong [7], nhóm tác giả đã chỉ ra thông qua mô phỏng, semi-GBSM là phương pháp cho tốc độ bit lớn nhất tương đương với mô hình kênh truyền xác định qua đo đạc, do đây cũng là hai phương pháp gần nhất với môi trường truyền thông không dây thực tế. Điểm hạn chế của cả hai phương pháp xác định và semi-GBSM nằm ở độ phức tạp trong các bài toán nghiên cứu. Với trường hợp mô hình kênh như của 3GPP đã được trình bày ở trên, phải cần đến 12 bước theo như tài liệu gốc tại [1] để mô hình hoá một kênh truyền cho một điều kiện truyền thông cụ thể duy nhất. Do vậy, hai phương pháp này cũng khó có tính tổng quát trong các bài toán nghiên cứu khi cố gắng tìm các giải thuật ước lượng kênh truyền phục vụ nhiều loại môi trường truyền nhất có thể. Với phương pháp RT, công việc khó nhất đó là xây dựng mô hình 3D của môi trường truyền, không thể mô phỏng chính xác tuyệt đối, yêu cầu môi trường truyền bất biến trong quá trình mô phỏng, và yêu cầu tài nguyên tính toán khổng lồ nên khó để đưa vào các hệ thống thời gian thực. Các phương pháp như CBSM hay parametric dù cho dung năng kênh giảm so với hai phương pháp kể trên, tuy nhiên lại đáp ứng được hai yêu cầu về độ phức tạp và tính tổng quát của một mô hình kênh truyền. Từ các lí do kể trên, trong luận văn này, hai mô hình kênh truyền ngẫu nhiên CBSM trong trường hợp i.i.d Rayleigh và GBSM với parametric được sử dụng, phù hợp với các nghiên cứu lí thuyết về ước lượng kênh truyền.

1.2. Các phương pháp nhận dạng hệ thống MIMO kích thước lớn

Ngày nay, các thuật toán ước lượng kênh truyền không dây đã đạt được các bước tiến đáng kể về độ chính xác. Dựa trên đặc điểm của các thuật toán có thể chia thành bốn hướng tiếp cận chính như trên hình 1.4. Bao gồm phương pháp không mù (Non-blind), mù (B), bán mù (SB), và dựa trên học máy, học sâu (AI-based) [60]. Với mỗi phương pháp, rất nhiều thuật toán đã được đề xuất và cho hiệu quả trong các tình huống cụ thể như các trích dẫn trên hình 1.4. Từ cách phân loại kể trên, đôi nét cơ bản về các phương pháp ước lượng kênh truyền này sẽ được trình bày, trong số đó, một số thuật toán được dùng để so sánh kết quả trong các chương sau của luận văn sẽ được trình bày chi tiết.

1.2.1. Nhận dạng kênh không mù

Như trên hình 1.4, các phương pháp nhận dạng kênh không mù có thể chia làm hai nhóm chính, bao gồm các phương pháp sử dụng dữ liệu (Data-aided) [60] và các phương pháp dựa trên hướng quyết định (Decision-directed) [42]. Các thuật toán sử dụng dữ liệu có thể chia làm hai loại nhỏ hơn, gồm có các phương pháp dựa trên việc đào tạo (Training-based) và các phương pháp dựa trên tín hiệu hoa tiêu (Pilot-assisted). Khác biệt chính giữa hai phương pháp là loại tín hiệu được dùng để ước lượng kênh truyền.



Hình 1.4. Phân loại các phương pháp ước lượng kênh truyền viễn thông.

Với Training-based, bên phát sẽ gửi các dữ liệu huấn luyện gốc, bên thu chỉ biết thời điểm dữ liệu huấn luyện này được truyền nhưng không biết trước thông tin của dữ liệu. Tín hiệu bên thu nhận được gồm tín hiệu gốc và tín hiệu đã bị méo, từ đó, một mô hình ước lượng được huấn luyện bằng cách tối ưu hóa một hàm mất mát (loss function) giữa kết quả ước lượng kênh truyền và giá trị thực tế của kênh truyền. Với Pilot-assisted, các ký hiệu pilot được chèn trực tiếp vào khung dữ liệu gửi đi, và bên phát biết cả thời gian, vị trí, và giá trị gốc của các ký hiệu pilot này. Từ đó, bên thu có thể ước lượng ra ảnh hưởng của kênh truyền đến các tín hiệu pilot và nội suy ra ảnh hưởng của kênh truyền đến toàn bộ khung dữ liệu. Các giải thuật phổ biến được sử dụng cho phương pháp Data-aided có thể kể đến như bộ phát hiện ép về không (ZF - Zero forcing), lỗi trung bình phương sai tối thiểu (MMSE - Minimum mean square error) [19]. Hai giải thuật này là giải thuật tuyến tính và sẽ được trình bày chi tiết hơn ở mục 1.2.1.1 và 1.2.1.2. Tuy phổ biến và được áp dụng trong các hệ truyền thông thực tế, nhưng các phương pháp sử dụng dữ liệu để ước lượng kênh truyền có một nhược điểm đó là giảm hiệu quả sử dụng phổ do một phần băng thông bị lãng phí để truyền tải các dữ liệu huấn luyện hoặc pilot.

Phương pháp hướng quyết định (DDCE - Decision-directed channel estimation) cũng dựa trên việc sử dụng dữ liệu, tuy nhiên, thay vì ước lượng kênh truyền chỉ trong một bước DDCE có thêm một bước nữa [60]. Cụ thể, tại bước một, DDCE vẫn ước lượng kênh truyền dựa trên một trong hai phương pháp Training-based hoặc Pilot-assisted như Data-aided. Sau đó, khôi phục các tín hiệu dựa trên trạng thái kênh truyền vừa ước lượng

được. Ở bước tiếp theo, các dữ liệu mới được khôi phục sẽ tiếp tục được đưa vào thuật toán ước lượng nhằm cập nhật trạng thái thông tin về kênh truyền cho đến khi các ký hiệu trong một phiên được truyền hết. Chi tiết hơn, bộ thu sẽ so sánh ký tự đã nhận được với ký tự được dự đoán dựa trên ký tự trước đó và ước lượng kênh truyền hiện tại. Nếu có sai sót giữa ký tự đã nhận và ký tự dự đoán, bộ thu sẽ điều chỉnh lại giá ước lượng để cải thiện độ chính xác của dự đoán ký tự tiếp theo. Quá trình này được lặp lại cho mỗi ký tự nhận được. Việc ra quyết định bit là 0 hay 1 trong DDCE sẽ được thực hiện theo hai giải thuật gồm quyết định mềm (soft) và quyết định cứng (hard). Quyết định mềm [44] sẽ xác định giá trị của các bit dữ liệu bằng cách tính toán xác suất bit đó được truyền qua kênh truyền. Ngược lại, với quyết định cứng [53], một ngưỡng xác định được đưa ra, nếu lớn hơn ngưỡng này sẽ là bit 1, ngược lại là bit 0. Tuy nhiên, phương pháp DDCE có điểm hạn chế đó là quá trình ước lượng lại bị phụ thuộc vào các dữ liệu cũ, dẫn đến việc, có thể kênh truyền hiện tại không còn tương ứng với các dữ liệu từ thời điểm quá khứ. Điều này dẫn đến các lỗi tích lũy và làm giảm hiệu năng của hệ thống nhận dạng.

1.2.1.1. Zero Forcing

Các giải thuật nhận dạng tuyến tính dựa trên các phép biến đổi tuyến tính của các tính hiệu nhận được \mathbf{x} . Các giải thuật này thường có độ phức tạp thấp, hoặc trung bình. Nhưng độ phức tạp sẽ tăng lên nếu hệ thống có số chiều lớn, ví dụ số lượng ăng-ten T hay L rất lớn trong mMIMO dẫn đến phép nghịch đảo ma trận tiêu tốn nhiều tài nguyên tính toán hơn. Một bộ nhận dạng tuyến tính có thể biểu diễn như bên dưới

$$\mathbf{s} = \mathbf{G}\mathbf{x} \quad (1.9)$$

ZF là phương pháp đơn giản nhất của các thuật toán nhận dạng tuyến tính. Trong đó, ma trận kênh truyền \mathbf{H} sẽ được nghịch đảo để loại bỏ ảnh hưởng của kênh truyền. Ma trận làm bằng (equalizer matrix) \mathbf{G}_{ZF} của bộ nhận dạng ZF như sau

$$\mathbf{G}_{ZF} = \left(\mathbf{H}^H \mathbf{H} \right)^{-1} \mathbf{H}^H \quad (1.10)$$

Với \mathbf{G}_{ZF} , tín hiệu gốc được khôi phục/ước lượng bằng cách

$$\hat{\mathbf{s}}_{ZF} = \left(\mathbf{H}^H \mathbf{H} \right)^{-1} \mathbf{H}^H \mathbf{x} \quad (1.11)$$

1.2.1.2. Minimum Mean Square Error

Hiệu năng của bộ nhận dạng ZF thường bị ảnh hưởng bởi tạp âm AWGN. Do vậy, MMSE kết hợp thêm thông tin phương sai của nhiễu trước khi nghịch đảo ma trận để

cho ra độ chính xác cao hơn. Ma trận làm bằng \mathbf{G}_{MMSE} của bộ nhận dạng MMSE được biểu diễn dưới dạng

$$\mathbf{G}_{MMSE} = \left(\mathbf{H}^H \mathbf{H} + \frac{\sigma^2}{\mathbb{E}(\mathbf{s})} \mathbf{I} \right)^{-1} \mathbf{H}^H \quad (1.12)$$

với σ^2 là phương sai của nhiễu AWGN, $\mathbb{E}(\mathbf{s})$ là công suất trung bình của mỗi ký hiệu gửi đi, và \mathbf{I} là ma trận đơn vị. với \mathbf{G}_{MMSE} , tín hiệu gốc được khôi phục như sau

$$\hat{\mathbf{s}}_{MMSE} = \left(\mathbf{H}^H \mathbf{H} + \frac{\sigma^2}{\mathbb{E}(\mathbf{s})} \mathbf{I} \right)^{-1} \mathbf{H}^H \mathbf{x} \quad (1.13)$$

Ưu điểm của MMSE, các giá trị thấp trong quá trình nghịch đảo có thể dẫn đến hiện tượng khuếch đại tạp âm (deep null) khi sử dụng ZF, được khắc phục bởi công suất tạp âm khác không. Tuy nhiên, có thể nhận thấy cả hai giải thuật ZF và MMSE cần các chuỗi pilot để ước lượng kênh truyền, sau đó nội suy ra ma trận \mathbf{H} .

1.2.2. Nhận dạng kênh mù

Các kỹ thuật nhận dạng hệ thống mù (hoặc tương tự như giải mã mù hoặc cân bằng mù) đã được biết đến từ đầu những năm 1980. Theo [60], có thể chia các thuật toán mù vào hai nhóm chính. Thứ nhất là các kỹ thuật ước lượng kênh truyền dựa trên đặc tính thống kê của tín hiệu thu được, có thể là đặc tính thống kê bậc hai (SOS - Second-order statistics), hoặc bậc cao (HOS - Higher-order statistics). Cách tiếp cận SOS được đề xuất trong [57] yêu cầu các tín hiệu có đặc tính chu kỳ hoặc đa dạng kênh (channel diversity) với các hệ thống đơn đầu đơn vào đầu ra (SISO - Single-input single-output). SOS có ưu điểm là yêu cầu lượng dữ liệu ít hơn để có được các ước tính thống kê đáng tin cậy tương đương với phương pháp HOS. Một ví dụ của SOS là thuật toán MRE [11] khai thác đặc trưng tham chiếu của hệ thống gồm nhiều cảm biến thu (sensor) hay được hiểu là đa ăng-ten trong một hệ đơn đầu vào đa đầu ra (SIMO - Single-input multi-output). Gasbert và các cộng sự đề xuất phương pháp ước lượng một tập K bộ lọc để làm bằng kênh. Khác với SOS, HOS [12] có lợi thế là cung cấp thông tin về pha mà không cần đa dạng kênh với đánh đổi là cần một lượng lớn dữ liệu lấy mẫu và khả năng tính toán cao hơn. Có thể kể đến như thuật toán mô-đun không đổi (CMA - Constant modulus algorithm) [58] khai thác đặc trưng là giá trị mô-đun không đổi của các tín hiệu phức khi sử dụng các bộ điều chế như điều chế pha (PSK - Phase-shift keying), điều chế biên độ cầu phương 4 điểm (4-QAM - Quadrature amplitude modulation). Từ đó, Treichler và các đồng tác giả đã đề xuất cân bằng kênh truyền bằng một bộ lọc thích nghi (adaptive filter) để đưa mô-đun của tín hiệu thu được về các giá trị chuẩn của PSK hay 4-QAM.

Nhóm kỹ thuật thứ hai đó là khai thác các thông tin đã xác định (deterministic) của tín hiệu hoặc hệ thống. Trong [5], nhóm tác giả sử dụng đặc trưng thưa (sparsity) của tín hiệu thường xuất hiện nhiều trong các kênh truyền mMIMO hay bước sóng mi-li-mét (mmWave - Millimeter wave) hiện nay. Bằng cách sử dụng tính chất thưa, các tín hiệu gốc có thể được khôi phục trong trường hợp hệ thống dưới mức xác định (underdetermined). Trong một số điều kiện cụ thể, việc áp dụng ràng buộc thưa có thể làm cải thiện hiệu năng của việc nhận dạng hệ thống mù.

1.2.3. Nhận dạng kênh bán mù

Các phương pháp nhận dạng kênh bán mù được sinh ra từ sự kết hợp của các kỹ thuật Non-blind (NB) và Blind. Giải pháp lai (hybrid) này được kỳ vọng sẽ giảm đi lượng pilot cần thiết mà vẫn bù đắp lại được độ chính xác bằng các thông tin từ kỹ thuật mù mang lại. Các tiếp cận đơn giản nhất đó là kết hợp trực tiếp các bộ nhận dạng như ZF, MMSE với các thông tin thống kê SOS, HOS đã được trình bày ở trên. Các công bố [23, 46, 61] đi theo hướng tiếp cận này đều cho ra các kết quả vượt trội khi so với với NB truyền thống trong một số điều kiện nhất định. Ngoài ra, việc kết hợp các thông tin xác định của các bộ cân bằng mù như được trình bày ở mục 1.2.2 cũng là các hướng nghiên cứu tiềm năng trong tương lai.

Tiếp đến, ngoài các đặc trưng của tín hiệu, các thông tin bên lề (side-information) của hệ thống thu phát cũng có thể được xem xét để cải thiện khả năng nhận dạng kênh truyền. Có thể kể đến như sử dụng thêm thông tin hướng sóng đến/đi (DoA/DoD) như trong [63], nhóm tác giả đã đề xuất sử dụng DoA của các người dùng khác nhau để giảm thiểu/loại bỏ sự ảnh hưởng của ô nhiễm pilot (PC - Pilot contamination), qua đó hiệu suất của việc nhận dạng kênh truyền đã được cải thiện. Tiếp đến, [30] đề xuất sử dụng thông tin về toạ độ/vị trí (location) người dùng để đánh giá đáp ứng tần số kênh truyền mmWave. Kết quả mô phỏng cho thấy cả độ chính xác và độ phức tạp của mô hình ước lượng đều giảm đi khi có thêm loại thông tin bên lề này.

1.2.4. Nhận dạng kênh sử dụng học máy

Nhận dạng kênh truyền sử dụng ML/DL là hướng tiếp cận bùng nổ trong các năm trở lại đây. Do các bước tiến đã đạt được trước đó của việc xử lý các loại tín hiệu âm thanh, hình ảnh sử dụng các mạng học sâu. Việc chuyển tiếp các kỹ thuật sẵn có này sang viễn thông xảy ra nhanh chóng và bước đầu các nghiên cứu đã chỉ ra các kết quả tiềm năng. Điểm khác biệt của hướng tiếp cận này đó là nó bao hàm được lý thuyết của cả ba hướng tiếp cận kể trên bao gồm mù, bán mù, và không mù. Tuy nhiên, thay vì việc tìm các phương pháp tối ưu và nghiệm chính xác, ML/DL sử dụng các thuật toán ML

cơ bản, mạng nơ-ron (NN - Neural network), hay học tăng cường (RL - Reinforcement learning) với các đầu vào của hệ thống nhận dạng với B, SB, NB.

Các phương pháp sử dụng học máy cổ điển để nhận dạng kênh truyền được phát triển trước tiên, do độ phức tạp ở mức thấp. Trong [54], việc ước lượng ma trận làm bằng G_{MMSE} được thay thế bằng thuật toán hồi quy Gaussian (GPR - Gaussian process regression). Các ưu điểm của GPR như (i) tỷ lệ lỗi bit (BER - Bit error rate) thấp hơn MMSE truyền thống; (ii) nội suy chính xác hơn ước tính kênh ở giữa các ký hiệu pilot so với kỹ thuật nội suy tuyến tính. Ngoài phương pháp hồi quy, các giải thuật cổ điển của học máy như giảm số chiều của dữ liệu (PCA - Principal components analysis, ICA - Independent component analysis), học Bayesian cũng được đề xuất và cho thấy sự hiệu quả [60].

Các phương pháp nhận dạng sử dụng mạng nơ-ron còn có những bước tiến rõ ràng hơn, khi NN phức tạp hơn và số lượng tham số đào tạo cũng là rất lớn để đáp ứng được các mô hình kênh phức tạp. Các nghiên cứu trong mục 1.2.3 như [30, 61] cũng sử dụng các thông tin bên lề cho SB nhưng thay vì phương pháp tối ưu đại số, các mạng nơ-ron sâu (DNN - Deep-neural network) đã được đề xuất để ước lượng kênh truyền. Một trong những mạng DNN đầu tiên được đề xuất cho việc nhận dạng hệ thống MIMO/mMIMO đó là mạng phát hiện (DetNet - Detection network) [51]. Với kiến trúc là các phép lặp của thuật toán giảm độ dốc dự kiến kết (gradient descent) hợp thành một mạng. DetNet đã cho kết quả về độ chính xác vượt trội các phương pháp nhận dạng tuyến tính ở mức BER đạt 10^{-3} dB tại tỷ số tín trên tạp (SNR - Signal noise ratio) 10 dB. Tuy nhiên, do số lượng tham số cần đào tạo là lớn nên quá trình đào tạo có thể tốn chi phí, từ đó một số mạng dựa trên ISD khác đã được đề xuất [29, 34], với độ chính xác tốt hơn DetNet nhưng số lượng tham số đào tạo chỉ dưới 100. Ngoài ra, rất nhiều các mô hình mạng NN khác đã được đề xuất, như mạng trí nhớ dài hạn/ngắn hạn (LSTM - Long/short-term memory), bộ tự mã hoá (Autoencoders), ... [60].

Tương tự như hai phương pháp kể trên, RL cũng được đưa sang ứng dụng cho nhận dạng kênh truyền. Trong nghiên cứu [41], nhóm các tác giả đã trình bày một phương pháp khử nhiễu trên miền tần số dựa trên RL không cần kiến thức kênh tiền nghiệm và dữ liệu được dán nhãn trước. Cụ thể, thuật toán cung cấp một cải tiến đáng kể so với phương pháp ước lượng bình phương nhỏ nhất (LS - Least squares) và mang lại hiệu suất tiệm cận với ước lượng lỗi bình phương trung bình nhỏ nhất tuyến tính (LMMSE - Linear MMSE) lý tưởng với toàn bộ thông tin về trạng thái kênh (CSI - Channel state information).

CHƯƠNG 2

ĐÁNH GIÁ ẢNH HƯỞNG CỦA CẤU HÌNH MẢNG ĂNG-TEN TRONG NHẬN DẠNG HỆ THỐNG MIMO KÍCH THƯỚC LỚN

Trong chương này, tác giả sẽ xem xét sự ảnh hưởng của các cấu hình mảng ăng-ten khác nhau đến sai số của việc ước lượng kênh truyền trong các hệ thống mMIMO. Đầu tiên, mô hình kênh truyền có cấu trúc cho hai cấu hình mảng ăng-ten 1D và 3D sẽ được trình bày. Sau đó là sơ lược về việc sử dụng CRB để ước lượng sai số trong các bộ nhận dạng sử dụng pilot và SB. Các kết quả mô phỏng được đưa ra để cho thấy hiệu quả của mô hình kênh truyền có cấu trúc, mảng ăng-ten 3D (UCyA), và phương pháp ước lượng SB trong việc nhận dạng hệ thống viễn thông.

2.1. Mô hình kênh truyền có cấu trúc với các cấu hình mảng ăng-ten khác nhau

Trong chương này, mô hình toán của một hệ thống mMIMO sử dụng điều chế ghép kênh phân chia theo tần số trực giao (OFDM - Orthogonal frequency-division multiplexing) với K sub-carriers trong kênh đường lên gồm T ăng-ten (mono-antenna) phát, và L ăng-ten thu. Mỗi ký hiệu OFDM bao gồm K ký hiệu dữ liệu và một phần tiền tố vòng (CP - Cyclic Prefix). Giả sử độ dài của CP lớn hơn hoặc bằng thời gian trễ tối đa của kênh truyền (coherent time). Tại ăng-ten thu thứ l , sau khi đã loại bỏ thành phần CP và thực hiện biến đổi Fourier (FFT - Fast fourier transform), tín hiệu đầu ra tại ăng-ten thu thứ l (\mathbf{x}_l) được biểu diễn như sau [24]

$$\mathbf{x}_l = \sum_{t=0}^{T-1} \mathcal{F} \mathcal{T}(h_{l,t}) \frac{\mathcal{F}}{K} \mathbf{s}_j + \mathbf{w}_l \quad (2.1)$$

trong đó \mathcal{F} đại diện cho ma trận Fourier rời rạc, gồm K điểm và \mathcal{T} là ma trận lặp lại của $h_{l,t}$. Tiếp đến, \mathbf{s}_j là ký hiệu OFDM thứ k có độ dài K và $\mathbf{w}_l \in \mathbb{C}^{K \times 1}$ là tạp âm AWGN có dạng i.i.d phân bố Gaussian $\mathcal{CN}(0, \sigma_w^2 \mathbf{I}_L)$. Cuối cùng, $h_{l,t}$ là hệ số kênh thuộc ma trận kênh truyền được biểu diễn dưới dạng véc-tơ $\mathbf{h} \in \mathbb{C}^{LT \times 1}$.

$$\mathbf{h} = \left[\mathbf{h}_0^\top, \mathbf{h}_1^\top, \dots, \mathbf{h}_{L-1}^\top \right]^\top, \quad (2.2)$$

$$\mathbf{h}_l = [h_{l,0}, h_{l,1}, \dots, h_{l,T-1}]^\top$$

với $(.)^\top$ là phép chuyển vị (transpose) ma trận. Giả sử rằng M là số lượng đường truyền giữa một cặp ăng-ten phát, thu. Dựa trên hướng tiếp cận mô hình kênh truyền có cấu trúc, $h_{l,t}$ được mô hình tương ứng với M đường truyền, mỗi đường truyền bao gồm hệ số khuếch đại phức (complex gain) và véc-tơ lái (φ - steering) như sau

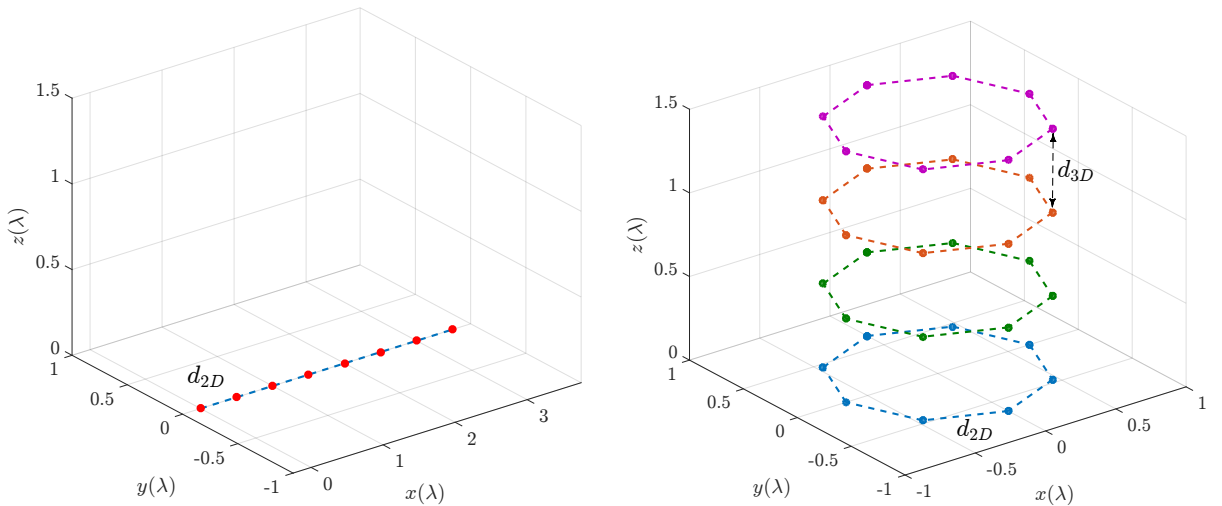
$$h_{l,t} = \sum_{m=0}^{M-1} \beta_{m,t} e^{\varphi_{m,t}} = \sum_{m=0}^{M-1} \beta_{m,t} \cdot e^{-ik_s c_l(\theta_{m,t}, \phi_{m,t})} \quad (2.3)$$

tại đường truyền thứ m , hệ số β biểu diễn cho hệ số khuếch đại phức. Góc ngẩng (zenith) và góc phương vị (azimuth) của hướng sóng đến (DoA)¹ lần lượt là θ, ϕ . Ký hiệu (\cdot) tương ứng là phép nhân vô hướng.

Các ký hiệu còn lại trong phương trình (2.3) lần lượt là

$$\begin{aligned} k_s &= 2\pi/\lambda \\ c_l(\theta_{m,t}, \phi_{m,t}) &= \hat{\mathbf{c}} \cdot \mathbf{c}_l \\ \hat{\mathbf{c}} &= \sin \theta_{m,t} \cos \phi_{m,t} \hat{\mathbf{x}} + \sin \theta_{m,t} \sin \phi_{m,t} \hat{\mathbf{y}} + \cos \theta_{m,t} \hat{\mathbf{z}} \\ \mathbf{c}_l &= x_l \hat{\mathbf{x}} + y_l \hat{\mathbf{y}} + z_l \hat{\mathbf{z}} \end{aligned} \quad (2.4)$$

trong đó, λ là bước sóng; $\hat{\mathbf{c}}$ là véc-tơ đơn vị trong hệ toạ độ đề-các (Descartes) ba chiều của DoA; và \mathbf{c}_l là vị trí của phần tử thứ l trong mảng ăng-ten bên thu ứng với toạ độ (x_l, y_l, z_l) .



(a) Mảng thẳng cách đều (ULA - Uniform Linear Array)

(b) Mảng trụ đồng nhất (UCyA - Uniform Cylindrical Array)

Hình 2.1. Minh hoạ hai cấu hình 1D và 3D của các mảng ăng-ten.

Trong luận văn này, hai cấu hình của các mảng ăng-ten được biểu diễn trên hình 2.1

¹Để đơn giản hoá, thông tin về DoD được coi là không biết trước tại bên thu của kênh đường lên.

được khảo sát, gồm mảng 1D và 3D [45]. Ứng với mảng ăng-ten 1D, cấu hình đơn giản nhất là mảng thẳng cách đều (ULA) gồm N_{ULA} phần tử được xem xét. Trong đó khoảng cách giữa các phần tử ăng-ten trong mảng ULA là d_{2D} . Ứng với mảng ăng-ten 3D, cấu hình mảng trụ đồng nhất (UCyA) gồm N_{3D} lớp của N_{UCA} phần tử thuộc một mảng tròn cách đều (UCA - Uniform circle array) được xem xét. Với khoảng cách giữa các phần tử trong mảng UCA cũng là d_{2D} và khoảng cách giữa các lớp của UCyA là d_{3D} theo hướng của trục z . Từ đó, bán kính (r) của các mảng vòng UCA được tính như sau

$$r = \frac{1/2 \cdot d_{2D}}{\sin(\pi/N_{UCA})} \quad (2.5)$$

Toạ độ (s_l) của các phần tử trong hai cấu hình mảng ăng-ten nêu trên là

$$s_l(\text{ULA}) = \begin{cases} x_l = n_{ULA} * d_{2D} \\ y_l = 0 \\ z_l = 0 \end{cases} \quad (2.6)$$

$$s_l(\text{UCyA}) = \begin{cases} x_l = r * \sin(n_{UCA} * \frac{2\pi}{N_{UCA}}) \\ y_l = r * \cos(n_{UCA} * \frac{2\pi}{N_{UCA}}) \\ z_l = n_{3D} * d_{3D} \end{cases} \quad (2.7)$$

với $n_{ULA} = 0, 1, \dots, N_{ULA} - 1$; $n_{UCA} = 0, 1, \dots, N_{UCA} - 1$, và $n_{3D} = 0, 1, \dots, N_{3D} - 1$.

2.2. Cramér Rao Bound cho giải thuật nhận dạng hệ thống không mù và bán mù

Trong mục này, tác giả sẽ trình bày về phương pháp đánh giá hiệu năng sử dụng CRB cho cả hai mô hình kênh truyền có cấu trúc (structured) và phi cấu trúc (unstructured) trong các hệ thống mMIMO. Sau đó, hai giải hướng tiếp cận: (i) sử dụng pilot (OP - Only pilot); (ii) bán mù (SB) sử dụng thêm một phần thông tin từ đặc trưng thống kê của dữ liệu, sẽ được so sánh hiệu năng cũng dựa trên CRB.

2.2.1. CRB trong trường hợp chỉ sử dụng pilot

Như đã trình bày ở phần mở đầu của luận văn, việc sử dụng các ký hiệu pilot hay tín hiệu tham chiếu để ước lượng sự ảnh hưởng của kênh truyền vô tuyến là phương pháp mà WiFi hay 5G đang sử dụng. Về cơ bản, trong các bộ truyền nhận OFDM, K_p ký hiệu pilot sẽ được chèn vào đoạn dữ liệu truyền đi và cả bên thu và phát đều biết trước các giá trị của các ký hiệu pilot này. Bên thu khai thác các ký hiệu pilot thu được để ước lượng ra sự ảnh hưởng của kênh truyền, từ đó nghịch đảo để khôi phục lại tín hiệu gốc.

Tuy nhiên, không có giải thuật nào có thể cho độ chính xác tuyệt đối trong việc nhận dạng kênh truyền vô tuyến thực. Do đó, các chuẩn truyền thông chỉ đưa ra phương pháp là tăng/giảm số lượng pilot khi kênh truyền ở các trạng thái khác nhau. Vậy nên, để so sánh hiệu suất làm việc của các giải thuật, CRB [21] có thể được sử dụng. Đây là phương pháp được sử dụng rộng rãi để ước lượng độ chính xác tối đa của các bộ nhận dạng không thiên vị (unbias). Biểu diễn của CRB như sau

$$\text{CRB}(\Theta) = \mathbf{J}_{\Theta\Theta}^{-1} \quad (2.8)$$

trong đó, $\mathbf{J}_{\Theta\Theta}$ là ma trận thông tin Fisher (FIM - Fisher information matrix) với Θ là các véc-tơ tham số không biết trước cần được ước lượng. Trong mô hình kênh phi cấu trúc, $\Theta \simeq \mathbf{h}$ [24], FIM chỉ phụ thuộc vào các ký hiệu pilot nên sẽ được ký hiệu là $\mathbf{J}_{\Theta\Theta}^p$. Từ đó, các tham số cần được ước lượng sẽ được biểu diễn như sau [36]²:

$$\Theta = [\mathbf{h}^\top, (\mathbf{h}^*)^\top] \quad (2.9)$$

Trong các hệ thống mMIMO-OFDM, K_p ký hiệu pilot sẽ được sắp xếp trong các ký hiệu OFDM [10] và do giả thiết tạp âm là một quá trình ngẫu nhiên i.i.d., FIM trong trường hợp OP thu được như sau

$$\mathbf{J}_{\Theta\Theta}^p = \sum_{i=1}^{K_p} \mathbf{J}_{\Theta\Theta}^{p_i} \quad (2.10)$$

với $\mathbf{J}_{\Theta\Theta}^{p_i}$ là FIM tương ứng với pilot thứ i [21] được cho bởi

$$\mathbf{J}_{\Theta\Theta}^{p_i} = \mathbb{E} \left\{ \left(\frac{\partial \ln p(\mathbf{x}(i), \mathbf{h})}{\partial \Theta^*} \right) \left(\frac{\partial \ln p(\mathbf{x}(i), \mathbf{h})}{\partial \Theta^*} \right)^H \right\} \quad (2.11)$$

trong đó \mathbb{E} là toán tử kỳ vọng; $p(\mathbf{x}(i), \mathbf{h})$ là hàm mật độ xác suất (Pdf - Probability density function) của tín hiệu nhận được đã biết \mathbf{h} . Phương trình (2.11) gồm các phép đạo hàm số phức, nên có thể biểu diễn dưới dạng

$$\mathbf{J}_{\Theta\Theta}^{p_i} = \frac{\mathbf{s}(i)^H \mathbf{s}(i)}{\sigma_w^2} \quad (2.12)$$

Tiếp đến, mô hình kênh có cấu trúc như trên phương trình (2.3) được xem xét.

²Công suất nhiễu được bỏ qua (σ_w^2) do lỗi ước lượng của tạp âm không ảnh hưởng đến \mathbf{h} .

Véc-tơ tham số có kích thước $4T \times M$ cần được ước lượng là

$$\Theta = [\beta^\top, (\beta^*)^\top, \theta^\top, \phi^\top] \quad (2.13)$$

với $\beta = [\beta_{0,0}, \dots, \beta_{M-1,T-1}]^\top$, $\beta^* = [\beta_{0,0}^*, \dots, \beta_{M-1,T-1}^*]^\top$, $\theta = [\theta_{0,0}, \dots, \theta_{M-1,T-1}]^\top$, và $\phi = [\phi_{0,0}, \dots, \phi_{M-1,T-1}]^\top$ lần lượt tương ứng là các véc-tơ có kích thước $T \times M$ của hệ số khuếch đại phức, liên hợp phức của hệ số khuếch đại phức, góc ngẩng, và góc phương vị của DoA. Dựa trên phép chuyển đổi của việc đạo hàm theo các tham số kể trên, FIM (\mathbf{J}_{hh}^p) của kênh truyền h trên (2.2) sẽ là

$$\mathbf{J}_{\text{hh}}^p = \frac{\partial \mathbf{h}}{\partial \Theta} \mathbf{J}_{\Theta\Theta}^p \frac{\partial \mathbf{h}^H}{\partial \Theta} \quad (2.14)$$

với

$$\frac{\partial \mathbf{h}}{\partial \Theta} = \left[\frac{\partial \mathbf{h}}{\partial \beta}, \frac{\partial \mathbf{h}}{\partial \beta^*}, \frac{\partial \mathbf{h}}{\partial \theta}, \frac{\partial \mathbf{h}}{\partial \phi} \right] \quad (2.15)$$

Chi tiết hơn, ví dụ đạo hàm riêng theo β trên phương trình (2.15) có dạng như sau

$$\frac{\partial \mathbf{h}}{\partial \beta} = \left[\mathbf{B}_0^\top, \mathbf{B}_1^\top, \dots, \mathbf{B}_{L-1}^\top \right]^\top \quad (2.16a)$$

$$\mathbf{B}_l = \text{diag}([\mathbf{B}_{l,0}, \mathbf{B}_{l,1}, \dots, \mathbf{B}_{l,T-1}]) \quad (2.16b)$$

$$\mathbf{B}_{l,t} = \left[\frac{\partial h_{l,t}}{\partial \beta_{0,t}}, \frac{\partial h_{l,t}}{\partial \beta_{1,t}}, \dots, \frac{\partial h_{l,t}}{\partial \beta_{M-1,t}} \right]^\top \quad (2.16c)$$

Các đạo hàm riêng của $h_{l,t}$ theo $\beta_{m,t}, \beta_{m,t}^*, \theta_{m,t}, \phi_{m,t}$ được biểu diễn chi tiết trên các phương trình (2.17).

$$\frac{\partial h_{l,t}}{\partial \beta_{m,t}} = \frac{1}{2}(1-i) \cdot e^{-ik_s c_l(\theta_{m,t}, \phi_{m,t})} \quad (2.17a)$$

$$\frac{\partial h_{l,t}}{\partial \beta_{m,t}^*} = \frac{1}{2}(1+i) \cdot e^{-ik_s c_l(\theta_{m,t}, \phi_{m,t})} \quad (2.17b)$$

$$\frac{\partial h_{l,t}}{\partial \theta_{m,t}} = \beta_{m,t} [-ik_s (\cos \theta_{m,t} \cos \phi_{m,t} x_l + \cos \theta_{m,t} \sin \phi_{m,t} y_l - \sin \theta_{m,t} z_l)] \cdot e^{-jk_s c_l(\theta_{m,t}, \phi_{m,t})} \quad (2.17c)$$

$$\frac{\partial h_{l,t}}{\partial \phi_{m,t}} = \beta_{m,t} [-ik_s (-\sin \theta_{m,t} \sin \phi_{m,t} x_l + \sin \theta_{m,t} \cos \phi_{m,t} y_l + \cos \theta_{m,t} z_l)] \cdot e^{-ik_s c_l(\theta_{m,t}, \phi_{m,t})} \quad (2.17d)$$

2.2.2. CRB trong trường hợp bán mù

Theo hướng tiếp cận SB, ngoài sử dụng các ký hiệu pilot, các bộ nhận dạng còn sử dụng thêm thông tin từ các ký hiệu dữ liệu (data) không biết trước trong việc ước

lượng kênh truyền. Trong phần này, giả thiết rằng các ký hiệu pilot và data trong ký hiệu OFDM là độc lập về mặt thống kê. Từ đó, FIM của phương pháp SB có thể được biểu diễn đơn giản là tổng của FIM từ các ký hiệu pilot và FIM từ các ký hiệu data như dưới đây

$$\mathbf{J}_{\Theta\Theta}^{SB} = \mathbf{J}_{\Theta\Theta}^p + \mathbf{J}_{\Theta\Theta}^d \quad (2.18)$$

với $\mathbf{J}_{\Theta\Theta}^d$ tương ứng là FIM của các ký hiệu data chưa biết trước và $\mathbf{J}_{\Theta\Theta}^p$ FIM của các ký hiệu pilot như đã được trình bày trên phương trình (2.10). Giả sử K_d ký hiệu data là i.i.d với trung bình thống kê là 0 và thông tin SOS là ma trận hiệp phương sai $\mathbf{C}_s = \text{diag}(\sigma_s^2)$. Trong đó, $\sigma_s^2 \stackrel{\text{def}}{=} [\sigma_{s_0}^2, \dots, \sigma_{s_{T-1}}^2]^\top$ với $\sigma_{s_t}^2$ là công suất truyền tại ăng-ten thứ t . Ma trận hiệp phương sai \mathbf{C}_x là:

$$\mathbf{C}_x = \sum_{t=0}^{T-1} \sigma_{s_t}^2 \boldsymbol{\lambda}_t \boldsymbol{\lambda}_t^H + \sigma_w^2 \mathbf{I}_{KL} \quad (2.19)$$

trong đó, \mathbf{I}_{KL} là ma trận đơn vị có kích thước KL và $\boldsymbol{\lambda}$ được định nghĩa là:

$$\begin{aligned} \boldsymbol{\lambda} &= [\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_{T-1}] \\ \boldsymbol{\lambda}_t &= [\boldsymbol{\lambda}_{0,t}, \boldsymbol{\lambda}_{1,t}, \dots, \boldsymbol{\lambda}_{L-1,t}]^\top \end{aligned} \quad (2.20)$$

với $\boldsymbol{\lambda}_{l,t} = \text{diag}(\mathcal{F}_0 h_{l,t})$ trong đó, \mathcal{F}_0 là cột đầu tiên của ma trận \mathcal{F} . FIM của các ký hiệu data có dạng như sau:

$$\mathbf{J}_{\Theta\Theta}^d = K_d \begin{bmatrix} \mathbf{J}_{hh}^d & \mathbf{J}_{hh^*}^d \\ \mathbf{J}_{h^*h}^d & \mathbf{J}_{h^*h^*}^d \end{bmatrix} \quad (2.21)$$

Theo [21], FIM $\mathbf{J}_{\Theta\Theta}^d$ của các ký hiệu data sẽ được biến đổi về dạng cuối cùng là

$$\mathbf{J}_{\Theta\Theta}^d = \text{tr} \left\{ \mathbf{C}_x^{-1} \frac{\partial \mathbf{C}_x}{\partial \mathbf{h}^*} \mathbf{C}_x^{-1} \left(\frac{\partial \mathbf{C}_x}{\partial \mathbf{h}^*} \right)^H \right\} \quad (2.22)$$

trong đó, $\frac{\partial \mathbf{C}_x}{\partial \mathbf{h}^*} = \boldsymbol{\lambda} \mathbf{C}_s \frac{\partial \boldsymbol{\lambda}^H}{\partial \mathbf{h}^*}$ và tr là toán tử tính tổng các thành phần trên đường chéo của một ma trận. Nếu sử dụng mô hình kênh phi cấu trúc, CRB của phương pháp SB sẽ là nghịch đảo của phương trình (2.18). Ngược lại, nếu sử dụng mô hình kênh có cấu trúc, CRB của phương pháp SB thu được bằng cách đưa FIM ở phương trình (2.18) qua phép chuyển đổi đã trình bày tại phương trình (2.14).

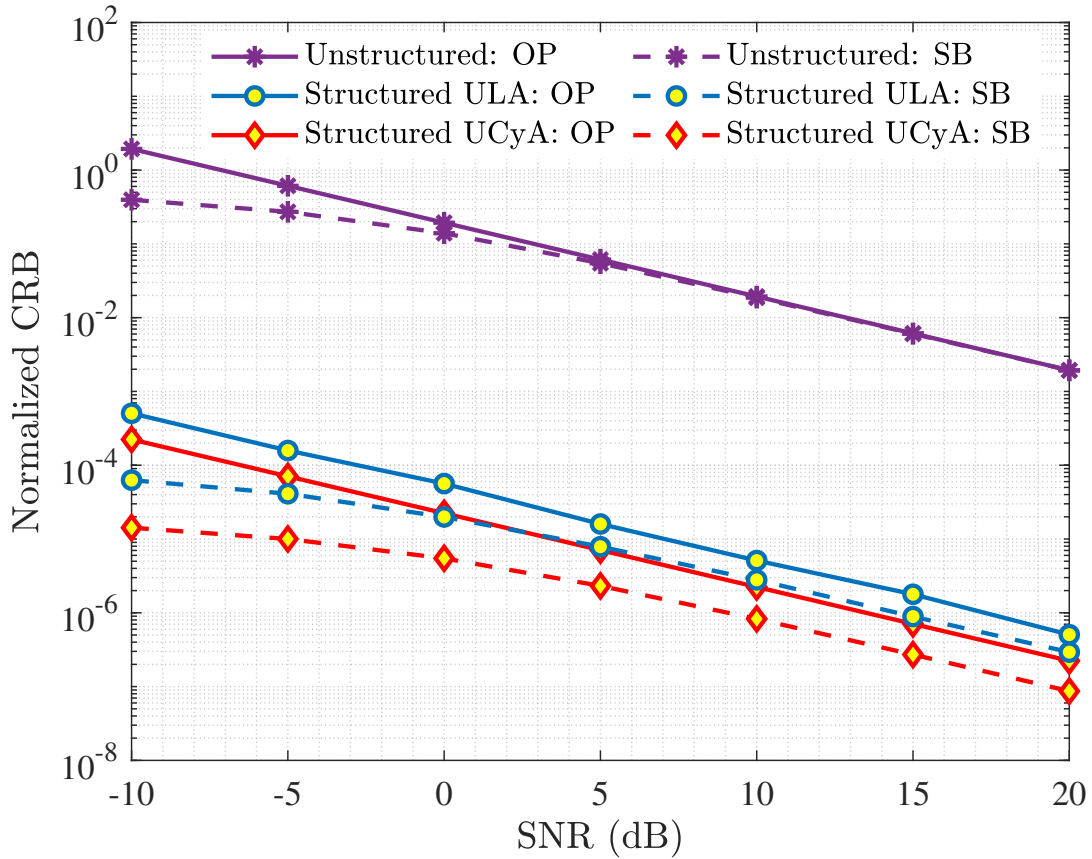
2.3. Mô phỏng và đánh giá

Để xem xét ảnh hưởng của cấu hình mảng ăng-ten trong hệ thống mMIMO, ba kịch bản mô phỏng sẽ được xem xét. Cụ thể, CRB của việc ước lượng kênh truyền khi: (i) SNR thay đổi; (ii) số lượng các lớp N_{3D} của cấu hình UCyA thay đổi; (iii) số lượng phần tử N_{UCA} của một mảng tròn UCA thay đổi. Các thông số mô phỏng của hệ thống

truyền thông MIMO kích thước lớn được sử dụng có tại bảng 2.1 [56]. Do các mô phỏng yêu cầu tài nguyên tính toán lớn (đặc biệt là dung lượng RAM), nên hệ mMIMO trong bảng 2.1 tuy có số lượng phần tử ăng-ten thu có thể lên đến 256 nhưng chưa đáp ứng được yêu cầu về số lượng người dùng tại một thời điểm như trong [25]. Kết quả mô phỏng được lấy trung bình của 1.000 lần chạy và các CRB được chuẩn hoá dưới dạng $\log_{10}(\text{CRB})$.

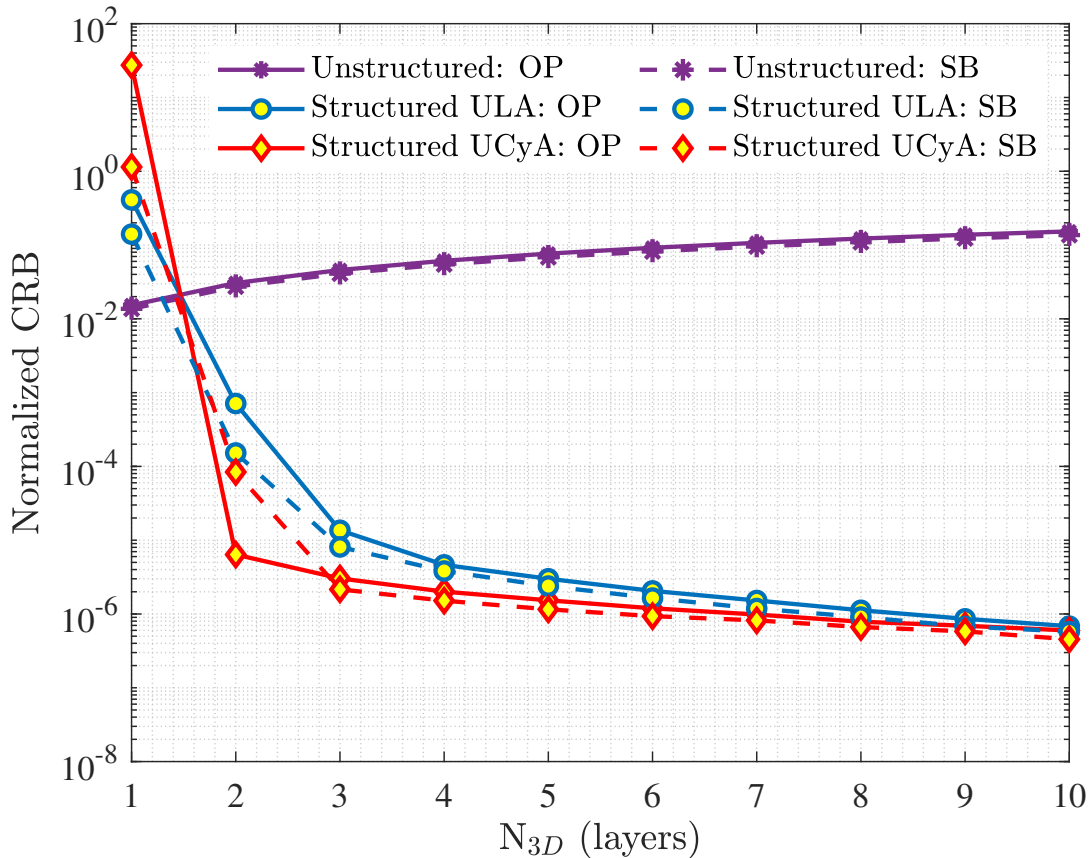
Bảng 2.1. Các tham số mô phỏng hệ thống truyền thông không dây để ước lượng CRB.

Thông số mô phỏng	Giá trị
Số ăng-ten phát	$T = 2$
Khoảng cách giữa các phần tử ăng-ten	$d_{2D} = d_{3D} = \lambda/2$
Số lượng các đường truyền	$M = 4$
Số sub-carriers	$K = 64$
Số ký hiệu pilot, data	$K_p = 16, K_d = 48$
Hệ số khuếch đại phức	$\beta \sim \mathcal{CN}(0, 1)$
Góc phương vị của DoA	$\phi^\circ \sim \mathcal{U}(-\pi/2, \pi/2)$
Góc ngẩng của DoA	$\theta^\circ \sim \mathcal{U}(-\pi/2, \pi/2)$



Hình 2.2. CRB của hai cấu hình ULA và UCyA ứng với mô hình kênh truyền có cấu trúc (structured) và phi cấu trúc (unstructured). Cấu hình của mảng ăng-ten như sau $N_{ULA} = 96$, $N_{UCA} = 24$, và $N_{3D} = 4$.

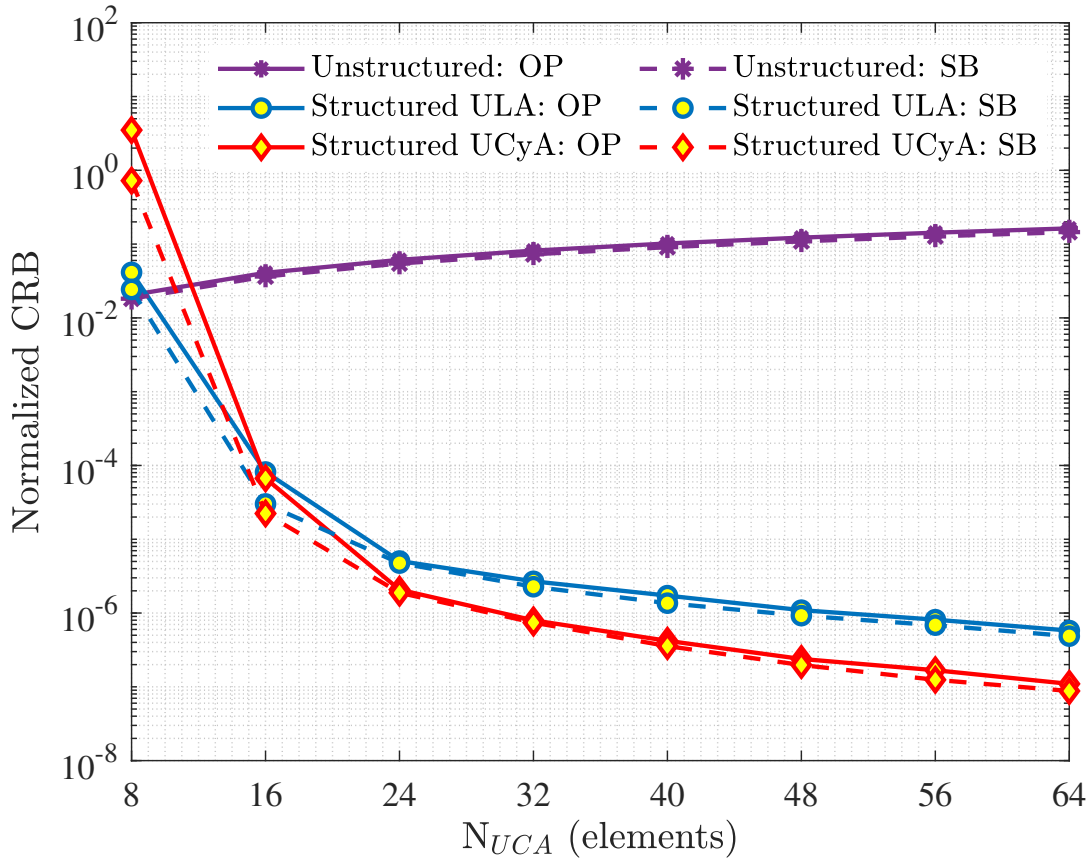
Trên hình 2.2, số lượng phần tử của mảng thu MIMO kích thước lớn là 96 trong đó $N_{ULA} = 96$, $N_{UCA} = 24$, và $N_{3D} = 4$. Nhìn chung, sai số ước lượng của mô hình kênh truyền có cấu trúc vượt trội khi so sánh với mô hình kênh truyền phi cấu trúc với độ lợi khoảng 10^3 dB. Khi so sánh hai phương pháp ước lượng SB và OP, ở các mức SNR thấp ($\text{SNR} \leq 5$ dB), phương pháp SB áp dụng cho mô hình kênh truyền phi cấu trúc có thể cho sai số thấp hơn một chút khi so sánh với việc chỉ sử dụng pilot của OP. Với mô hình kênh structured, phương pháp SB vẫn cho CRB tốt hơn ở các mức SNR thấp, và giữ ở mức ổn định khi $\text{SNR} \geq 5$ dB. Tiếp đến là so sánh về ảnh hưởng của cấu hình mảng ăng-ten, cấu hình UCyA cho độ chính xác cao hơn so với ULA ở cả hai phương pháp ước lượng OP và SB. Có thể rút ra nhận xét, việc sử dụng mô hình kênh truyền có cấu trúc, phương pháp ước lượng SB, và cấu hình mảng ăng-ten 3D (UCyA) có thể cho ra độ chính xác cao hơn cho các bộ nhận dạng của hệ thống mMIMO.



Hình 2.3. CRB của hai cấu hình ULA và UCyA khi thay đổi N_{3D} . Các thông số mô phỏng như sau $N_{UCA} = 24$, $N_{ULA} = 24 * N_{3D}$, và $\text{SNR} = 5$ dB.

Trên hình 2.3, số lượng các lớp N_{3D} của cấu hình UCyA được khảo sát bằng cách giữ nguyên số phần tử thuộc mảng vòng $N_{UCA} = 24$ và $\text{SNR} = 5$ dB. Một lần nữa, các CRB chỉ ra rằng với N_{3D} khác nhau, mô hình kênh truyền có cấu trúc hầu như có thể cho

ra độ chính xác cao hơn mô hình kênh truyền phi cấu trúc. Xét riêng mô hình kênh truyền unstructured, số N_{3D} tăng lên kéo theo sai số ước lượng cũng tăng lên dù không quá lớn. Ngay cả khi sử dụng phương pháp ước lượng SB, độ chính xác thu được với unstructured gần như không được cải thiện. Ngược lại, với mô hình kênh truyền structured, CRB có xu hướng đi xuống khi số lớp của mảng UCyA tăng cho đến khi tất cả hội tụ tại sai số khoảng 10^{-6} . Tại các giá trị N_{3D} nhỏ, trong khoảng từ 2 đến 6 lớp, việc sử dụng cấu hình UCyA nhìn chung vẫn cho hiệu quả đáng kể so với ULA. Khi xem xét sự ảnh hưởng của phương pháp OP hay SB trong kịch bản này, rõ ràng không có sự khác biệt quá rõ ràng nếu $N_{3D} \geq 3$. Có thể rút ra nhận xét thứ hai, sử dụng mô hình kênh truyền có cấu trúc và cấu hình mảng UCyA có thể cho hiệu suất ước lượng kênh truyền tốt hơn khi N_{3D} nhỏ. Tuy nhiên, lưu ý rằng, ngoài lợi thế về độ chính xác, khi số lượng phần tử trong mảng lên đến 240 như trong mô phỏng, cấu hình mảng UCyA sẽ giúp tiết kiệm được rất nhiều diện tích lắp đặt mảng ăng-ten.



Hình 2.4. CRB của hai cấu hình ULA và UCyA khi thay đổi N_{UCA} . Các thông số mô phỏng như sau $N_{3D} = 4$, $N_{ULA} = 4 * N_{UCA}$, và $SNR = 5$ dB.

Cuối cùng, trên hình 2.4, số lượng phần tử của mảng tròn UCA được thay đổi trong khoảng từ 8 đến 64 phần tử, trong khi giữ nguyên $N_{3D} = 4$, $N_{ULA} = 4 \times N_{UCA}$, và $SNR = 5$ dB. Do CRB của mô hình kênh truyền phi cấu trúc chỉ bị ảnh hưởng bởi số lượng

phần tử ăng-ten nên các CRB vẫn giữ như trên hình 2.3. Với mô hình kênh truyền có cấu trúc, CRB thay vì hội tụ tại một điểm sẽ có xu hướng giảm dần cùng với N_{UCA} . Khi số phần tử N_{UCA} đủ lớn ($N_{UCA} \geq 24$), các CRB của UCyA cho độ chính xác tốt hơn một cách tuyến tính khi so sánh với mảng ULA. Phương pháp SB trong kịch bản mô phỏng này nhìn chung chỉ cho độ chính xác tốt hơn nhưng không nhiều so với OP. Có thể rút ra nhận xét thứ ba, sử dụng mô hình kênh truyền có cấu trúc và cấu hình mảng UCyA có thể cho hiệu suất ước lượng kênh truyền tốt hơn và tăng dần với N_{UCA} . Tuy nhiên, cần lưu ý dù cho độ chính xác cao hơn, nhưng khi N_{UCA} tăng cao, bán kính của mảng tròn UCA lớn dần lại dẫn đến lãng phí diện tích lắp đặt mảng ăng-ten.

2.4. Kết luận chương

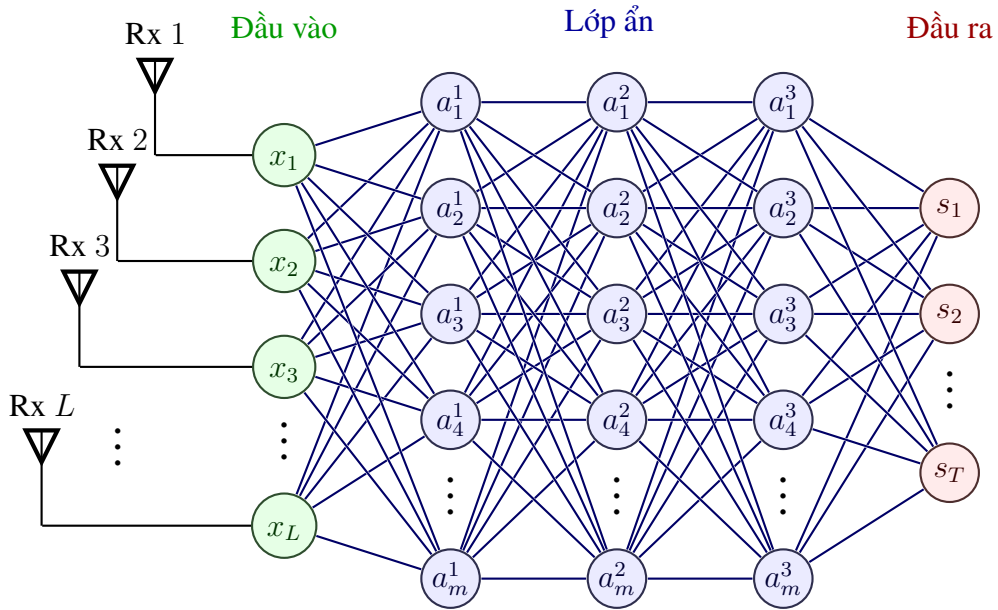
Trong chương này, Cramér Rao Bound đã được sử dụng để xem xét hiệu quả của các cấu hình mảng ăng-ten khác nhau đến sai số ước lượng kênh truyền trong các hệ thống mMIMO. Lý thuyết về CRB cho việc ước lượng kênh truyền được trình bày trong hai trường hợp, OP và SB. Các kết quả mô phỏng đã chỉ ra hiệu năng của việc ước lượng kênh truyền có thể được cải thiện rõ rệt nếu sử dụng mô hình kênh truyền có cấu trúc. Ngoài ra, cấu hình mảng ăng-ten UCyA và phương pháp SB cũng sẽ góp phần cải thiện độ chính xác khi so sánh với cấu hình ULA và phương pháp OP truyền thống.

CHƯƠNG 3

NHẬN DẠNG HỆ THỐNG SỬ DỤNG MẠNG HỌC SÂU

Trong chương này, tác giả đề xuất một kiến trúc mạng học sâu ISDNN sử dụng cho nhận dạng hệ thống viễn thông mMIMO. Đầu tiên, sơ lược về hai hướng tiếp cận sử dụng mạng nơ-ron sâu sẽ được giới thiệu. Kế đến là khái niệm về kỹ thuật mở rộng sâu (Deep unfolding). Một kiến trúc được mở rộng sâu từ bộ tối ưu MLE là DetNet được trình bày để so sánh ở mục các kết quả mô phỏng. Tiếp đến, từ một giải thuật ISD đã được đề xuất trong [34], kết hợp với cách tiếp cận mở rộng sâu tại [29], một mạng nơ-ron sâu tên ISDNN được đề xuất để nhận dạng hệ thống. Kiến trúc mạng ISDNN được đề xuất cho cả hai mô hình kênh truyền có cấu trúc và phi cấu trúc đã được trình bày tại chương 2. Các bước mô phỏng và đánh giá sẽ được đưa ra để cho thấy tiềm năng của phương pháp đề xuất và kết luận của chương.

3.1. Giới thiệu về mạng nơ-ron sâu và mở rộng sâu



Hình 3.1. Minh họa sử dụng DNN để nhận dạng hệ thống viễn thông.

Trong chương 1, các phương pháp nhận dạng hệ thống sử dụng các phương pháp ML/DL đã được chia làm ba loại, trong đó phương pháp sử dụng các mạng nơ-ron đang được quan tâm nghiên cứu. Các mạng nơ-ron sâu đã được sử dụng rộng rãi trong các ứng dụng như xử lý tiếng nói, ngôn ngữ tự nhiên, hình ảnh, thị giác máy, trò chơi trực tuyến [49]. Mười năm trở lại đây, đã có nhiều nghiên cứu ứng dụng các mạng DNN

khác nhau cho vấn đề nhận dạng hệ thống viễn thông không dây. Trên hình 3.1 là một mô hình minh họa việc sử dụng DNN để ước lượng kênh truyền và khôi phục tín hiệu gốc. Có thể chia các phương pháp này thành hai hướng tiếp cận, bao gồm hướng dữ liệu (data-driven) và hướng mô hình (model-driven) [29]. Các phương pháp data-driven trực tiếp học các đặc trưng từ một tập lớn các dữ liệu (dataset) để phục vụ cho các mục đích như ước lượng kênh truyền, phản hồi CSI, Tuy các phương pháp data-driven được đề xuất đều cho độ chính xác cao nhưng vẫn có những thách thức khi yêu cầu số lượng mẫu rất lớn và kéo theo đó là thời gian/chi phí cho việc đào tạo lớn. Các phương pháp model-driven [16] có thể một phần khắc phục các hạn chế này bằng việc tối ưu/đưa thêm các tham số học vào các mô hình có sẵn để kết hợp ưu điểm của data-driven và các mô hình toán học truyền thống.

Trong các năm gần đây, kỹ thuật mở rộng sâu [65] là một giải pháp tiềm năng để chuyển các giải thuật truyền thống thành các kiến trúc mạng DNN theo hướng tiếp cận model-driven. Chi tiết về mở rộng sâu tại [17], các phương pháp yêu cầu các vòng lặp đi lặp lại (iterative inference) có thể dễ dàng chuyển đổi sang các lớp của một mạng NN. Sau đó, sử dụng các giải thuật giảm dần độ dốc (GD - gradient descent) để đào tạo tham số trên các lớp mạng. Sau K lớp đào tạo tương tự như K vòng lặp trong thuật toán gốc, mô hình có thể đạt được mục tiêu mong muốn. Ví dụ, DetNet [50] là một mạng DNN dựa trên việc mở rộng sâu bộ nhận dạng MLE và sử dụng giảm dần độ dốc dự kiến (PGD - projected gradient descent) [6]. Trong mục tiếp theo, mạng nơ-ron sâu DetNet sẽ được giới thiệu ngắn gọn và kết quả của DetNet sẽ so sánh với kiến trúc mạng đề xuất.

3.2. Mạng nơ-ron sâu DetNet

Vẫn sử dụng mô hình hệ thống mMIMO đã trình bày ở hình 1.2.

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{w} \quad (3.1)$$

Các phần tử trong ma trận kênh truyền \mathbf{H} được biểu diễn dưới dạng số phức, tức đại diện cho cả ảnh hưởng về biên độ và pha gây ra bởi kênh. Cách biểu diễn này phù hợp trên lý thuyết, các phương pháp giải tối ưu, và các phần mềm mô phỏng như Matlab. Tuy nhiên, trong học máy/học sâu, các giá trị số phức thường sẽ được tách riêng rẽ thành phần thực (\Re) và ảo (\Im) riêng biệt. Điều này giúp quá trình học của thành phần thực và ảo thuộc một giá trị số phức cũng được tách riêng biệt, qua đó phát huy lợi thế của ML/DL. Hơn nữa các phép toán phổ biến trong xử lý tín hiệu như chuyển vị, liên hợp phức, hay nghịch đảo cũng sẽ được thực hiện dễ dàng hơn nếu tách riêng thành phần thực và ảo. Từ các lý do kể trên, các phương pháp sử dụng ML/DL nhằm nhận dạng kênh sẽ chuyển đổi tất

cả các giá trị phức sang dạng thực trước khi đưa vào quá trình đào tạo trong mạng. Các mạng này thường được phát triển trên ngôn ngữ Python và các thư viện nền tảng thông dụng như Tensorflow¹ của Google hay Pytorch² của Facebook. Do vậy, không làm mất đi tính tổng quát, trong chương này, các ma trận và véc-tơ trên mô hình kênh kể trên sẽ được biểu diễn dưới dạng các thành phần thực (\Re) và ảo (\Im) như sau

$$\mathbf{s} = \begin{bmatrix} \Re(\mathbf{s}) \\ \Im(\mathbf{s}) \end{bmatrix}; \mathbf{x} = \begin{bmatrix} \Re(\mathbf{x}) \\ \Im(\mathbf{x}) \end{bmatrix}; \mathbf{w} = \begin{bmatrix} \Re(\mathbf{w}) \\ \Im(\mathbf{w}) \end{bmatrix} \quad (3.2)$$

$$\mathbf{H} = \begin{bmatrix} \Re(\mathbf{H}) & -\Im(\mathbf{H}) \\ \Im(\mathbf{H}) & \Re(\mathbf{H}) \end{bmatrix} \quad (3.3)$$

trong đó, $\mathbf{H} \in \mathbb{R}^{2L \times 2T}$, $\mathbf{s} \in \mathbb{R}^{2T \times 1}$, $\mathbf{w} \in \mathbb{R}^{2L \times 1}$, và $\mathbf{x} \in \mathbb{R}^{2L \times 1}$. Thông thường, các kiến trúc mạng ML/DL sử dụng để ước lượng kênh truyền giả sử rằng ma trận kênh truyền \mathbf{H} được mô hình hoá dưới dạng phi cấu trúc (unstructured), hay các hệ số của \mathbf{H} được chọn ngẫu nhiên và không bị ảnh hưởng bởi các thông số vật lý khác như DoA, cấu hình mảng ăng-ten, Để tìm bộ nhận dạng cho hệ thống kể trên, định nghĩa hàm mất mát $\mathcal{L}(\mathbf{s}; \hat{\mathbf{s}}_{\Theta}(\mathbf{H}, \mathbf{x}))$ là khoảng cách giữa véc-tơ ký hiệu gốc và ký hiệu được ước lượng. Tìm giá trị Θ bằng cách tối thiểu hoá hàm mất mát kể trên.

$$\min_{\Theta} \mathbb{E} \{ \mathcal{L}(\mathbf{s}; \hat{\mathbf{s}}_{\Theta}(\mathbf{H}, \mathbf{x})) \} \quad (3.4)$$

Giải thuật tối ưu để giải quyết (3.4) là MLE như sau

$$\hat{\mathbf{s}}_{\Theta}(\mathbf{H}, \mathbf{x}) = \arg \min_{\mathbf{s} \in \mathbb{R}^{2T}} \|\mathbf{x} - \mathbf{H}\mathbf{s}\|^2 \quad (3.5)$$

Tuy nhiên, độ phức tạp của MLE sẽ tăng theo cấp số mũ $\mathcal{O}(2^T)$ nên khó để triển khai trong các hệ mMIMO. Do vậy, DetNet được đề xuất nhằm tạo ra một kiến trúc mạng DNN tiệm cận độ chính xác với MLE. Trong nghiên cứu gốc, thay vì tạo ra một mạng nơ-ron nhằm ánh xạ trực tiếp từ \mathbf{x} về \mathbf{s} , việc phân tách \mathbf{x} thành các thành phần \mathbf{H} , \mathbf{s} , và \mathbf{w} sẽ cho hiệu quả cao hơn.

$$\mathbf{H}^{\top} \mathbf{x} = \mathbf{H}^{\top} \mathbf{H} \mathbf{s} + \mathbf{H}^{\top} \mathbf{w} \quad (3.6)$$

Kiến trúc DetNet dựa trên phương pháp PGD [6] cho việc tối ưu MLE như trên phương trình (3.5). Đạo hàm riêng được tách như trên (3.7) sử dụng luật chuỗi (chain

¹<https://github.com/tensorflow/tensorflow>

²<https://github.com/pytorch/pytorch>

rule) [37].

$$\begin{aligned}\hat{\mathbf{s}}_{k+1} &= \Pi \left[\hat{\mathbf{s}}_k - \delta_k \frac{\partial \|\mathbf{x} - \mathbf{H}\mathbf{s}\|^2}{\partial \mathbf{s}} \Big|_{\mathbf{s}=\hat{\mathbf{s}}_k} \right] \\ &= \Pi \left[\hat{\mathbf{s}}_k - \delta_k \mathbf{H}^\top \mathbf{x} + \delta_k \mathbf{H}^\top \mathbf{H} \hat{\mathbf{s}}_k \right]\end{aligned}\quad (3.7)$$

với $\hat{\mathbf{s}}_k$ là véc-tơ giá trị nguồn ước lượng tại lớp thứ k , $\Pi[\cdot]$ là một phép biến đổi phi tuyến tính, và δ_k là độ dài bước (step size) của quá trình học. Kiến trúc của một lớp mạng DetNet đề xuất trong [50] được minh hoạ như trên hình 3.2 và cách biểu diễn dưới dạng ma trận như sau

$$\mathbf{z}_k = \rho \left(\mathbf{W}_k^1 \begin{bmatrix} \mathbf{H}^\top \mathbf{x} \\ \hat{\mathbf{s}}_k \\ \mathbf{H}^\top \mathbf{H} \hat{\mathbf{s}}_k \\ \mathbf{v}_k \end{bmatrix} + \mathbf{b}_k^1 \right) \quad (3.8a)$$

$$\hat{\mathbf{s}}_{k+1} = \psi_{t_k} (\mathbf{W}_k^2 \mathbf{z}_k + \mathbf{b}_k^2) \quad (3.8b)$$

$$\hat{\mathbf{v}}_{k+1} = \mathbf{W}_k^3 \mathbf{z}_k + \mathbf{b}_k^3 \quad (3.8c)$$

$$\hat{\mathbf{s}}_1 = \mathbf{0} \quad (3.8d)$$

trong đó, $k = 1, \dots, K$ là số các lớp của mạng DetNet, ρ là một toán tử tuyến tính. ψ_{t_k} ký hiệu cho phép biến đổi phi tuyến tính phân đoạn, ở các mức t khác nhau, $\psi_{t_k}(s)$ được minh hoạ trên hình 3.3 và có biểu diễn toán học như sau

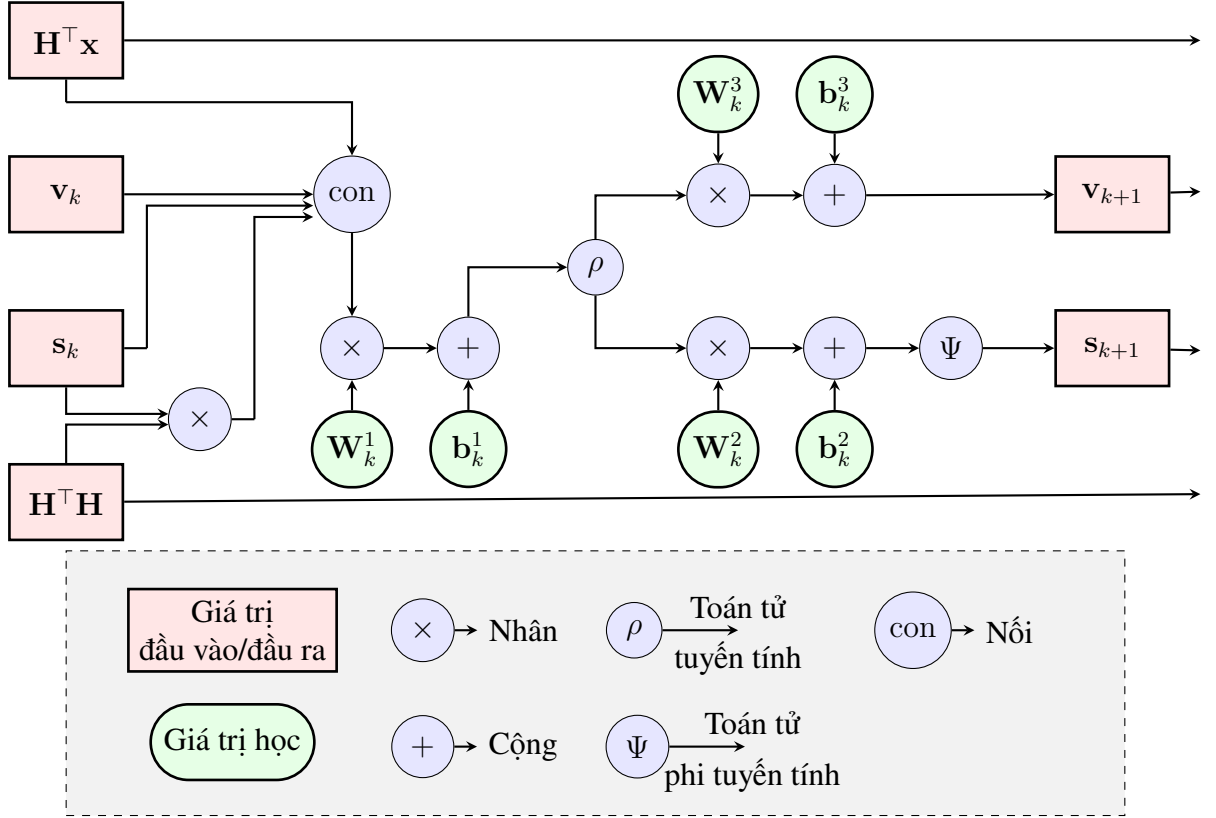
$$\psi_{t_k}(s) = -1 + \frac{\rho(s + t_k)}{|t_k|} - \frac{\rho(s - t_k)}{|t_k|} \quad (3.9)$$

Các tham số của việc học sẽ bao gồm

$$\Theta = \{\mathbf{W}_k^1, \mathbf{b}_k^1, \mathbf{W}_k^2, \mathbf{b}_k^2, \mathbf{W}_k^3, \mathbf{b}_k^3, t_k\}_{k=1}^K \quad (3.10)$$

Sau mỗi iteration, một hàm mất mát có dạng sai số toàn phương trung bình (MSE - Mean Squared Error) sẽ tính toán sự sai khác của véc-tơ các ký hiệu ước lượng so với các ký hiệu gốc. Kết quả này được sử dụng để xem xét sự hội tụ của việc đào tạo, và được trả về cho giải thuật tối ưu (ví dụ: Adam) để tối ưu các tham số học của mạng. Hàm mất mát này được định nghĩa như dưới đây

$$\mathcal{L}(\mathbf{s}; \hat{\mathbf{s}}_\Theta(\mathbf{H}, \mathbf{x})) = \frac{1}{2T} \sum_{t=1}^{2T} \|s_t - \hat{s}_t\|^2 \quad (3.11)$$



Hình 3.2. Kiến trúc của một lớp trong kiến trúc mạng DetNet [50].

3.3. Đề xuất mạng nơ-ron sâu ISDNN cho nhận dạng kênh truyền

Trong phần này, giải thuật của bộ nhận dạng ISD công bố tại [34] sẽ được trình bày. Từ đó, một mạng nơ-ron sâu ISDNN với mô hình kênh truyền có và phi cấu trúc được đề xuất dựa trên kỹ thuật mở rộng sâu cho giải thuật ISD trước đó.

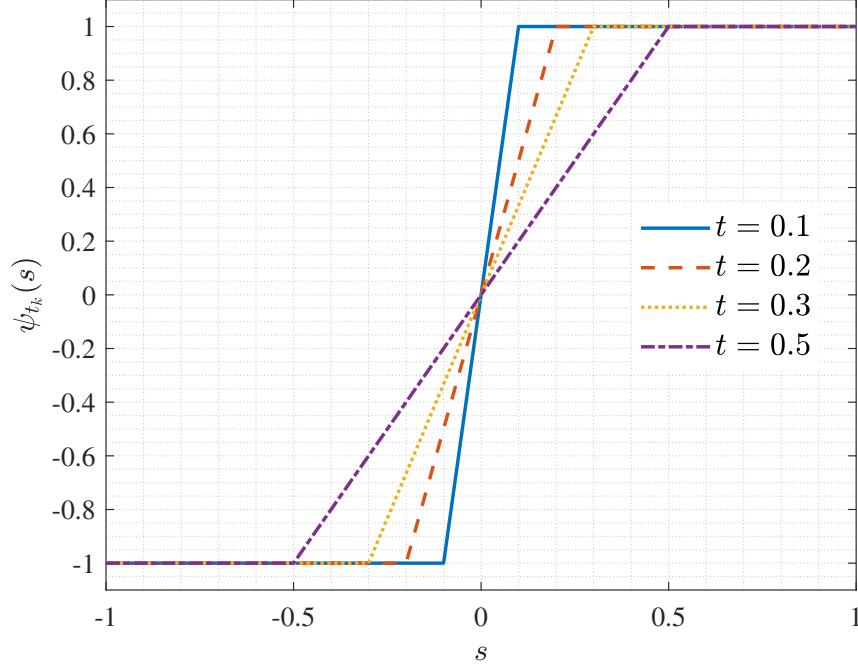
3.3.1. Bộ nhận dạng ISD cho hệ thống mMIMO

Giải thuật gốc tại [34] đã đề xuất một bộ nhận dạng kênh truyền tuần tự lặp lại gọi tắt là ISD để đạt được hiệu suất của MMSE với độ phức tạp thấp. Trong đó, bộ nhận dạng MMSE đã được chứng minh [47] có thể đạt được độ chính xác tiệm cận của MLE cho kênh đường lên cho các hệ mMIMO với $L/T \geq 10$. Lưu ý rằng, do \mathbf{H} là ma trận của các số thực nên phép chuyển vị liên hợp phức $(.)^H$ sẽ được chuyển thành chuyển vị $(.)^T$.

$$\hat{\mathbf{s}}_{MMSE} = \left(\mathbf{H}^T \mathbf{H} + \frac{\sigma^2}{\mathbb{E}(\mathbf{s})} \mathbf{I}_{2T} \right)^{-1} \mathbf{H}^T \mathbf{x} = \mathbf{P}^{-1} \mathbf{q} \quad (3.12)$$

ký hiệu $\mathbf{G}_H = \mathbf{H}^T \mathbf{H}$, $\mathbf{P} = \mathbf{H}^T \mathbf{H} + \frac{\sigma^2}{\mathbb{E}(\mathbf{s})} \mathbf{I}_{2T}$, và $\mathbf{q} = \mathbf{H}^T \mathbf{x}$. Các thành phần đường chéo (diagonal component) của ma trận \mathbf{P} tạo thành ma trận $\mathbf{D} = \text{diag}(\mathbf{P})$. Lưu ý, độ phức tạp của việc nghịch đảo \mathbf{P} là $\mathcal{O}(TL^3)$, sẽ tăng nhanh khi L lớn.

Để đạt được hiệu năng cao hơn với ít số lần lặp lại, [34] đề xuất khởi tạo véc-tơ của



Hình 3.3. Hàm phi tuyến tính phân đoạn $\psi_{t_k}(s)$ được sử dụng trong DetNet.

các ký hiệu ước lượng (\hat{s}_1) như trên phương trình (3.13) [9] thay vì đặt tất cả bằng 0.

$$\hat{s}_1 = \mathbf{D}^{-1}\mathbf{q} = [s_1(1), s_1(2), \dots, s_1(2T)] \quad (3.13)$$

Từ véc-tơ tín hiệu thu, tín hiệu của ăng-ten/người dùng thứ j thu được bằng cách loại bỏ tạp âm từ các ăng-ten/người dùng khác.

$$\hat{\mathbf{x}}(j) = \mathbf{x} - \sum_{t=1, t \neq j}^{2T} \mathbf{h}_t \hat{s}_k(t) \quad (3.14)$$

với $\hat{\mathbf{x}}(j)$ thu được, ký hiệu được gửi từ người dùng thứ j được ước lượng như sau

$$\begin{aligned} \hat{s}_{k+1}(j) &= \frac{\mathbf{h}_j^\top}{\|\mathbf{h}_j\|^2} \hat{\mathbf{x}}(j) \\ &= \hat{s}_k(j) + \frac{1}{\mathbf{G}_{\mathbf{H}}(j, j)} \left(\mathbf{q}(j) - \sum_{t=1}^{2T} \mathbf{G}_{\mathbf{H}}(j, t) s_k(t) \right) \end{aligned} \quad (3.15)$$

trong đó, \mathbf{h}_j là cột thứ j của ma trận \mathbf{H} , $\mathbf{G}_{\mathbf{H}}(i, j)$ là phần tử thứ (i, j) của ma trận $\mathbf{G}_{\mathbf{H}}$, và $\mathbf{q}(j)$ là phần tử thứ j của véc-tơ \mathbf{q} . Véc-tơ các ký hiệu ước lượng \hat{s} được cập nhật như trong thuật toán 3.1 của giải thuật ISD [34].

Để chứng minh giải thuật ISD là hiệu quả cho việc ước lượng kênh truyền, véc-tơ phần dư (sai số) sẽ được sử dụng. Cụ thể, véc-tơ phần dư thu được sau khi khởi tạo với

Thuật toán 3.1 Bộ nhận dạng Iterative Sequential [34].

Input: $\mathbf{x}, \mathbf{H}, L, T, K, \sigma^2, \mathbb{E}(\mathbf{s})$

Output: $\hat{\mathbf{s}}_{out} = \hat{\mathbf{s}}_K^{2T}$

```

1:  $\mathbf{G}_H \leftarrow \mathbf{H}^\top \mathbf{H}$ 
2:  $\mathbf{A} \leftarrow \mathbf{G}_H + \frac{\sigma^2}{\mathbb{E}_x} \mathbf{I}_{2T}$ 
3:  $\mathbf{s}_0 \leftarrow \mathbf{s}_{in} = \mathbf{D}^{-1} \mathbf{q}$ 
4:
5: for  $k = 0, k < K$  do
6:   for  $j = 1, j \leq 2T$  do
7:      $\hat{s}_k(j+1) \leftarrow \hat{s}_k(j) + \frac{1}{\mathbf{G}_H(j,j)} \left( \mathbf{q}(j) - \sum_{t=1}^{2T} \mathbf{G}_H(j,t) \hat{s}_k(t) \right)$ 
8:
9:      $\hat{\mathbf{s}}_{k+1}^j \leftarrow [\hat{s}_{k+1}(1), \dots, \hat{s}_{k+1}(j), \hat{s}_k(j+1), \dots, \hat{s}_k(2T)]$ 
10:     $j \leftarrow j + 1$ 
11:   end for
12:    $k \leftarrow k + 1$ 
13: end for

```

các giá trị $\hat{\mathbf{s}}_1$ là

$$\mathbf{e}_1 = \mathbf{x} - \mathbf{H}\hat{\mathbf{s}}_1 \quad (3.16)$$

từ đó, véc-tơ phần dư sau khi cập nhật ký hiệu thứ j tại lớp thứ k sẽ được biểu diễn như sau

$$\mathbf{e}_k^j = \mathbf{x} - \mathbf{H}\hat{\mathbf{s}}_k^j \quad (3.17)$$

thay $\hat{\mathbf{s}}_k^j$ bằng các biểu diễn hồi quy như trong giải thuật 3.1 thu được

$$\begin{aligned}
\mathbf{e}_k^j &= \mathbf{x} - \mathbf{h}_j \left(\hat{s}_k(j-1) + \frac{1}{\mathbf{G}_H(j-1, j-1)} \left(\mathbf{q}(j-1) - \sum_{t=1}^{2T} \mathbf{G}_H(j-1, t) \hat{s}_k(t) \right) \right) \\
&= \mathbf{x} - \mathbf{h}_j \left(\hat{s}_k(j-1) + \frac{1}{\mathbf{h}_{j-1}^\top \mathbf{h}_{j-1}} \left(\mathbf{h}_{j-1}^\top \mathbf{x} - \sum_{t=1}^{2T} \mathbf{G}_H(j-1, t) \hat{s}_k(t) \right) \right) \\
&= \mathbf{e}_k^{j-1} - \mathbf{h}_j \frac{\mathbf{h}_j^\top}{\|\mathbf{h}_j\|^2} \mathbf{e}_k^{j-1}
\end{aligned} \quad (3.18)$$

Trong [34], Mandloi M. và các cộng sự đã chứng minh rằng $\|\mathbf{e}_k^j\|^2 < \|\mathbf{e}_k^{j-1}\|^2$. Điều đó chỉ ra rằng mỗi khi ký hiệu thứ j được cập nhật, véc-tơ phần dư sẽ được chiếu lên mặt phẳng ‘null’ của cột thứ j thuộc ma trận \mathbf{H} . Hay véc-tơ phần dư sẽ trực giao với \mathbf{h}_j , do đó l_2 – norm bình phương của véc-tơ lỗi sẽ giảm sau mỗi lần ký hiệu j được cập nhật cho đến khi véc-tơ \mathbf{e} trực giao với không gian con kéo dài bởi cột của ma trận \mathbf{H} .

3.3.2. Đề xuất mạng nơ-ron sâu ISDNN cho mô hình kênh truyền phi cấu trúc

Từ giải thuật ISD được trình bày ở trên, theo hướng tiếp cận model-driven và kỹ thuật deep unfolding, một kiến trúc mạng nơ-ron sâu có tên ISDNN (Iterative sequential

deep-neural network) tương ứng cho mô hình kênh truyền phi cấu trúc (gọi tắt là **ISDNN phi cấu trúc**) được đề xuất. Đầu tiên, việc cập nhật các ký hiệu s tại dòng 7 của giải thuật 3.1 được viết lại dưới dạng ma trận như sau

$$\hat{s}_{k+1} = \hat{s}_k + e_{k+1} \quad (3.19)$$

trong đó, e_{k+1} là véc-tơ phần dư cũng được viết dưới dạng ma trận là

$$e_{k+1} = D^{-1} \left(H^T x - H^T H \hat{s}_k \right) \quad (3.20)$$

với D được đơn giản hoá lấy ý tưởng từ bộ nhận dạng ZF khi không có thông tin về SNR tại bên thu, tức nghịch đảo của ma trận Gram G_H , $D = \text{diag}(H^T H)$. Nhận thấy rằng, \hat{s}_{k+1} không chỉ chịu ảnh hưởng trực tiếp bởi e_{k+1} mà còn tất cả các véc-tơ phần dư trước đó e_k, e_{k-1}, \dots, e_1 như biểu diễn ở công thức (3.18). Do vậy, để đạt được hiệu quả cao hơn trong việc loại bỏ tạp âm từ các người dùng khác, các tham số học được thêm vào α^1 vào mỗi lớp (layer) của mạng nơ-ron.

$$\hat{s}_{k+1} = \hat{s}_k + e_{k+1} + \alpha_k^1 e_k + \alpha_{k-1}^1 e_{k-1} + \dots + \alpha_1^1 e_1 \quad (3.21)$$

Tuy nhiên, do mối tương quan giữa các véc-tơ phần dư liên kế là lớn nhất, nên trong mạng ISDNN chỉ xem xét ảnh hưởng của e_k ở lớp thứ k nhằm đơn giản hoá kiến trúc mạng. Phương trình (3.21) trở thành

$$\mu_k = \hat{s}_k + e_{k+1} + \alpha_k^1 e_k \quad (3.22)$$

Thay vì gán trực tiếp $\hat{s}_{k+1} = \mu_k$, sự tương quan giữa μ_k và \hat{s}_k được xem xét trước khi đưa làm đầu vào của lớp tiếp theo. Sử dụng kết hợp lồi (convex combination) [14] của \hat{s}_k và μ_k với hệ số α^2 . Do đó, \hat{s}_{k+1} chịu ảnh hưởng bởi cả \hat{s}_k và μ_k theo tỷ lệ α^2 . Trong đó, α_k^2 là tham số có thể học, $\sum_{i=k}^{k+1} \alpha_i^2 \hat{s}_i$ với $\sum_{i=k}^{k+1} \alpha_i^2 = 1$, tại mỗi lớp. Kết hợp tuyến tính của \hat{s}_k và μ_k có dạng như sau

$$\hat{s}_{k+1} = (1 - \alpha_k^2) \mu_k + \alpha_k^2 \hat{s}_k \quad (3.23)$$

Ngoài ra, để đạt được độ chính xác cao hơn ở các loại điều chế bậc cao như (16-QAM, 64-QAM, ...), véc-tơ phần dư sẽ được điều chỉnh linh hoạt hơn bằng cách thêm hai bộ biến đổi tuyến tính vào kiến trúc mạng ISDNN để cập nhật e_k trước khi nhân với

α_k^1 .

$$\mathbf{e}_k \leftarrow w_k^2 (w_k^1 \mathbf{e}_k + b_k^1) + b_k^2 \quad (3.24)$$

Bộ nhận dạng trong các các nghiên cứu [29, 34, 51] tuân tự đưa các dữ liệu huấn luyện chỉ ứng với một lần tạo dữ liệu độc lập qua mạng, tức $\mathbf{H} \in \mathbb{R}^{2L \times 2T}$, $\mathbf{s} \in \mathbb{R}^{2T \times 1}$, và $\mathbf{x} \in \mathbb{R}^{2L \times 1}$. Điều này làm giảm đáng kể tốc độ học của một mạng DNN do không tận dụng được lợi thế của kiểu dữ liệu ten-sơ (tensor) cho phép thao tác với các biến đa chiều. Do đó, điểm cải tiến tiếp theo của kiến trúc mạng ISDNN đó là sử dụng ‘Batch size’ lớn hơn rất nhiều. Với ‘bs’ (Batch size) là lượng dữ liệu được sử dụng trong một vòng lặp. Các dữ liệu đầu vào của mạng ISDNN sẽ được chuyển sang dạng ten-sơ 3 chiều như sau

$$\mathbf{E}_k \leftarrow [\mathbf{e}_k^1, \mathbf{e}_k^2, \dots, \mathbf{e}_k^{bs}] \quad (3.25a)$$

$$\mathbf{S}_k \leftarrow [\mathbf{s}_k^1, \mathbf{s}_k^2, \dots, \mathbf{s}_k^{bs}] \quad (3.25b)$$

$$\mathbf{X}_k \leftarrow [\mathbf{x}_k^1, \mathbf{x}_k^2, \dots, \mathbf{x}_k^{bs}] \quad (3.25c)$$

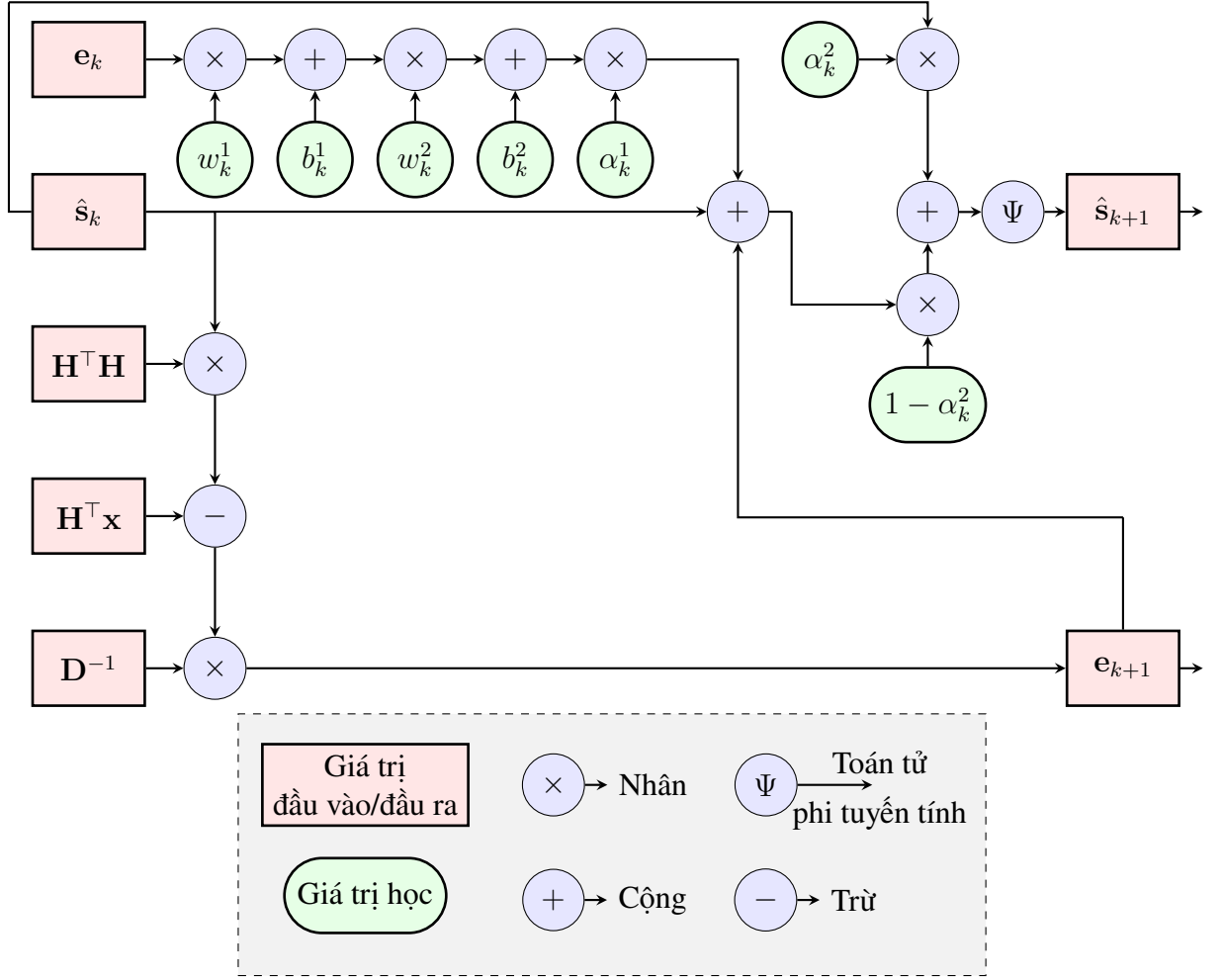
$$\mathbf{H} \leftarrow [\mathbf{H}^1, \mathbf{H}^2, \dots, \mathbf{H}^{bs}] \quad (3.25d)$$

Các phép toán biến đổi ten-sơ, ví dụ, phép nhân các ma trận được chuyển đổi thành nhân các ten-sơ 3 chiều được thực hiện bằng hàm ‘*torch.bmm*’³ của thư viện nền tảng Pytorch. Tuy nhiên, do đây là đề xuất về mặt kĩ thuật lập trình nên trong các phần tiếp theo, các kí hiệu vẫn giữ ở dạng nhân ma trận tương tự như $bs = 1$ để giảm sự nhầm lẫn.

Kiến trúc cuối cùng của mạng ISDNN cho mô hình kênh truyền phi cấu trúc được đề xuất trong luận văn như trên hình 3.4. So với giải thuật ISD được đề xuất trước đó, mạng nơ-ron sâu ISDNN được đề xuất có sự cải tiến bằng việc: (i) thêm véc-tơ phần dư của lớp trước đó và tham số học α^1 để ước lượng $\hat{\mathbf{s}}$; (ii) tham số học α^2 được thêm vào để tăng tính chính xác của việc học; (iii) véc-tơ phần dư được đưa qua hai bộ biến đổi để có được tính linh hoạt cho các loại điều chế bậc cao; (iv) sử dụng Batch size lớn để giảm thời gian học.

Các tham số khởi tạo của mạng ISDNN được đề xuất như sau để nhanh chóng đạt được sự hội tụ [38]: $\hat{\mathbf{s}}_1 = \mathbf{D}^{-1} \mathbf{q}$; α_1^1 được chọn ngẫu nhiên theo phân bố đều $\alpha_1^1 \in \mathcal{U}[0 \ 1]$; $\alpha_1^2 = 0,5$; véc-tơ \mathbf{e}_1 được chọn lựa ngẫu nhiên theo phân bố đều $\mathbf{e}_1 \in \mathcal{U}[0 \ 1]$. Do các đầu vào cho lớp tiếp theo $\hat{\mathbf{s}}_{k+1}$ cần được ánh xạ về khoảng giá trị $[-1.0 \ 1.0]$, một hàm kích hoạt (activation function) sẽ được sử dụng. Trong DL, có nhiều hàm kích hoạt được sử dụng rộng rãi như ReLu, Tanh, Sigmoid, ... như được biểu diễn trên hình 3.5. Cụ thể,

³<https://pytorch.org/docs/stable/generated/torch.bmm.html>



Hình 3.4. Kiến trúc của một lớp trong mạng nơ-ron sâu ISDNN đề xuất cho mô hình kênh truyền phi cấu trúc.

mạng ISDNN lựa chọn sử dụng hàm Tanh có biểu diễn toán học như sau

$$\Psi(s) = \text{Tanh}(s) = \frac{e^s - e^{-s}}{e^s + e^{-s}} \quad (3.26)$$

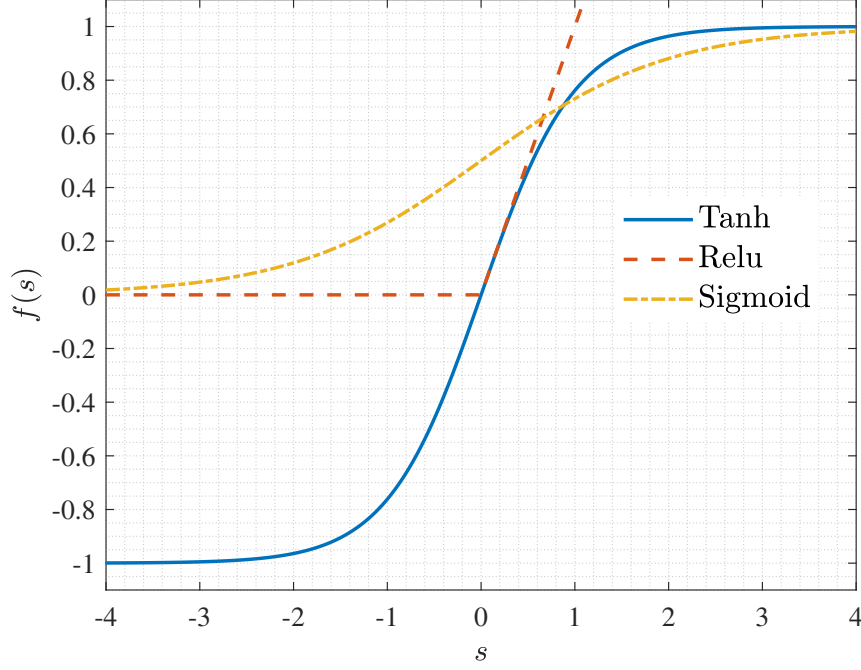
Các tham số của việc học sẽ bao gồm

$$\Theta = \{w_k^1, b_k^1, w_k^2, b_k^2, \alpha_k^1, \alpha_k^2\}_{k=1}^K \quad (3.27)$$

Một hàm mất mát MSE cũng được định nghĩa như trên phương trình (3.11) của DetNet để biểu diễn sự hội tụ của mạng học sâu ISDNN và bước back-propagate [27] của một mạng NN.

Bốn bước của một vòng lặp (iteration) trong quá trình học như sau

1. Khởi tạo các tham số ban đầu và véc-tơ phần dư của mạng ISDNN: $s_0, e_0, \alpha_0^1, \alpha_0^2$.



Hình 3.5. Minh hoạ một số hàm kích hoạt được dùng trong kiến trúc đề xuất.

2. Bộ dữ liệu được đưa qua K lớp của mạng (forward propagation), sau đó ước lượng sự mất mát qua hàm $\mathcal{L}(s; \hat{s}_{\Theta}(\mathbf{H}, \mathbf{x}))$.
3. Back-propagate $\mathcal{L}(s; \hat{s}_{\Theta}(\mathbf{H}, \mathbf{x}))$ để thu được độ dốc (gradient).
4. Từ gradient thu được, sử dụng một thuật toán tối ưu, ví dụ là Adam [22] (xem thêm tại phụ lục A), cập nhật các tham số học $\Theta = \{w_k^1, b_k^1, w_k^2, b_k^2, \alpha_k^1, \alpha_k^2\}_{k=1}^K$.

3.3.3. Đề xuất mạng nơ-ron sâu ISDNN cho mô hình kênh truyền có cấu trúc

Theo hướng tiếp cận khi mô hình kênh truyền theo dạng có cấu trúc như đã trình bày tại chương 2, một phần thông tin bên lề gồm DoA và cấu hình mảng ăng-ten tại bên thu được đề xuất sử dụng cho việc học của mạng ISDNN kênh phi cấu trúc tại mục 3.3.2. Lý do chọn thông tin DoA là do trong các hệ mMIMO, kỹ thuật định hướng bức sóng (beamforming) [52] có vai trò đặc biệt quan trọng trong việc tăng công suất truyền và giảm tỷ lệ tạp âm liên người dùng. Trước khi kỹ thuật beamforming có thể được sử dụng, việc biết hướng cần phát, tức hướng của người dùng (UE - User equipment) là điều kiện cần có. Góc phát này được ước lượng thông qua tín hiệu từ các phiên truyền đường lên trước đó. Các thuật toán phổ biến được sử dụng để ước lượng hướng sóng đến như phương pháp trễ và cộng (CBF - Conventional beamforming method), Capon, hay MUSIC [63].

$$h_{l,t} = \beta_{l,t} e^{j\varphi_{l,t}} = \beta_{l,t} e^{-jk_s c_l(\theta_{l,t}, \phi_{l,t})} \quad (3.28)$$

Với giả thiết rằng DoA, tức các véc-tơ θ, ϕ , của các UE đã được trạm cơ sở ước lượng và khả dụng trước khi ước lượng kênh truyền, kiến trúc ISDNN sẽ được sửa đổi để phù hợp hơn với mô hình kênh có cấu trúc (gọi tắt là **ISDNN có cấu trúc**). Trước hết, thay vì phải ước lượng ma trận \mathbf{H} , do biết trước θ, ϕ cũng như cấu hình của mảng ăng-ten tại trạm cơ sở, ma trận \mathbf{H} được giản ước về chỉ còn thành phần hệ số khuếch đại phức của các đường truyền ($\hat{\beta}$).

$$\hat{\beta}_{l,t} = \frac{h_{l,t}}{\varphi_{l,t}} = \frac{h_{l,t}}{e^{-ik_s c_l(\theta_{l,t}, \phi_{l,t})}} \quad (3.29)$$

Dùng phép chia vô hướng ở đây là do giả thiết các hệ số trong mô hình kênh truyền trên (3.28) là dạng rời rạc và vô hướng. Làm điều tương tự với véc-tơ tín hiệu thu được \mathbf{x} . Giả thiết rằng, với toàn bộ thông tin véc-tơ lái φ , véc-tơ \mathbf{x} có thể được biến đổi về dạng $\hat{\mathbf{x}}$ khi kênh truyền chỉ được đại diện bởi các hệ số β như trên phương trình (3.29) thông qua phép biến đổi f_1 .

$$\hat{\mathbf{x}} \leftarrow f_1(\mathbf{x}, \varphi) \quad (3.30)$$

Từ ma trận $\hat{\beta}$ và véc-tơ $\hat{\mathbf{x}}$ thu được, các dữ liệu đầu vào còn lại mạng ISDNN cũng được thay đổi thông qua một tập các hàm xử lý tín hiệu đơn giản gọi tắt là f_2 . Trong đó, hai giá trị \mathbf{D} và s_{in} có biểu diễn là

$$\begin{aligned} \mathbf{D} &= \text{diag}(\hat{\beta}^\top \hat{\beta}) \\ \hat{s}_1 &= \mathbf{D}^{-1} \hat{\beta}^\top \hat{\mathbf{x}} \end{aligned} \quad (3.31)$$

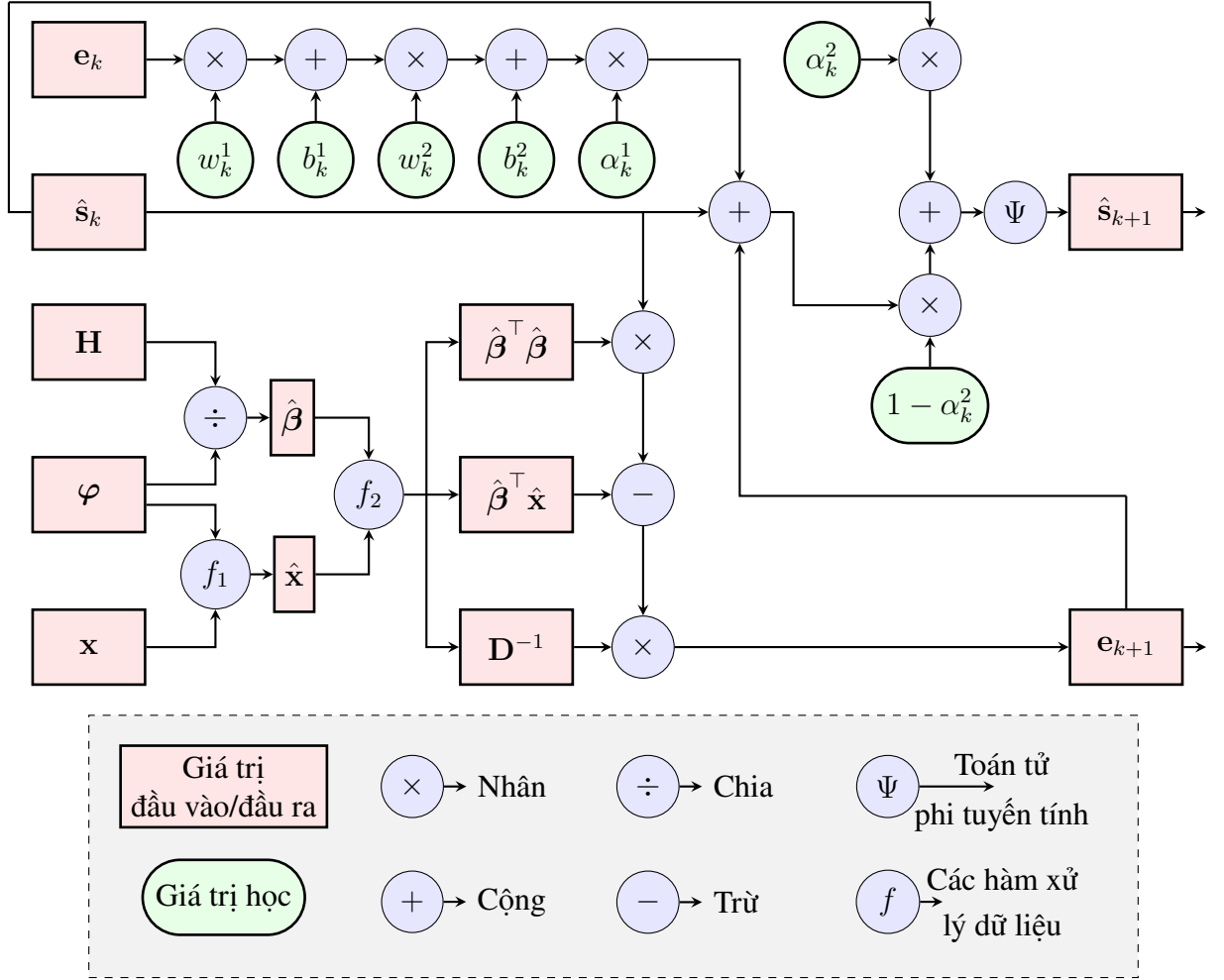
Kiến trúc một lớp mạng ISDNN cho kênh truyền có cấu trúc với thông tin bên lề là DoA tại bên thu được biểu diễn như trên hình 3.6. Các tham số khởi tạo và quá trình đào tạo của mạng vẫn tương tự như kiến trúc ISDNN đã được trình bày sử dụng cho mô hình kênh truyền phi cấu trúc tại mục 3.3.2.

3.4. Mô phỏng và đánh giá

Trong phần này, tác giả sẽ trình bày phương pháp tạo bộ dữ liệu đào tạo cho các mạng nơ-ron sâu DetNet và ISDNN. Sau khi được đào tạo, các kết quả so sánh và đánh giá về độ chính xác, độ phức tạp, và tính chống chịu lỗi sẽ được đưa ra.

3.4.1. Tạo bộ dữ liệu

Trong bảng 3.1, các tham số mô phỏng của hệ thống mMIMO cũng như kiến trúc mạng DetNet và ISDNN được đưa ra. Chi tiết, các tập dữ liệu được tạo cho việc đào tạo (training)/thử nghiệm (testing) sẽ độc lập với nhau nhưng cùng chung phân bố. Mỗi tín hiệu của bên phát s sẽ được gieo ngẫu nhiên theo phân bố đều và sử dụng chung một loại



Hình 3.6. Kiến trúc của một lớp trong mạng nơ-ron sâu ISDNN đề xuất cho mô hình kênh truyền có cấu trúc. Giả sử biết thông tin của DoA và cấu hình mảng ăng-ten tại bên thu $\varphi(\theta, \phi)$.

điều chế. Tùy thuộc vào loại điều chế, mà một nhóm gồm 1, 2, 4, 8, ... bit sẽ được nhóm thành ký hiệu. Trong mô phỏng, điều chế 16-QAM được lựa chọn, với 4 bit liên nhau sẽ được gộp lại, $u_i \in \mathbb{R}^4$, tạo thành ký hiệu s_i . Trên bảng 3.2 và hình 3.7 là biểu diễn ánh xạ các nhóm 4 bit thành các ký hiệu s_i .

Ma trận kênh truyền H lấy theo mô hình kênh CBSM i.i.d Rayleigh đã trình bày trong chương 1, trong đó, các hệ số phức của kênh truyền được gieo ngẫu nhiên độc lập và cùng phân bố Gaussian với giá trị trung bình μ và độ lệch chuẩn σ^2 như trên hình 3.8a.

$$\Re(h_{l,t}) = f(x | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \text{với } x \in \mathbb{R} \quad (3.32)$$

Trong mô hình kênh truyền có cấu trúc, vẫn giữ nguyên cách gieo ngẫu như ma trận H như mô hình phi cấu trúc. Sau đó, từ các thông tin bên lề là DoA và cấu hình mảng ăng-ten thu, các giá trị thuộc ma trận $\hat{\beta}$ được tính ngược lại theo công thức (3.29).

Bảng 3.1. Các tham số mô phỏng hệ thống truyền thông không dây của mạng nơ-ron sâu ISDNN được đề xuất.

Thông số mô phỏng	Giá trị
Kích thước hệ thống mMIMO	$T = 8, L = 64$
Loại điều chế	16-QAM
Các mức SNR của dataset	$[0, 5, 10, 15, 20]$ dB
Số mẫu đào tạo	20.000
Số mẫu thử nghiệm	5.000
Số lớp mạng của DetNet	$K_{DetNet} = 4; 10$
Số lớp mạng của ISDNN	$K_{ISDNN} = 4$
Số lượng mẫu trong một lần huấn luyện	Batch size = 100.000
Thuật toán tối ưu	Adam [22]
Giá trị khởi tạo của tốc độ học	$\delta = 0,0001$
Số vòng lặp đào tạo	20.000

Bảng 3.2. Ánh xạ các nhóm 4 bit thành các ký hiệu sử dụng điều chế 16-QAM.

Ký hiệu	$s_0 = -1 - 1i$	$s_1 = -1 - 0,33i$	$s_2 = -1 + 0,33i$	$s_3 = -1 + 1i$
	$s_4 = -0,33 - 1i$	$s_5 = -0,33 - 0,33i$	$s_6 = -0,33 + 0,33i$	$s_7 = -0,33 + 1i$
	$s_8 = 0,33 - 1i$	$s_9 = 0,33 - 0,33i$	$s_{10} = 0,33 + 0,33i$	$s_{11} = 0,33 + 1i$
	$s_{12} = 1 - 1i$	$s_{13} = 1 - 0,33i$	$s_{14} = 1 + 0,33i$	$s_{15} = 1 + 1i$
Nhóm bit	$\mathbf{u}_0 = [0, 0, 0, 0]$	$\mathbf{u}_1 = [0, 0, 0, 1]$	$\mathbf{u}_2 = [0, 0, 1, 0]$	$\mathbf{u}_3 = [0, 0, 1, 1]$
	$\mathbf{u}_4 = [0, 1, 0, 0]$	$\mathbf{u}_5 = [0, 1, 0, 1]$	$\mathbf{u}_6 = [0, 1, 1, 0]$	$\mathbf{u}_7 = [0, 1, 1, 1]$
	$\mathbf{u}_8 = [1, 0, 0, 0]$	$\mathbf{u}_9 = [1, 0, 0, 1]$	$\mathbf{u}_{10} = [1, 0, 1, 0]$	$\mathbf{u}_{11} = [1, 0, 1, 1]$
	$\mathbf{u}_{12} = [1, 1, 0, 0]$	$\mathbf{u}_{13} = [1, 1, 0, 1]$	$\mathbf{u}_{14} = [1, 1, 1, 0]$	$\mathbf{u}_{15} = [1, 1, 1, 1]$

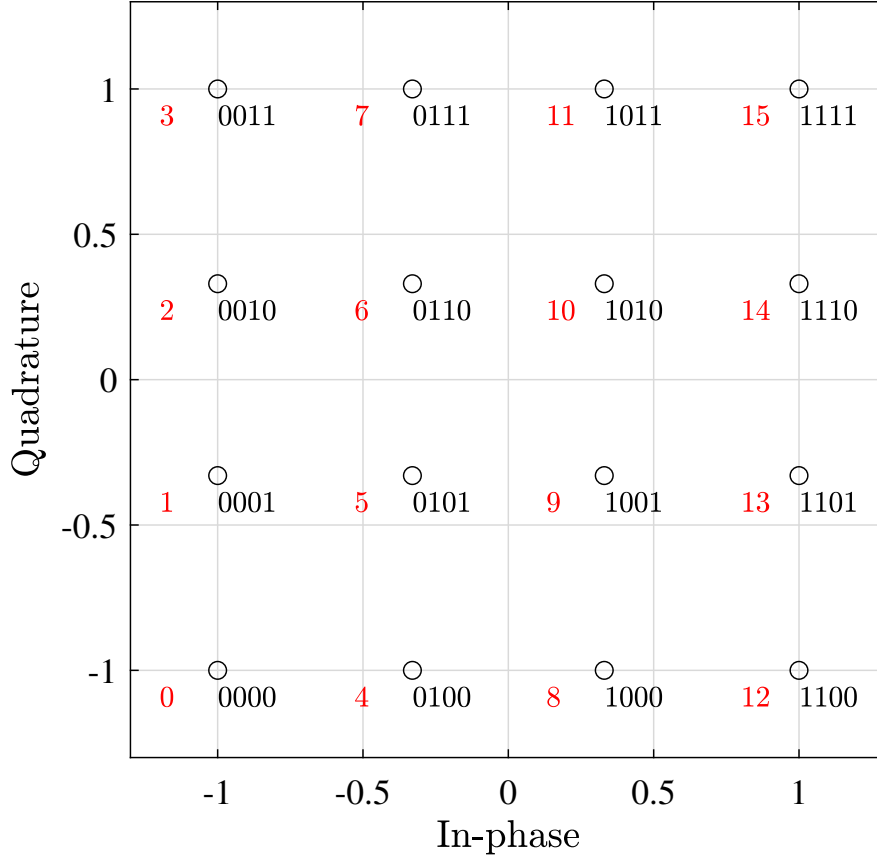
Trên hình 3.8b là biểu diễn histogram các giá trị phần thực $\hat{\beta}_{l,t}$ thuộc ma trận $\hat{\beta}$. Có thể nhận thấy các phần tử trong ma trận \mathbf{H} và $\hat{\beta}$ cùng có phân bố chuẩn với kỳ vọng sấp xỉ bằng 0. Sự khác nhau nằm ở phương sai của các phân bố này, lần lượt là $\approx \mathcal{N}(0; \frac{1}{\sqrt{2}})$, $\mathcal{N}(0; 0,644)$ ứng với $\Re(h_{l,t})$ và $\Re(\hat{\beta}_{l,t})$. Với phương sai nhỏ hơn, có thể dự đoán rằng các mạng học sâu như ISDNN có cấu trúc có thể cho ra các hệ số của ma trận $\hat{\beta}$ nhanh và chính xác hơn phi cấu trúc \mathbf{H} .

Ngoài việc đào tạo dựa trên các thông tin kênh truyền \mathbf{H} và $\hat{\beta}$ chính xác. Trước hết, xem xét việc đào tạo ISDNN cho mô hình kênh phi cấu trúc trong trường hợp thông tin \mathbf{H} không chính xác (im - imperfect) để kiểm tra khả năng chịu lỗi của kiến trúc mạng nơ-ron sâu ISDNN đề xuất. Lý do là trong các điều kiện thực tế, các ma trận đầu vào để đào tạo được đo lường không thể có được sự chính xác hoàn hảo. Hai mức sai số sẽ được xem xét đó là 1% và 5%.

$$\mathbf{H}_{im} = \mathbf{H} \pm 0,01\mathbf{H} \quad (3.33)$$

$$\mathbf{H}_{im} = \mathbf{H} \pm 0,05\mathbf{H}$$

Tương tự, xem xét việc đào tạo kiến trúc mạng ISDNN cho mô hình kênh có cấu trúc trong trường hợp cả thông tin \mathbf{H} và φ đều có sai số. Sai số từ \mathbf{H} vẫn tương tự như giả



Hình 3.7. Ánh xạ các nhóm 4 bit thành các ký hiệu sử dụng điều chế 16-QAM.

thiết của mô hình kênh phi cấu trúc. Sai số của dữ liệu đầu vào φ đến từ việc ước lượng sai các giá trị (θ, ϕ) . Tương tự như trên, hai mức sai số của φ sẽ được xem xét đó là 1% và 5%. Sai số của φ được gọi là “**sai số của thông tin bên lề**”.

$$\begin{aligned}\varphi_{im} &= \varphi \pm 0,01\varphi \\ \varphi_{im} &= \varphi \pm 0,05\varphi\end{aligned}\tag{3.34}$$

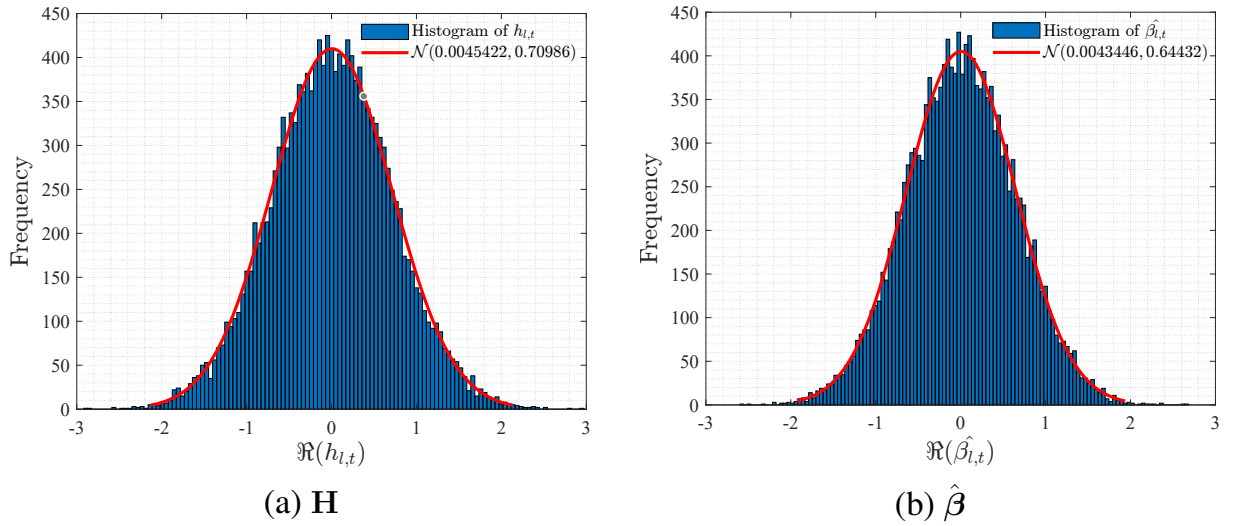
Sau khi đi qua kênh truyền H , các ký hiệu sẽ được cộng thêm với AWGN ở các giá trị SNR khác nhau tính theo thang dB (decibel).

$$\text{SNR} = 10 \log \left(\frac{\mathbb{E} (\|\mathbf{H}\mathbf{s}\|_2^2)}{\mathbb{E} (\|\mathbf{w}\|_2^2)} \right) \text{ (dB)}\tag{3.35}$$

3.4.2. Đào tạo và đánh giá kiến trúc mạng nơ-ron sâu đề xuất

3.4.2.1. Phương pháp đánh giá

Sau khi đã tạo được các bộ dữ liệu, việc đào tạo được triển khai trên máy tính với cấu hình: vi xử lý Intel Core i9-10900, 64 GB RAM. Ngôn ngữ lập trình Python được



Hình 3.8. Phân bố khi gieo ngẫu nhiên của các hệ số phần thực của $h_{l,t}$ và $\hat{\beta}_{l,t}$ trong hai ma trận \mathbf{H} , $\hat{\beta}$.

lựa chọn để xây dựng các mô phỏng của ISDNN và DetNet. Thư viện nền tảng Pytorch được sử dụng cho ISDNN, và Tensorflow cho DetNet. Mã nguồn của DetNet lấy từ kho lưu trữ công khai của nhóm tác giả trên bài báo gốc tại Github⁴. Sai số của các mạng nơ-ron này được đánh giá thông qua thông số BER tương ứng là số bit ước lượng sai chia cho tổng số bit. Ở bước thử nghiệm, 100 bộ dữ liệu thử nghiệm, mỗi bộ gồm 5.000 mẫu được tạo ra, kết quả ước lượng của các mô hình thu được sau quá trình đào tạo được tính bằng BER trung bình của 100 lần thử nghiệm.

$$\text{BER} = \frac{1}{100} \sum_{K=1}^{100} \frac{N_e}{5000} \quad (3.36)$$

3.4.2.2. So sánh độ phức tạp của các phương pháp

Trước hết, độ phức tạp của các thuật toán sẽ được so sánh như trên bảng 3.3. Trong đó, hai bộ nhận dạng truyền thống ZF và MMSE đều có độ phức tạp $\mathcal{O}(TL^3)$ do phép nghịch đảo của ma trận \mathbf{H} với kích thước đầy đủ [43]. Tiếp đến, kiến trúc mạng DetNet cho độ phức tạp $\mathcal{O}(TL^2)$ do không phải nghịch đảo ma trận $\mathbf{G}_{\mathbf{H}}$ nên thành phần phức tạp nhất trong DetNet là các phép nhân ma trận $\mathbf{H}^T \mathbf{H}$ và $\mathbf{H}^T \mathbf{H} \mathbf{s}_k$. Cuối cùng là độ phức tạp của kiến trúc mạng ISDNN được đề xuất cho cả hai mô hình kênh truyền có và phi cấu trúc cũng ở mức $\mathcal{O}(TL^2)$. Dù có phép nghịch đảo ma trận \mathbf{D}^{-1} ở đầu vào, tuy nhiên như đã trình bày ở trên, ma trận \mathbf{D} chỉ gồm các phần tử trên đường chéo chính của ma trận Gram. Do vậy, việc nghịch đảo ma trận này chỉ có độ phức tạp $\mathcal{O}(TL)$, vì chỉ cần sử dụng phép biến đổi tuyến tính. Vậy nên, độ phức tạp tổng thể của ISDNN vẫn tương tự

⁴<https://github.com/neevesamuel/DeepMIMODetection>

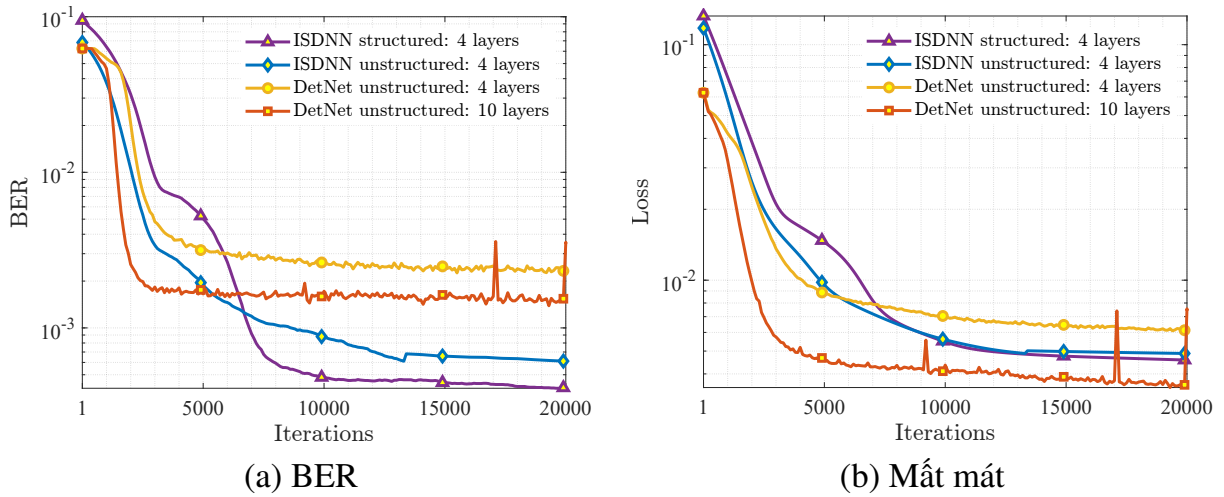
Bảng 3.3. So sánh độ phức tạp của các thuật toán nhận dạng kênh truyền.

Bộ nhận dạng	Độ phức tạp	Số giá trị học
ZF	$\mathcal{O}(TL^3)$	
MMSE	$\mathcal{O}(TL^3)$	
DetNet: 4 layers [51]	$\mathcal{O}(TL^2)$	105.416
ISDNN unstructured	$\mathcal{O}(TL^2)$	24
ISDNN structured	$\mathcal{O}(TL^2)$	24

như DetNet chỉ dừng ở các phép nhân ma trận. Có thể kết luận rằng, các phương pháp sử dụng học sâu đã giảm thiểu độ phức tạp đi $\mathcal{O}(L)$ so với các bộ ước lượng tuyến tính truyền thống. Đây là khoảng cách rất lớn, vì trong các hệ mMIMO, giá trị của L có thể lên đến trên 100. So sánh riêng hai kiến trúc mạng DNN là DetNet và ISDNN, dù có chung độ phức tạp nhưng nhận thấy số lượng giá trị học của ISDNN là không đáng kể khi so sánh với DetNet. Điều này có được là do các bộ biến đổi tuyến tính (\mathbf{W} , \mathbf{b}) trong DetNet ở dưới dạng ma trận và véc-tơ có kích thước lớn. Trong khi đó, ISDNN chỉ yêu cầu hai tham số học vô hướng (w , b) cho mỗi bộ biến đổi tại mỗi lớp mạng. Do vậy, chỉ 24 tham số học cần được sử dụng trong ISDNN, dẫn đến mô hình thu được sau quá trình đào tạo chỉ có kích thước 7 KB so với 1,236 KB của DetNet với cùng số lớp mạng là 4. Đây là lợi thế rất lớn, khi kích thước nhỏ và độ phức tạp thấp giúp làm tăng khả năng ứng dụng trên cả các thiết bị có giá thành thấp.

3.4.2.3. So sánh độ chính xác của mạng nơ-ron sâu đề xuất với các phương pháp khác

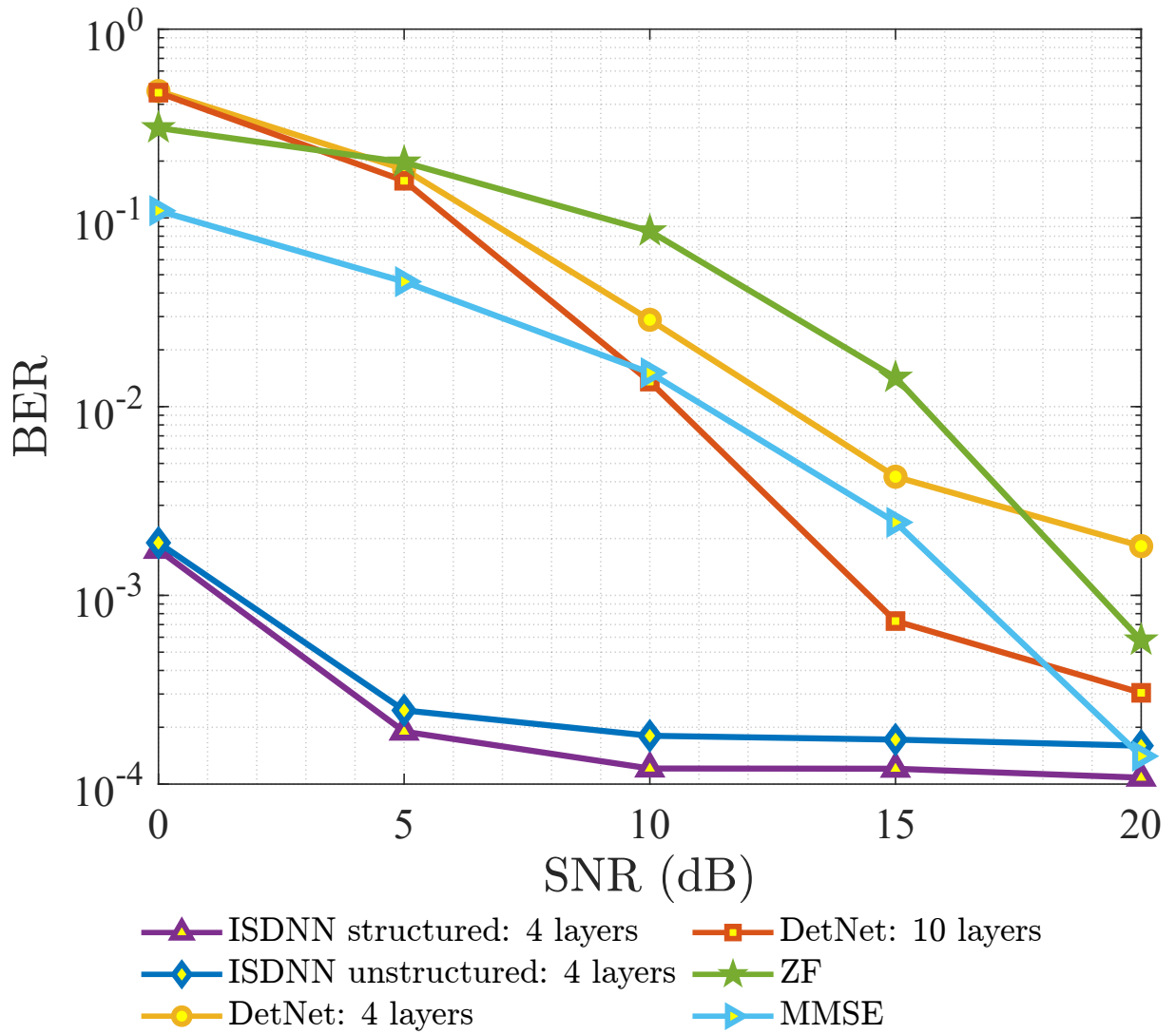
Trên hình 3.9 là quá trình đào tạo của hai mạng nơ-ron sâu ISDNN và DetNet với số lớp mạng, mô hình kênh khác nhau. Hình 3.9b xem xét về thời gian hội tụ thông qua đầu ra của hàm mất mát giữa véc-tơ ký hiệu gốc \mathbf{s} và véc-tơ ký hiệu ước lượng $\hat{\mathbf{s}}$. Nhận thấy, thời gian hội tụ của mạng DetNet với 4 lớp mạng có phần nhanh hơn so với ISDNN có và phi cấu trúc cùng số lớp mạng. Tuy nhiên, xét về tổng thể, đầu ra hàm mất mát của ISDNN cả có và phi cấu trúc cho kết quả tốt hơn so với DetNet cùng số lớp mạng dù phải cần đến vòng đào tạo thứ 12.000. Xét riêng hai mô hình có cấu trúc và phi cấu trúc của ISDNN, tốc độ hội tụ và giá trị của hàm mất mát tại vòng lặp cuối cùng khá tương đồng nhau khoảng 4×10^{-3} . Nếu tăng số lớp mạng của DetNet lên 10, do số lượng tham số học tăng lên đáng kể, thời gian hội tụ và đầu ra mất mát cuối cùng cũng cho kết quả tốt hơn ISDNN chỉ 4 lớp mạng. Tuy nhiên, đánh đổi ở đây là số lượng tham số học sẽ lên đến 316.244. Xét về sự hội tụ dựa trên tỷ lệ sai số bit của các kiến trúc như trên hình 3.9a. Trước hết, BER của ISDNN chỉ với 4 lớp mạng sau 20.000 vòng đào tạo là vượt trội so với DetNet dù 4 hay 10 lớp mạng, hội tụ ở giá trị BER $\approx 4 \times 10^{-4}$ và 6×10^{-4} lần lượt với



Hình 3.9. Sự hội tụ của quá trình đào tạo mạng ISDNN và DetNet.

mô hình có và phi cấu trúc. So sánh với DetNet dù với 10 lớp mạng và lượng tham số học khổng lồ cũng chỉ có đạt được sai số chưa đến 10^{-3} . Tuy nhiên, từ mô phỏng cũng cho thấy rằng, độ chính xác của DetNet cho thời gian hội tụ là nhanh hơn nhiều so với ISDNN khi chỉ cần đến khoảng 5.000 vòng đào tạo. Trong quá trình đào tạo, thông qua cả hai thông số BER và hàm mất mát, có thể rút ra nhận xét rằng, kiến trúc ISDNN có cấu trúc cho độ chính xác trong quá trình học là tốt hơn ISDNN phi cấu trúc và DetNet, tuy nhiên, DetNet lại là mạng có thời gian hội tụ nhanh nhất.

Sau quá trình đào tạo, kiểm tra mô hình thu được trên các bộ dữ liệu thử nghiệm độc lập với tập dữ liệu huấn luyện. Kết quả thu được khi so sánh độ chính xác của các phương pháp nhận dạng kênh truyền gồm ZF, MMSE, DetNet, và ISDNN khi SNR thay đổi được biểu diễn trên hình 3.10. Trước hết, có thể kết luận, độ chính xác của mạng ISDNN đề xuất là vượt trội so với các phương pháp còn lại. Khi so sánh với hai phương pháp tuyến tính là ZF và MMSE, đường BER của ISDNN và DetNet đều cho thấy sự khác biệt, khi độ dốc của BER từ các mạng DNN giảm dần theo SNR còn ZF và MMSE thì ngược lại. Phải cần đến mức SNR = 20 dB, phương pháp MMSE mới đạt đến độ chính xác của ISDNN phi cấu trúc tức $BER \approx 1,5 \times 10^{-4}$, do giải thuật gốc ISD cũng xuất phát từ MMSE nên có thể coi đây là giá trị tối ưu của ISD. Khi so sánh với mạng nơ-ron sâu DetNet gồm 4 lớp mạng, ISDNN phi cấu trúc cho độ lợi về BER đạt 10^2 tại các mức SNR thấp, và 10^1 tại SNR cao. ISDNN có cấu trúc cho tỷ lệ sai số bit thấp hơn mạng phi cấu trúc, tuy nhiên độ lợi và chưa thực sự vượt trội. Khi tăng số lớp của DetNet lên 10, cũng tương tự như quá trình huấn luyện, độ chính xác đã được cải thiện, tuy nhiên dù SNR ở mức cao như 20 dB, BER của DetNet cũng chỉ tiệm cận được đến độ chính xác của ISDNN phi cấu trúc với 4 lớp mạng. Có thể rút ra nhận xét, kiến trúc ISDNN ngoài việc cho độ chính xác vượt trội so với ZF, MMSE, và DetNet còn có ưu



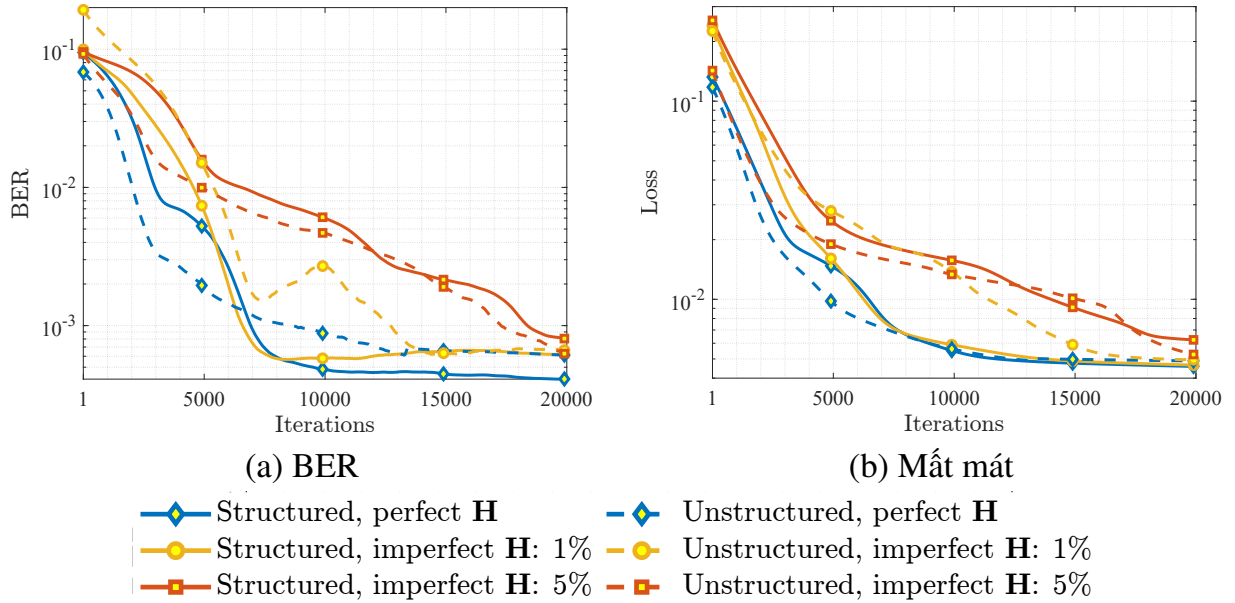
Hình 3.10. Độ chính xác của mạng ISDNN so sánh với DetNet và các bộ nhận dạng tuyến tính.

điểm là BER không có sự biến đổi quá lớn ở các mức SNR khác nhau. Đây là ưu điểm quan trọng của việc sử dụng DNN cho việc nhận dạng kênh truyền, khi tạp âm/công suất phát luôn là một vấn đề mà các thế hệ mạng viễn thông thế hệ mới như 5G quan tâm. Nếu độ chính xác của việc nhận dạng không bị phụ thuộc nhiều vào SNR thì mật độ bao phủ, cũng như hiệu quả về năng lượng là những lợi ích có thể khai thác.

3.4.2.4. Xem xét độ chính xác của mạng nơ-ron sâu ISDNN khi có sai số trong tập dữ liệu đào tạo

Tiếp theo, xem xét đến tính chống chịu lỗi của mạng nơ-ron sâu ISDNN. Như đã trình bày ở trên, để có thể được áp dụng thực tế, các bộ dữ liệu cần được thu thập từ các hệ thống viễn thông thực. Tuy nhiên, sai số khi đo lường các đầu vào cho việc đào tạo là không thể tránh khỏi. Trong phần này: (i), ma trận kênh truyền H được giả sử là có sự sai khác 1% và 5% so với H hoàn hảo; (ii), cả ma trận kênh truyền H và ma trận thông

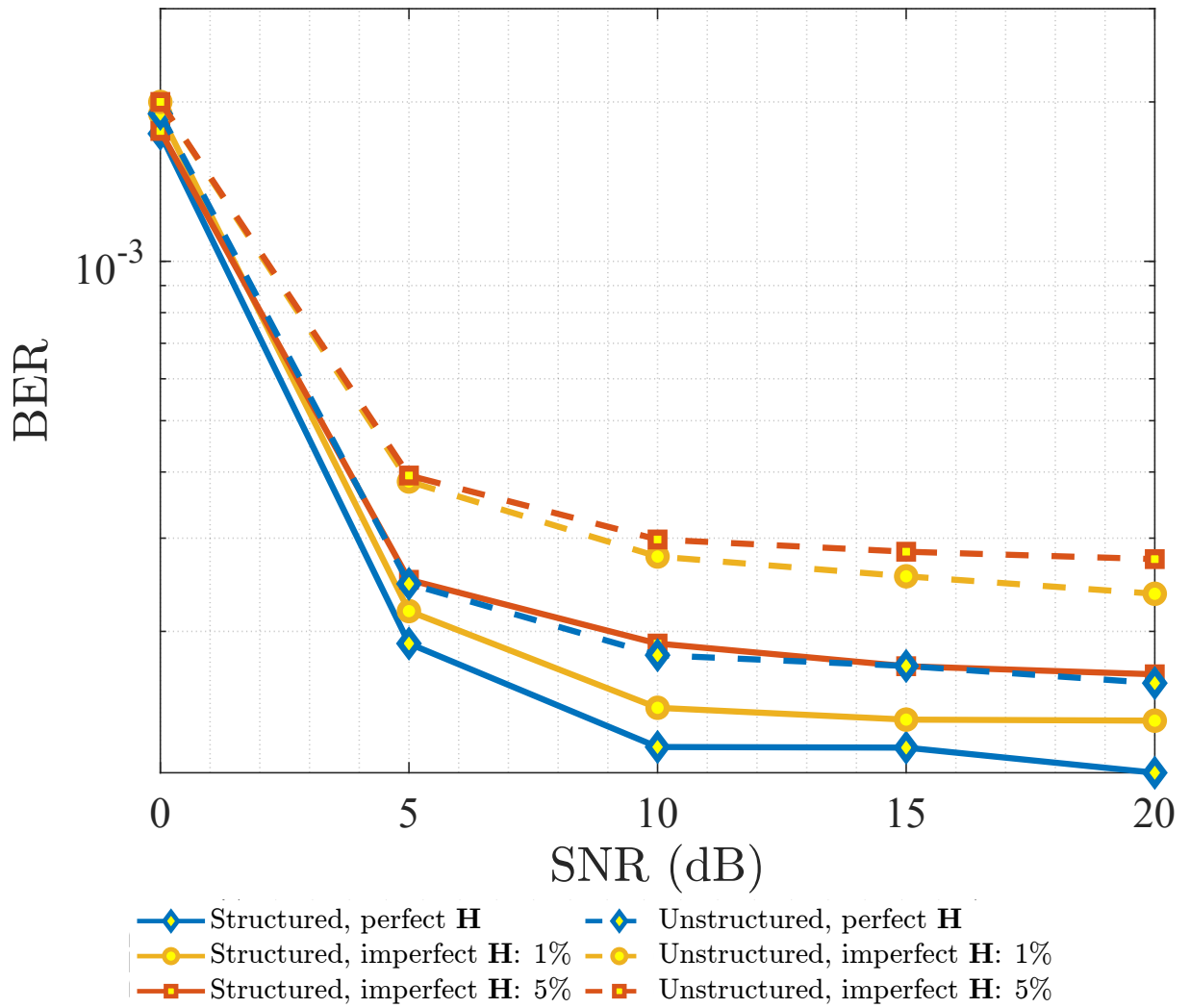
tin bên lề φ đều có sai số là 1% và 5% so với giá trị thực tế.



Hình 3.11. Sự hội tụ của quá trình đào tạo các mạng ISDNN với các sai số kênh truyền đầu vào khác nhau.

Trên hình 3.11 là kết quả của việc đào tạo mạng ISDNN có và phi cấu trúc với 3 bộ dữ liệu có các mức sai số kênh truyền khác nhau. Đầu tiên, hình 3.11b cho thấy rõ ràng sai số của dữ liệu đầu vào ảnh hưởng trực tiếp độ sự hội tụ của một mạng DNN. Cả hai kiến trúc ISDNN với kênh truyền chính xác cho tốc độ hội tụ về hàm mất mát là nhanh hơn đáng kể khi so với trường hợp kênh truyền có sai số. Với sai số 1% cần đến 14.000 và 18.000 vòng đào tạo lần lượt với mạng có và phi cấu trúc để đầu ra của hàm mất mát hội tụ. Tương tự, khi sai số là 5%, sau 20.000 vòng đào tạo vẫn chưa có được sự hội tụ của hàm mất mát. Tuy có sự khác nhau về tốc độ, nhưng giá trị cuối cùng của hàm mất mát tại cả 6 trường hợp đều khá tương đồng khoảng 5×10^{-3} . Tiếp theo, về độ chính xác trong quá trình đào tạo cũng cho kết quả tương đồng, biểu diễn trên hình 3.11a. Khi dữ liệu kênh truyền có sai số, đường BER trong quá trình học của các mạng ISDNN có sự không ổn định và cần đến thêm ít nhất 5.000 vòng đào tạo để đạt được giá trị hội tụ. Do đó, xét cả hàm mất mát và BER, kiến trúc ISDNN vẫn có thể thích nghi với sai số 1% của ma trận kênh truyền dù ảnh hưởng đến thời gian huấn luyện cần để hội tụ nhưng vẫn sẽ hội tụ ở giá trị khá tương đương với kênh truyền hoàn hảo.

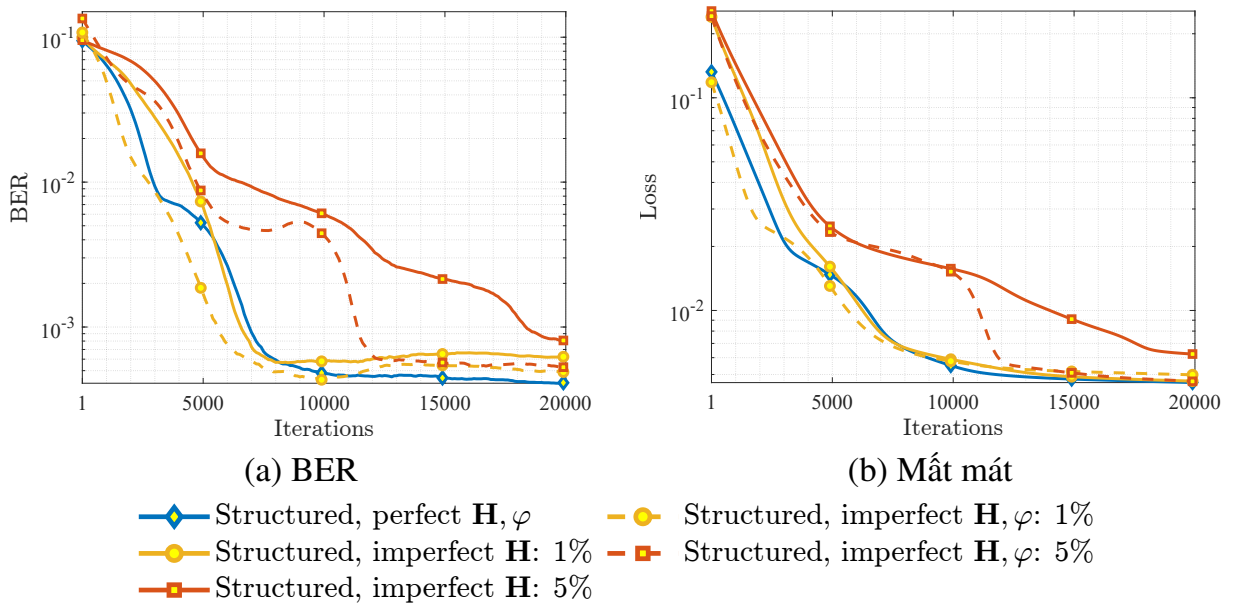
Tương tự, mô hình thu được sau quá trình đào tạo ở cả ba trường hợp sẽ được đánh giá trên các bộ dữ liệu thử nghiệm, kết quả thu được như trên hình 3.12. Để nhận thấy, sự tương quan của sai số ma trận kênh truyền đầu vào với BER đầu ra của mô hình đã huấn luyện. Với ISDNN phi cấu trúc, ở các giá trị $\text{SNR} \geq 5 \text{ dB}$, BER trong trường hợp kênh truyền chính xác và có sai số khá ổn định, lần lượt ở mức $\approx 1,7 \times 10^{-4}$, $2,5 \times 10^{-4}$,



Hình 3.12. Độ chính xác của mạng ISDNN với các sai số kênh truyền đầu vào khác nhau.

và $2,8 \times 10^{-4}$. Điều tương tự với ISDNN có cấu trúc khi BER tỷ lệ nghịch với sai số của tập dữ liệu đầu vào. Khi $\text{SNR} \geq 5 \text{ dB}$, các giá trị BER trung bình lần lượt của ISDNN có cấu trúc là $\approx 1,1 \times 10^{-4}$, $1,3 \times 10^{-4}$, và $1,7 \times 10^{-4}$. Nếu so sánh mức sai số này với các kết quả thu được trên hình 3.10, kể cả ở mức sai số 5% của ma trận kênh truyền, ISDNN phi cấu trúc vẫn cho sai số tương đương với DetNet 10 lớp mạng và vượt trội DetNet nếu chỉ 4 lớp mạng. Khi so sánh với hai bộ ước lượng tuyến tính, MMSE (ISD gốc) sẽ tốt hơn ISDNN phi cấu trúc với sai số dữ liệu 5% nếu SNR ở các giá trị $\geq 15 \text{ dB}$. Ngược lại, nếu sử dụng ZF hoặc SNR thấp hơn thì ISDNN vẫn cho sai số tốt hơn. Từ các kết quả trên, có thể thấy sự ảnh hưởng của dữ liệu đầu vào tới ISDNN nói riêng và các mạng DNN nói chung, tuy nhiên ở các mức sai số nhỏ, mô hình đầu ra vẫn sẽ cho độ chính xác ở mức chấp nhận được, và vẫn sẽ hơn các giải thuật tuyến tính hay mạng DetNet. Đây chính là tiềm năng để triển khai ISDNN với các bộ dữ liệu thực và trên các hệ thống viễn thông thực tế trong tương lai.

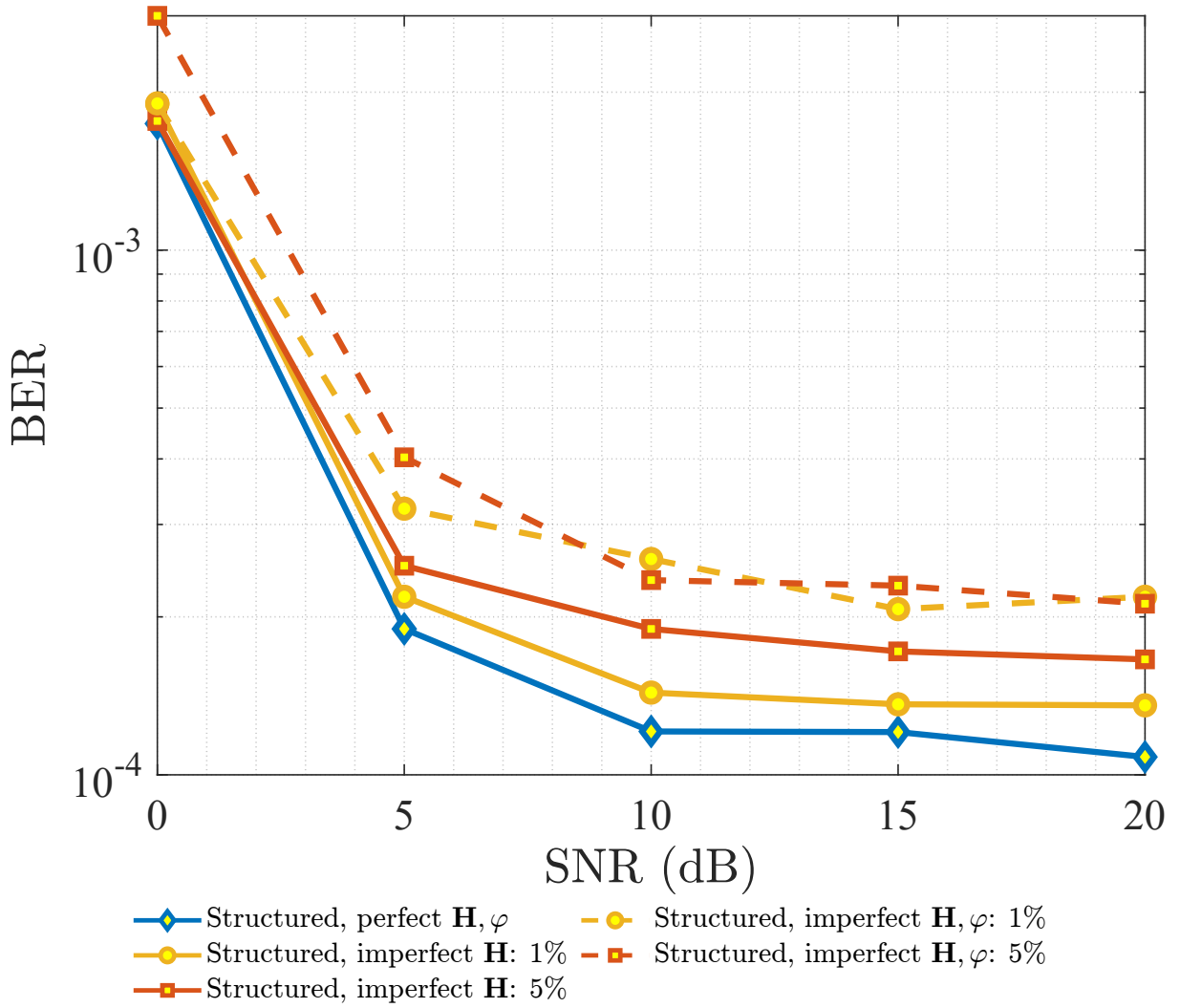
Trên hình 3.13 là kết quả đào tạo của mạng ISDNN có cấu trúc khi các bộ dữ liệu



Hình 3.13. Sự hội tụ của quá trình đào tạo các mạng ISDNN có cấu trúc với các sai số kênh truyền và thông tin bên lề đầu vào khác nhau.

huấn luyện có cả sai số trong \mathbf{H} và φ . Với việc cả hai ma trận trên đều có sai số sẽ hình thành sai số tích lũy cho cả các đầu vào như $\hat{\mathbf{s}}, \hat{\mathbf{x}}, \mathbf{D}, \dots$ có thể gây khó khăn cho bộ nhận dạng. Trước hết trên hình 3.13b là đầu ra của hàm mất mát thu được trong suốt quá trình đào tạo. Tuy có thêm sai số đầu vào từ DoA, nhưng giá trị đầu ra của hàm mất mát lại cho thấy điều ngược lại, khi có thêm sai số này, mạng ISDNN hội tụ nhanh hơn cả trường hợp khi chỉ có sai số của kênh truyền \mathbf{H} . Đặc biệt, với trường hợp cả \mathbf{H} và φ đều có sai số 5%, sau 12.000 vòng đào tạo, giá trị hàm mất mát đã đạt được giá trị hội tụ và ổn định như trường hợp không có sai số. Tương tự với kết quả của BER trong quá trình đào tạo trên hình 3.13a, khi có thêm sai số của φ , đường BER còn có độ lớn hơn trong trường hợp chỉ có sai số của ma trận kênh truyền.

Mạng ISDNN có cấu trúc được lưu tại vòng lặp cuối cùng của việc học và thử nghiệm với các giá trị SNR khác nhau như trên hình 3.14. Khi có sai số về ma trận φ , đường BER đầu ra sẽ cho kết quả kém hơn một chút so với khi chỉ có sai số về \mathbf{H} . Tuy nhiên, cũng như trong quá trình học, đường BER khi sai số của \mathbf{H} và φ ở mức 5% cho kết quả không ổn định. Tại SNR = 15 dB, BER của ISDNN còn tốt hơn trường hợp chỉ có sai số 1% của bộ dữ liệu. Nguyên nhân có thể xuất phát từ việc gieo ngẫu nhiên những sai số trong bộ dữ liệu đầu vào, tạo nên ma trận $\hat{\beta}$ tốt hơn ở giá trị SNR này. Trong trường hợp này, so sánh với hình 3.12, kết quả mô hình đầu ra trong hai trường hợp có sai số từ \mathbf{H} và thông tin bên lề của mạng ISDNN có cấu trúc là tương đương với mạng ISDNN phi cấu trúc chỉ có sai số về ma trận kênh truyền.



Hình 3.14. Độ chính xác của mạng ISDNN có cấu trúc với các sai số kênh truyền và thông tin bên lề đầu vào khác nhau.

3.5. Kết luận chương

Trong chương này, tác giả đã trình bày khái quát về việc sử dụng DNN và mở rộng sâu cho việc nhận dạng kênh truyền. Tiếp đến, một kiến trúc mạng nơ-ron sâu tên DetNet đã được đề xuất trước đây được trình bày ngắn gọn. Từ một phương pháp nhận dạng không mù với độ phức tạp thấp ISD đã được công bố trước đó, kiến trúc ISDNN mới được đề xuất cho cả hai mô hình kênh truyền có cấu trúc và phi cấu trúc sử dụng phương pháp mở rộng sâu. Kết quả mô phỏng đã chỉ ra hiệu năng về độ chính xác và độ phức tạp của kiến trúc ISDNN được đề xuất. Trước hết, về độ chính xác, ISDNN cho kết quả vượt trội so với các thuật toán nhận dạng tuyến tính không mù như ZF, MMSE, và mạng nơ-ron sâu DetNet dù với số lượng lớp mạng ít hơn. Về độ phức tạp, so với các giải thuật tuyến tính, ISDNN cho độ lợi $\mathcal{O}(L)$ tương tự như DetNet. Hơn nữa, ISDNN chỉ cần 24 tham số học cho 4 lớp mạng khi so sánh với 105.416 tham số của mạng DetNet cùng số lớp mạng. Từ hai khía cạnh trên, kết luận mạng ISDNN đã giải quyết cả hai vấn đề là

độ phức tạp, và chính xác đã được đề ra ở phần Mở đầu. Kiến trúc ISDNN được đề xuất cho mô hình kênh truyền có cấu trúc cũng cho ra độ chính xác là tốt hơn so với ISDNN phi cấu trúc như đã chỉ ra trong chương 2. Ngoài ra, để xem xét khả năng ứng dụng vào thực tế, hiệu suất của ISDNN được xem xét nếu có sai số trong các bộ dữ liệu đầu vào, có thể xảy ra bởi sai số đo lường. Các kết quả mô phỏng chỉ ra sự ảnh hưởng của tập dữ liệu huấn luyện đến kết quả đào tạo. Tuy nhiên, sai số cũng ở mức chấp nhận được và vẫn là tốt hơn nếu so sánh với các phương pháp đã kể trên.

KẾT LUẬN

Trong luận văn, tác giả tập trung giải quyết các thách thức về chi phí và độ phức tạp của các phương pháp nhận dạng trong các thể hệ mạng di động mới sử dụng mMIMO. Trước hết, các phương pháp mô hình kênh truyền trong mMIMO được khảo sát để chọn ra phương pháp phù hợp cho nghiên cứu trong luận văn. Tiếp đến, một khảo sát về bốn phương pháp nhận dạng hệ thống viễn thông không dây được trình bày. Qua đó, chỉ ra được sự cần thiết của việc ứng dụng chi thức mới vào bài toán nhận dạng hệ thống thông qua hai hướng tiếp cận là bán mù và sử dụng học sâu. Từ đó, trong chương 2, tác giả đề xem xét sự ảnh hưởng của các cấu hình mảng ăng-ten khác nhau và giải thuật SB đến độ chính xác của việc ước lượng kênh truyền dựa trên CRB. Kết quả chỉ ra rằng việc mô hình kênh truyền có cấu trúc, sử dụng các mảng ăng-ten 3D như UCyA, hay phương pháp SB đều có thể giúp giảm sai số của việc ước lượng kênh truyền đi đáng kể. Ngoài ra việc sử dụng các mảng ăng-ten 3D sẽ giúp tiết kiệm diện tích lắp đặt cho các trạm cơ sở đi đáng kể khi so sánh với các cấu hình ULA truyền thống. Trong chương 3, hướng tiếp cận học sâu, sử dụng thêm thông tin bên lề về hướng sóng đến và cấu hình mảng ăng-ten, cũng được tác giả xem xét để nhận dạng kênh truyền cho hệ thống mMIMO. Cụ thể, một mạng ISDNN được đề xuất nhằm giảm thiểu độ phức tạp và chi phí so với thuật toán ISD gốc dựa trên bộ nhận dạng MMSE. Kiến trúc mạng nơ-ron sâu được đề xuất chỉ yêu cầu 24 tham số học và 7 KB cho mô hình được đào tạo với cấu hình gồm 4 lớp mạng. Đây là số lượng rất nhỏ và hoàn toàn vượt trội khi so sánh với một mạng nơ-ron sâu khác cũng với cách tiếp cận tương tự là DetNet. Từ các kết quả mô phỏng, hiệu suất về thời gian đào tạo và độ chính xác của ISDNN cũng được kiểm chứng là vượt trội cả các phương pháp tuyến tính và mạng nơ-ron sâu DetNet. Ngoài ra, tác giả cũng xem xét đến hiệu suất của mạng nếu dữ liệu đầu vào xuất hiện sai số trong việc đo lường. Kết quả thu được cho thấy sự ảnh hưởng rõ ràng của sai số từ dữ liệu huấn luyện đến mô hình. Tuy nhiên sau khi đào tạo, độ chính xác của mô hình được đề xuất vẫn giữ được dạng gốc và vẫn có phần vượt trội so với các phương pháp khác tương tự như khi không có sai số trong dữ liệu.

Dù đạt được kết quả về mặt mô phỏng và số học, tuy nhiên, vẫn còn các điểm hạn chế có thể được cải thiện trong tương lai của luận văn. Thứ nhất là các mô phỏng trong luận văn còn hạn chế về cấu hình mMIMO, đặc biệt là số lượng ăng-ten phát/người dùng do hạn chế về mặt kỹ thuật. Thứ hai, các phương pháp như ISDNN hay DetNet được đề

xuất hiện chỉ được thử nghiệm hoạt động với các kênh truyền có phân bố xác định. Tuy nhiên, điều này ít xảy ra trên thực tế, việc cải tiến, kiểm nghiệm mô hình ISDNN để đáp ứng cho trường hợp kênh truyền có phân bố thay đổi liên tục là cần thiết để đề xuất có tính ứng dụng thực tiễn hơn nữa. Thứ ba, kiến trúc mạng ISDNN cho mô hình kênh có cấu trúc được giả sử rằng số lượng đường truyền giữa một cặp ăng-ten thu/phát đang là 1, tức tầm nhìn thẳng, vẫn chưa tiến sát với thực tế. Đòi hỏi cần có thêm các nghiên cứu để giải các vấn đề kể trên.

DANH MỤC CÔNG TRÌNH KHOA HỌC CỦA TÁC GIẢ LIÊN QUAN ĐẾN LUẬN VĂN

Kết quả nghiên cứu của chương 2 đã được công bố:

Do Hai Son and Tran Thi Thuy Quynh (2023), “Impact Analysis of Antenna Array Geometry on Performance of Semi-blind Structured Channel Estimation for massive MIMO-OFDM systems,” in *IEEE Statistical Signal Processing Workshop (SSP)*, Hanoi, Vietnam, July. [accepted]

TÀI LIỆU THAM KHẢO

- [1] 3GPP (2020), 5g; nr; physical channels and modulation, Standard 3GPP TS 38.211 version 16.2.0 Release 16, ETSI.
- [2] *802.11ac Technology Introduction White Paper* (2009), Standard, Rohde & Schwarz.
- [3] Abed-Meraim K., Qiu W., and Hua Y. (1997), “Blind system identification”, *Proceedings of the IEEE* 85(8), pp. 1310–1322.
- [4] Al Mtawa Y., Haque A., and Bitar B. (2019), “The mammoth internet: Are we ready?”, *IEEE Access* 7, pp. 132894–132908.
- [5] Aïssa-El-Bey A., Abed-Meraim K., and Laot C. (2011), “Adaptive blind estimation of sparse simo channels”, in *International Workshop on Systems, Signal Processing and their Applications (WOSSPA)*, Tipaza, Algeria, pp. 348–351.
- [6] Chen Y., and Wainwright M. J. (2015), “Fast low-rank estimation by projected gradient descent: General statistical and algorithmic guarantees”, *arXiv:1509.03025* v1, pp. 1–63.
- [7] Feng R., Wang C.-X., Huang J., Gao X., Salous S., and Haas H. (2022), “Classification and comparison of massive mimo propagation channel models”, *IEEE Internet of Things Journal* 9(23), pp. 23452–23471.
- [8] Fleury B., Tschudin M., Heddergott R., Dahlhaus D., and Ingeman Pedersen K. (1999), “Channel parameter estimation in mobile radio environments using the sage algorithm”, *IEEE Journal on Selected Areas in Communications* 17(3), pp. 434–450.
- [9] Gao X., Dai L., Ma Y., and Wang Z. (2014), “Low-complexity near-optimal signal detection for uplink large-scale mimo systems”, *Electronics Letters* 50(18), pp. 1326–1328.
- [10] Garro E., Fuentes M., Carcel J. L., Chen H., Mi D., Tesema F., Gimenez J. J., and Gomez-Barquero D. (2020), “5g mixed mode: Nr multicast-broadcast services”, *IEEE Transactions on Broadcasting* 66(2), pp. 390–403.
- [11] Gesbert D., Duhamel P., and Mayrargue S. (1997), “On-line blind multichannel equalization based on mutually referenced filters”, *IEEE Transactions on Signal Processing* 45(9), pp. 2307–2317.
- [12] Giannakis G. (1997), “Filterbanks for blind channel identification and equalization”, *IEEE Signal Processing Letters* 4(6), pp. 184–187.

- [13] Hajji Z., Aïssa-El-Bey A., and Amis K. (2018), “Simplicity-based recovery of finite-alphabet signals for large-scale mimo systems”, *Digital Signal Processing* 80, pp. 70–82.
- [14] Hammad M. M., and Yahia M. M. (2023), *Mathematics for Machine Learning and Data Science: Optimization with Mathematica Applications*, arXiv 2302.05964, Egypt.
- [15] Hao L., Rodríguez-Piñeiro J., Yin X., and Wang H. (2019), “Measurement-based massive mimo polarimetric channel characterization in outdoor environment”, *IEEE Access* 7, pp. 171285–171296.
- [16] He H., Jin S., Wen C.-K., Gao F., Li G. Y., and Xu Z. (2019), “Model-driven deep learning for physical layer communications”, *IEEE Wireless Communications* 26(5), pp. 77–83.
- [17] Hershey J. R., Roux J. L., and Weninger F. (2014), “Deep unfolding: Model-based inspiration of novel deep architectures”, *arXiv:1409.2574 v1*, pp. 1–27.
- [18] *IST-4-027756 WINNER II: Channel Models* (2007), Standard D1.1.2 V1.2, WINNER.
- [19] Jiang Y., Varanasi M. K., and Li J. (2011), “Performance analysis of zf and mmse equalizers for mimo systems: An in-depth study of the high snr regime”, *IEEE Transactions on Information Theory* 57(4), pp. 2008–2026.
- [20] Ju S., Xing Y., Kanhere O., and Rappaport T. S. (2021), “Millimeter wave and sub-terahertz spatial statistical channel model for an indoor office building”, *IEEE Journal on Selected Areas in Communications* 39(6), pp. 1561–1575.
- [21] Kay S. M. (1993), *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, Inc., United States.
- [22] Kingma D. P., and Ba J. (2015), “Adam: A method for stochastic optimization”, in *3rd International Conference on Learning Representations*, San Diego, CA, USA, pp. 1–15.
- [23] Ladacyia A., Belouchrani A., Abed-Meraim K., and Mokraoui A. (2019), “Semi-blind mimo-ofdm channel estimation using expectation maximisation like techniques”, *IET Communications* 13(20), pp. 3452–3462.
- [24] Ladacyia A., Mokraoui A., Abed-Meraim K., and Belouchrani A. (2017), “Performance bounds analysis for semi-blind channel estimation in mimo-ofdm communications systems”, *IEEE Transactions on Wireless Communications* 16(9), pp. 5925–5938.
- [25] Larsson E. G., Edfors O., Tufvesson F., and Marzetta T. L. (2014), “Massive mimo for next generation wireless systems”, *IEEE Communications Magazine* 52(2), pp. 186–195.

- [26] Le Magoarou L., and Paquelet S. (2018), “Parametric channel estimation for massive mimo”, in *IEEE Statistical Signal Processing Workshop (SSP)*, Freiburg im Breisgau, Germany, pp. 30–34.
- [27] LeCun Y., Bengio Y., and Hinton G. (2015), “Deep learning”, *Nature* 521(7553), pp. 436–444.
- [28] Li J., Ai B., He R., Yang M., Zhang Y., Liu X., Zhong Z., and Hao Y. (2018), “Directional analysis of massive mimo channels at 11 ghz in theater environment”, in *IEEE 88th Vehicular Technology Conference (VTC Fall)*, Chicago, IL, USA, pp. 1–5.
- [29] Liao J., Zhao J., Gao F., and Li G. Y. (2020), “A model-driven deep learning method for massive mimo detection”, *IEEE Communications Letters* 24(8), pp. 1724–1728.
- [30] Lin X., Zhang L., and Jiang Y. (2020), “Location aided intelligent deep learning channel estimation for millimeter wave communications”, in *2020 IEEE/CIC International Conference on Communications in China (ICCC)*, Chongqing, China, pp. 489–494.
- [31] Liu G., She J., Lu W., Zhang M., and Bo Y. (2020), “3d deterministic ray tracing method for massive mimo channel modelling and parameters extraction”, *IET Communications* 14(18), pp. 3169–3174.
- [32] Liu L., Tao C., Matolak D. W., Lu Y., Ai B., and Chen H. (2015), “Stationarity investigation of a los massive mimo channel in stadium scenarios”, in *IEEE 82nd Vehicular Technology Conference (VTC Fall)*, Boston, MA, USA, pp. 1–5.
- [33] Ljung L. (1999), *System Identification: Theory for the User*, Prentice Hall information and system sciences series, Prentice Hall PTR, United States.
- [34] Mandloi M., and Bhatia V. (2017), “Low-complexity near-optimal iterative sequential detection for uplink massive mimo systems”, *IEEE Communications Letters* 21(3), pp. 568–571.
- [35] MatLab (2023), “Cdl channel model customization with ray tracing”. Accessed: Feb. 14, 2023. [Online]. Available: <https://www.mathworks.com/help/5g/ug/cdl-channel-model-customization-with-ray-tracing.html>.
- [36] Menni T., Chaumette E., Larzabal P., and Barbot J. P. (2012), “New results on deterministic cramer–rao bounds for real and complex parameters”, *IEEE Transactions on Signal Processing* 60(3), pp. 1032–1049.
- [37] Minka T. P. (2000), “Old and new matrix algebra useful for statistics”.
- [38] Narasimhan T. L., and Chockalingam A. (2014), “Channel hardening-exploiting message passing (chemp) receiver in large-scale mimo systems”, *IEEE Journal of Selected Topics in Signal Processing* 8(5), pp. 847–860.

- [39] Ngo H. Q., Larsson E. G., and Marzetta T. L. (2014), “Aspects of favorable propagation in massive mimo”, in *22nd European Signal Processing Conference (EUSIPCO)*, Lisbon, Portugal, pp. 76–80.
- [40] Nguyễn Linh Trung (2019). Đề tài NAFOSTED về “Kỹ thuật nhận dạng hệ thống: từ xử lý mù truyền thống sang xử lý có tri thức mới,” Trường Đại học Công Nghệ - ĐHQGHN, mã số: 01/2019/TN.
- [41] Oh M. S., Hosseinalipour S., Kim T., Brinton C. G., and Love D. J. (2021), “Channel estimation via successive denoising in mimo ofdm systems: A reinforcement learning approach”, in *IEEE International Conference on Communications (ICC)*, Montreal, QC, Canada, pp. 1–6.
- [42] Ozdemir M. K., and Arslan H. (2007), “Channel estimation for wireless ofdm systems”, *IEEE Communications Surveys & Tutorials* 9(2), pp. 18–48.
- [43] Pan V. (1992), “Complexity of computations with matrices and polynomials”, *Society for Industrial and Applied Mathematics* 34(2), pp. 225–262.
- [44] Park S., Shim B., and Choi J. W. (2015), “Iterative channel estimation using virtual pilot signals for mimo-ofdm systems”, *IEEE Transactions on Signal Processing* 63(12), pp. 3032–3045.
- [45] Poormohammad S., and Farzaneh F. (2017), “Precision of direction of arrival (doa) estimation using novel three dimensional array geometries”, *AEU - International Journal of Electronics and Communications* 75, pp. 35–45.
- [46] Rekik O., Mokraoui A., Thuy Quynh T. T., Le T.-T., and Abed-Meraim K. (2021), “Side information effect on semi-blind channel identification for mimo-ofdm communications systems”, in *55th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, pp. 443–448.
- [47] Rusek F., Persson D., Lau B. K., Larsson E. G., Marzetta T. L., Edfors O., and Tufvesson F. (2013), “Scaling up mimo: Opportunities and challenges with very large arrays”, *IEEE Signal Processing Magazine* 30(1), pp. 40–60.
- [48] Saleh A., and Valenzuela R. (1987), “A statistical model for indoor multipath propagation”, *IEEE Journal on Selected Areas in Communications* 5(2), pp. 128–137.
- [49] Samek W., Montavon G., Lapuschkin S., Anders C. J., and Müller K.-R. (2021), “Explaining deep neural networks and beyond: A review of methods and applications”, *Proceedings of the IEEE* 109(3), pp. 247–278.
- [50] Samuel N., Diskin T., and Wiesel A. (2017), “Deep mimo detection”, *arXiv:1706.01151 v1*, pp. 1–5.
- [51] Samuel N., Diskin T., and Wiesel A. (2019), “Learning to detect”, *IEEE Transactions on Signal Processing* 67(10), pp. 2554–2564.

- [52] Shaik N., and Malik P. K. (2021), “A comprehensive survey 5g wireless communication systems: open issues, research challenges, channel estimation, multi carrier modulation and 5g applications”, *Multimedia Tools and Applications* 80(19), pp. 28789–28827.
- [53] Shi K., Serpedin E., and Ciblat P. (2005), “Decision-directed fine synchronization in ofdm systems”, *IEEE Transactions on Communications* 53(3), pp. 408–412.
- [54] Simeon R., Kim T., and Perrins E. (2022), “Machine learning with gaussian process regression for time-varying channel estimation”, in *IEEE International Conference on Communications (ICC)*, Seoul, Korea, pp. 3400–3405.
- [55] Singh V. K., Flanagan M. F., and Cardiff B. (2019), “Generalized least squares based channel estimation for fbmc-oqam”, *IEEE Access* 7, pp. 129411–129420.
- [56] Swindlehurst A. L., Zhou G., Liu R., Pan C., and Li M. (2022), “Channel estimation with reconfigurable intelligent surfaces—a general framework”, *Proceedings of the IEEE* 110(9), pp. 1312–1338.
- [57] Tong L., Xu G., and Kailath T. (1994), “Blind identification and equalization based on second-order statistics: a time domain approach”, *IEEE Transactions on Information Theory* 40(2), pp. 340–349.
- [58] Treichler J., and Agee B. (1983), “A new approach to multipath correction of constant modulus signals”, *IEEE Transactions on Acoustics, Speech, and Signal Processing* 31(2), pp. 459–472.
- [59] Tse D., and Viswanath P. (2005), *Fundamentals of Wireless Communication*, Cambridge University Press, United States.
- [60] Vilas Boas E. C., de Figueiredo F. A., Mendes L. L., de Souza R. A. et al. (2022), “Artificial intelligence for channel estimation in multicarrier systems for b5g/6g communications: a survey”, *EURASIP Journal on Wireless Communications and Networking* 2022(1), pp. 1–63.
- [61] Wan F., Zhu W.-P., and Swamy M. N. S. (2008), “A semiblind channel estimation approach for mimo-ofdm systems”, *IEEE Transactions on Signal Processing* 56(7), pp. 2821–2834.
- [62] Wang C., Zhang J., Tian L., Liu M., and Wu Y. (2017), “The spatial evolution of clusters in massive mimo mobile measurement at 3.5 ghz”, in *IEEE 85th Vehicular Technology Conference (VTC Spring)*, Sydney, NSW, Australia, pp. 1–6.
- [63] Wang Z., Zhao P., Qian C., and Chen S. (2016), “Location-aware channel estimation enhanced tdd based massive mimo”, *IEEE Access* 4, pp. 7828–7840.
- [64] Willhammar S., Flordelis J., Van Der Perre L., and Tufvesson F. (2020), “Channel hardening in massive mimo: Model parameters and experimental assessment”, *IEEE Open Journal of the Communications Society* 1, pp. 501–512.

- [65] Wisdom S., Hershey J., Le Roux J., and Watanabe S. (2016), “Deep unfolding for multichannel source separation”, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, pp. 121–125.
- [66] Yang S., and Hanzo L. (2015), “Fifty years of mimo detection: The road to large-scale mimos”, *IEEE Communications Surveys & Tutorials* 17(4), pp. 1941–1988.
- [67] Zhang C., Patras P., and Haddadi H. (2019), “Deep learning in mobile and wireless networking: A survey”, *IEEE Communications Surveys & Tutorials* 21(3), pp. 2224–2287.

PHỤ LỤC

A. Thuật toán tối ưu Adam

Thuật toán tối ưu Adam là phương pháp tối ưu hơn so với giải thuật tối ưu giảm dần độ dốc ngẫu nhiên (SGD - Stochastic gradient descent). Adam áp dụng các tốc độ học tập thích nghi (δ - Adaptive learning rate) khác nhau cho mỗi tham số học. Điều này mang lại lợi thế lớn khi các mô hình mạng nơ-ron với kiến trúc phức tạp. Một số phần trong mạng nơ-ron nhạy cảm với sự thay đổi của trọng số theo các cách riêng biệt. Do vậy, các phần này sẽ cần tốc độ học nhỏ hơn với các vùng khác. Trong luận văn này, tác giả chỉ đưa ra một số biểu diễn toán học quan trọng của Adam, chi tiết về giải thuật tối ưu này có tại [22].

$$m_t = \beta_1 m_{t-1} - (1 - \beta_1) g_t \quad (\text{A.a})$$

$$v_t = \beta_2 v_{t-1} - (1 - \beta_2) g_t^2 \quad (\text{A.b})$$

$$\hat{m}_t = \frac{m_t}{1 + \beta_1^t} \quad (\text{A.c})$$

$$\hat{v}_t = \frac{v_t}{1 + \beta_2^t} \quad (\text{A.d})$$

$$w_t = w_{t-1} - \delta \frac{m_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (\text{A.e})$$

trong đó:

- δ : tốc độ học.
- β_1, β_2 : tỉ lệ giảm dần theo cấp số nhân cho ước lượng moment thứ nhất và hai.
- m_t : giá trị trung bình của ước lượng moment thứ nhất.
- v_t : phương sai của ước lượng moment thứ hai.
- g_t : gradient.
- \hat{m}_t : các công cụ ước lượng hiệu chỉnh bias cho moment thứ nhất.
- \hat{v}_t : các công cụ ước lượng hiệu chỉnh bias cho moment thứ hai.
- w_t : trọng số của mô hình.