Artificial Intelligence Aided Physical and MAC Layer Optimization Schemes for Three Dimensional NOMA Networks

アハマド ハサン エルセイド マンソア ジェンディア

九州大学
KYUSHU UNIVERSITY

Artificial Intelligence Aided Physical and MAC Layer Optimization Schemes for Three Dimensional NOMA Networks

By

Ahmad Hasan Elsayed Mansour Gendia

A Thesis Submitted to the
Graduate School of
Information Science and Electrical Engineering,

Kyushu University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in

Electronics and Communications Engineering

September 2023

DECLARATION AND CERTIFICATE OF ORIGINALITY

I certify that in the preparation of this thesis, I have observed the provisions of Kyushu University Code of Ethics. Further; I certify that this work is free of plagiarism and all materials appearing in this thesis have been properly quoted and attributed. I certify that all copyrighted material incorporated into this thesis is in compliance with the international copyright law and that I have received written permission from the copyright owners for my use of their work, which is beyond the scope of the law. I agree to indemnify and save harmless Kyushu University from any and all claims that may be asserted or that may arise from any copyright violation. I hereby certify that the research work in this thesis is my original work and it does not include any copied parts without the appropriate citation.

Fukuoka, Japan,

July 18, 2023

Ahmad Hasan Elsayed Mansour Gendia.

SUMMARY

Non-orthogonal multiple access (NOMA) allows multiple user equipment (UE) to simultaneously share the same resource blocks using varying levels of transmit power at the base station (BS) side. This powerful feature of NOMA makes it a suitable candidate for future 6G mobile networks multiple access in order to meet the high connection density and data rates expectations set forth for 6G applications that require stringent operational demands such as haptic internet, 16K real-time video streaming, and bandwidth-hungry holographic AR/VR applications. In this thesis, we designed novel artificial intelligence (AI)-based frameworks to solve challenging PHY and MAC layers problems for three-dimensional NOMA (3D-NOMA) networks utilizing both terrestrial and non-terrestrial based communications. To optimize the PHY layer operation of 3D-NOMA networks, neural network (NN)-based orthogonal frequency division multiplexing (OFDM) signal processing is considered, whereas MAC layer operation optimization is accomplished via reinforcement learning (RL)-based resource management operation for both terrestrial and non-terrestrial NOMA downlink communications.

For OFDM signaling, high peak-to-average power ratio (PAPR) is a typical characteristic inherent in the multiplexed signals and has been a long-standing critical issue for mobile networks PHY layer optimization. The ever-increasing demand for low-latency operation calls for the development of low-complexity novel solutions to the PAPR problem. To address this issue while providing an enhanced PAPR reduction performance, in this thesis we propose a synchronous NN-based solution to achieve PAPR reduction performance exceeding the limits of conventional clipping and filtering (CF)-based schemes with lower computational complexity. The proposed scheme trains a neural network module using hybrid collections of samples from multiple OFDM symbols to arrive at a signal mapping with desirable characteristics.

On the other hand, to unlock the full potential of NOMA and reap its benefits, MAC layer-level resource management operation optimization is crucial. Specifically, proper allocation of transmission power and selection of candidate users for pairing over the same resource block are critical for an efficient

utilization of the available resources. Reinforcement learning deploying double deep-Q networks (DDQN) is a promising framework that can be adopted for tackling the problem. In this thesis, an RL-based DDQN scheme is proposed for opportunistic user pairing to access downlink NOMA systems with capacity-limited backhaul link connections. The proposed algorithm relies on proactive data caching to alleviate the throttling caused by backhaul bottlenecks, and optimized UE selection and power allocation are accomplished through the continuous interaction between an RL agent and the NOMA environment to increase the overall system throughput.

Moreover, for non-terrestrial 3D-NOMA operation, this thesis proposes two energy-efficient RL-based algorithms for millimeter wave (mmWave)-enabled unmanned aerial vehicle (UAV) communications toward 6G. This can be especially useful in ad-hoc communication scenarios within a neighborhood with main-network connectivity problems such as in areas affected by natural disasters. This is important since 6G mobile network design must accommodate use cases for mission-critical communications during emergencies. Dynamic selection of suitable hovering spots within the target zone where the battery-constrained UAV needs to be positioned as well as calibrated NOMA power control with proper device pairing are critical for optimized performance. We propose cost-subsidized multiarmed bandit (CS-MAB) and DDQN-based solutions to address these problems jointly.

For 3D-NOMA PHY layer, the proposed AI-aided OFDM PAPR reduction scheme operates at 16.7% faster signal processing speed along with about 12% improvement in the cubic metric measure of the OFDM signal compared to conventional clipping and filtering (SCF) for a 256-subcarrier OFDM PHY layer with QPSK modulation. On the other hand, at the MAC layer level, an 18.4% faster data transfer rates can be accommodated by the proposed DDQN RL agent for cache-enabled NOMA downlink resource management over a capacity-limited backhaul link when operating with 38-GHz mmWaves and 500 MHz system bandwidth. Moreover, significant UAV battery energy savings exceeding 90% can be accomplished for mission-critical, non-terrestrial NOMA downlink transmissions.

To my .....

ACKNOWLEDGMENT

First of all, countless thanks to ALLAH Almighty

I wish to express my sincere appreciation to my supervisor: Assoc. Prof. Dr. Osuma Muta for his continuous support, excellent suggestions, discussions and valuable comments, and patience to improve the quality of this work. He provided me a tailored research environment fostering high degrees of comfort and flexibility. He also provided much-needed moral support during hard turns. The work presented in this thesis was not possible to emerge in its fine-tuned final form without his help and expertise. For this and more, I remain grateful to him.

I would also like to immensely thank Prof. Kohei Hatano, Assoc. Prof. Dr. Sherief Hashima, and Dr. Ahmed Nasser for their valuable contributions and engaging discussions that helped shaping this work. I would also like to thank and express my great appreciation for Prof. Koji Inoue, Prof. Koji Okamura, and Prof. Yutaka Arakawa for their valuable critique and insightful suggestions that helped me to improve this thesis greatly. In addition, I would like to thank the administration of Kyushu University for their efforts to provide a distinctive research university model.

Last but not least, I would like to show my sincere appreciation and gratitude to my colleagues and friends. Thanks to my beloved family for their self-sacrifice, consistent love, support, understanding and encouragement.

Finally, this thesis was supported in part by the Center for Japan-Egypt Cooperation in Science and Technology, Kyushu University Platform of Inter/Transdisciplinary Energy Research (Q-PIT), Support Center for Advanced Telecommunications Technology Research Foundation (SCAT), and Kyushu University. Moreover, the author is supported by a scholarship from the Egyptian Ministry of Higher Education (MoHE) which is gratefully acknowledged.

Ahmad Gendia

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

NOMENCLATURE

$\lvert . \rvert$	Absolute value operator.
$[.]^T$	Transpose operator.
$\mathcal{R}\{.\}$	Real part operator.
$\mathcal{I}\{.\}$	Imaginary part operator.
$\angle(.)$	Radian angle operator.
$\overline{(.)}$	Expectation operator.
$0_{a,b,c}$	An a-by-b-by-c matrix of all zeros.
η	OFDM signals oversampling factor.
N	OFDM sub-carriers.
$x[n]$	n-th sample of time-domain OFDM signal.
$X[k]$	k-th frequency-domain data-bearing modulated symbol.
ω_k	OFDM k-th sub-carrier center frequency.
Ω	Achievable cubic metric level of OFDM signal.
Ω_0	Cubic metric threshold for CCDF calculation.
$\Gamma^{[dB]}$	Raw cubic metric of time-domain OFDM signal.
$\Gamma_{ref}^{[dB]}$	Raw cubic metric reference level.
κ	Cubic metric empirical factor.
U	OFDM selected mapping block size.
$x_u[n]$	n-th sample of the u-th SLM candidate for time-domain OFDM signal.
$x^*[n]$	n-th sample of the SLM chosen candidate for time-domain OFDM signal.
$\mathcal{S} \triangleq \{1, \ldots, U\}$	Set of indices of all candidate sequences within SLM generated block.
R	CF clipping amplitude.
L	Number of operated iterations of the ICF algorithm.

$\hat{x}[n]$	n-th sample of clipped time-domain OFDM signal.
$d[n]$	n-th time-domain sample of CF clipping noise.
$D[k]$	CF clipping noise frequency-domain representation.
$\hat{D}[k]$	Filtered CF clipping noise.
$\tilde{X}[k]$	Post-clipping k-th frequency-domain denoised data symbol.
$\tilde{x}[n]$	n-th sample of CF-processed time-domain OFDM signal.
ρ	SCF noise scaling coefficient.
λ	SCF-ICF equivalency parameter.
σ	Standard deviation of unclipped OFDM signal.
$s_k^{(0)} \triangleq I_k$	The k-th entry in the input layer of a neural network.
$s_j^{(o)}$	The response of the j-th neuron in the output layer of a neural network.
$f_i(.)$	The activation function of the i-th hidden layer.
$s_j^{(i)}$	The response of the j-th neuron in the i-th hidden layer.
$w_{j,k}^{(i)}$	The learnable weight connecting the k-th neuron in the $(i-1)$-th layer to the j-th neuron in the i-th layer.
$b_j^{(i)}$	The learnable bias of the j-th neuron in the i-th layer.
$f_o(.)$	The activation function of the output layer of a neural network.
$\mathcal{D}$	3D training dataset for the SNN-CF algorithm.
$\mathcal{D}(i,.,.)$	The i-th 2D slice of the SNN-CF training dataset.
M	SNN-CF input dimensionality.
P	SNN-CF output dimensionality.
V	Depth of SNN-CF neural module.
h_v	Size of the v-th hidden layer.
$x_i[n]$	Time-domain OFDM input training template for SNN-CF.
$x_o[n]$	Time-domain OFDM output training vector of SNN-CF.
S	Number of OFDM training slices for SNN-CF.
ϕ	SNN-CF neural network parameter set.
$\mathcal{E}_t$	BER threshold for SNN-CF training.
$E_b/N_0\|_t$	SNN-CF training control for E_b/N_0.
Net_b	Optimal network found through SNN-CF training.

$\mathcal{F}_b$	Optimal CCDF found through SNN-CF training.
$\mathcal{R}_{max}$	Total number of scanned mappings in the neighborhood of SCF or ICF.
$Net^{(r)}$	SNN-CF training network at iteration r.
$\mathcal{F}^{(r)}$	Cubic metric CCDF of OFDM signal generated by $Net^{(r)}$.
$E^{(r)}$	BER level achievable by $Net^{(r)}$.
Ω_{0d}	Cubic metric design threshold for SNN-CF scanning.
$f(E)$	Transfer function of the HPA.
E_0	HPA output limiting parameter.
g	HPA small-signal gain.
c	HPA parameter controlling the transition rate from the linear region to the limiting region.
γ	Achievable normalized out-of-band leakage power level.
γ_0	Threshold level for the normalized out-of-band leakage power.
$\mathcal{F}(\gamma)$	CCDF of the normalized out-of-band leakage power.
P_l	Total leakage power.
P_{in}	Power residing in the in-band.
N	Candidates for cache-equipped NOMA downlink system.
M	Active NOMA UE participants.
μ_m	m-th NOMA UE transmit power allotment.
x_m	m-th NOMA UE information-bearing signal.
P_T	Total BS transmit power budget.
R	Sum-rate of activated NOMA UEs.
C	Per-user backhaul link capacity.
C_T	Total backhaul link capacity.
B	Cache-equipped NOMA system transmission bandwidth.
h_m	m-th NOMA UE channel gain.
σ^2	AWGN noise power.
s	Cache-equipped NOMA environment state.
a	Reinforcement learning agent action.
r	Reinforcement learning agent reward.
θ	DDQN agent policy.

Y_j^T	An RL agent's learning target at the j-th interactive experience.
$\bar{\theta}$	DDQN agent target network.
$\mathcal{B}$	DDQN agent experience-gathering container.
τ	Update period of DDQN agent's target network.
π	Generic Q-learning policy.
$Q_\pi(s,a)$	Expected sum of immediate and discounted future rewards gained by taking action a in state s and following the policy π thereafter.
$Q_*(s,a)$	Optimal action value.
$Q_\pi(s,a;\bar{\theta})$	Q-learning approximate parametric representation of $Q_*(s,a)$.
γ	Q-learning discount factor.
β	Q-learning standard update step size.
∇_{θ_i}	Gradient operator with respect to θ_i.
Y_i^Q	Standard Q-learning target value at the i-th update.
$\hat{Q}(.)$	DDQN agent target network response.
Y_i^{DDQN}	DDQN agent learning target at the i-th agent-environment interaction.
$\triangle$	QoS control parameter for reliable SIC operation.
$\mathcal{G}_{max}$	Total number of mini games played by the DDQN agent within the NOMA environment.
$\mathcal{I}_{max}$	The number of interactions per game.
ϵ_s	The starting value of the probability of executing a random action in an agent-environment interaction within the game.
ϵ_f	The final value of random action selection.
ϵ_d	Decay factor to gradually phase out randomized decisioning.
α	Smoothing parameter for DDQN agent's main policy update.
T	Cache status transition matrix.
T_p	The probability that data persists in the cache.
T_c	The probability that cached data is cleared.
T_o	The probability that non-cached data remains out-of-storage.
T_a	The probability that non-cached data is added to the cache.

$u_n(i)$	Binary indicator for the caching status of the data requested by the n-th candidate UE at the i-th agent-environment interaction.
$a_n(i)$	Binary indicator for the activation status of the n-th candidate UE following the i-th agent-environment interaction.
$r_n(i)$	Immediate reward achieved by the n-th candidate UE at the i-th interaction between the DDQN agent and the cache-equipped NOMA environment.
$R(I)$	System overall cumulative reward over interaction step horizon I.
P_b	mmWave blocking probability.
ρ	mmWave blockers density parameter relative to the BS.
$\bar{D}$	Average mmWave blocker length.
δ	BS distribution density parameter.
W	NOMA-UAV-MBS system bandwidth.
g_d	CSI of ground user d in NOMA-UAV-MBS system.
p_d	UAV-MBS power allocation to ground user d.
$X(i)$	UAV-MBS X-coordinate on the grid at step i.
$Y(i)$	UAV-MBS Y-coordinate on the grid at step i.
χ	Battery energy capacity of the UAV-MBS.
$L(i)$	UAV-MBS battery level at time step i.
η	UAV-MBS energy expense per unit distance.
$Z(i)$	Distance traversed by the UAV-MBS at step i.
F_r	Rician channel parameter from UAV-MBS to ground users.
$\bar{g}_d$	LOS channel from UAV-MBS to ground user d.
$\tilde{g}_d$	NLOS channel from UAV-MBS to ground user d.
$\mathcal{L}_i$	NOMA-UAV-MBS system state at the i-th interactive step.
$s_k(i)$	Binary selection indicator of the k-th candidate ground user at step i.
$\mathcal{A}_i$	Action applied by the UAV-MBS agent at step i.
V_q^π	Mean discounted summation of long-term rewards acquired by a UAV-MBS DDQN agent following policy π.

β	UAV-MBS DDQN agent's rewards discount factor.
Ψ_i	UAV-MBS DDQN agent's main policy DNN.
$\bar{\Psi}_i$	UAV-MBS DDQN agent's target network.
$\hat{V}_q$	Target network response of the UAV-MBS DDQN agent.
$\Gamma_k(i)$	SINR at ground user k for the i-th interaction in a NOMA-UAV-MBS system.
μ	SIC detection threshold in a NOMA-UAV-MBS system.
$\mathbf{1}[.]$	Indicator function.
$\Omega(i)$	Filtered candidate set of the UAV-MBS CS-MAB agent at step i.
λ	Cost-subsidizing control parameter of the UAV-MBS CS-MAB agent.
f_i	Utility assigned to arm i by the UAV-MBS CS-MAB agent.
$\mathcal{H}$	Interaction horizon for the UAV-MBS CS-MAB agent.

ABBREVIATIONS

$1G$	First Generation.
$2G$	Second Generation.
$3G$	Third Generation.
$3GPP$	Third Generation Partnership Project.
$3D-NOMA$	Three Dimensional Non-Orthogonal Multiple Access.
$4G$	Fourth Generation.
$5G$	Fifth Generation.
$6G$	Sixth Generation.
ACE	Active Constellation Extension.
ADC	Analog-to-Digital Converter.
AI	Artificial Intelligence.
AM/AM	Amplitude to Amplitude.
ANN	Asynchronous Neural Network.
AR	Augmented Reality.
$AWGN$	Additive White Gaussian Noise.
$B5G$	Beyond Fifth Generation.
BER	Bit Error Rate.
BS	Base Station.
$CCDF$	Complementary Cumulative Distribution Function.
$CDMA$	Code-Division Multiple Access.
CF	Clipping and Filtering.
CI	Close-in (model).
CM	Cubic Metric.
CNN	Convolutional Neural Network.
CS	Compressive Sensing.
CSI	Channel State Information.

$CS-MAB$	Cost-Subsidized Multiarmed Bandit.
$CS-MOSS$	Cost-Subsidized Minimax Optimal Stochastic Strategy.
CSP	Channel State Pairing.
$CS-UCB$	Cost-Subsidized Upper Confidence Bound.
$D2D$	Device to Device.
DAC	Digital-to-Analog Converter.
$DDQN$	Double Deep Q-Network.
$DIPN$	Deep Image Prior Network.
DL	Deep Learning.
DLM	Deep Learning Model.
DNN	Deep Neural Network.
DSP	Digital Signal Processing.
EB	Exabytes.
EVM	Error Vector Magnitude.
FFT	Fast Fourier Transform.
$FFMLP$	Feed-Forward Multi-Layer Perceptron.
$FDMA$	Frequency-Division Multiple Access.
FSI	Full State Information.
GPS	Global Positioning System.
GSM	Generalized Spatial Modulation.
$HAPS$	High-Altitude Platform Station.
HNN	Hopfield Neural Network.
HPA	High Power Amplifier.
ICF	Iterative Clipping and Filtering.
$IFFT$	Inverse Fast Fourier Transform.
IoE	Internet of Everything.
IoT	Internet of Things.
ISI	Intersymbol Interference.
ITU	International Telecommunication Union.
$ITU-R$	ITU Radio-communication sector.
LOS	Line-of-Sight.
LSE	Least Square Estimation.

$mMIMO$	Massive Multiple-Input Multiple-Output.
$MMSE$	Minimum Mean Square Error.
$mmWaves$	MilliMeter Waves.
MAB	Multiarmed Bandit.
MAC	Medium Access Control (Layer).
MAP	Maximum A Posteriori.
$MIMO$	Multiple-Input Multiple-Output.
ML	Machine Learning.
MLD	Maximum Likelihood Detection.
MLP	Multi-Layer Perceptron.
$MOSS$	Minimax Optimal Stochastic Strategy.
MSE	Mean Square Error.
NN	Neural Network.
$NLOS$	Non-Line-of-Sight.
$NOLP$	Normalized Out-of-band Leakage Power.
$NOMA$	Non Orthogonal Multiple Access.
NR	New Radio.
OoB	Out-of-Band (radiations).
$OFDM$	Orthogonal Frequency-Division Multiplexing.
$OFDMA$	Orthogonal Frequency-Division Multiple Access.
OMA	Orthogonal Multiple Access.
$OTFS$	Orthogonal Time Frequency Space.
$PAPR$	Peak-to-Average Power Ratio.
PHY	Physical (Layer).
PSD	Power Spectral Density.
PSI	Partial State Information.
PTS	Partial Transmit Sequence.
QAM	Quadrature Amplitude Modulation.
QoS	Quality of Service.
$QPSK$	Quadrature Phase Shift Keying.
RAN	Radio-Access Network.
RB	Resource Block.

$RBFN$	Radial Basis Function Network.
RCM	Raw Cubic Metric.
RD	Receiving Device.
$ReLU$	Rectified Linear Unit.
RF	Radio Frequency.
RL	Reinforcement Learning.
RNN	Recurrent Neural Network.
RPA	Random Pairing Algorithm.
SCF	Simplified Clipping and Filtering.
$SGDM$	Stochastic Gradient Descent with Momentum.
SIC	Successive Interference Cancellation.
$SIMO$	Single-Input Multiple-Output.
$SINR$	Signal-to-Interference-plus-Noise Ratio.
SLM	Selected Mapping.
$SNN-CF$	Synchronous Neural Network Clipping and Filtering.
SNR	Signal-to-Noise Ratio.
$TT-DNN$	Tensor-Train Deep Neural Network.
TDD	Time Division Duplex.
$TDMA$	Time-Division Multiple Access.
$TFNN$	Time-Frequency Neural Network.
TR	Tone Reservation.
UAV	Unmanned Aerial Vehicle.
$UAV-MBS$	Unmanned Aerial Vehicle Mobile Base Station.
UCB	Upper Confidence Bound.
UE	User Equipment.
UP	User Plane.
VR	Virtual Reality.
ZF	Zero Forcing.

Chapter 1

INTRODUCTION

1.1 6G Trends, Requirements, and Candidate Technologies

For decades, both developed and developing nations have been relying on telecommunications technologies which contribute significantly to the economic evolution of countries and the advancement of their social well-being. Today, wireless communication systems are woven seamlessly into the life of many people around the world. In the near future, widespread adoption of wireless technologies is only expected to expand further. This situation is invariably coupled with great expectations and challenging requirements facing the 6th generation (6G) of telecommunication systems.

1.1.1 Wireless Market: Trending Mobile Services

Figure 1.1 illustrates the two major trends of mobile and wireless services:

- The internet of everything (IoE) [1–4] is being devised to establish wireless connectivity of all things to expedite information acquisition and data tracking processes, thus giving us comprehensive control over our devices and equipment anytime, anywhere. Emergent IoE services cover device-to-device (D2D), and internet of things (IoT) applications [5–8] including robots and mobile sensory equipment, online vehicles and connected smart houses.

- Always-on wireless services delivering deeply-rich contents in real-time over secured wireless links. Emergent applications for such services cover augmented and virtual

reality (AR/VR) [9–18], seamless streaming of realistic 3D holographic video content at extreme resolutions (e.g., 8K and 16K) [19–25], traffic safety applications, haptic internet [26–28], and multimedia-enriched social networking.

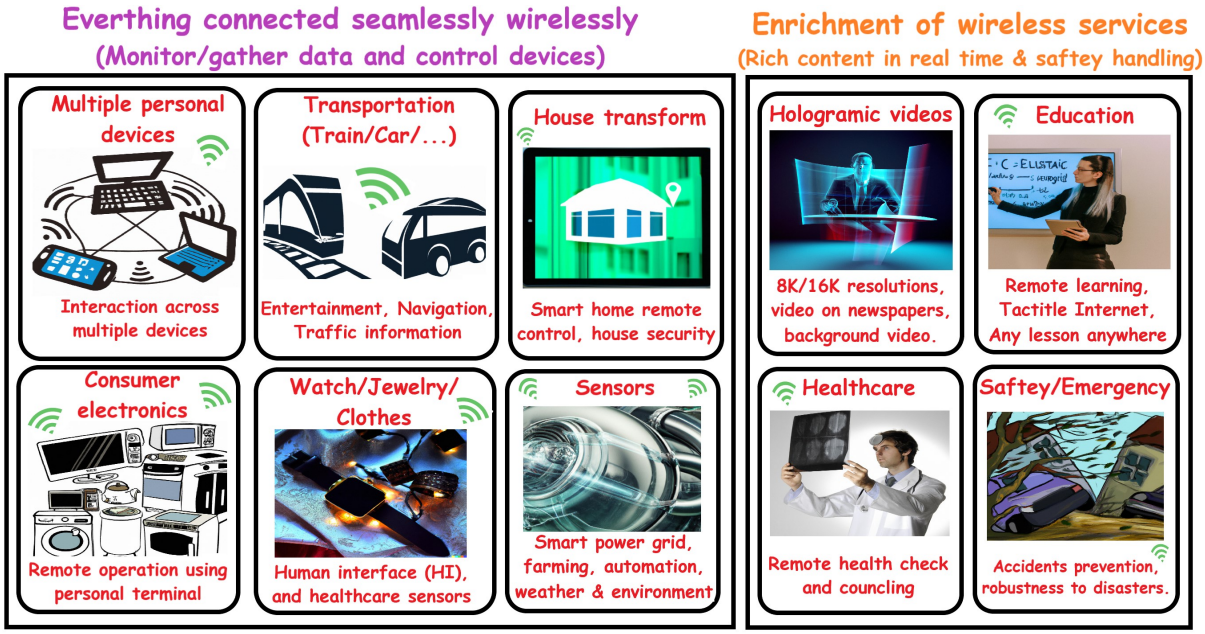


Figure 1.1: Major trends for mobile services evolution towards 2030 and beyond.

1.1.2 6G Requirements

Global data traffic on mobile networks has undergone more than a 10-fold volume uptake between 2017 and 2022, with video content dominating at around 70% of the overall traffic generated in 2022 [29]. Under current trends of the mobile sector, a resultant uptake in data traffic rates is projected to increase steadily within the decade where 5 billion active 5G mobile subscriptions are anticipated by 2028. Moreover, according to the international telecommunication union radio communication sector (ITU-R) report, *ITU-R M.2370-0* [30], total mobile subscriptions by 2028 is estimated at over 15 billion. The report also indicates over a thousand-fold uptake in global mobile broadband traffic by 2030 compared to 2010. In addition, as demonstrated in Fig. 1.1, diversifying services and applications is a key requisite of future 6G mobile networks wherein the system must support a wide range of applications and services from small-packet services (e.g., low-rate D2D communications and real-time remote control) to applications requiring rich contents (e.g., AR/VR, beyond ultra high definition streaming, and tactile internet). Beside the exponential uptake in traffic volume, large variations are anticipated as a result of the skewed distribution of traffic data across mobile networks: traffic data reach peak

volumes in high-density regions (e.g., stations, shopping malls, and stadiums). Therefore, to handle dynamic scenarios for future 6G mobile systems and support end users with suitable quality of service (QoS) levels, 6G networks must possess inherently adaptive high-level design methodology which can fulfill the associated challenging requirements illustrated in Fig. 1.2:

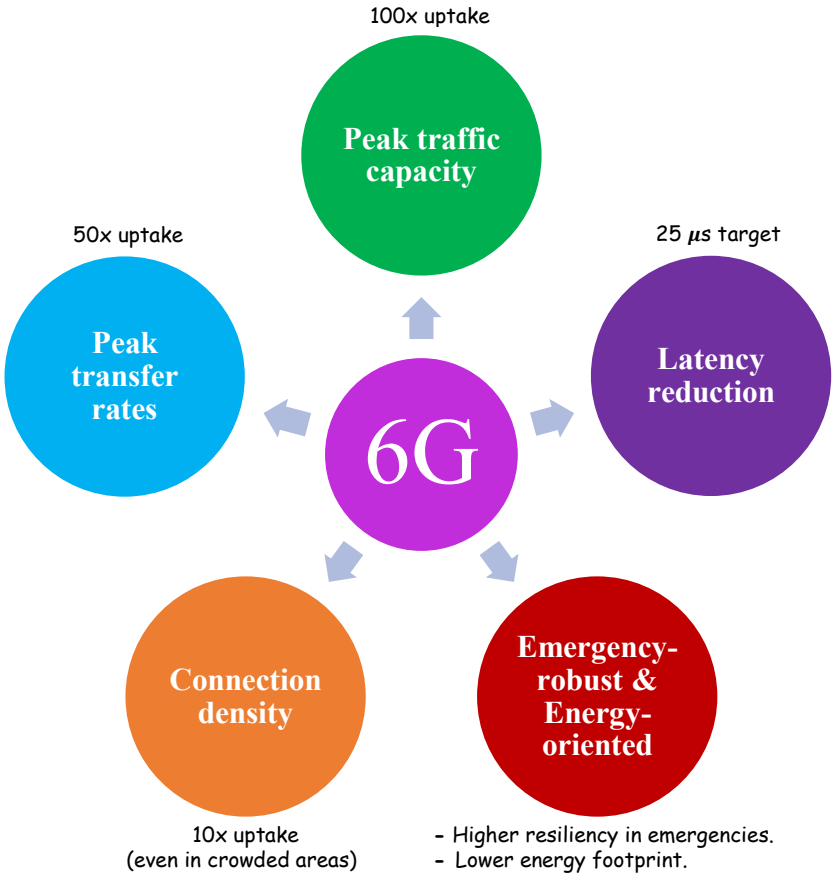


Figure 1.2: 6G Targets [31, 32].

- **Peak traffic capacity:** One important yet challenging requirement for future 6G systems is that 6G high-level design must incorporate the capability to process extremely large volumes of intense traffic which will normally be orders of magnitude larger than the usual volumes flowing through previous generations of mobile networks. A target of hundred-fold uptake is set compared to 5G area traffic capacity/m^2 [31]. In this regard, novel smart solutions to alleviate traffic congestion is a key enabler for 6G.

- **Peak transfer rates:** 6G high-level design must accommodate practical data transfer speeds at much higher levels compared to what older generations' technologies could achieve. This is in line with emerging applications that require

high-speed real-time communications such as tactile internet and holographic communications. A requirement of at least 50-fold increase in peak rates compared to 5G networks is set [31]. It will be crucial to reach such high communication speeds in certain transmission scenarios such as those incorporating mobile backhauling.

- **Connection density:** 6G high-level mobile network design has to prepare for a surge in the number of supported simultaneous connections to facilitate support for cloud-based applications that are always-connected while providing IoT connections with a wider variety of connected machines. A target of 10x uptake from 5G is set in order to support about 10^7 active devices per squared-km [31].

- **Latency reduction for radio-access network (RAN) interface:** Beside providing faster connections, 6G high-level mobile network design is required to operate at as low latency levels as 25 μs over the user plane RAN (UP-RAN) interface [31]. Such a reduction is a large leap from the operation point of 5G systems which is typically around 1 ms. This stringent requirement on extremely low latency levels is essential for the adequate establishment and operation of new latency-sensitive application services including adaptive instant control of real-time D2D operations, tactile internet, AR/VR, and holographic-based communications.

- **Emergency-robust energy-oriented design:** 6G high-level mobile network design must adopt a robust, cost-effective methodology that is primarily energy-oriented while being resilient and flexible enough to withstand and handle natural disasters more gracefully. A 6G target of 10$\sim$100-fold improvement upon 5G in terms of energy efficiency is set [32]. Such design approach is of great importance to give future 6G networks the ability to operate in a wide range of diversified communication environments with embedded support for multiple kinds of applications and services. In addition, densifying network components distribution further via increased mirco/nano-cell deployments with the ability to quickly dispatch ready-to-deploy unmanned aerial vehicles (UAVs) in conjunction with millimeter waves (mmWaves) utilization is key element in meeting the flexibility and high-throughput expectations of 6G systems [33].

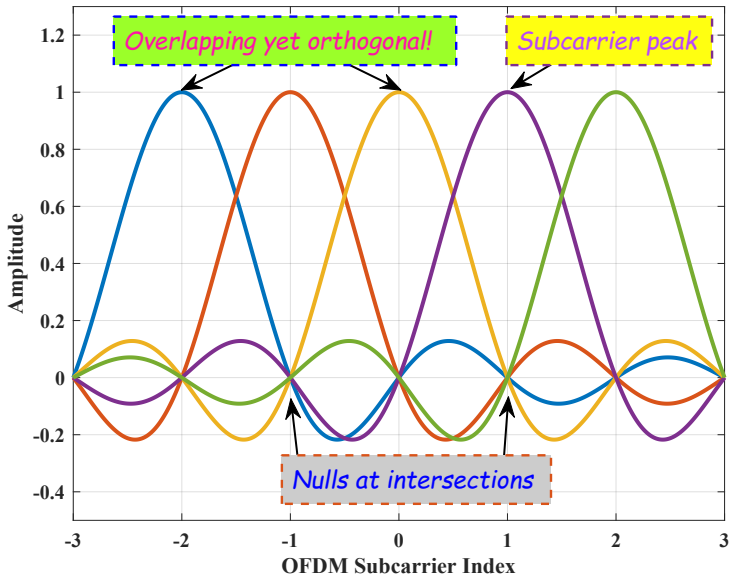


Figure 1.3: Multicarrier modulation: OFDM signaling.

1.1.3 6G Candidate Technologies

Future 6G networks can conform to the aforementioned requirements through the aggregate deployment of an array of various prominent candidate solutions. These technologies include:

- **OFDM:** Orthogonal frequency division multiplexing (OFDM)-based transmissions have long been the standard backbone of communication systems since its adoption in 4G networks. In its 5G new radio (NR) specifications, 3GPP has extended OFDM lifetime by choosing it as the standard underlying baseband formatting scheme for all NR channels [34]. OFDM systems offer a myriad of benefits including high spectral efficiency as well as low-cost receiver implementation. In OFDM-based transmissions, densely-packed multicarrier signaling is employed where subcarriers are designed to overlap in an orthogonal manner as depicted in Fig. 1.3. From a link-level perspective, OFDM can provide an efficient underlying PHY framework for the realization of 6G wireless networks. Nonetheless, in order to reap the benefits of OFDM-based communications, proper handling of its PHY layer characteristics must be adopted. Specifically, the long-standing issue of high peak-to-average power ratio (PAPR) must be addressed efficiently at the transmitter side prior to transmission.
- **Millimeter Waves:** For decades, sub-6 GHz operation dominated almost every wireless transmission mode (e.g., GPS, Wi-Fi, AM/FM radio, high-definition TV,

cellular and satellite communications). Wireless commercial applications favored this particular range within the RF spectrum range (known as the *sweet spot*) since it possesses attractive propagation characteristics. Nonetheless, confining the entirety of wireless communications applications to an extremely narrow part of the entire RF spectrum is poor resource utilization, and, more importantly, will no longer be a viable option going forward given the ever-mounting demand for more simultaneous connections as well as increased user-experienced transfer rates. Therefore, the authors in [35] were first to explore operation at higher frequencies using millimeter waves (mmWaves)-based transmission for a proposed mobile broadband design to operate in 5G systems. To utilize the abundant RF spectrum more efficiently, the authors made the suggestion to go beyond 3 GHz for wireless commercial applications operation. The huge 30-300 GHz bandwidth of mmWaves and sub-terahertz signals can unlock unrivaled support for user-experienced data rates as well as carry extreme amounts of mobile networks traffic [36–40].

- **Non-Orthogonal Multiple Access:** Multiple access scheme format is a key technology that has always been a distinguishing feature of mobile systems generations (e.g., 1G adopting FDMA, 2G adopting TDMA, 3G adopting CDMA, 4G adopting OFDMA, and 5G adopting an upgraded OFDMA version that has multiple numerologies to provide higher degrees of flexibility and accommodate more emerging use cases). Going forward, transmit base stations of future 6G networks (e.g., satellites, high-altitude platform station (HAPS) transmitters, UAVs, and terrestrial-based BSs in Fig. 1.4) can adopt a three-dimensional, intracell, multiuser non-orthogonal multiple access (3D-NOMA) format to support even denser connectivity requirements and boost the system's spectral efficiency by exploiting the received signals variations to leverage the power gaps at the receiving terminals for both terrestrial as well as non-terrestrial transmissions in Fig. 1.4. This exploitation of the underutilized power gaps has not been implemented in previous FDMA-, TDMA-, CDMA-, and OFDMA-based generations. A key concept of NOMA downlink communications is that non-orthogonality is injected intentionally by aligning a stack of receiving equipment onto unified resource blocks in time, frequency, and code domains, while using distinctive separable levels within the available BS power budget as shown in Fig. 1.5. Process reversal and extraction of

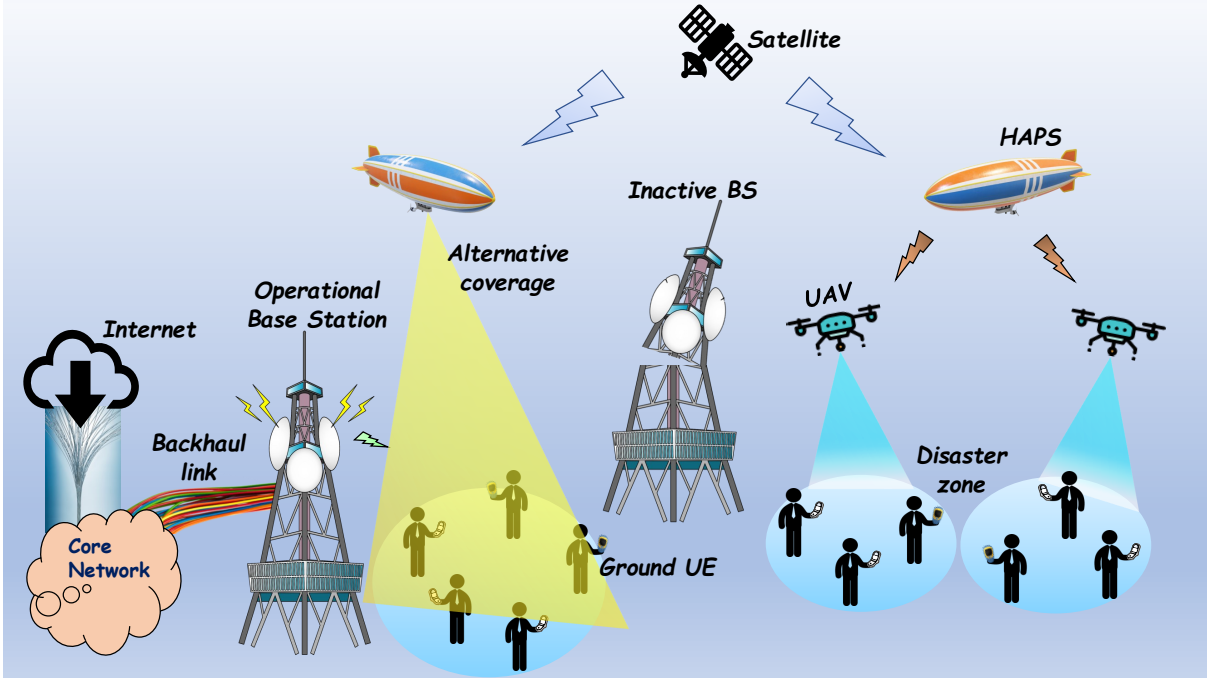


Figure 1.4: 3D-NOMA 6G system top-level view.

each device's intended signal can then be carried out at the respective receivers by employing successive interference cancellation (SIC) methods [41, 42]. The key idea is to transform the wireless channels' gain gaps among grouped users to multiplexing gains through the superposition of their signals with distinct shares of the transmit BS power. Evidently, multiplexed users can then reap the main benefits of experiencing scheduling more frequently as well as enjoy larger aggregated bandwidth [43]. System overall capacity as well as user density can consequently be noticeably improved. However, the ability of 3D-NOMA to deliver on its great potential is largely determined by the degree of success in optimizing NOMA user selection and power allocation within various deployment scenarios. A highly successful, system-level NOMA resource management strategy design is thus imperative for 3D-NOMA to be considered for future network access frameworks in order to meet the challenging requirements facing 6G. This is particularly pressing for future 6G networks where intense loads are expected which might cause throttling bottlenecks at the backhaul interface connecting the base station to the core network if the operation is not finely-tuned. In addition, energy-efficient dynamic trajectory planning is important for mission-critical, UAV-assisted emergency communication scenarios such as in disaster zones as shown in Fig. 1.4. In such scenarios where the main serving base station is out-of-commission due to sustaining damage, the UAV mo-

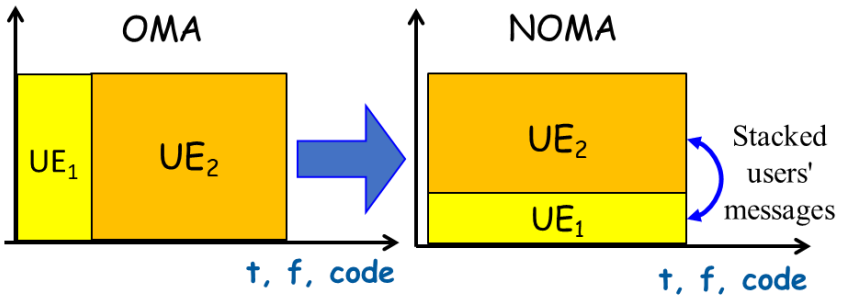


Figure 1.5: NOMA: Intentional inter-user non-orthogonality.

bile base station can accurately obtain its location information within the afflicted area through HAPS or satellite links. Moreover, optimizing the underlying PHY layer OFDM signal processing shown in Fig. 1.6 is key for successful realization of 3D-NOMA.

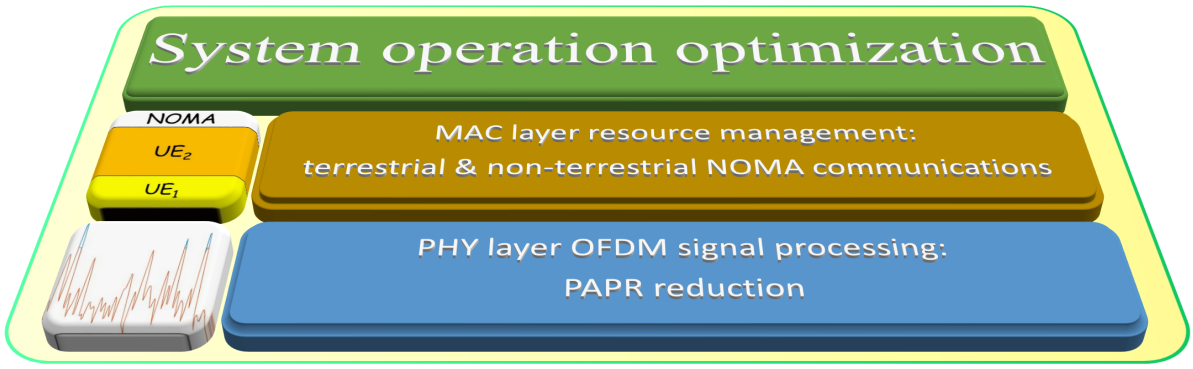


Figure 1.6: System layers optimization: MAC layer user selection and power management & underlying PHY layer processing.

1.2 Motivation

Owing to their transformative technological power, future 6G wireless networks are at the heart of the current scientific research enterprise. To reach its potential, the research and development phase of 6G must explore and utilize novel key technologies (e.g., 3D-NOMA) efficiently both at the PHY layer as well as the MAC layer levels. OFDM is an efficient backbone realization scheme for NOMA in wideband channels. Link-level optimization of 3D-NOMA can be achieved by optimizing the underlying OFDM PHY layer operation. On the other hand, system-level operation tuning of 3D-NOMA can be accomplished via MAC layer design optimization for both terrestrial and non-terrestrial (e.g., UAV-based) communications modes. AI-based operation is a very promising framework to revolutionize the way 6G wireless communications systems are designed and operated, including 3D-NOMA networks. However, in order to successfully reap the benefits of

AI-based operation, challenging problems facing the aforementioned potential 6G candidate technologies in general and 3D-NOMA networks in particular must be handled properly and efficiently. Specifically, novel AI-based PAPR reduction approaches for OFDM-based transmissions is important to optimize PHY layer operation. On the other hand, to optimize MAC layer operation for mmWave-enabled NOMA-based downlink communications, novel AI-based algorithms are needed to handle user pairing and power control design issues as well as dynamic UAV path planning for UAV-based transmissions. This is of great importance to satisfy the high-throughput and energy efficiency operation requirements set forth for future wireless systems.

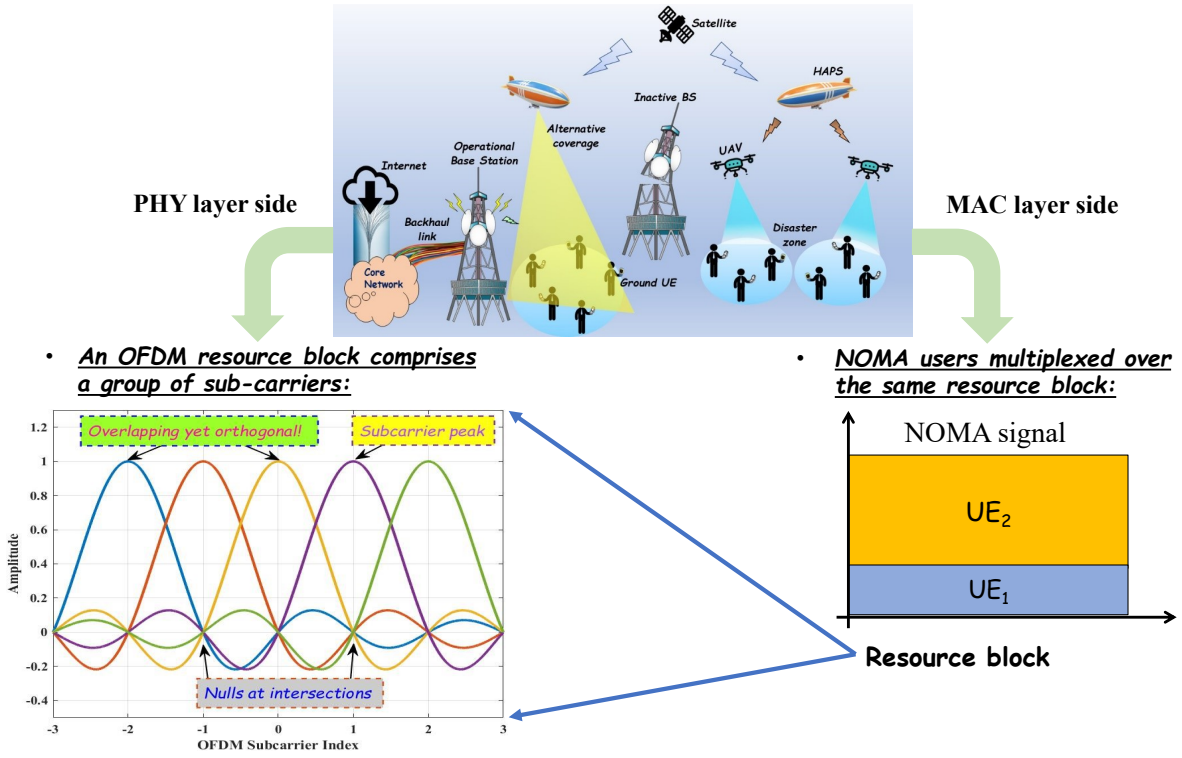


Figure 1.7: 3D-NOMA layers interplay: The MAC layer selects candidate users to grant access to the same underlying OFDM resource block of the PHY layer which processes the signals of the selected users prior to transmission.

1.3 Scope and Objectives

This thesis focuses on the challenges facing the successful design and operation of 3D-NOMA networks. Specifically, both the PHY and MAC layers must be properly optimized. Figure 1.7 demonstrates the relationship among the two layers in 3D-NOMA. At the MAC layer level, users are granted non-orthogonal access to the same underlying PHY layer resource block. As such, the selection of users and their shares of the transmit

power budget must be properly tuned at the MAC layer side. On the other hand, prior to transmission, the signal of each multiplexed user must be processed at the PHY layer to ensure desirable transmission characteristics. This is especially important for NOMA-based systems where multiple users are assigned the same underlying OFDM resource block of the PHY layer as shown in Fig. 1.7. Therefore, for the PHY layer, the scope is OFDM-based signal formatting at the transmitter side where noticeable PAPR reduction is the main design focus. On the other hand, MAC layer optimization focuses on tackling the resource management aspects of mmWave-enabled NOMA-based downlink transmissions. Moreover, to accommodate mission-critical use cases of 3D-NOMA, we extend the resource management optimization to cover non-terrestrial, UAV-based communications where battery-friendly operation is important for energy-efficient UAV deployment.

We, therefore, identify the specific research objectives of the thesis as follows:

- Designing low-complexity neural networks to format OFDM signals with desirable PAPR characteristics for optimized PHY layer operation of 3D-NOMA networks.
- Designing a powerful deep reinforcement learning based solution at the MAC layer level to handle the NOMA user selection and power allocation problem for sum-rate maximization of 3D-NOMA.
- Designing multiarmed bandit and deep reinforcement learning algorithms for battery-friendly dynamic UAV trajectory planning and joint resource management for deployment in NOMA-based downlink emergency communications.

1.4 Original Contributions

We describe the main contributions of the thesis which can help realize the aforementioned design objectives and pave the way towards a more efficient operation of 3D-NOMA networks toward future 6G systems. The generation and processing of PHY layer transmitted signals with low-complexity is achieved in chapter 3 to enable reduced latency services. Chapter 4 targets the provision of higher data rates to facilitate for higher-resolution video streaming and the adoption of immersive AR/VR 6G applications. Chapter 5 contributes battery-friendly design for energy efficiency handling in non-terrestrial 3D-NOMA emergency communications. A detailed breakdown of the contributions of the research work presented in this thesis is given:

1. Chapter 3 contributions focus on a low-complexity PHY layer signal processing scheme to accelerate the adoption of 6G applications which have tight low-latency operation demands such as anytime/anywhere remote learning, and transportation control and real-time traffic monitoring:

    - As an efficient PHY layer, low-complexity PAPR reduction approach, we propose a synchronous neural network-based OFDM PAPR solution surpassing conventional clipping-and-filtering (CF)-based approaches. The proposed method accurately generates CF-like output signals by synchronously learning both real and imaginary parts of raw OFDM input signals and mapping them to the desired target output. Hence, higher PAPR reduction performance and lower OoB power leakage can be achieved compared to conventional asynchronous NN solutions while operating in a low-complexity mode, thus making it more suitable for 6G applications with low-latency requirements.
    - To find a neural network with desirable characteristics, we propose a network-selection algorithm to explore a wide variety of candidate solutions during the training stage. The algorithm scans the neighborhood of conventional SCF or ICF searching for the mapping network with the highest PAPR reduction for a maximum allowable BER threshold.
    - To prevent overfitting traps and arrive at a network model that generalizes well to new data during deployment, we propose an algorithm for the compilation of a 3-D training dataset that randomly scatters the training samples of the same OFDM symbol across multiple different 2-D slices within the dataset. Thus, hybrid OFDM symbols are constructed from the time-domain samples of multiple different raw OFDM symbols during the training stage.
    - Through the numerical evaluation of the system performance, we show that the proposed SNN-CF scheme achieves superior performance in terms of PAPR reduction capability, while exhibiting less OoB radiation for both QPSK and 16-QAM compared with ANN. Moreover, the proposed scheme maintains adequate BER levels for higher order modulations such as 16-QAM, where the asynchronous approach shows rapid BER degradation.

2. Chapter 4 contributions focus on a reinforcement learning (RL)-based MAC layer

resource management design to deliver boosted data rates and alleviate data traffic congestion at the core network which can have a positive impact for 6G applications with increased speed requirements such as extremely high-definition video transmissions and AR/VR services:

- We propose a cache-enabled RL-based dynamic power allocation and user selection scheme for opportunistic access to downlink NOMA networks with capacity-limited backhaul connections where selected UEs are activated opportunistically based on rate maximization. The scheme utilizes a double deep Q network (DDQN) agent to interact with the NOMA environment to gain the experience needed to attain higher sum-rate. The proposed agent monitors the state of the NOMA environment and learns to update its action policy towards optimal power allocation and user selection decisions. This RL-based design can boost downlink speeds considerably while reducing the backhaul link bottlenecks at the core network side to pave the way for 6G applications facing transfer rate challenges (e.g., holographic video services and accurate weather data acquisition and environment control).
- To demonstrate the merit of utilizing cache state information by the proposed scheme, we study the comparative performance of the proposed RL agent with the availability of full state information of the NOMA environment vs partial state information availability where only the channel-gain conditions of candidate users are available to the agent.
- We study the robustness of the proposed scheme in a variety of settings including cache-enabled and cache-disabled communications, different levels of backhaul throttling, and the blockage effects of *mmWaves* transmissions. Numerical analysis of the achievable performance shows that the proposed RL-based scheme can consistently achieve near-optimal sum-rate performance. Moreover, the proposed agent exhibits adequate convergence performance towards the optimal long-term mapping policy.

3. Chapter 5 contributions focus on 3D non-terrestrial NOMA network operation extension to accommodate 6G emergency use cases where ready-to-dispatch UAVs can be quickly deployed to support fast downlink channels with battery-optimized UAV

trajectory planning towards an environment-friendly, energy-efficient 3D-NOMA network operation:

- We propose two RL-based schemes for energy-efficient UAV trajectory course planning and joint downlink NOMA power allocation and receiver selection. Both schemes are operated within battery-constrained UAV-NOMA environments with dynamic wireless channels. This is in line with the broader vision of a more comprehensive 6G system where reliable, energy-efficient non-terrestrial communication availability is of great importance.
- The DDQN RL agent is trained to absorb the underlying characteristics of the UAV-NOMA environment within the DNN it uses to implement its action-selection policy. We define the appropriate UAV-NOMA states, actions, and rewards so that the trained agent can achieve energy-efficient near-optimal sum-rate performance when deployed for operation.
- The MAB-based agent is configurable with either CS-UCB or CS-MOSS operation modes and can learn to quickly converge to a highly-rewarding long-term operation policy by scanning the search space while balancing the exploration-vs-exploitation issue through the dynamic evaluation of various arms' utilities. The agent makes on-the-fly decisions and update them as needed.
- We operate and test the proposed solutions within *mmWave*-enabled UAV-NOMA environments having line-of-sight (LOS) signals of variable strength.

Thus, chapter 3 provides an optimized underlying PHY layer foundation for the overall design of 3D-NOMA network, whereas chapter 4 calibrates its MAC layer design for optimized transmission speeds, and chapter 5 accommodates non-terrestrial communications mode where an energy-efficient UAV path design is incorporated for environment-friendly deployments.

1.5 Thesis Structure

The aforementioned objectives of the thesis are tracked in the subsequent chapters. Firstly, a literature review on AI-aided wireless communications is presented in chapter 2. Afterwards, OFDM PHY layer optimization is tackled in Chapter 3 whereas MAC

layer resource management operation optimization is addressed in chapters 4 and 5 for NOMA-based downlink systems:

- Chapter 2 gives a holistic idea on various AI-aided solutions for PHY and MAC layer optimization of wireless communication systems. The chapter presents PHY layer research work on DNN-based channel estimation and signal detection at the receiving devices as well as multicarrier OFDM-based signal processing at the transmitting side. In addition, the chapter presents MAC layer research work on computation and communication resource management operations for optimized utilization of the available resources.

- Chapter 3 describes in detail a proposed synchronous neural network-based OFDM PAPR solution surpassing conventional CF-based approaches. The proposed method accurately generates CF-like output signals by synchronously learning both real and imaginary parts of raw OFDM input signals and mapping them to the desired target output.

- Chapter 4 describes in detail a proposed cache-enabled RL-based dynamic power allocation and user selection scheme for opportunistic access to downlink NOMA networks with capacity-limited backhaul connections where selected UEs are activated opportunistically based on rate maximization.

- Chapter 5 describes in detail two proposed RL-based schemes for energy-efficient UAV trajectory course planning and joint downlink NOMA power allocation and receiver selection. Both schemes are operated within battery-constrained UAV-NOMA environments with dynamic wireless channels.

- Chapter 6 draws conclusions on the work presented in the thesis based on the results obtained in previous chapters. The chapter also outlines potential directions for related future research work.

Chapter 2

AI-Aided Wireless Communications

2.1 Introduction

AI integration into future 6G systems will enable widespread adoption and realization of a variety of diverse applications and services. As 5G is gradually unrolling over the globe, both the academic and industrial sectors are steadily shifting focus towards conceptualizing a 6G roadmap with AI-based technologies and solutions at its core. AI-driven operation management and system design as well as protocol optimization are key features of future wireless systems to enable ubiquitous AI applications throughout the network from the core and all the way to the terminal equipment [44].

In this regard, promising studies on AI-aided operation have been proposed to revolutionize the design and operation of various components within the wireless communication system to surpass the limits of conventional technologies and fulfill the ultimate vision of future 6G networks. These studies cover various important PHY layer operations such as channel estimation modeling, signal detection, and multicarrier modulation as well as critical MAC layer operations such as efficient utilization and management of compute and communications resources as shown in Fig. 2.1. In this chapter, we present an overview of relevant prominent AI-based solutions. In addition, we also present related works to our core goal of optimizing resource management operation in 3D-NOMA systems for maximized achievable sum-rate performance in addition to the optimization of the underlying PHY layer signal processing for more efficient implementation of 3D-NOMA towards 6G.

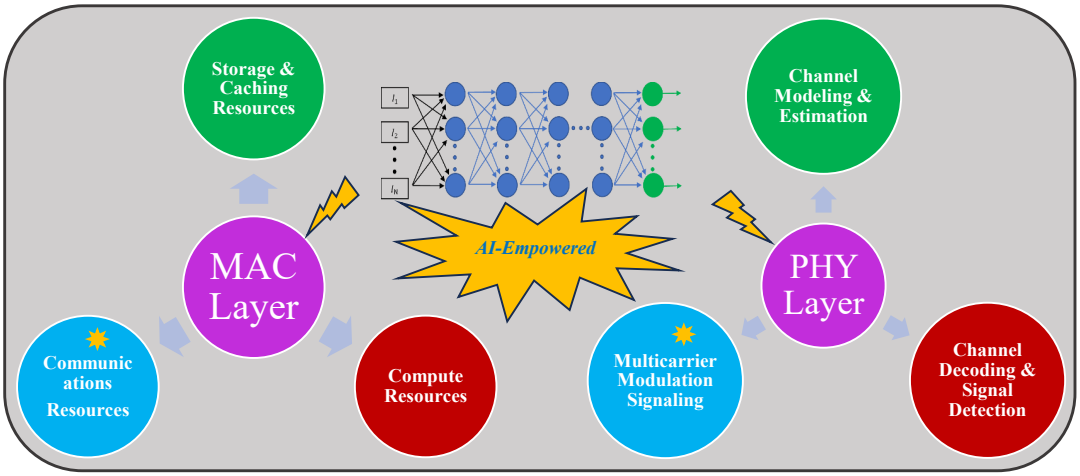


Figure 2.1: AI-empowered wireless communications: PHY & MAC layers operation optimization.

2.2 AI-Driven PHY Layer Operation

The physical layer of wireless communications networks is suffering numerous impediments which include hardware-related issues (e.g., oscillators leakages and amplifier-induced nonlinear distortions) as well as channel imperfections (e.g., interference, jamming, and fading). In the presence of such challenging impairments, efficient and reliable communications require careful designing of a large number of system-wide control parameters that must be optimized and fine-tuned. 6G PHY layer design is envisioned to incorporate AI technologies to self-learn and self-optimize with automated data-collection and sensing mechanisms, and advanced low-complexity signal processing techniques.

2.2.1 AI-Aided Channel Estimation

Realistic wireless communications channels usually comprise both deterministic and random components, varying in time and also depending on the environment. Therefore, conventional channel models construct stochastic models to represent the transmission medium and benchmark various transmit-receive approaches. These stochastic models of the wireless channel are a kind of abstraction that results from measurements taken in a variety of transmission scenarios. They cannot accurately predict the true realistic channel state of a given environment at an arbitrary given moment [45].

Going beyond 5G, the goal of AI-aided channel modeling is to absorb the knowledge of the transmission environment characteristics to enable powerful channel predication and state assessment capability by means of well-trained deep neural networks (DNNs)

for example. The idea is to firstly collect and build a large dataset of true channel measurements from actual field tests. The nest step is to utilize DNNs to learn and extract the unique features of the channel from the training data. Lastly, inferences can be made for important tasks such as channel feedback and predication [46, 47].

The authors in [48] developed a pre-training self-supervised DNN technique that captures the state of the wireless channel during multiple tasks without compromising the personal data of the users as training labels. The process is automated for label samples generation and the DNN model learned to resolve the wireless channel's features. Subsequently, the learned model of the channel can be embedded and further tuned for a specific task such as system optimization for downlink communications. Various AI models can be employed including auto-regression models [49] and DNN-based models [50]. The authors in [51] proposed an online DNN-based approach for channel estimation in doubly selective fading environments. The technique utilizes a DNN to extract the specific properties of the varying channel given previous estimations and properly selected data while also harnessing the extra feature information extracted from received signals and pilots. The proposed DNN method exploits least-square estimations to enhance the estimation performance. In the proposed method, the DNN trains on simulated datapoints at first during offline time. Then, the network starts tracking the dynamic behavior of the channel during online usage. A pre-training method is used for the refinement of the initialization settings of the DNN parameters. The technique presented does not require prior information on the channel's statistics making it a good choice for scenarios having modeling errors or mobile vehicular environments. The authors in [52] proposed a channel estimation approach based on DL for massive multiple-input multiple-output (mMIMO) communications. The technique focuses on scenarios where the number of transmitting antennas is larger than the pilot length. The technique performs a two-phase process to estimate the channel: firstly, a DL-aided pilot-based estimation of the wireless channel, then a DL-aided data-based estimation. For the initial phase, the channel estimator is designed jointly with the pilot via a 2-layer NN as well as a DNN. Subsequently, the next phase is used to improve the estimation accuracy via a second DNN using an iterative approach. The authors in [53] proposed a low complexity, DL-assisted channel estimation mechanism for RIS-aided single-input multiple-output (SIMO) OFDM systems suffering from hardware-related impairment. The technique uses an untrained DNN relying on

DIPN to reduce the noise in the system's pilot-based effective channel which is based on least-square estimation (LSE) with the target to arrive at more accurate estimations. The authors in [54] proposed a 2-phase DNN-aided channel estimation approach for mMIMO wireless communication networks experiencing nonlinear amplification distortion. The method can learn both a DNN-based estimation of the wireless channel and the nonlinear transfer characteristic of the amplifier in an online fashion by harnessing the information gleaned from real-time measurements of the received pilots, while simultaneously constructing the corresponding channel estimation. The authors accomplish this task by designing an online-based DNN loss function independent from the base channel while incorporating the nonlinear distortion into account. The approach's accuracy was higher compared to conventional compressive sensing (CS) techniques while operating within lower time frames for channel inference. The authors in [55] proposed a tensor-train DNN (TT-DNN) approach for time-varying MIMO channel estimation. A centralized estimator based on DNN is constructed using parallel-path distributed TT-DNN. The TT-DNN method provided a compressed model by unraveling the DNN layers in a TT form that utilizes a smaller parameter count. The proposal is adapted for the number of antennas, pilot density, and block structure. To reduce the input dimensionality and the parameter count, the technique performs the estimation for one block and one antenna at a time. Additionally, an initial pre-training phase is used for the TT-DNN based estimation to improve the representation accuracy. The model is also fitted to the semi-blind scenario in which only the preamble data is available.

2.2.2 AI-Aided Reception

The basic key idea regarding the application of AI-based solutions in general and DNNs in particular to wireless communications systems is the replacement of conventional building blocks within the network with intelligent, adaptable DNNs. Two examples are illustrated in this section: signal detection and channel decoding. In both cases, DNN-based optimization attempts are presented wherein the ability of low-latency DNNs is harnessed to accurately model the system's inherent non-linearities and imperfections.

2.2.2.1 Signal Detection

The authors in [56] proposed a DNN-based detection scheme for signals modulated using the orthogonal time frequency space (OTFS) scheme. The authors considered an

architectural DNN structure that assigns a dedicated DNN for each multiplexed symbol within the delay-Doppler net. This is opposed to the alternative format where the DNN jointly handles all the symbols of the OTFS frame under processing. The symbol-oriented DNN structuring requires lower processing complexity as it harnesses a reduced number of parameters. The authors in [57] proposed a DNN-based block detection scheme for RIS-assisted generalized spatial modulation (GSM) MIMO communications to determine active antennae and retrieve the transmit signals at the receiving side. Standard signal detection schemes (e.g., zero-forcing (ZF), maximum likelihood detection (MLD), and minimum mean square error (MMSE) detector) have also been considered. The proposed work managed to achieve similar BER levels as the MLD while operating at lower complexity. The authors in [58] proposed a joint channel estimation and signal detection scheme for OFDM networks. The scheme applies DNN-based end-to-end processing for OFDM channels to produce an implicit estimation of the CSI whilst directly detecting the transmitted signals. To handle channel-induced distortions, the authors used simulated data that they generated using the statistics of the wireless channel to train a DLM in an offline setting, and employed the trained model to directly retrieve the transmit symbols during online operation. The proposed scheme exhibited similar performance to the MMSE detection approach.

2.2.2.2 Channel Decoding

The authors in [59] proposed a one-shot multi-layer perceptron (MLP)-based channel decoding algorithm. A black-box style was adopted for the MLP training where no prior knowledge utilization is assumed as is usually the case for conventional decoders. The received corrupt codeword is decoded by feeding it as an input to a pre-trained MLP decoding network. The MLP network processes the input codeword to map it onto one of the valid codewords at the MLP output. The proposed approached was evaluated and has shown promising decoding capabilities of DNN-based operation which can be on bar with the maximum a posteriori (MAP) algorithm. The authors in [60] investigated three NN decoding architectures: an MLP, a recurrent neural network (RNN), and a convolution neural network (CNN). The same parameter count was fixed for the three decoders. RNN-based decoding provided the best decoding performance among the three tested approaches at the expense of more computational demand. For all three types, the authors also reported a saturation length for the input codeword beyond which DNN

underfitting and overfitting problems started to arise due to their restricted learning abilities. The authors in [61] proposed TurboNet, a model-driven DL-based turbo decoding scheme which integrated DNN-based processing into conventional MAP decoding. TurboNet was designed by unraveling the originally iterative format in terms of DNN blocks. Learnable parameters were incorporated into the structure of the MAP decoder, and subsequently tuned via the supervised learning approach. TurboNet was subsequently pruned to produce TurboNet+, a more compact version with comparable performance. TurboNet+ showed reduced computational cost which helps to alleviate decoding overhead. The evaluation of the proposed work showed promising error-correction capability and adaptability to a variety of transmission test cases.

2.2.3 Multicarrier Modulation (OFDM) Signaling

OFDM is an efficient multicarrier modulation scheme that densely packs a group of overlapping yet orthogonal subcarriers into resource blocks to achieve higher utilization efficiency of the available spectrum bandwidth, while also combating undesirable inter-symbol interference (ISI) issues caused by the multipath nature of realistic wireless communications channels which induces frequency-selective fading. This is accomplished by converting the single-stream, information bearing signal to multiple parallel streams each with smaller bandwidth requirements while maintaining the overall original bandwidth fixed. This effectively renders the frequency-selective channel into a flat-fading channel over each individual subcarrier. DSP-based operation facilitates for an easy implementation of OFDM modulation using FFT operations. Although OFDM systems enjoy a wide range of benefits, the generated time-domain OFDM signals suffer from a typically high peak-to-average power ratio (PAPR). Due to the nonlinear nature of the amplification stage at the transmitter side, a high PAPR would cause severe in-band signal distortion as well as induce out-of-band (OoB) emissions.

To reap the benefits of OFDM systems, an efficient PAPR reduction scheme is needed to address this long-standing, critical issue. To this end, various clipping and filtering PAPR reduction approaches have been proposed in the literature [62–68]. The iterative CF (ICF)-based technique is effective in reducing the PAPR level of the OFDM signal [62]. The technique relies on time-domain clipping of the signal amplitude whenever it exceeds a certain threshold which then gives rise to high-frequency components. Therefore, ICF

then applies a frequency-domain filtering operation to suppress the generated OoB emissions. The ICF algorithm applies IFFT and FFT operations to switch to the time and frequency domains, respectively. Unfortunately, when the OoB emissions are suppressed, the time-domain signal amplitude experiences some regrowth. Thus, ICF keeps switching between the time and frequency domains in an iterative manner to clip the amplitudes exceeding the clipping threshold while ensuring that OoB emissions are suppressed. This repeated application of IFFT/FFT operations is computationally expensive. The authors in [65] thus proposed a single-shot, simplified CF-based (SCF) processing approach to arrive at the same PAPR reduction performance of ICF in a single iteration. In SCF, the time-domain OFDM signal is generated by applying the standard IFFT operation which is then clipped according the limiting threshold. Afterwards, the clipping noise is generated (as the difference between the clipped and unclipped signals) and converted to the frequency domain where it is filtered, scaled by a factor depending on the clipping amplitude and the OFDM signal variance, and then subtracted from the original frequency-domain input signal before producing the SCF scheme's final output by applying an IFFT operation. This cascade of operations provides the same PAPR reduction capability of ICF at relatively lower complexity. However, the processing complexity of SCF remains relatively high. Neural network-based processing was proposed to reduce the system's complexity while providing the same PAPR reduction capability [69]. Two NN modules were utilized to process the in-phase (I) and quadrature (Q) parts of the input OFDM signals independently. The two network modules are trained to output the I and Q components of the output signal, respectively. The scheme managed to provide the same PAPR reduction capability of conventional SCF. However, the asynchronous nature of the NN-based processing approach can lead to underfitting or overfitting issues, resulting in underoptimized OoB operation due to sensitivity to mapping errors. In addition, although conventional NN-based operation can achieve the same PAPR reduction performance, it does not provide enhanced PAPR reduction compared to conventional CF-based approaches.

2.3 AI-Driven MAC Layer Operation

MAC layer resource management framework optimization is expected to undergo a paradigm shift from *network of entities* to *network of intelligent functions* [70]. Proper,

adaptable utilization of the available RAN communications, computing, and cloud caching resources is indispensable for future 6G networks to meet the expected application-driven demands. Given the wide variety of emerging applications with pressing needs for increased access to communication, compute, and data storage services, the highly dynamic and heterogeneous nature of future 6G networks calls for the development of efficient resource management approaches breaking beyond the limits of conventional solutions across all three sectors. 6G MAC layer design is envisioned to incorporate novel, AI-aided resource management approaches that can deliver excellent performance and fulfill the stringent requirements set forth for 6G systems.

2.3.1 AI-Aided Compute Resource Management

DNN-based tasks can be computation-hungry thus creating a challenge for small mobile equipment that do not have direct local access to abundant computation resources. Therefore, active research has been undergoing for efficient edge and cloud-based computing. The authors in [71] studied binary computation offloading with the objective to minimize energy consumption. The authors proposed an offloading mechanism relying on distributed DNN operation where a group of DNNs are used in parallel to arrive at suitable offloading actions. The authors also presented a hierarchical model for offloading to balance delay and energy consumption. A weighted sum optimization problem was then formulated. The proposed method was able to score noticeable compute energy savings while maintaining reasonable delay and offloading accuracy. The authors in [72] proposed Edgent, an edge-computing solution utilizing edge-device synergy for collaborative DNN-based inference. The proposed framework adaptively leveraged DNN partitioning where computations are partitioned among the edge and device sides to exploit the power of cloud resources through the proximity of edge nodes to achieve real-time inferences. The technique also employ DNN resizing to improve computation latency by appropriately choosing an intermediate stage (i.e., DNN layer) to exit the inference operation. Edgent was designed with the aim to address static as well as dynamic environments for practical suitability. Fog computing at the edge of networks is a strong candidate approach to reduce the latency associated with offloading computation tasks to cloud-based servers while providing compute services to resource-limited mobile equipment. Inspired by fog computing potential, the authors in [73] proposed a DNN-based partitioning solution

for cooperative, wireless fog computation offloading. The authors considered dividing the DNN to form a layer-based partitions. They also partitioned the layers' input data across multiple users to parallelize the computation process. The authors focused on minimizing the consumption of energy by different cooperative agents through parameter optimization and workload balancing for the partaking devices. Compared to uniform workload division among participants, the proposed DDN-baded solution managed to achieve lower compute energy footprint. Since edge-caching of DNN-based models provides numerous advantages in terms of security, privacy, and efficiency, the authors in [74] tackled the problem of cache model selection and edge-server processor allocation with an aim to bring down the total cost of the system. The authors focused on user-experienced delay as well as compute energy consumption under constrained resources. The authors developed an online Gibbs sampling technique for compute resource allocation and cashing of DNN models which they evaluated through simulations to demonstrate its effectiveness.

2.3.2 AI-Aided Communications Resource Management

Proper management of the available RAN communications resources is of utmost importance for 6G networks to fully utilize and allocate the limited transmission power and spectrum resources among multiple users simultaneously. This is especially important for future networks where higher connection density levels are expected. The network must accommodate the expected surge in data traffic while providing users with high transfer rates suitable for AI-driven, speed-demanding applications. To this end, various research contributions have been made to optimize the MAC layer level communications resource utilization in wireless systems. This include the optimization of the transmission power allocation among users as well as the optimization of user pairing over unified resource blocks (e.g., time slots, OFDM frequency channels, and orthogonal code assignment). To this end, the authors in [75] considered a group of access points in a cell-free, multiuser mMIMO network. The authors proposed a distributed DNN-based power allocation scheme with the aim to maximize the system's spectral efficiency. The proposed DNN was trained using the fading parameters to feed the DNN to achieve an adequate dynamic range and simplify the DNN structure. The proposed scheme managed to achieve higher performance than the heuristic-based solution. In [76], the authors proposed a DNN-based per-antenna power allocation scheme for MIMO systems using

six layers (two convolutional followed by four fully-connected). The DNN is fed collected fading data to generate and output the per-antenna and per-user shares of the transmission power. Intensive pre-training of the employed DNN was carried out and Sub-6GHz TDD-based operation was assumed. The proposed DNN-based approach showed comparable performance to the bisection method. Although these solutions can perform adequately, optimized performance can only be achieved by taking user scheduling into account. Therefore, the authors in [77] proposed a DNN-based power allocation and link scheduling algorithm for multi- small cell operation. A suitable user scheduling is selected followed by power allocation. Simulation of the proposed technique showed faster operation compared to scheduling via an exhaustive scan and geometric programming power control. Although the proposed work considered both power allocation and user selection, performance tuning for non-orthogonal transmissions is not considered. Performance tuning can be especially sensitive for power allocation in non-orthogonal based multiuser multiplexing schemes such as NOMA-based systems where the interuser interference must be carefully handled for optimized transfer rates. The authors in [78] thus proposed a DNN-based power allocation and user selection approach for NOMA systems. The presented solution operated in a two-stage manner. In the first stage, the transmission power is allocated by training a DNN model to emulate the interior point algorithm output. The objective was to increase the sum-rate of the system while maintaining adequate processing complexity. Once power allocation is done, a user selection mechanism is applied as a second stage to enhance the achievable performance. Although the proposed solution considered power allocation and user selection for NOMA systems, incorporation of mmWaves blockage effects was not taken into account. Future 6G networks will heavily rely on mmWaves for their great potential of unlocking ultra-high speeds. The authors in [79] considered DNN-based NOMA resource management incorporating mmWave-enabled operation. The proposed scheme aimed to maximize the system's energy efficiency under QoS constraints. Lagrangian dual decomposing was firstly applied to solve the user association problem, then a semi-supervised DNN-based approach was applied to tackle subchannel and power allocation.

Although the proposed works showed promising performance, backhaul link capacity constraints were not considered. This is an important aspect since future 6G networks are expected to primarily carry bandwidth hungry video content (about 75% of total mobile

network traffic by 2030 according to the ITU [30]) for various AI-based applications and services. The content will usually be transferred from the source point to the core network via backhaul interfaces experiencing huge, real-time traffic surges. Cache-enabled operation utilizing the popularity of rich multimedia contents in the locale of the serving base stations can alleviate backhaul traffic congestion and provide an effective way to optimize the system's performance.

In addition, non-terrestrial UAV-assisted communications have been gaining a lot of attention due to their potential to be quickly dispatched to establish and provide wireless coverage in a variety of important use cases including remote areas with no sufficient telecommunications infrastructure availability or emergency use cases (e.g., in areas affected by natural disasters where the main serving network might be sustaining some damage) [80]. The authors in [81] studied a UAV-based NOMA system for emergency communications. The authors proposed to establish an uplink channel so the UAV can gather information related to the IoT equipment within the emergency zone. Afterwards, coverage is provided for IoT users via a joint UAV dispatching and power control approach. Nonetheless, NOMA downlink user pairing was not considered. To overcome this, the authors in [82] tackled both NOMA user pairing and power distribution along with UAV course planning with the goal to optimize the transfer rate. Energy-oriented operation, however, was not optimized as battery-aware design was not taken into account.

2.4 Target Research Position

Figure 2.2 illustrates the position of the target research of this thesis within both the PHY and MAC layers of the wireless communications system. From a high-level viewpoint, various aspects of the wireless communication system can be considered at both layers. For example, MAC layer storage and compute resource management and PHY layer channel estimation and signal detection operations have been considered in the literature as discussed earlier. Of particular importance to the proposed 3D-NOMA network architecture introduced in Fig. 1.4 is the optimization of communications resource management operation at the MAC layer side to maximize the achievable communication speeds and accommodate the demanding high-rate and connection density requirements of future 6G networks. On the other hand, optimizing the multicarrier modulation scheme

at the transmitter side via AI-empowered OFDM signaling is of great importance to provide fast processing capabilities at the PHY layer level, which is extremely crucial for the successful realization of future 6G applications and services that require low-latency, real-time operation. Therefore, in this thesis we pursue AI-aided optimized operation of the MAC layer communications resources as well as AI-driven low-complexity OFDM signal processing for enhanced PHY layer operation.

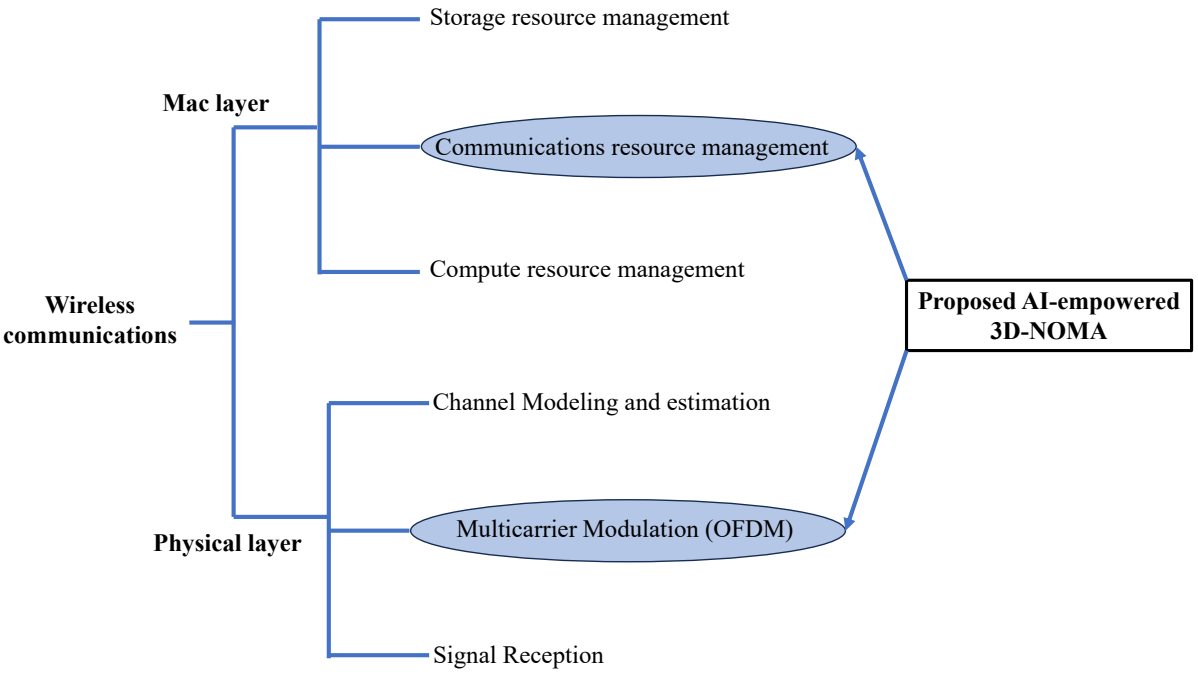


Figure 2.2: Proposed 3D-NOMA PHY and MAC layers research focus.

2.4.1 PHY Layer Side

Efficient operation of the PHY layer block chain in Fig. 2.3 hinges to a large extent on a highly optimized multicarrier modulation scheme to process the signals prior to transmission over the wireless channel. As pointed out earlier, OFDM multicarrier signaling is a very promising framework that provides a wide range of benefits from both theoretical and practical standpoints. For example, the scheme provides high spectral efficiency and enjoys very powerful resilience to ISI issues induced by the multipath nature of the wireless channel. Moreover, practical OFDM receivers are easy to implement using DSP-based operations. Given its potential and robustness to channel impairments, OFDM has been the choice of 3GPP standardization committees for 5G NR modulation.

However, powerful AI-aided OFDM signal formatting approaches are needed to improve raw OFDM signals and ensure the transmitted signals exhibit desirable character-

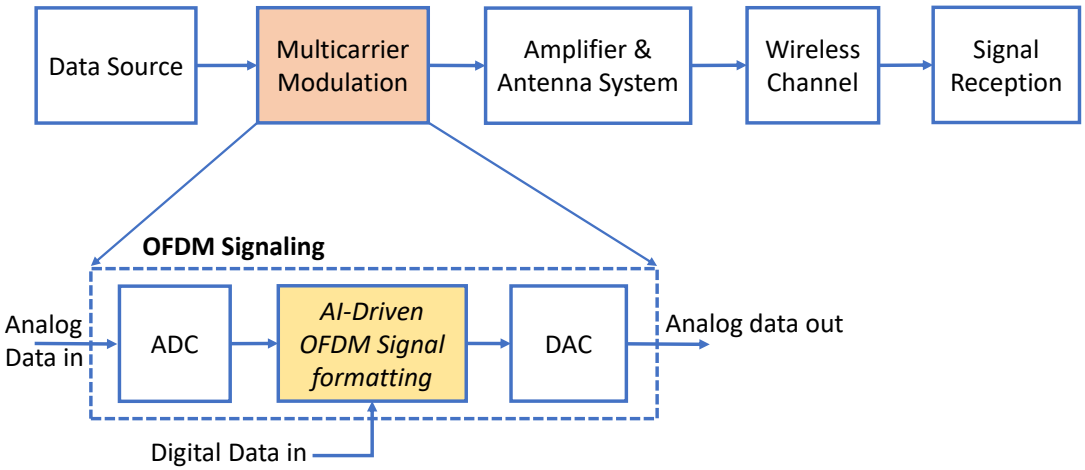


Figure 2.3: Top-level view of the PHY layer block chain.

istics. This is particularly pressing going beyond 5G where low-complexity, AI-oriented OFDM signal formatting alternatives will be desirable to provide enhanced underlying PHY layer operation for novel candidate technologies such as 3D-NOMA networks that can ultimately drive and enable latency-demanding 6G services and applications. To this important goal, in this thesis we focus on AI-empowered, low-complexity OFDM signal processing operation. In chapter 3, we present a detailed treatment of the proposed AI-aided OFDM signal formatting scheme.

2.4.2 MAC Layer Side

The main MAC layer operations of wireless communication networks include resource management operations of the available storage, communications, and compute resources as depicted in Fig. 2.4. Of special interest is the efficient utilization of the communications spectrum and transmit power budget resources due to the ever-growing demands for higher data transfer rates and more simultaneous connections. This is particularly important in order for future 6G networks to deliver an excellent user experience and meet the challenging requirements of the emerging applications and services as pointed out earlier. As such, promising candidate technologies that can maximize the utilization of the system's communications resources are vital. 3D-NOMA networks employing the NOMA communications protocol can provide an excellent communications architecture for increased user density and boosted data rates.

3D-NOMA networks must therefore incorporate robust and adaptive user selection and transmit power control mechanisms to efficiently manage the available communications resources. To this end, highly-optimized AI-aided communications resource man-

agement schemes are the research focus of this thesis at the MAC layer level for both terrestrial and non-terrestrial communications in order to achieve seamless operation of the proposed 3D-NOMA architecture in Fig. 1.4. In chapter 4, we describe the details of the proposed AI-aided communications resource management for terrestrial based NOMA downlink communications. In addition, chapter 5 extends the proposed operation for non-terrestrial UAV-based 3D-NOMA networks.

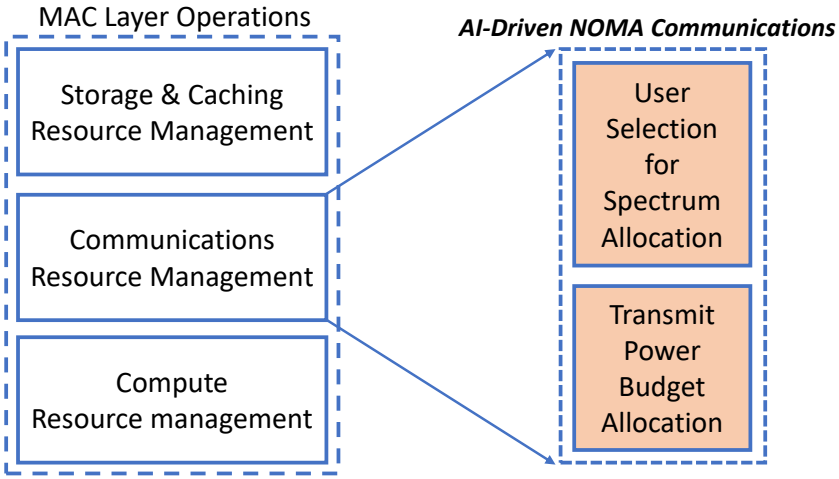


Figure 2.4: MAC layer operations.

Chapter 3

OFDM PAPR Reduction via Synchronous Neural Networks

3.1 Introduction

The high peak-to-average power ratio (PAPR) is a typical characteristic of orthogonal frequency division multiplexing (OFDM) signals. Prior to transmission through the channel, OFDM signals are amplified, and to achieve high efficiency, the power amplifier is driven near its saturation region. This, however, causes undesired non-linear amplification effects including signal distortion and out-of-band (OoB) radiation, especially if the signal input to the amplifier has high PAPR. Therefore, it is imperative to tackle the high PAPR problem inherent in OFDM signals. Although the PAPR reduction is not the ultimate goal of the communication system, it provides an instrument for maintaining the bit error rate (BER) while keeping the spectral mask at the transmitter output according to standard requirements.

To address the PAPR issue, many classical solutions have been proposed over the years, including partial transmit sequence (PTS) [83], tone reservation (TR) [84], companding [85], selected mapping (SLM) [86], and clipping and filtering (CF) techniques [62–68]. CF is a simple but powerful approach that does not generally require the use of additional resource blocks or the transmission of side information, which make it of particular interest in the literature. In CF, the time-domain OFDM signal amplitude is limited to a predefined threshold level to keep the resulting PAPR within acceptable ranges. However, clipping the signal amplitude causes OoB power leakage which may not be tolerable

by adjacent channels. The signal is therefore filtered to eliminate the OoB emissions. Nonetheless, filtering is invariably coupled with amplitude regrowth of the time-domain OFDM signal and this in turn increases the PAPR beyond the acceptable threshold. Consequently, iterative-based clipping and filtering (ICF) operation is needed to repeatedly clip the signal amplitude and filter the leaked emissions. However, ICF techniques are computationally expensive in general, especially when the number of sub-carriers is large. Simplified clipping and filtering (SCF) was introduced to achieve similar PAPR reduction performance in a single iteration of modified processing at a relatively cheaper cost. However, the required computational complexity is still high even for SCF.

To solve this problem, the authors in [69] proposed an alternative SCF-based method which uses asynchronous neural network (ANN) modules at the transmitter side to achieve PAPR reduction performance comparable to existing SCF methods while significantly reducing the required complexity. Since the ANN technique focused on modeling the output signal's real part as a function of only the input's real part whereas the output's imaginary part was similarly modeled as a function of only the input's imaginary part, synchronous operation was, therefore, not used. Thus, the method in [69] is asynchronous in nature due to the separate training of two independent NN modules using correlated real and imaginary components, respectively. This can potentially cause convergence mismatch between the two NN modules leading, thereby, to one module under or overfitting with respect to the other. This ultimately results in increased OoB radiations and degrading BER performance for higher order modulations. In particular, OoB radiation is very sensitive to non-linear distortion due to peak limiting. The above inaccurate convergence in conventional ANN may result in unacceptable OoB emissions, which needs a novel solution.

In this chapter, we propose a synchronous neural network CF-based (SNN-CF) approach where the two modules of ANN are replaced by a single NN module which is simultaneously fed by both real and imaginary components to take the effect of their interdependence into account at the training stage.

3.2 Related Works

The authors in [87] investigated the modeling of the PAPR reduction problem as a combinatorial search problem and proposed to use a Hopfield neural network (HNN)

optimizer [88] to solve it. Although the technique showed promising results, it would require the transmission of a large amount of side information and suffer high sensitivity to phase noise. To achieve PAPR reduction performance close to that provided by HNN while avoiding the transmission of side information, an SLM-based radial basis function network (RBFN) solution was introduced in [89]. RBFN is essentially a multivariate interpolator that generates a phase rotation vector which is fed to an SLM encoder for phase selection and OFDM signal generation. Although no explicit side information is needed, the technique requires the adaptive updating of the weights of the network's output layer as well as the activation functions of the network's hidden layers.

The authors in [90] proposed an active constellation extension-based (ACE) method that deploys a time-frequency neural network (TFNN) to accomplish PAPR reduction performance close to that obtainable using the original ACE technique with less complexity. However, TFNN modules operate in complex time-domain and frequency-domain modes and exhibit degraded BER performance in fading channels for high-order modulations. To overcome this problem, Ref. [91] proposed an ACE-based NN variation to achieve much better BER performance for high-order modulation transmissions over fading channels while maintaining almost the same PAPR reduction performance of the TFNN method. Although the technique requires fewer in-use computations than the TFNN approach, it would rely on the continuous training and updating of adaptive NN modules at the receiver side, based on the average received signal-to-noise (SNR) ratio, through the transmission of predefined training signals over dedicated spectral resources.

Recently, the authors in [92] proposed a deep neural network-based (DNN) encoders and decoders deploying multi-layer DNN modules for constellation mapping and demapping at the transmitter and receiver sides, respectively. The DNN-based approach can simultaneously improve the performance in terms of both PAPR reduction and BER. This comes, however, at the expense of highly expensive computational cost due to the deployment of complex multi-layer DNNs, rendering their usage unfeasible for low-latency demanding scenarios that require strict constraints on complexity.

Unlike the above works, the proposed SNN-CF is a low-complexity PAPR reduction approach which requires neither the active adaptation of the learned network nor the transmission of side information or the dedication of additional resource blocks. In addition, the actual merit of the proposed scheme is expected to be observed for

higher-order modulations where conventional ANN schemes start to exhibit severe BER performance deterioration.

3.3 OFDM Model

3.3.1 Signal Model

Consider an OFDM system utilizing N sub-carriers. In order to closely follow the continuous amplitude envelope of the time-domain OFDM signal, an oversampling factor $\eta \geq 4$ is customarily assumed. The time-domain OFDM signal can be written as

$$x[n] = \frac{1}{\sqrt{\eta N}} \sum_{k=1}^{\eta N} X[k] e^{\frac{\omega_k n}{\eta N}} \; \forall \; n = 1, 2, \ldots, \eta N \,, \tag{3.1}$$

where $\omega_k = 2\pi k$ is the k-th sub-carrier carrying the data-bearing modulated symbol $X[k]$. Quadrature phase shift keying (QPSK) modulation and quadrature amplitude modulation (16-QAM) are both considered subsequently.

3.3.2 Performance Metric

As mentioned earlier, high-power amplifiers (HPAs) have non-linear responses causing undesired distortion effects which are poorly captured by the classical peak-to-average power ratio metric. The cubic metric (CM) [93] uses higher-order statistical information gleaned from the OFDM signal to appropriately account for the signal's envelope fluctuations. The third-generation partnership project (3GPP) defines the CM, Ω, as

$$\Omega = \frac{\Gamma^{[dB]} - \Gamma_{ref}^{[dB]}}{\kappa} dB \,, \tag{3.2}$$

where $\Gamma^{[dB]}$ is the raw cubic metric (RCM) of the time-domain OFDM signal in (3.1) and is defined as

$$\Gamma^{[dB]} = 20 \log_{10} \left\{ \sqrt{\overline{\left(\left[\frac{|x[n]|}{\sqrt{\overline{x[n]}}} \right]^3 \right)}} \right\} dB \,, \tag{3.3}$$

where the factor κ is empirically determined and $\Gamma_{ref}^{[dB]}$ is a fixed RCM reference level. Typical values for these two parameters are used later in the numerical evaluation of the system performance.

3.4 Benchmark Techniques

3.4.1 Selected Mapping

Undistorting PAPR reduction schemes, such as SLM, sacrifice complexity or the utilization of additional resources to help maintain the same BER performance of the original OFDM. In SLM [86], a large block of U different OFDM sequences, $x_u[n]\ \forall\ u = 1, 2, \ldots, U$, $n = 1, 2, \ldots, \eta N$, representing alternative candidates to map the same information-bearing symbol stream, is generated. The candidate sequence, $x^*[n]$, with the lowest CM metric, is selected for transmission. Thus, for SLM:

$$x^*[n] = \underset{\substack{x_u[n] \\ u \in \mathcal{S}}}{arg\ min} \quad \frac{20\log_{10}\left\{\sqrt{\left(\left[\frac{|x_u[n]|}{\sqrt{x_u[n]}}\right]^3\right)}\right\} - \Gamma_{ref}^{[dB]}}{\kappa}, \tag{3.4}$$

where $\mathcal{S} \triangleq \{1, \ldots, U\}$ is the set of indices of all candidate sequences within the generated block.

3.4.2 Clipping-Based Techniques

3.4.2.1 Iterative Clipping and Filtering

The ICF algorithm is outlined in Fig. 3.1. The signal amplitude is kept below a preconfigured limiting threshold by repeatedly clipping amplitudes exceeding the threshold and filtering the resulting OoB radiations. The process therefore zigzags back and forth between time and frequency domain representations through the iterative computation of IFFT and FFT operations for a preset number of times, L. Fortunately, amplitude regrowth decays with the increasing number of iterations, thereby enabling the technique to effectively reduce the PAPR level as more iterations are applied.

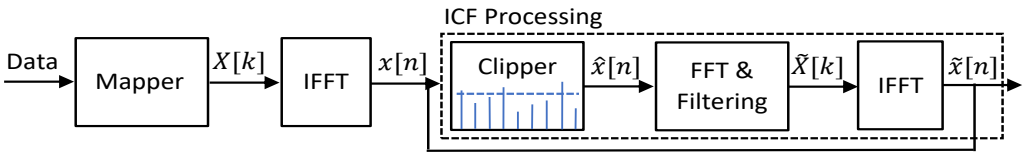


Figure 3.1: Top-level view of ICF.

3.4.2.2 Simplified Clipping and Filtering

SCF mimics the powerful PAPR reduction capability of the ICF-based mechanism while dispensing the need for iterative operation. As in ICF, SCF algorithm [65] starts by applying IFFT operation to generate the time-domain OFDM signal, $x[n]$, from the modulated symbol stream sequence, $X[k]$, and then proceeds to compute a clipped version, $\hat{x}[n]$, with maximum amplitude R, according to the same criterion. However, the operation is then modified to achieve almost the same performance in a single iteration of signal processing. The clipped signal, $\hat{x}[n]$, is then subtracted from the unclipped OFDM signal, $x[n]$, to produce a clipping noise signal, $d[n]$, which is then converted to frequency-domain representation, $D[k]$, by applying FFT operation as shown in Fig. 3.2. To suppress OoB emissions, $D[k]$ is then filtered to yield $\hat{D}[k]$ with nulled $(\eta - 1)N$ trailing components. A crucial step giving SCF its ability to come very close to ICF performance without the need for multiple iterations was then performed. The frequency-domain clipping of the original modulated sequence, $X[k]$, is accomplished by scaling the filtered noise signal, $\hat{D}[k]$, and subtracting it from $X[k]$, to obtain $\tilde{X}[k]$ which is finally converted to the time domain sequence $\tilde{x}[n]$ by applying IFFT operation. The clipped signal, $\tilde{X}[k]$, was computed as

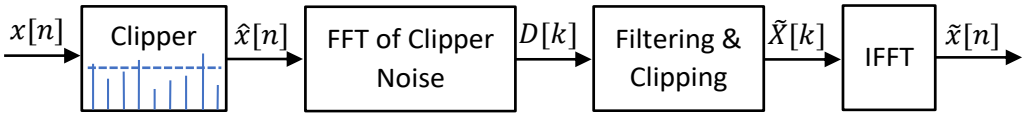


Figure 3.2: Top-level view of SCF.

$$\tilde{X}[k] = X[k] - \rho\hat{D}[k]\,, \tag{3.5}$$

where the scaling coefficient, ρ, is calculated as

$$\rho = \frac{1 - (1-\mu)^{1.5\lambda}}{1 - (1-\mu)^{1.5}}\,, \tag{3.6}$$

where the factor λ was used to establish an equivalency for the iterative nature of ICF and takes the same value as the number of iterations used by ICF. The parameter, μ, is a function of the maximum allowable amplitude, R, and the standard deviation, σ, of the

unclipped time-domain OFDM signal, $x[n]$, and is computed as

$$\mu = \sqrt{\frac{8}{3\pi}} \left(\frac{\sigma}{R} \right). \tag{3.7}$$

3.5 Proposed Neural Network Solution

3.5.1 Network Architecture

The objective is to train a neural network to extract the identifying characteristics inherent in time-domain OFDM signals processed by conventional CF-based solutions. A feed-forward multi-layer perceptron (FFMLP) neural network is pretrained offline before it is deployed for the real-time processing of original OFDM signals prior to transmission. As shown in Fig. 3.3 (a), the general structure of FFMLP consists of three main sections: an input layer defining the input format (type, dimensionality, etc.,) followed by a set of hidden layers, each including a configurable number of neurons with a unified activation function for all neurons of each layer. The last layer is the output layer, which is a convergence point collecting the data traversing the network from the input layer through the hidden layers. The data are weighted and summed by the output layer's neurons and passed through an activation function to produce the final output. The input to the FFMLP is M-dimensional. There is a total of V hidden layers. The i-th hidden layer contains h_i processing neurons. The final output of the FFMLP network module is P-dimensional. The response of the j-th neuron in the i-th hidden layer is:

$$s_j^{(i)} = f_i \left(\sum_{k=1}^{h_{i-1}} w_{j,k}^{(i)} s_k^{(i-1)} + b_j^{(i)} \right), \tag{3.8}$$

where $f_i(.)$ is the activation function of the i-th hidden layer and $s_k^{(0)} \triangleq I_k$ is the k-th entry in the input layer. $w_{j,k}^{(i)}$ is the learnable weight connecting the k-th neuron in the $(i-1)$-th layer to the j-th neuron in the i-th layer. $b_j^{(i)}$ is the learnable bias of the j-th neuron in the i-th layer. Similarly, the response of the j-th neuron in the output layer is:

$$s_j^{(o)} = f_o \left(\sum_{k=1}^{h_V} w_{j,k}^{(o)} s_k^{(V)} + b_j^{(o)} \right). \tag{3.9}$$

Figure 3.3 (b) illustrates the application of an FFMLP module to process raw OFDM

input signals. The input layer extracts the in-phase and quadrature components of the input OFDM signal. The two parts are then simultaneously processed by the subsequent core of the pretrained neural network module to generate the corresponding in-phase and quadrature components of a CF-like output signal with the desired characteristics.

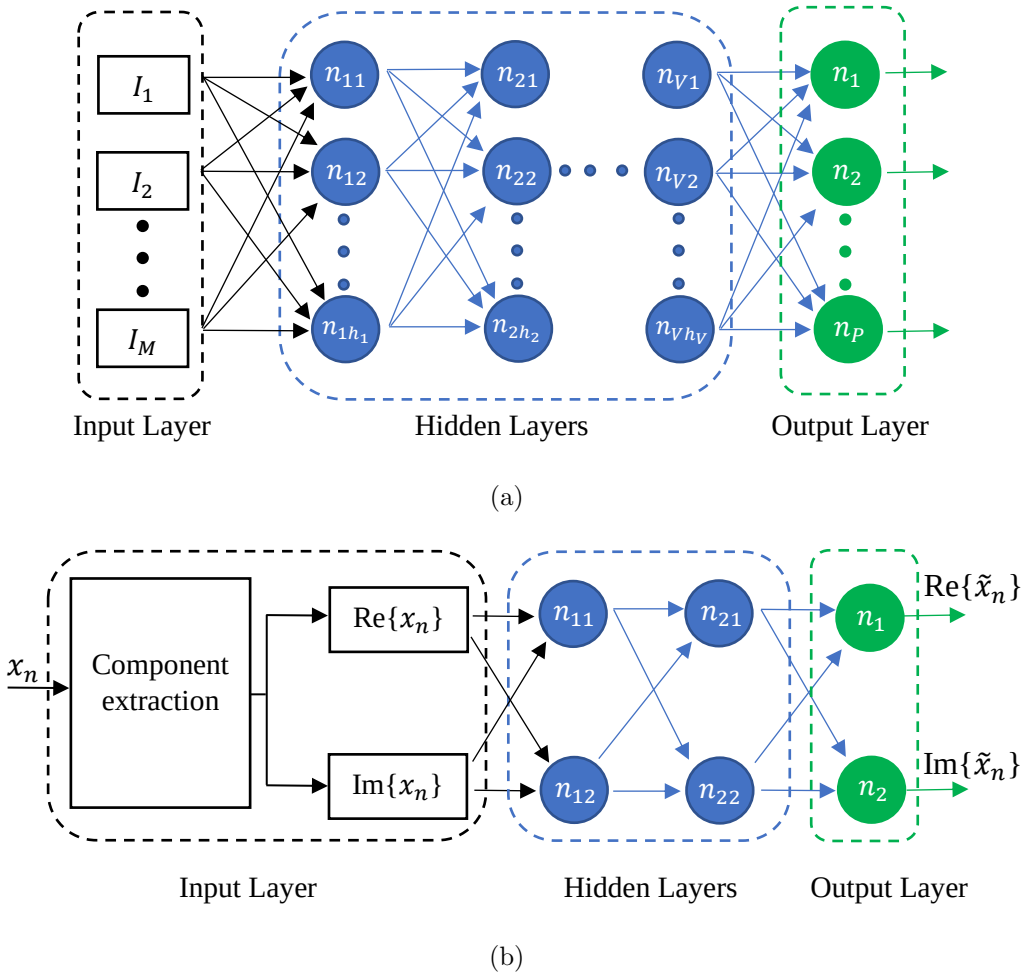


Figure 3.3: FFMLP neural networks: (a) general structure; and (b) OFDM signal processing.

3.5.2 Proposed Dataset Compilation and Network Selection

A 3-D training dataset, $\mathcal{D}$, is formed by running Algorithm 1. Initially, $\mathcal{D}$ is propagated by zeros. The dataset, $\mathcal{D}$, is comprised of multiple 2-D slices as shown in Fig. 3.4. The i-th 2-D slice, $\mathcal{D}(i,.,.)\,\forall\, i = 1, 2, \ldots, S$, is filled with a random selection of a group of ηN time-domain samples. The samples are obtained by generating a vector of samples of an original OFDM symbol, $x_i[n]\,\forall\, n = 1, 2, \ldots, \eta N$, which is then fed to an ICF-based clipping algorithm to generate the corresponding output training vector, $x_o[n]\,\forall\, n = 1, 2, \ldots, \eta N$. SCF-based training can be applied alternatively. The samples of the two generated vectors are jointly scattered across the slices in a random fashion

as shown in Fig. 3.4. The process is repeated until all the slices within $\mathcal{D}$ are filled with the training data.

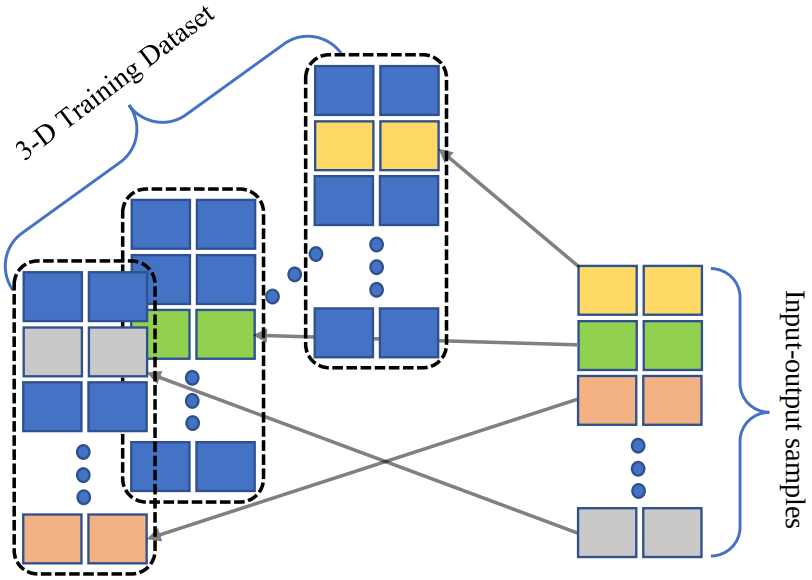


Figure 3.4: Proposed dataset compilation mechanism.

Figure 3.5 illustrates the concept of the proposed SNN-CF scheme. Unlike conventional NN solutions, we propose using a synchronous FFMLP-based network structure to achieve a more accurate performance with a slight increase in computational cost. In this scheme, instead of processing the correlated (I) and (Q) parts independently, only one FFMLP module is employed to process both parts simultaneously, using 2-dimensional (2-D) input and output sequences ($M = 2$ and $P = 2$ in Fig. 3.3 (a)). This synchronous approach can emulate SCF with higher flexibility and handle underfitting and overfitting problems more gracefully than ANN where two modules are trained separately. Therefore, SNN-CF essentially models $\mathcal{R}\{\tilde{x}[n]\}$ as a function of both $\mathcal{R}\{x[n]\}$ and $\mathcal{I}\{x[n]\}$. Similarly, $\mathcal{I}\{\tilde{x}[n]\}$ is modeled as a function of both $\mathcal{R}\{x[n]\}$ and $\mathcal{I}\{x[n]\}$. Thus, SNN-CF has potentially higher chances of capturing the essence of the SCF peak-limiting mapping algorithm.

Algorithm 1 Proposed Dataset Compilation.

- **Initialize** $s = 0, \mathcal{D} = 0_{S,\eta N,2}$
- **While** $s < S$ repeat the following steps:
  1. Generate training samples:
     - Apply IFFT to compute the time-domain OFDM signal, $x[n]$, of (3.1):
       - Append $(\eta - 1)N$ trailing zeros to $X[k]$ for oversampling.
       - Compute IFFT of $X[k]$ with size ηN.
     - Save the generated OFDM signal:

$$x_i[n] \leftarrow x[n]$$

     - while $l < L$, repeat the following steps:
       - Generate clipped OFDM signal, $\hat{x}[n]$, by replacing all samples of $x[n]$ with amplitudes exceeding the limiting value, R, according to,

$$\hat{x}[n] = Re^{j\angle x[n]} \; \forall \; n \; \ni |x[n]| > R \,,$$

         where $R = C\sqrt{\overline{x[n]}}$ and C is the desired clipping ratio.
       - Filter the resulting OoB radiations:
         - Compute FFT of $\hat{x}[n]$ with size ηN to generate $\hat{X}[k]$.
         - Zero out the trailing $(\eta - 1)N$ frequency-domain samples of $\hat{X}[k]$ to generate filtered OFDM signal, $\tilde{X}[k]$.
         - Recompute the corresponding time-domain OFDM signal, $\tilde{x}[n]$, using IFFT with size ηN.
       - $x[n] \leftarrow \tilde{x}[n]$, $l \leftarrow l + 1$
     - end while
     - Save the generated OFDM signal:

$$x_o[n] \leftarrow \tilde{x}[n]$$

  2. Scatter the ηN input and output OFDM samples across ηN random empty locations within $\mathcal{D}$:

$$\mathcal{D}\,(i, j, 1) = x_i[n],$$
$$\mathcal{D}\,(i, j, 2) = x_o[n]$$

  3. $s \leftarrow s + 1$
- **end While**

Where

- $x_i[n]$ and $x_o[n]$ $\forall$ $n = 1, 2, \ldots, \eta N$ are the OFDM input and output training samples.
- i and j, $1 \leq i \leq S$, $1 \leq j \leq \eta N$, are random indices of empty storing locations within the dataset $\mathcal{D}$.

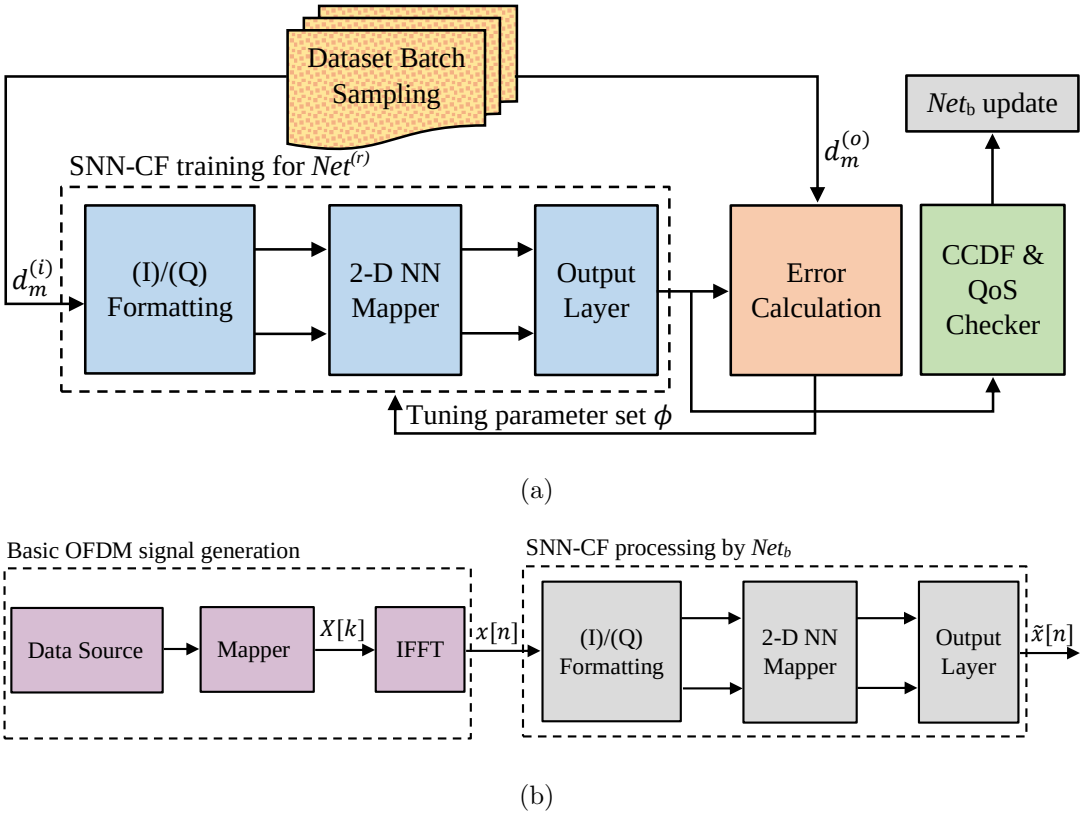


Figure 3.5: The proposed scheme: (a) training; and (b) deployment.

During training the network model in Fig. 3.5, each training data point is formed from $\mathcal{D}$ by constructing an input training vector, $[\mathcal{R}\{\mathcal{D}(i,j,1)\}, \mathcal{I}\{\mathcal{D}(i,j,1)\}]^T$, and a corresponding target vector, $[\mathcal{R}\{\mathcal{D}(i,j,2)\}, \mathcal{I}\{\mathcal{D}(i,j,2)\}]^T$. The training process goes on until the mean square error (MSE) between the actual output of the NN module and the target output is minimized. An SNN-CF network is found once the training process halts. The network is then ready and can be deployed for the processing of raw OFDM signals that it has not processed before while learning SCF patterns. Compared to ANN, the proposed SNN-CF scheme is able to converge within fewer iterations during the training stage.

When deployed for operation, SNN-CF starts by reformatting the original OFDM input $x[n]$ to produce $[\mathcal{R}\{x[n]\}, \mathcal{I}\{x[n]\}]^T$, a 2-D vector representation of the in-phase and quadrature components of $x[n]$. The input vector is then applied to the network's Hidden Layers section which is preconfigured during the training stage to implement a map $f : \mathbb{R}^2 \to \mathbb{R}^2$ in the neighborhood of conventional ICF or SCF. The output layer processes the 2-D signal vector produced by the vector mapper f to generate a close estimate to $\tilde{x}[n]$, which would be the output of conventional ICF or SCF for the same input $x[n]$.

Attaining a higher performance than conventional CF-based schemes is accomplished by exploring the space of different maps in the neighborhood of ICF or SCF. This is performed by repeatedly training different SNN-CF modules until a network providing the highest positive performance gap is found under certain quality of service (QoS) requirements. Thus, to ensure the fidelity of the solution, we impose a threshold BER, $\mathcal{E}_t$, for the trained SNN-CF network at an arbitrary $E_b/N_0|_t$. Therefore, in order to find a mapper f with desirable characteristics, we explore the neighborhood of a conventional CF map during the offline training stage. The outline of the proposed network-selection algorithm is given in Algorithm 2. The basic concept of the algorithm is to optimize the network structure to achieve the lowest possible CM value while maintaining a given QoS requirement. Although a maximum allowable BER level is used as a metric to enforce the required QoS, other metrics can be applied, such as the sum-rate and error vector magnitude (EVM). In fact, EVM is used as a metric to achieve the target BER in [94]. The algorithm starts by initializing an SNN-CF network, $Net^{(0)}$, with a random parameter set ϕ and updates it using a random batch of training data points from the dataset $\mathcal{D}$ in order to find a mapping solution that minimizes the MSE mentioned previously. For a preset number of iterations, different SNN-CF networks are generated. In each iteration, r, a different non-linear map in the neighborhood of ICF or SCF is learned by training a new candidate SNN-CF network, $Net^{(r)}$. Upon training, the performance of $Net^{(r)}$ is measured by evaluating $\mathcal{F}^{(r)} \triangleq Pr\{\Omega > \Omega_{0d}\}$, the complementary cumulative distribution function (CCDF) of the OFDM signal generated by $Net^{(r)}$. As the iterations advance, the algorithm keeps track of Net_b, the SNN-CF network providing the best CCDF value, $\mathcal{F}_b$. The best-performing network is thus updated and saved as the algorithm progresses. Finally, the algorithm outputs the best-performing SNN-CF network found when it stops after the last iteration.

Algorithm 2: Proposed SNN-CF Network Selection.

- **Initialize** $r = 0, \mathcal{F}_b = 1$
- **Set** Ω_{0d} , $E_b/N_0|_t$, $\mathcal{E}_t$, $\mathcal{R}_{max}$
- **While** $r < \mathcal{R}_{max}$ repeat the following steps:
  1. Initialize a random SNN-CF network, $Net^{(r)}$, with parameter set ϕ and apply a hybrid batch of random samples $\left(d_m^{(i)}, d_m^{(o)}\right)$ from $\mathcal{D}$ to update ϕ.
  2. Compute $\mathcal{F}^{(r)}$, the CCDF of the OFDM signal generated by $Net^{(r)}$, at the design threshold Ω_{0d}
  3. Compute $E^{(r)}$, the achievable BER of $Net^{(r)}$, at the design target $E_b/N_0|_t$
  4. Update the best CCDF, $\mathcal{F}_b$, and save its corresponding SNN-CF network, Net_b:

$$\mathcal{F}_b = \begin{cases} \mathcal{F}^{(r)} & \mathcal{F}^{(r)} < \mathcal{F}_b \\ & E^{(r)} < \mathcal{E}_t \\ \mathcal{F}_b & \text{Otherwise} \end{cases}$$

$$Net_b = \begin{cases} Net^{(r)} & \mathcal{F}^{(r)} < \mathcal{F}_b \\ & E^{(r)} < \mathcal{E}_t \\ Net_b & \text{Otherwise} \end{cases}$$

  5. $r \leftarrow r + 1$
- **end While**

Where

- $\mathcal{R}_{max}$ is the total number of scanned mappings in the neighborhood of conventional SCF or ICF.
- $\left(d_m^{(i)}, d_m^{(o)}\right) \triangleq (\mathcal{D}(s, j, 1), \mathcal{D}(s, j, 2))$ is an input-output training data point sampled from $\mathcal{D}$ at a random entry j within a random slice s.
- $\mathcal{E}_t$ is the maximum allowable BER at the arbitrary input design target $E_b/N_0|_t$.

3.6 Complexity Analysis

In this section, we discuss the online computational requirements of the proposed scheme compared to the conventional solutions as well as ANN approach. We focus on the on-line computational demand since the training algorithms are offline-based, rendering the deployment phase as the critical stage for computation. SLM performs the highest amount of computations and generally requires U IFFT operations, U CM computations and U check operations for the selection of the sequence having the low-

est CM value. For all other schemes, the time-domain OFDM signal is first generated through the application of an IFFT operation to the modulated data sequence, $X[k] \; \forall \; k = 1, 2, \ldots, \eta N$, which typically requires $(N/2)log_2(N)$ complex multiplications and $Nlog_2(N)$ complex additions.

Each iteration of the ICF algorithm requires one clipping operation followed by an FFT operation to zero out the frequency-domain OoB radiations and finally an IFFT operation to convert back to the time domain. ICF thus requires L clipping operations as well as a total of $2LNlog_2(N)$ and $LNlog_2(N)$ complex addition and multiplication operations, respectively. SCF processes the signal in a single iteration and requires one clipping operation and a total of $N(2log_2(N) + 2)$ and $N(log_2(N) + 1)$ complex addition and multiplication operations, respectively.

Each module of ANN performs $2N$ real multiplications and $2N$ real additions in each hidden layer as well as N real multiplications and N real additions in the output layer. Thus, ANN requires a total of $10N$ real multiplications and $10N$ real additions. The output layer as well as each hidden layer of the proposed SNN-CF scheme performs $4N$ real multiplications and $4N$ real additions, resulting in a total of $12N$ real multiplications and $12N$ real additions. Finally, unlike conventional approaches, neither ANN nor SNN-CF requires additional checking or clipping operations. Therefore, neural network-based approaches provide significantly less computation requirements in general, rendering the real-time operation of such schemes much more favorable.

The computational cost of the investigated schemes are summarized in Table 3.1. The complexity of both ANN and the proposed SNN-CF schemes are based on computations in $\mathbb{R}$, the set of real numbers. However, the remaining schemes operate in the complex domain. Since one complex addition is equivalent to two real additions and one complex multiplication is equivalent to four real multiplications and two real additions, the equivalent complexities of ANN and SNN-CF are $\mathcal{O}(6.67N)$ and $\mathcal{O}(8N)$, respectively. Although compared to ANN the proposed SNN-CF scheme requires a slight increase in computational cost within the same asymptotic bound on complexity, it can potentially boost the system performance in terms of PAPR reduction and better controlled OoB emissions while simultaneously providing a lower BER for high-order modulations.

Table 3.1: Complexity requirements of the investigated schemes.

Scheme	Complexity
SLM [86]	$\mathcal{O}(U(1.5Nlog_2N + N + 1))$
ICF [62]	$\mathcal{O}(LN(3log_2N + 1))$
SCF [65]	$\mathcal{O}(N(3log_2N + 4))$
ANN [69]	$\mathcal{O}(20N)$
Proposed SNN-CF	$\mathcal{O}(24N)$

3.7 Numerical Results and Discussion

3.7.1 Simulation Setup

The parameters of the simulation environment are configured as in Table 3.2. We assume an OFDM system utilizing $N = 256$ sub-carriers loaded with QPSK or 16-QAM data. An oversampling factor $\eta = 4$ was chosen for the accurate tracking of the time-domain signal envelope. The RCM of the reference signal is simulated at 1.52 dB and the empirical factor κ is fixed at the value of 1.56 for all simulated schemes. The clipping ratio defining the maximum allowable amplitude of the time-domain OFDM signal processed by ICF and SCF schemes is set to 6 dB. Three ICF iterations are used and the equivalency factor used in SCF is set to the same value. The signal's standard deviation parameter σ is set to 1/2 for the proper estimation of the frequency-domain scaling coefficient ρ. The block size representing the total number of alternative sequences in SLM is set to 16.

The feed-forward multi-layer perceptron NN architecture model is adopted in both ANN and the proposed SNN-CF scheme due to its simplicity and robustness in accurately implementing various mappings in the processing of wireless communication signals. ANN uses two NN modules with two hidden layers in each module. The first layer has two neurons and the second layer has one neuron. The activation function used in both hidden layers of ANN is the triangular function defined as $f(x) \triangleq 1 - |x|, |x| \leqslant 1$, and $f(x) \triangleq 0$ otherwise.

A design target $E_b/N_0|_t$ is defined at the input of the PAPR reduction where the achievable BER is confined to a limiting threshold $\mathcal{E}_t$. The proposed SNN-CF scheme deploys a single NN module which has two hidden layers. Both layers have two neurons each. The activation function used to transfer the weighted input plus bias from the

input terminal of a neuron to its output terminal is the hyperbolic tan-sigmoid transfer function defined as $g(x) \triangleq 2/(1+e^{-2x}) - 1 \ \forall \ x \in \mathbb{R}$. Both ANN and the proposed SNN-CF scheme use the *purelin* activation function, $h(x) \triangleq x \ \forall \ x \in \mathbb{R}$, for the output layer's neurons. The design threshold Ω_{0d} used in Algorithm 2 for SNN-CF network selection is arbitrary and is set to 3 dB for both QPSK and 16 QAM cases. In addition, the maximum allowable BER, $\mathcal{E}_t$, is set to 3×10^{-4} and the design target $E_b/N_0|_t$ is set to 8 dB for QPSK signals. For the 16-QAM case, $\mathcal{E}_t$ is set to 4×10^{-3} and target $E_b/N_0|_t$ is set to 10 dB.

Table 3.2: Simulation settings: OFDM PAPR reduction.

Scheme	Parameter	Value
Common parameters	Underlying modulation	QPSK/16-QAM
	Sub-carriers number, N	256
	Oversampling factor, η	4
	Reference signal RCM, Γ_{ref}	1.52 dB
	Cubic metric empirical factor, κ	1.56
	Cubic metric threshold, Ω_0	[2.5,6] dB
	E_b/N_0	[0,9] dB
ICF, SCF	Clipping ratio, C	6 dB
ICF	Number of iterations, L	3
SCF	Equivalency factor, λ	3
	Standard deviation, σ	1/2
SLM	Block size, U	16
ANN, SNN-CF	Network Architecture	FFMLP
	Output layer activation type	Linear
	Normalization	Enabled
ANN	Number of NN modules	2
	Number of hidden layers, V	2
	Number of neurons in 1st layer, h_1	2
	Number of neurons in 2nd layer, h_2	1
	Hidden layers activation type	Triangular
	Training algorithm	Levenberg–Marquardt
	Training epochs	100
SNN-CF	Number of NN modules	1
	Number of hidden layers, V	2
	Number of neurons in 1st layer, h_1	2
	Number of neurons in 2nd layer, h_2	2
	Hidden layers activation type	Hyperbolic tangent
	Training algorithm	Proposed algorithm 2
	Training epochs	50

The Levenberg–Marquardt algorithm [95] is used to train ANN to emulate the time-domain OFDM signal produced by conventional SCF when fed by the original unmodified

OFDM signal at its input. The training epochs defining the number of training iterations performed for each input–output sample in the training dataset is configured to the value of 100 iterations for ANN, whereas for the proposed SNN-CF method, we only use 50 updating iterations which is facilitated by virtue of its potentially faster convergence capability.

The dataset batch used for learning the SCF scheme behavioral patterns is complied from a 100 input–output samples. The dataset is split into three parts: (a) the training set, containing 70% of the dataset samples which are only used during the training stage in which the weights and biases of all layers are found and updated to minimize the MSE over the training set; (b) the testing set, consisting of 15% of the dataset samples which are used to test the network found by applying the training set and possibly fine-tuning the weights to account for the testing set; and (c) the validation set containing the remaining 15% of the dataset and is used to validate the trained network's ability to generalize to new data it has never seen before.

3.7.2 Results and Discussion

For QPSK-OFDM signals, the CM reduction performance was investigated for original unmodified OFDM, SLM, SCF, ICF, ANN, and the proposed SNN-CF schemes in terms of the complementary cumulative distribution function, $\mathcal{F}(\Omega) = Pr\{\Omega > \Omega_0\}$, for different cubic metric threshold (Ω_0) levels ranging from 2.5 to 6 dB as shown in Fig. 3.6 (a). The proposed SNN-CF scheme outperforms the best of the three conventional approaches and shows more than a 0.5 dB performance improvement compared to ANN at CCDF levels as low as 10^{-3}. Similarly, Figure 3.6 (b) shows that the proposed SNN-CF scheme maintains its superior CM reduction performance for the higher order 16-QAM-OFDM case. This is accomplished by scanning the neighborhood of SCF during the training stage to find an SNN-CF network with the highest possible performance.

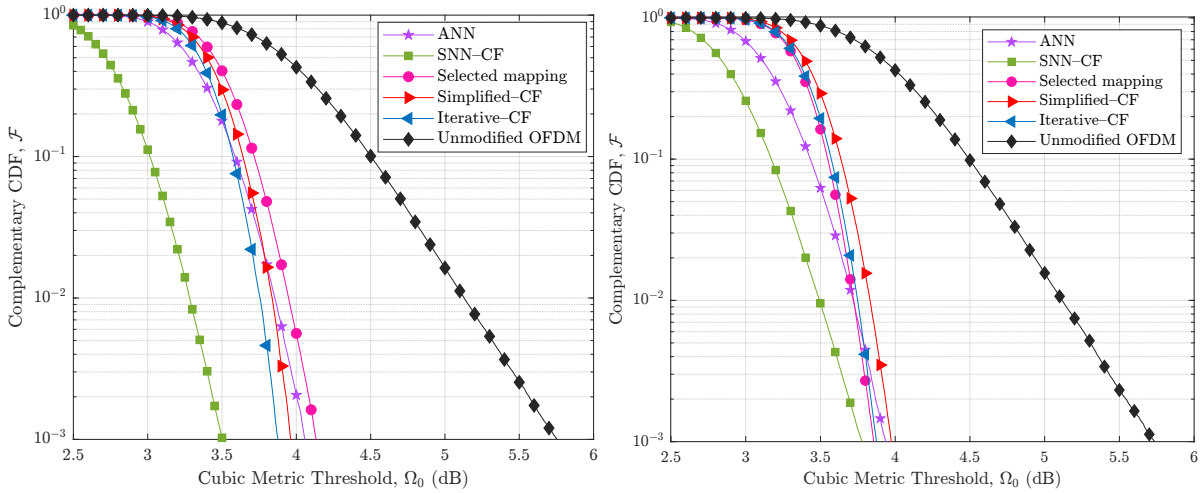


Figure 3.6: CM reduction performance of the proposed SNN-CF vs. ANN [69], selected mapping [86], simplified-CF [65], iterative-CF [62], and unmodified OFDM for (a) QPSK-OFDM, and (b) 16-QAM-OFDM.

The proposed SNN-CF exhibits lower OoB power leakage than ANN for both QPSK and 16-QAM modulations, as shown in Fig. 3.7. This figure plots the power spectral density (PSD) curves of OFDM signals as a function of normalized frequency. Only original OFDM is included as a baseline for comparison since SLM has identical power distribution as the original OFDM and conventional ICF and SCF are both assumed ideal, rendering their PSD performance matching with the original unprocessed OFDM. In the neighborhood of the in-band, both ANN and the proposed SNN-CF show similar leakage for QPSK-OFDM as shown in Fig. 3.7 (a). However, the OoB spectrum radiation caused by the proposed SNN-CF scheme is rapidly diminished compared to ANN, facilitating more relaxed guard-band requirements between different operators. For 16-QAM-OFDM, OoB power leakage of the proposed SNN-CF scheme is consistently lower than ANN, as shown in Fig. 3.7 (b). This is mainly because the proposed SNN-CF is able to learn the amplitude variations of SCF more accurately than ANN.

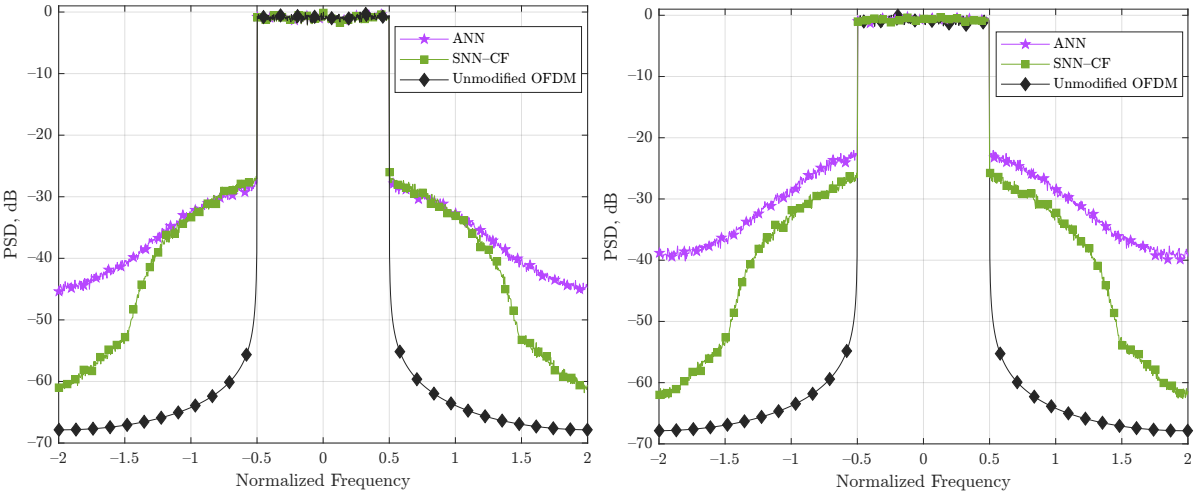
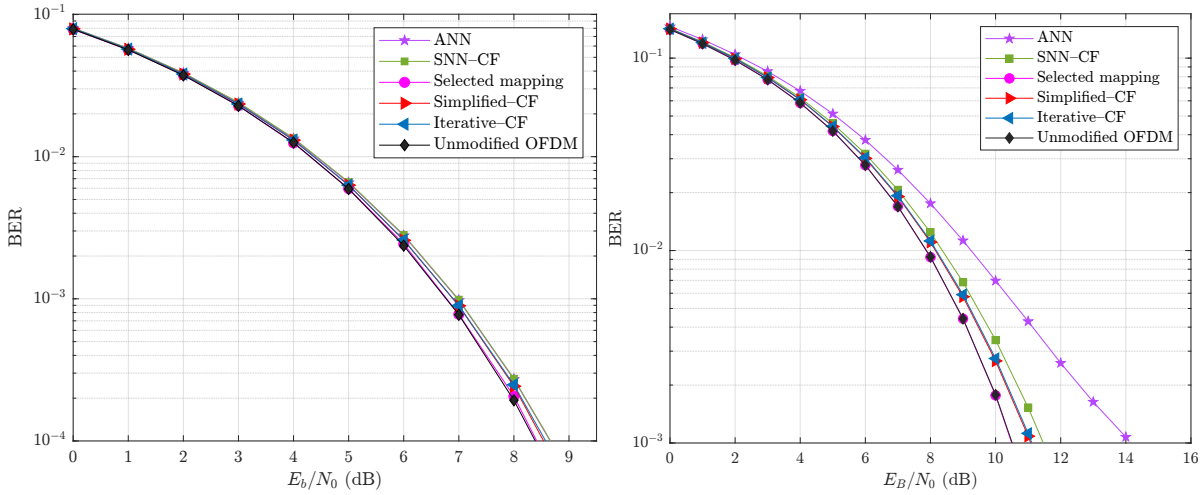


Figure 3.7: Power leakage performance of the proposed SNN-CF vs. ANN [69] and unmodified OFDM in terms of normalized PSD vs. normalized frequency for (a) QPSK-OFDM; and (b) 16-QAM-OFDM modulation.

The BER performance of all QPSK-based schemes are compared in Fig. 3.8 (a) for E_b/N_0 ranging from 0 to 9 dB. As expected, SLM preserves the signal integrity and results in the lowest possible BER which is identical to the BER performance of original OFDM. SCF and ICF come next and are closely followed by ANN and the proposed SNN-CF scheme. For 16-QAM-OFDM transmissions, the proposed SNN-CF scheme remains remarkably close to conventional SCF in terms of BER performance, as shown in Fig. 3.8 (b). Although ANN maintains its CM reduction performance, it starts to exhibit a deteriorated BER performance resulting in about a 3 dB performance gap from conventional SCF at the 10^{-3} BER level. In contrast, the proposed SNN-CF scheme remains within approximately 0.3 dB from conventional SCF at the same BER level. This is facilitated by the more accurate representation of SNN-CF to SCF signals.



(a) (b)

Figure 3.8: BER performance of the proposed SNN-CF vs. ANN [69], selected mapping [86], simplified-CF [65], iterative-CF [62], and unmodified OFDM for (a) QPSK-OFDM; and (b) 16-QAM-OFDM.

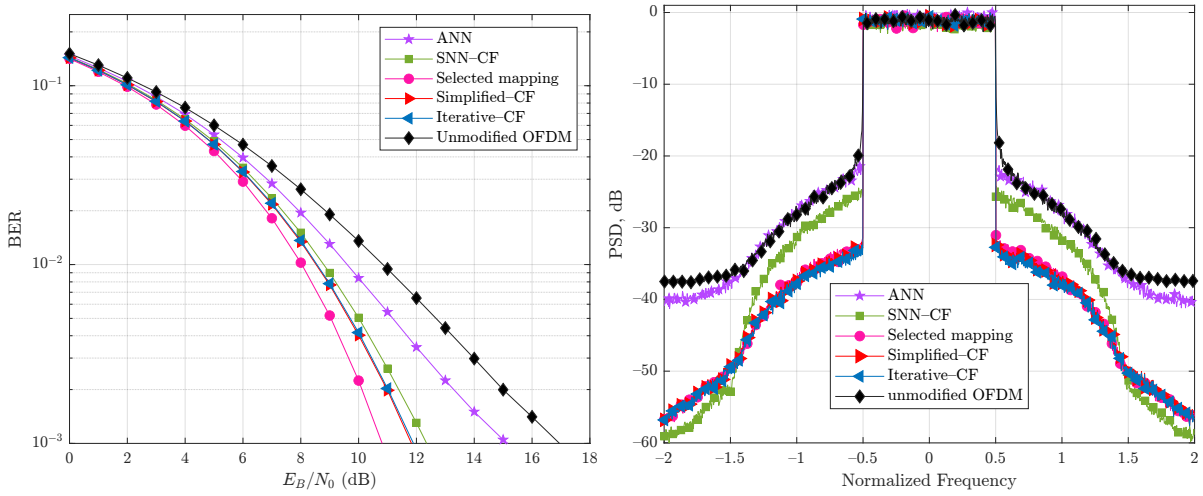
Figure 3.9 (a) plots the BER performance of the proposed SNN-CF scheme vs. ANN, SLM, SCF, ICF, and raw OFDM signals for 16-QAM-based OFDM transmissions deploying a non-linear solid state HPA with the signal envelope amplitude, E, undergoing an amplitude-to-amplitude (AM/AM) conversion accomplished by the transfer function [96]:

$$f(E) = \frac{gE}{\left(1 + \left(g\frac{E}{E_0}\right)^{2c}\right)^{1/2c}}, \tag{3.10}$$

where g is the small-signal gain, E_0 is the output limiting parameter, and c is a parameter that controls the transition rate from the linear operation region to the limiting operation region. These parameters are set to $g = 1$, $E_0 = 1$, and $c = 3$ as in [96]. Although ANN lags behind SCF and ICF with a performance gap exceeding 3 dB at the 10^{-3} mark, the proposed scheme manages to maintain a close proximity to within 0.4 dB of SCF and ICF at the same level.

Figure 3.9 (b) retraces the PSD performance of ANN and the proposed SNN-CF scheme along with the benchmark schemes and unmodified OFDM signals when the solid state HPA is deployed. SLM, SCF, and ICF provide more controlled OoB radiations in the neighborhood of the in-band, where the input back-off of HPA and clipping ratio are set to 3.5 and 6 dB, respectively. Both ANN and SNN-CF exhibit a steady roll-off decline in the neighborhood of the in-band with an almost consistent leakage gap of about

3 dB in favor of SNN-CF. In addition, as we move further away, the gap between ANN and the proposed SNN-CF scheme grows rapidly as ANN roll-off slows down whereas the OoB leakage of the proposed scheme decays significantly and goes below the benchmarks. Although SCF provides better overall PSD performance than the proposed SNN-CF, it requires a noticeably higher processing complexity.



(a) (b)

Figure 3.9: Performance evaluation for 16-QAM-OFDM transmissions deploying non-linear HPA with 3.5 dB input back-off and 6 dB clipping ratio: (a) BER; and (b) PSD.

Figure 3.10 shows an alternative view on OoB power leakage of ANN and the proposed SNN-CF for QPSK and 16-QAM. The normalized out-of-band leakage power (NOLP) γ is defined as

$$\gamma \triangleq 10\log_{10}\left(P_l/P_{in}\right)\, dB\,, \tag{3.11}$$

where P_l is the total leaked power and P_{in} is the power residing within the in-band. Lower NOLP levels are desirable for superior power leakage performance. Therefore, we use the complementary CDF on different threshold values of the NOLP: $\mathcal{F}(\gamma) \triangleq Pr\{\gamma > \gamma_0\}$. The threshold level γ_0 is varied within the ranges $[-42,-24]$ dB and $[-42,-22]$ dB for QPSK and 16-QAM, respectively, in order to capture the OoB leakage performance from CCDF unity level (corresponding to uncrossable threshold points) to levels as low as 10^{-3} (corresponding to crossable threshold points with 99.9% success rate). As shown in Fig. 3.10 (a), the proposed SNN-CF scheme generally outperforms ANN in terms of the reduction achieved on NOLP for QPSK modulation with a performance gap of about 4.25 dB at CCDF level of 10^{-3}. In addition, for the 16-QAM case, the proposed SNN-CF scheme exhibits a much more controlled power leakage than ANN does with an

approximately 7.25 dB performance gap at the 10^{-3} CCDF mark, as shown in Fig. 3.10 (b). This superior performance of SNN-CF compared to ANN is the result of the more accurate peak-tracking of SCF signals, as modeled by the non-linear SNN-CF mapper.

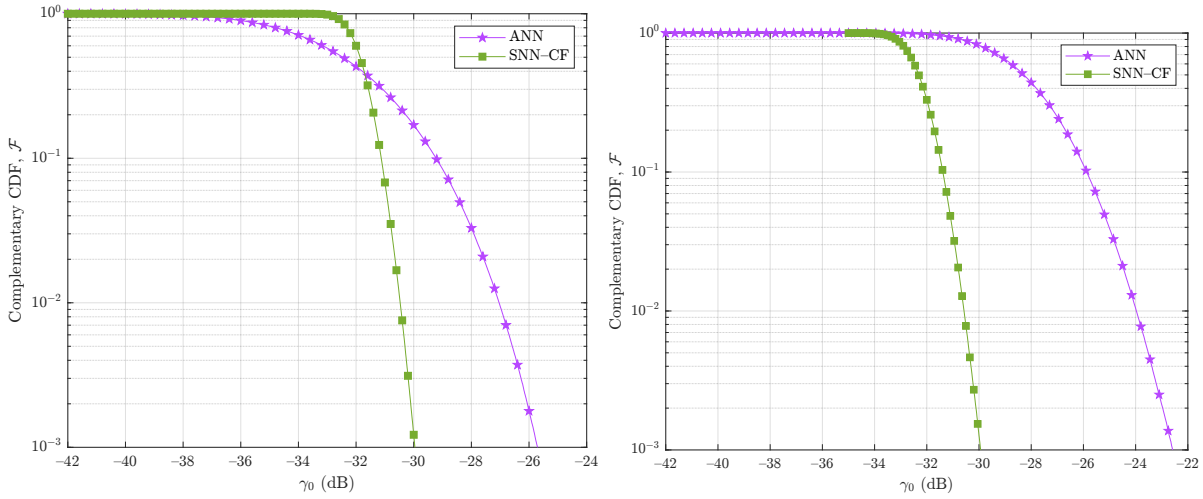


Figure 3.10: Power leakage performance of the proposed SNN-CF vs. ANN [69] in terms of the reduction achieved in terms of the outband-to-inband power ratio for (a) QPSK-OFDM; and (b) 16-QAM-OFDM.

Figure 3.11 investigates the BER performance of the proposed SNN-CF scheme for higher order QAM modulations. For all three cases of 64-QAM, 256-QAM, and 1024-QAM transmissions, the proposed SNN-CF scheme successfully remains within close performance to original OFDM. For example, the performance gap for 64-QAM-based operation is about 1.25 dB at a BER level of 10^{-3}. Similarly, for the 256-QAM and 1024-QAM cases, the performance gaps at the same BER level are about 1.5 and 1.75 dB, respectively. To attain even narrower gaps for higher modulation orders, the underlying neural network module can be expanded by slightly increasing the neurons within the layers, for example, to maintain the complexity within reasonable bounds.

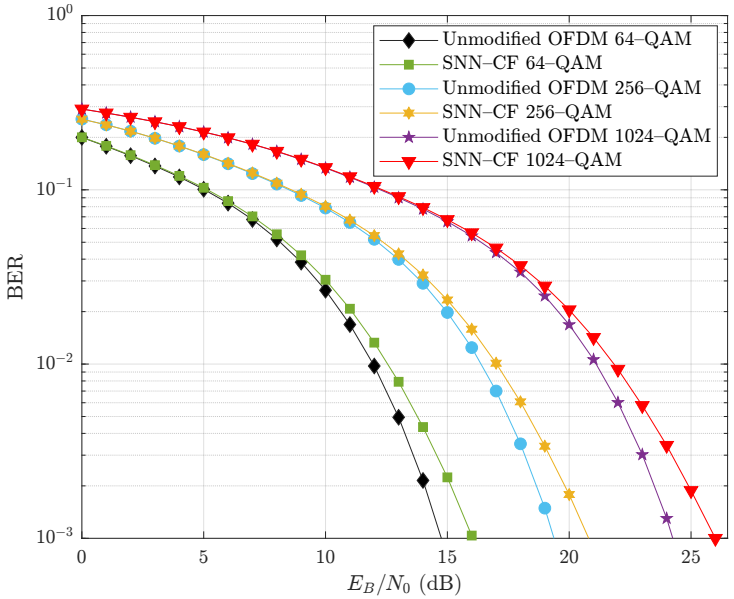


Figure 3.11: BER performance tracking of the proposed SNN-CF scheme for higher order modulations: 64-QAM, 256-QAM, and 1024-QAM.

Chapter 4

Cache-Enabled Reinforcement Learning based Resource Management for Downlink NOMA Systems

4.1 Introduction

Wireless communication is one of the most important technological achievements of the human race. It has gradually become an indispensable component of the modern society as it facilitates our social and business interactions, improves our lives, and contributes to societies' economic growth. It is estimated that by 2030 global mobile subscriptions will increase rapidly to about 17.1 billion, with monthly mobile data traffic reaching 5000 Exabytes (EB), of which 75% will be utilized for mobile video traffic [30]. Moreover, total cellular internet of things (IoT) connections are predicted to reach 5.9 billion as early as 2026 [97].

To accommodate this explosive demand on higher overall system capacity and meet the ever-increasing need for faster data transmission rates, beyond 5th generation (B5G) and future wireless systems must deploy an array of solutions to fulfill and sustain the expected requirements. In that regard, many promising candidate solutions have been put forward by both the industry and the academic community in order to increase the capabilities of wireless cellular networks and achieve a system-wide performance leap

from what 4G networks can provide. These solutions include millimeter wave ($mmWave$) transmissions which were proposed to facilitate very high communication speeds by operating at frequencies well beyond 6 GHz and providing users access to significantly wider transmission bandwidths [98].

Among the solutions with great potential for future wireless systems is a strong alternative to the conventional orthogonal multiple access (OMA) where, unlike OMA, intentional non-orthogonality is introduced to be able to schedule users more frequently and assign them wider bandwidths. In this non-orthogonal multiple access (NOMA) scheme, users are essentially multiplexed in power where the base station (BS) allocates different proportions of its transmitting power to the messages of different users conveyed by a NOMA signal. NOMA-based operation has been demonstrated to outperform OMA-based transmissions in a variety of communication environments [99, 100]. This superior performance of NOMA is mainly due to the aggregation of the data of multiple users over more resource blocks (RBs) than what each individual user would be granted in an OMA setting.

However, NOMA signals suffer from the inherent interference among the different messages of multiplexed user equipment (UEs) which can severely affect the achievable sum-rate levels and degrade the overall system performance. This can be particularly evident in cache-enabled transmission scenarios where the achievable performance becomes sensitive to the relevance of cached data to candidate users in the locale of BSs. Optimized power allocation and proper selection of users to share the same resource blocks (e.g. time slots and frequency channels) are therefore important to unlock the full potential of NOMA systems and reap their benefits.

Reinforcement learning (RL) is a very powerful machine learning (ML) approach that can be tailored to address the aforementioned power allocation and user selection challenge in NOMA systems. In this chapter, we propose an RL-based scheme to solve the problem. To find a state-to-action mapping policy with desirable characteristics, we train the agent by running a series of games during which the agent interacts dynamically with NOMA environments to absorb the various factors affecting its long-term performance. The agent initially executes random actions with high probability and gradually shifts its strategy to rely more often on its progressing policy as it gains more experience while learning. We form a reward function that incentivizes the agent to follow the user selection

and power allocation decisions resulting in higher sum-rate for the activated UEs. The agent aims to maximize the achievable sum-rate by acquiring immediate rewards over subsequent interactions with the NOMA environment. The immediate rewards gained by activating different candidate users is based on the caching status of their requested data.

4.2 Related Works

Optimal pairing of users in downlink NOMA transmissions requires the application of an exhaustive scanning of all selection possibilities, making it practically unrealistic due to the huge computational complexity involved [101]. Various user selection strategies have been proposed including the random pairing algorithm (RPA) and channel state pairing (CSP) [102] where strong users (i.e. users with high channel gains) are paired with weak users (i.e. users with low channel gains). In [103], the authors introduced an unsupervised learning technique for user clustering where an expectation maximization-based algorithm is developed to exploit the spatial correlation between different UEs to solve the user pairing problem. In addition, optimal power splitting among multiplexed users on a given resource block in power-domain NOMA can be accomplished by the method proposed in [104]. However, maximized performance can only be achieved through the joint optimization of user pairing and power allocation, which is a challenging NP-hard problem [105]. To address this issue, the authors in [105] proposed a sub-optimal user allocation scheme based on the correlative properties of the channels of different UEs, which relies on difference-of-convex programming for power allocation thereafter. In [106], the authors exploited the sparse properties of NOMA power allocation to form a convex relaxation of the user pairing and power allocation problem with the aim to minimize total power consumption. However, solving the resulting sequence of problems is computationally expensive. In [107], a relaxed l_1-norm problem is formulated to jointly tackle user pairing and power allocation, and then a compressive sensing-based solution is applied. In [108], the authors investigated maximizing the sum-rate in heterogeneous cloud radio access MIMO networks with fronthaul capacity constraints. Bandwidth allocation is fixed and the joint problem is divided into two problems that are solved separately for optimal power allocation of both macrocell users and remote radio head users. However, NOMA user paring is not considered. In [109], the authors converted the joint problem into a

single-objective formulation based on discretization of power and proposed a global optimal search for power allocation and UE matching within subchannels. However, due to the high computation cost inherent in the technique, it is more tailored towards providing limiting bounds on UE throughput performance.

The aforementioned works do not utilize content popularity in rich multimedia services which have been in high demand lately. Cache-enabled communication is a promising key technology that exploits the popularity of requested data to enable wireless networks to meet various challenges such as reducing latency, increasing spectral efficiency, and alleviating traffic congestion. In this regard, the authors in [110] proposed a vehicle-side caching architecture for vehicular NOMA networks in order to boost the spectral efficiency. Although cache-aided operation is exploited to increase the probability of successful decoding of NOMA signals, user pairing and power allocation are not considered in [110]. In [111], a cache replacement and content delivery joint optimization problem is formulated to proactively push files to a caching unit with the aim to minimize total power consumption in NOMA transmissions. Although cached information is regularly updated to ensure its popularity is up-to-date, backhaul capacity limitations are not considered. As a promising machine learning technique, the authors in [112] deployed reinforcement learning to tackle the power allocation problem in cache-enabled NOMA networks with the aim to maximize the joint decoding capability of the receiving terminals. NOMA user pairing, however, is not considered in [112]. In addition, reinforcement learning has been successfully applied in [113] to solve the user pairing problem and reduce the offloading latency in multi-user NOMA-based mobile edge computing systems. However, NOMA power allocation and backhaul capacity constraints are not considered.

Unlike the above works, the proposed RL-based power allocation and user selection scheme aims to maximize the achievable sum-rate in cache-enabled opportunistic access to NOMA systems suffering from capacity-limited backhaul link connections where requested data is of popular nature in the locale of the serving base station , which, to the best of our knowledge, has not been sufficiently addressed.

4.3 System Model

Consider a downlink NOMA system where N candidate UEs are requesting popular contents (e.g. popular internet data) for download. The serving base station (BS) has

to first fetch the requested information via capacity-limited backhaul link to the core network. Considering an M-user NOMA transmission over a certain resource block, where $M < N$, the BS then compiles and forwards a NOMA message, $x = \sum_{m=1}^{M} \sqrt{\mu_m} x_m$, containing the content of the M UEs with maximum potential contribution to the system throughput. x_m is the message of the m-th active UE and μ_m is the corresponding power allocation. Selected UEs are, therefore, granted active access opportunistically based on rate maximization. Deactivated users can be scheduled for transmission over other resource blocks. In addition, we assume that $\sum_{m=1}^{M} E\left[|x_m|^2\right] = 1$ and $\sum_{m=1}^{M} \mu_m = P_T$ where P_T is the BS transmit power, and that system-level UE-BS association is predetermined.

Assuming perfect successive interference cancellation (SIC) detection at the receiving terminals and that active UEs are ordered according to their channel gain conditions such that $h_1 > h_2 > \cdots > h_M$, then the achievable sum-rate of the active UEs is

$$R = \sum_{m=1}^{M} \min\left(C,\, B\, \log_2\left(1 + \frac{\mu_m h_m}{\sum_{i=1}^{m-1} \mu_i h_m + \sigma^2}\right)\right), \tag{4.1}$$

where C is the backhaul capacity allocated for an active UE, B is the transmission bandwidth, and σ^2 is the noise power. The total capacity of the backhaul, C_T, is divided evenly between active users. Therefore, the capacity per user is calculated as $C = C_T/M$.

4.4 Proposed RL-based Solution

A conceptual framework of the proposed DDQN RL-based user selection and power allocation scheme is depicted in Fig. 4.1. We use a caching unit at the BS side that stores high-demand contents to alleviate the data throttling caused by the backhaul link. Several methods are available to estimate the popularity of different kinds of data to control data storing and clearing operation [114]. As demonstrated in [114], the specific popularity distribution depends on the data content nature. For example, exponential or power-law distributions are appropriate to predict the popularity of online videos whereas log-normal-based distributions can describe online news popularity. However, data storing and clearing following certain distributions is not the scope of this work and we focus on whether data requested by candidate UEs is stored in the cache unit, which can be modeled as a two-state Markov chain to describe the cache status transition probabil-

ities. It is worth mentioning that from the perspective of real-world operation, actual data-caching mechanisms may be working in a way which is not completely captured by the two-state decision process. This may require the RL agent to readjust its policy to closely match the environment behavior and capture its characteristics more accurately if the deviation between the model and the actual environment becomes noticeable. Alternatively, the agent may adopt an empirical-based model which is custom-designed to the specific environment in which the agent will be deployed.

In this section, we propose an RL-based DDQN agent that interacts with the NOMA environment to learn the pairing and power-split decisions resulting in higher sum-rate for the activated UEs. Initially, the agent starts with a random policy that maps the state of the NOMA environment s to a random pairing and power-split action a. The action is executed and a corresponding reward r is observed. The agent then updates a DDQN policy θ based on the rewards observed over many interactions. This dynamic interaction enables the agent to absorb and adapt to the effects of the various factors (e.g. channel conditions of candidate UEs and the caching status of their respective data) within the NOMA environment that impact the achievable sum-rate. During the i-th interaction, the policy network θ is updated using random past experiences (s_j, a_j, r_j, s_{j+1}) towards a target Y_j^T which is generated from past experiences using a target network $\bar{\theta}$. Interaction data is stored in a buffer $\mathcal{B}$, and the target network is periodically updated using the policy network every τ iterations.

Outline of the proposed UE selection and power allocation algorithm is given in Algorithm 1. The basic idea is to learn a state-to-action mapping policy that has higher long-term rewards in terms of the aggregate sum-rate levels attained by various activated UEs. The algorithm is based on the fundamental principle of Q-learning which, under a given policy π, defines the true value of an action a in a state s as

$$Q_\pi(s,a) = E[r_1 + \gamma r_2 + \gamma^2 r_3 + ... | s_0 = s, a_0 = a]\,, \tag{4.2}$$

where $Q_\pi(s,a)$ is the expected sum of immediate and discounted future rewards gained by taking the action $a_0 = a$ in an initial state $s_0 = s$ and following the policy π thereafter. r_1 and $r_l \,\forall\, l > 1$ are, respectively, the immediate and future rewards. $\gamma \in [0,1]$ is a discount factor that trades off the importance of immediate and later rewards. Thus, the optimal value of an action is $Q_*(s,a) = \max_\pi Q_\pi(s,a)$. Evidently, an optimal policy is to

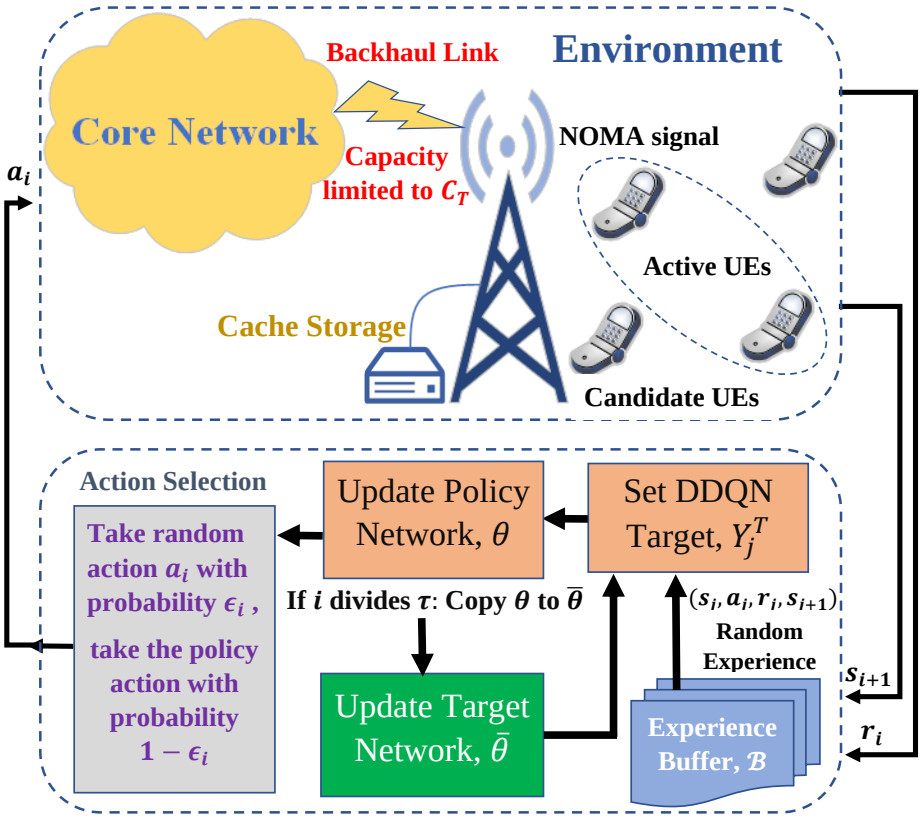


Figure 4.1: Top-level view of the proposed scheme.

select the highest-valued action $\arg\max_a Q_\pi(s,a)$ in each state. The Q-function $Q_*(s,a)$ representing the optimal value of each action can be implemented using a parameterized value function $Q(s,a;\theta_i) \approx Q_*(s,a)$. The parameter set θ_i can be iteratively updated to closely match the optimal value $Q_*(s,a)$. The Q-learning standard update of θ_i after taking action a_i in state s_i and observing the immediate reward r_i and the next state s_{i+1} is

$$\theta_{i+1} = \theta_i + \beta \left(Y_i^Q - Q(s_i, a_i; \theta_i) \right) \nabla_{\theta_i} Q(s_i, a_i; \theta_i) \,, \tag{4.3}$$

where β is a scalar step size, ∇_{θ_i} is the gradient operator with respect to θ_i, and the update target Y_i^Q is defined as

$$Y_i^Q = r_i + \gamma \max_a Q(s_{i+1}, a; \theta_i). \tag{4.4}$$

Double deep Q networks can be deployed to learn the parameterized function $Q(s,a,\theta_i)$. DDQNs utilize a multi-layered neural network to represent the parameters θ_i generating the optimal online policy $Q_*(s,a)$. A second deep neural network with parameter set $\bar{\theta}_i$ is used to generate the learning target Y_i^{DDQN} for DDQN agents. Following the Q-learning principle, Y_i^{DDQN} is evaluated as

Algorithm 1: Proposed RL-based UE Selection and Power Allocation.

– **Set** P_T, $\triangle$, $\mathcal{G}_{max}$, $\mathcal{I}_{max}$, τ, ϵ_s, ϵ_f, α, ϵ_d

– **Initialize** θ, $\bar{\theta} = \theta$, $g = 1$, $\epsilon_1 = \epsilon_s$, $\mathcal{B} = \text{NULL}$

– **While** $g \leq \mathcal{G}_{max}$ run the following mini game:

- Initialize the environment state s_1, $i = 1$
- **While** $i \leq \mathcal{I}_{max}$
  1. Select a random action a_i with probability ϵ_i, otherwise select $a_i = \arg\max_a Q(s_i, a; \theta)$
  2. If a_i violates condition (4.9):
     * For $l \in V$, $V \equiv$ set of UEs violating (4.9), enforce condition (4.9) on a_i by adjusting the power allocation of the l-th violating UE:
     $$\mu_l = \frac{\triangle\sigma^2}{h_{l-1}} + \sum_{k=1}^{l-1} \mu_k$$
     * Perform power normalization for all UEs:
     $$\mu_m \leftarrow \frac{\mu_m}{\sum_{k=1}^{M} \mu_k} P_T$$
  3. Apply a_i to the environment and get its immediate reward r_i and the next state s_{i+1}
  4. Store the interaction experience (s_i, a_i, r_i, s_{i+1}) in a buffer $\mathcal{B}$
  5. Select a random batch of experiences (s_j, a_j, r_j, s_{j+1}) from $\mathcal{B}$:
     * Set the target objective to
     $$Y_j^T = r_j + \gamma \hat{Q}(s_{j+1}, \arg\max_a Q(s_{j+1}, a; \theta); \bar{\theta})$$
     * Perform a gradient descent step w.r.t θ on
     $$(Y_j^T - Q(s_j, a_j; \theta))^2$$
  6. If i is an integer multiple of τ :Smoothe update $\bar{\theta}$ with smoothing factor α:
     $$\bar{\theta} \leftarrow \alpha\theta + (1-\alpha)\bar{\theta},\ 0 < \alpha < 1$$
  7. If $\epsilon_i > \epsilon_f$:Decay ϵ_i towards its final value ϵ_f with decay factor ϵ_d:
     $$\epsilon_{i+1} = \epsilon_i(1 - \epsilon_d), 0 < \epsilon_d < 1$$
  8. $i \leftarrow i + 1$
- **end While**
- Set random action probability for the next game:
  $$\epsilon_1 = \epsilon_{\mathcal{I}_{max}}$$
- $g \leftarrow g + 1$

– **end While**

Where

- $\mathcal{G}_{max}$ and $\mathcal{I}_{max}$ are the total number of mini games played and the number of interactions per game respectively.
- ϵ_s and ϵ_f are the starting and final values of the probability of executing a random action in an agent-environment interaction within the game.
- The parameter τ determines the frequency of updating the deep neural network implementing the target policy.

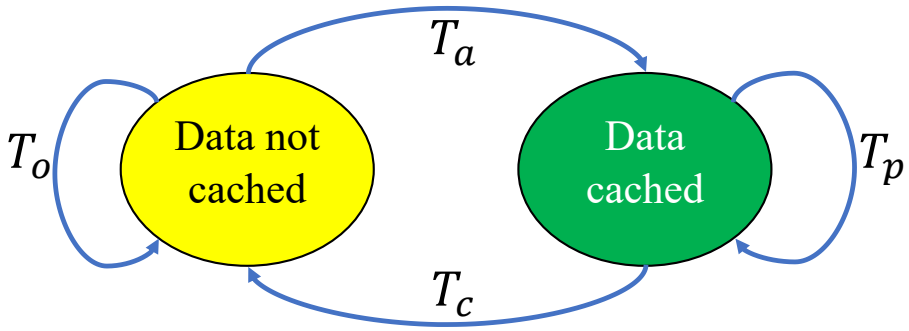


Figure 4.2: Markov chain model for cashed data status.

$$Y_i^{DDQN} = r_i + \gamma \hat{Q}(s_{i+1}, \arg\max_a Q(s_{i+1}, a; \theta_i); \bar{\theta}_i) \,, \tag{4.5}$$

where action selection is based on the online network θ_i whereas its evaluation is based on the target network $\bar{\theta}_i$, and $\hat{Q}(.)$ is the target network response.

NOMA: States, Actions, and Rewards

The state of the NOMA environment at step i is defined as

$$s_i = \{h_1(i), h_2(i), \cdots, h_N(i), u_1(i), u_2(i), \cdots, u_N(i)\} \,, \tag{4.6}$$

where $h_n(i)$ is the n-th candidate channel gain at time step i and $u_n(i) \in \{0, 1\}$ is a corresponding binary indicator to flag whether its requested data is cached or not. The two-state Markov chain defining the transitions of cached data status indicated by $u_n(i)$ is shown in Fig. 4.2. The corresponding cache status transition matrix is

$$T = \begin{bmatrix} T_p & T_c \\ T_o & T_a \end{bmatrix} , \tag{4.7}$$

where T_p is the probability that data persists in the cache and T_c is the probability that cached data is cleared. Similarly, T_o is the probability that non-cached data remains out-of-storage while T_a indicates the probability that it is added to the cache. These probabilities are assumed to be known and therefore can be configured to track different scenarios of cashed content availability (e.g. full, partial, and no-cashing scenarios).

The action a_i taken by the proposed agent on the NOMA state s_i is defined as

$$a_i = \{a_1(i), a_2(i), \cdots, a_N(i), \mu_1(i), \mu_2(i), \cdots, \mu_N(i)\} \,, \tag{4.8}$$

where the binary indicator $a_n(i) = 1$ denotes the activation of the n-th user and $\mu_n(i)$ defines its allocated portion of the transmit power. In addition, the condition

$$\left(\mu_m - \sum_{k=1}^{m-1} \mu_k\right) \frac{h_{m-1}}{\sigma^2} \geq \triangle \tag{4.9}$$

can be applied for the power allocation of the m-th activated UE, $m = 2, 3, \ldots, M$, to ensure efficient detection using SIC at the receiving terminals, where $\triangle$ is a threshold for reliable SIC operation [115]. If the action selected by the DDQN agent violates this condition, the condition can then be enforced before the action is applied to the environment by increasing the fraction of the transmit power allocated to the violating UEs. Power normalization can therefore be applied to ensure that the allocated power is constrained to the transmit power P_T.

The immediate reward of the interaction between the proposed agent and the NOMA environment at step i can be written as

$$r_i = \sum_{n=1}^{N} r_n(i)\,, \tag{4.10}$$

where the reward for the n-th user is given by

$$r_n(i) = \begin{cases} \min\left(C,\, \log_2\left(1 + SINR_n(i)\right)\right), & u_n(i) = 0 \\ \log_2\left(1 + SINR_n(i)\right), & \text{Otherwise} \end{cases} \tag{4.11}$$

where the signal-to-interference-plus-noise ratio (SINR) at the n-th user is computed as

$$SINR_n(i) = \frac{a_n(i)\mu_n(i)h_n(i)}{\sum_{k=1}^{n-1} a_k(i)\mu_k(i)h_n(i) + \sigma^2}. \tag{4.12}$$

The γ-discounted cumulative reward over interaction step horizon I can then be written as

$$R(I) = E[r_1 + \gamma r_2 + \gamma^2 r_3 + \cdots + \gamma^{I-1} r_I]. \tag{4.13}$$

In Algorithm 1, the proposed DDQN agent engages in a series of games with the NOMA environment to learn a state-to-action mapping policy that maximizes the long-term reward of (4.13). In the beginning, the policy network θ is randomly initialized and a duplicate copy is used as the target network $\bar{\theta}$. With probability ϵ_s, a random uniformly

distributed action is then applied to the environment. Otherwise, the action is based on the current policy θ with probability $1 - \epsilon_s$. The immediate reward and resulting state of the NOMA environment are then observed and saved in a buffer $\mathcal{B}$ that stores the information on the agent-environment ongoing interaction. This information forms the agent's experience of the NOMA environment and is used to update the policy network θ at each interaction step. A 4-tuple collection of a NOMA state, an action taken in that state, and the resulting reward and new state forms a training data point within the buffer. A random batch of such data points is used to update the network towards the DDQN learning target Y_j^T. For each data point j, the learning target is selected based on the main policy θ and evaluated through the current target network $\bar{\theta}$. The target network is updated periodically every τ interactions. The update is accomplished smoothly with a factor $\alpha \in (0, 1)$ through a fractional combination of the parameters of the updated main network θ and the current target network $\bar{\theta}$. At the end of each interaction step, the probability for taking a random action is annealed away from the starting value ϵ_s towards a small final value ϵ_f. This annealing process reflects the confidence of the agent to follow its own strategy more often as it gains more experience about the NOMA environment. The number of interaction steps required for the annealing process to complete is controlled by the use of a decay factor ϵ_d. Once ϵ_f is reached, the agent fixes the probability of executing random actions at ϵ_f for all subsequent interactions with the NOMA environment.

The training complexity of the agent depends on a lot of factors and is generally computationally expensive. However, it is of no big concern since it can be offloaded by training the agent on a simulated environment in an offline mode and then deploy the trained agent for online operation. In this work, we use the big $\mathcal{O}$ notation as a comparative metric for the floating-point computational demands of the proposed scheme and the benchmark approaches. For N candidate users, the deployment complexity of the proposed agent with L layers is then $\mathcal{O}(LN)$. This is comparable to the $\mathcal{O}(N \log N)$ complexity of CSP. On the other hand, optimal selection for the activation of M users over a given resource block is on the order of $\mathcal{O}\left(\binom{N}{M}\right)$, which is much more computationally demanding.

In a practical deployment of the proposed scheme to pair users over multiple resource blocks (RBs), the state of the NOMA environment, s_i, would depend on user-activation

actions adopted for previously allocated RBs. For example, if certain users have been selected for transmission over a given RB, then their corresponding channel gain information would be nulled during the allocation of the next RB among the remaining candidates, thus leading to selection choices impacting subsequent states of the overall environment. In addition, in realistic communication systems, activated users would remain active for some time during which the state of the transmission environment in (4.6) may change (possibly even multiple times), and thus the actions adopted at a certain point would have an impact on the system performance at later times, which makes it more suitable to consider the discounted cumulative reward since actions have long-term consequences. One final point regarding the use of Q-learning is that the dynamic process of cashing data at the base station (BS) side is partly driven by the adopted actions since activated users would impact the popularity distributions of different kinds of content within the locale of the BS, which in turn affects later states of the environment with respect to cached content availability as represented by the status indicators in (4.6), and therefore provides yet another good reason for the application of Q-learning. In addition, one important advantage of the proposed scheme is its fast processing speed during deployment. This can be important for practical systems with strict latency demand on the processing time. For example, the latency constraint on NOMA resource management solutions stems primarily from the fact that the operation must be completed within the coherence time of the wireless channel. Thus, a successful deployment of the proposed scheme in a realistic transmission scenario is contingent upon the proposed RL agent's ability to finish an inference of its DNN-based policy within the channel's coherence time. This is particularly important when the operation is mmWaves-enabled since the coherence time of the wireless channel is typically smaller when operating at the mmWaves frequencies. To verify the proposed scheme's ability to conform to the constraint, we tracked the DNN inference time during an agent-environment interaction by evaluating a total of 100,000 interactions. An inference of the agent's DNN-based policy requires a processing time of about 35 μsec. This is noticeably smaller than the coherence time of mmWaves channels which ranges from about 250 μsec to 1.5 $m$sec at 50% correlation when operating at a 60 GHz mmWaves carrier [116]. When operating at lower frequencies over non-mmWaves bands (e.g., using a 3.5 GHz carrier), the coherence time of the channel will be considerably larger and thus further relaxes the constraint on processing speed. For example,

Table 4.1: Simulation settings: Capacity-constrained backhaul operation.

Scheme	Parameter	Value
Common parameters	Carrier frequency	1.9 GHz
	Bandwidth, B	10 MHz
	Channel pathloss exponent	2.9
	Shadowing standard deviation	5.7 dB
	Channel fading type	Rayleigh
	Noise spectral density	-174 dBm/Hz
	BS-to-UE separation	$10 \sim 210$ m
	Cache status transition matrix	[0.6 0.4; 0.4 0.6]
	Backhaul link capacity, C_T	$4 \sim 8$ bps/Hz
	BS Tx power, P_T	$-30 \sim 10$ dBm
Proposed RL-based DDQN	Deep NN training algorithm	Adam optimizer
	Learning rate	0.001
	Target-network update frequency, τ	4
	Soft-update smoothing factor, α	0.001
	Experience buffer size	50000
	Training batch size	8
	Initial random-action probability, ϵ_s	1
	Final random-action probability, ϵ_f	0.1
	Random-action decay factor, ϵ_d	2.3×10^{-4}
	Number of mini games	1000
	Number of interactions	500

the average coherence time at 3.5 GHz is about 25 $m$sec [117]. Therefore, the proposed scheme can operate adequately within the time limit in both cases. By contrast, an exhaustive scan takes over 1 $m$sec to finish. While this may be enough for operation at sub-6 GHz frequencies, it violates the processing time limits when operating at mmWave carriers.

4.5 Simulation Analysis

4.5.1 Environment Setup

The parameters of the simulation environment are configured as in Table 4.1. Unless otherwise stated, the table lists the default values used for the simulation parameters across all studied scenarios. We assume a NOMA system with candidate UEs requesting the download of popular data. The carrier frequency is set to 1.9 GHz with a default transmission bandwidth of 10 MHz.

The close-in (CI) channel model is used with a pathloss exponent of 2.9 for non-line-of-sight (NLOS) propagation [118]. The shadowing standard deviation is set to 5.7 dB. The reference distance of the CI channel model is set to 1 meter since it provides high model accuracy and parameter stability for indoor as well as outdoor urban macro

and urban micro-channels across a wide span of *mmWave* and microwave frequencies. Although such a small value for the reference distance may lie within the near-field range of large arrays of antennas, the resulting error is negligible in the design of practical wireless systems [118, 119].

A typical value of -174 dBm/Hz is used for the noise spectral density. The separation distances from the BS to candidate UEs range from 10 to 210 meters. The probability that data persists in the cache is set to $T_p = 60\%$ and the probability that cached data is cleared is set to $T_c = 40\%$, whereas the probability that non-cached data is added to the cache is configured to $T_a = 60\%$ while the probability that it remains out-of-storage is $T_o = 40\%$. Backhaul link capacities of 4 and 8 bps/Hz are considered. The total BS transmit power is varied from -30 dBm to 10 dBm.

The deep neural network deployed by the DDQN agent consists of five fully-connected layers, with four hidden layers utilizing rectified linear unit (ReLU) activation (i.e., $f(x) = \max(0, x)$) followed by an output layer with linear activation (i.e., $f(x) = x$). The five layers have 100,95,90,85, and 100 neurons, respectively.

The Adam optimizer [120] is used to update the parameter set θ of the deep neural network implementing the state-to-action mapping policy of the RL agent. The corresponding training learning rate is set to 10^{-3}. The target-network parameter set $\bar{\theta}$ is soft-updated every 4 agent-environment interactions. Soft updating is accomplished through a smoothing factor of 10^{-3}. The buffer used by the agent to collect the acquired experiences over many interactions with the NOMA environment is configured to store a maximum of 50000 experiences.

The policy network is iteratively trained using batches of 8 experiences each. The initial probability of executing a random action in the NOMA environment is set to 100%. The probability of taking a random action in any given state is gradually annealed towards a final value of 10%. The annealing process is configured to last 10000 interactions by setting the decay factor ϵ_d to 2.3×10^{-4}. The maximum number of mini games played is configured to 1000 with each game consisting of a maximum of 500 interactions.

4.5.2 Results and Discussion

For cache-disabled transmissions, the achievable sum-rate performance of the proposed RL-based DDQN scheme is evaluated in Fig. 4.3 (a). Three benchmark approaches are

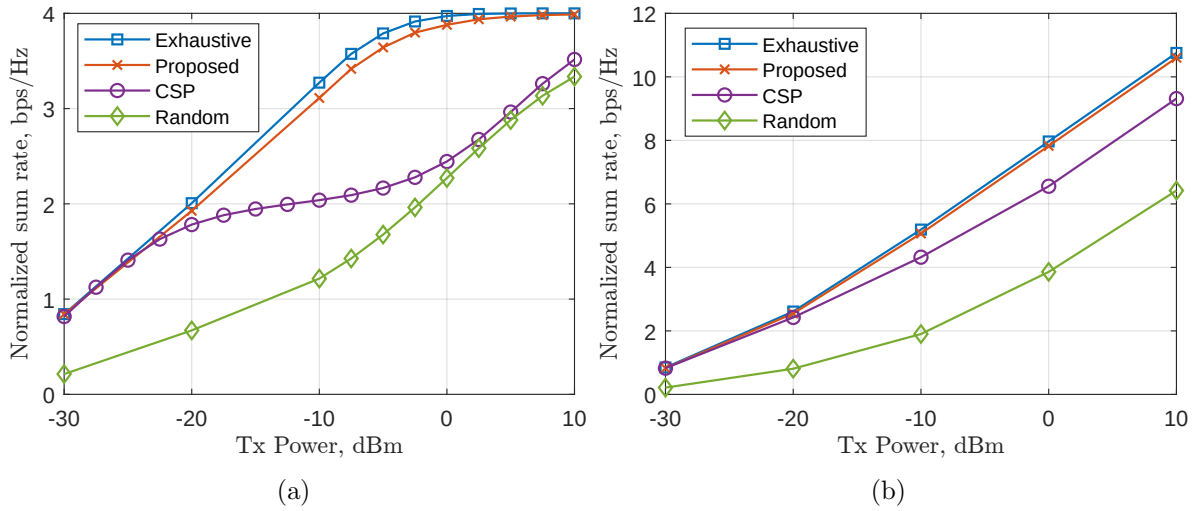


Figure 4.3: Performance of the proposed RL-based DDQN scheme vs optimal strategy, CSP algorithm [102], and random mapping policy in terms of normalized sum-rate vs transmit power with backhaul capacity of 4 bps/Hz for (a) cache-disabled, and (b) cache-enabled transmissions.

included: random and optimal mapping policies as well as the conventional CSP algorithm where candidate users are sorted according to the quality of their channel links and then users with high channel gain separation are paired [102]. The schemes are compared in terms of the normalized sum-rate for different BS transmit power levels ranging from -30 dBm to 10 dBm. The exhaustive search algorithm actively seeks the actions maximizing the achievable sum-rate in an online fashion during deployment. All possible alternatives are exhausted to ensure the highest-utility action is applied. The proposed agent manages to learn a near-optimal state-to-action mapping policy that converges to the capping limit of 4 bps/Hz at the 10 dBm transmit power level. At low transmit power values, optimal, proposed, and CSP algorithms exhibit similar performance since potential relative gains among candidate UEs are not noticeable. However, as transmit power is increased, potential contributions of different candidate UEs start to make a noticeable impact leading to a widened gap between more optimized selections (provided by exhaustive strategy and the trained agent) and the CSP algorithm. However, since the backhaul limits the maximum achievable sum-rate to 4 bps/Hz, further increase of the transmit power ensures that CSP eventually catches up and the performance gap starts to narrow again at high transmit power levels.

Similarly, Figure 4.3 (b) plots the normalized sum-rate vs BS transmit power for cache-enabled communications. The backhaul throttling is alleviated by virtue of cache unit deployment and normalized sum-rate levels as high as 10.75 bps/Hz can be reached

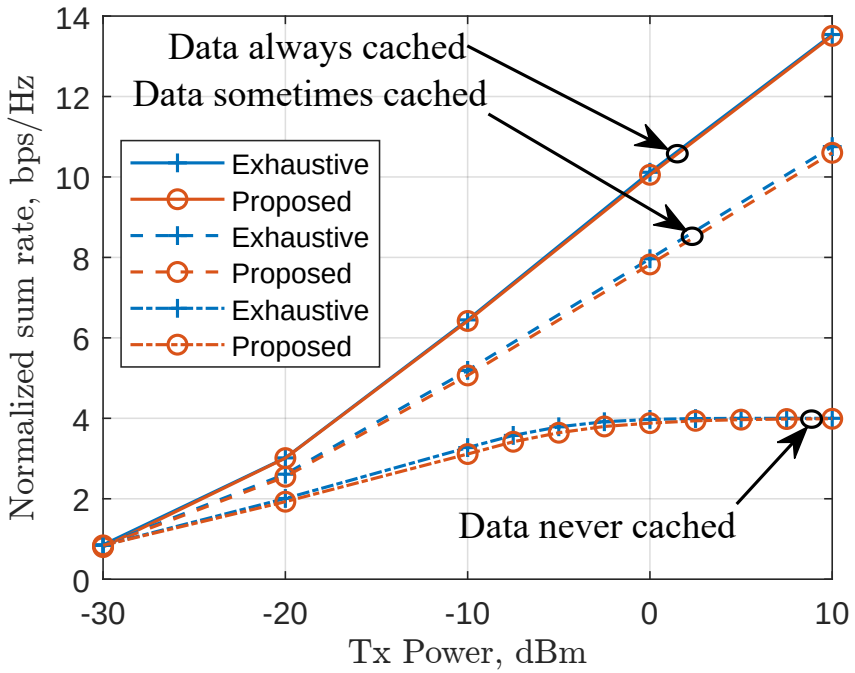


Figure 4.4: Sum-rate performance tracking of the proposed scheme for different data caching scenarios.

via optimal exhaustive strategy at transmit power of 10 dBm. The proposed RL agent closely follows the optimal policy and manages to achieve about 10.6 bps/Hz at the same transmit power level. Since the limiting ceil of 4 bps/Hz on the achievable sum-rate is broken whenever requested data by an activated UE is pre-cached, the gap between CSP and the proposed algorithm remains without shrinking at relatively high transmit power levels.

Figure 4.4 tracks the ability of the proposed RL agent to lock onto the optimal policy for three distinct data caching scenarios going from one extreme case where data is never cached to the other extreme scenario where local copies of requested data is always cached. The data-caching probability matrix for the first case, where requested data is consistently not available in the cache unit, is set to $T = [0 \;\; 1; \; 1 \;\; 0]$. The second scenario, where data is fully cached, is represented by the data-caching probability matrix $T = [1 \;\; 0; \; 0 \;\; 1]$. The last scenario of partial data caching is represented by $T = [0.6 \;\; 0.4; \; 0.4 \;\; 0.6]$. Clearly, the first scenario is equivalent to the case where the cache unit is not installed as evidenced by the saturating sum-rate performance that approaches the backhaul link capacity of 4 bps/Hz at the 10 dBm transmit power level. On the other extreme, however, the achievable sum-rate of the second scenario first exceeds the 4 bps/Hz barrier at power levels as low as -17 dBm with the gap increasing until it reaches about 3.4 folds at the 10 dBm power level compared to the case where requested data is never found in the cache.

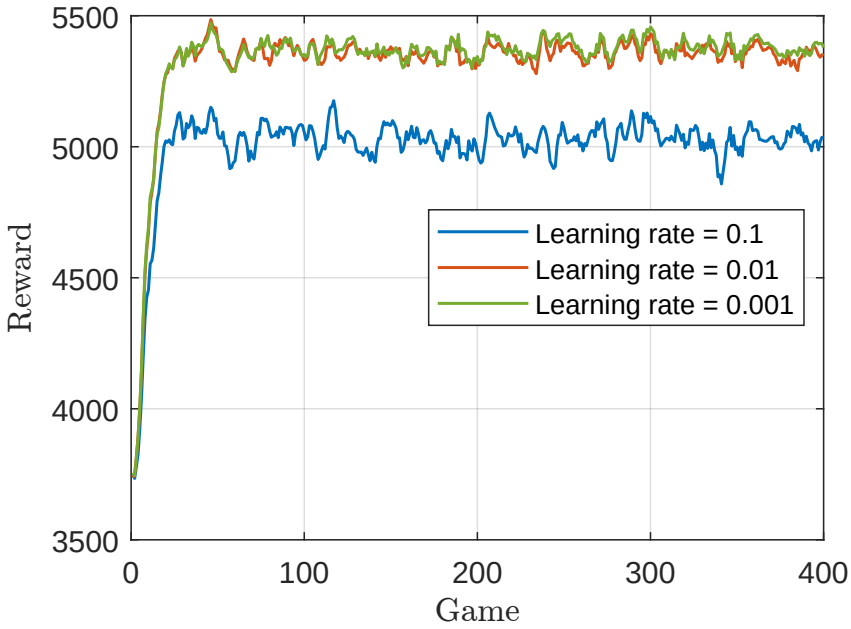


Figure 4.5: Training performance tracking of the proposed scheme for different learning rates.

In the last case representing the scenario where requested data is only partially cached, the achievable sum-rate is naturally lower than the ideal scenario where requested data is always available in the cache unit. However, it is a more realistic scenario to assume and it can still provide considerable improvement over the case where the cache unit is not utilized, especially for high levels of transmit power since it does not have performance floor by virtue of proactive data caching. For example, in this case the sum-rate first goes above the limiting threshold at around -14.2 dBm and achieves about 2.7-fold increase over the case where data is consistently unavailable in the cache unit at the 10 dBm power level. In all three case, the trained RL agent manages to successfully track the optimal policy.

Figure 4.5 examines the training progress of the proposed RL agent as it interacts with the NOMA environment to acquire a state-to-action mapping policy that can achieve higher long-term rewards. Specifically, the plot investigates the dependence of the learned policy on the learning rate parameter and how it can affect the obtained reward over subsequent game runs. Setting the learning rate value to 0.01 enhances the performance by about 7% compared to the case where a faster learning rate of 0.1 is used. Slowing the learning rate further to 0.001 boosts the performance by an extra 1% indicating a diminishing return effect on the rewards obtained over 400 games. Therefore, even though slower learning rates can achieve higher long-term rewards, careful consideration of an appropriate learning rate is important to balance reward performance and training speed

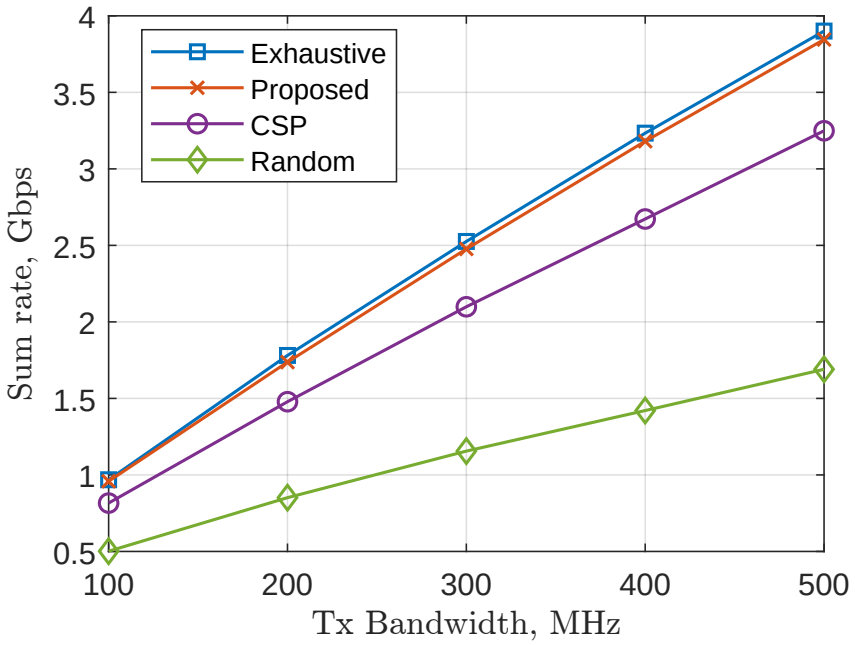


Figure 4.6: Sum-rate performance vs transmission bandwidth at carrier frequency of 38 GHz in the presence of *mmWave* blockers.

of the Adam optimizer. A learning rate of 0.001 is commonly used and is recommended by [120]. However, depending on the particular application requirements, a faster or slower rate may be more suitable.

Figure 4.6 plots the achievable sum-rate performance of the random and exhaustive search strategies as well as CSP and the proposed scheme for 38-GHz *mmWave* NOMA transmissions at 40 dBm transmit power level over a range of transmission bandwidths from 100 to 500 MHz with a backhaul connection limited to 8 bps/Hz. Transmissions at such high frequencies typically suffer blockage effects caused by objects of various sizes within the transmission medium. The blocking probability at a BS-UE separation d can be modeled as, $P_b = 1 - e^{-2d\rho\bar{D}\delta/\pi}$, where $\rho > 1$ is a parameter representing the density of blockers relative to the BS and is set to 5, $\bar{D}$ is the average blocker length and is set to 20 meters, and δ denotes the BS distribution density and is set to 1.92×10^{-5} to reflect the case when hexagonal cells of radius 200 meters are homogeneously deployed [121]. The proposed RL agent manages to adapt to the NOMA environment conditions and learns to closely follow the optimal policy to within 98.5% at the 500 MHz bandwidth level. By contrast, CSP reaches 83.3% of the optimal value and achieves a sum-rate of 3.25 Gbps utilizing the same amount of bandwidth. Finally, the random search approach falls behind at 43.3% of the optimal value as it lacks the ability to account for the presence of *mmWave* blockers.

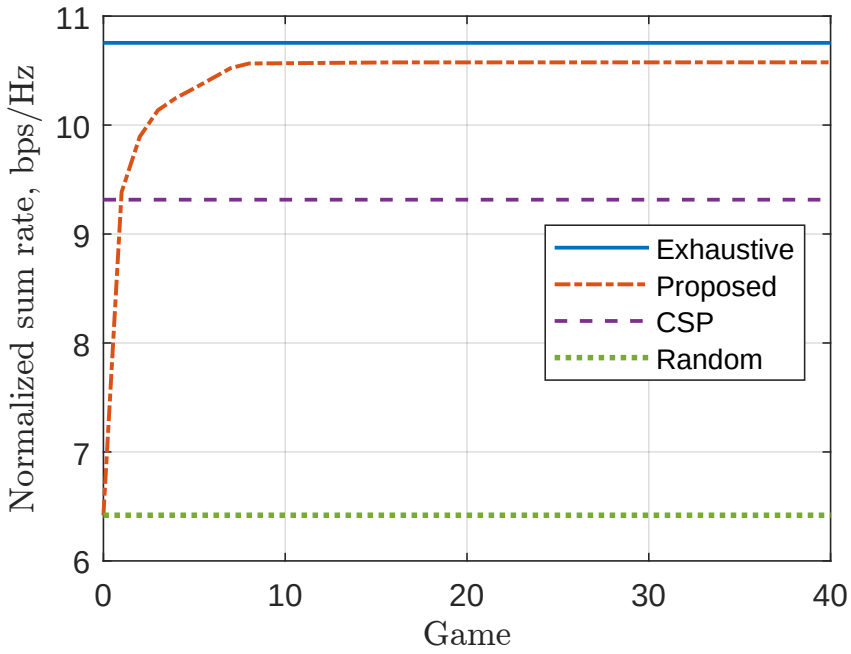


Figure 4.7: Sum-rate convergence performance of the proposed scheme with backhaul link capacity of 4 bps/Hz and BS transmit power of 10 dBm.

Figure 4.7 illustrates the convergence of the proposed RL agent to near-optimal sum-rate level at 10 dBm of transmission power and 4 bps/Hz backhaul connection. Unlike the benchmark schemes where the mapping policy does not evolve dynamically over iterated interactions with the NOMA environment, the proposed RL agent initially starts with a random policy and progresses towards the optimal policy as it interacts with the NOMA environment to learn more optimized decisions with higher long-term rewards. Therefore, the proposed RL agent initially exhibits low sum-rate performance that keeps on increasing as more games are run until it eventually reaches a saturated level of average normalized sum-rate exceeding 98% of the maximum possible level of the exhaustive search strategy. The proposed scheme firstly exceeds CSP by the second game played where the learned policy achieves an average sum-rate of 9.38 bps/Hz as opposed to 9.32 bps/Hz using the CSP algorithm. Starting from the third game, the sum-rate performance of the proposed RL agent rapidly increases to about 97.8% of the maximum level by the eighth game. After this point, the learned policy slowly improves to within 98.5% of the optimal policy.

To validate the robustness of the proposed scheme, we investigate its performance in Fig. 4.8 with full state information (FSI) vs partial state information (PSI) availabilities for both cache-enabled (Cache-ON) and cache-disabled (Cache-OFF) NOMA transmissions, with a backhaul link connection limited to 4 bps/Hz. In PSI, the cache state

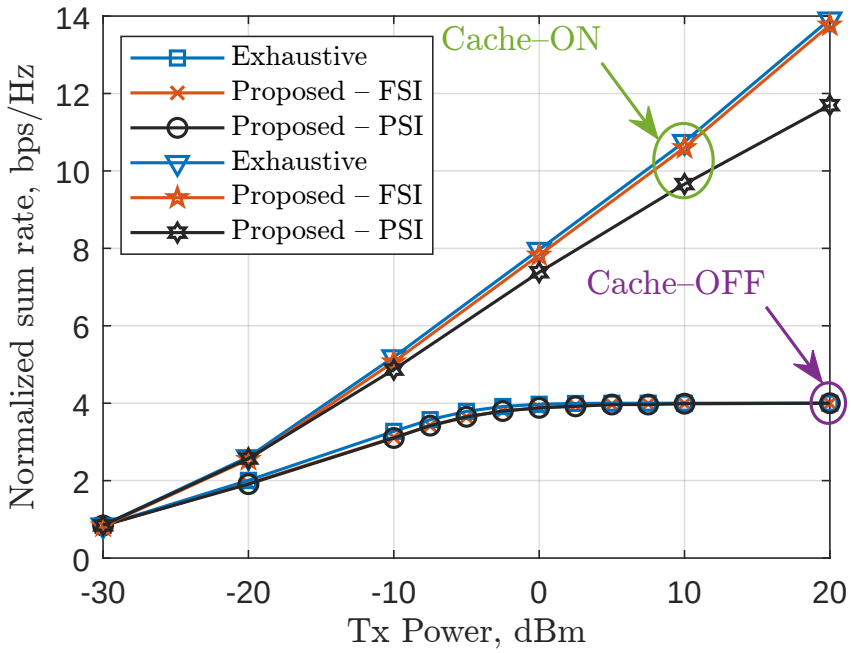


Figure 4.8: Performance tracking of the proposed scheme with full state information vs partial state information availabilities for both Cache-ON and Cache-OFF scenarios.

information is not provided to the RL agent. The normalized sum-rate is evaluated at different BS transmit power levels ranging from -30 dBm to 20 dBm. As expected, in the Cache-OFF scenario, since data caching is disabled, providing the proposed RL agent with only PSI does not have a degrading effect on its performance compared with FSI availability. Therefore, the proposed agent adapts to the NOMA environment to maintain close tracking of the optimal policy and approaches the maximum achievable sum-rate of 4 bps/Hz at 10 dBm of transmission power even when only PSI is available. However, the merit of providing the agent with FSI becomes apparent when caching local copies of trending popular content in the locale of the serving BS is utilized. For this Cache-ON scenario, the aggregate sum-rate can be boosted well above the limiting threshold of 4 bps/Hz to reach about 13.92 bps/Hz at the 20 dBm power point as achieved by the exhaustive strategy. The proposed RL agent with FSI comes in second at 13.77 bps/Hz whereas PSI availability causes a sum-rate performance degradation of about 15% at the same power level. From a transmit power perspective, an agent with PSI requires an extra 4.6 dB to reach the same 10.6 bps/Hz achieved by an agent operating at 10 dBm with FSI. This demonstrates the effectiveness of utilizing the cache state information by the proposed RL agent.

In Fig. 4.9, we examine the effect of the backhaul capping threshold on the achievable performance and verify the ability of the proposed RL-based agent to adapt its policy

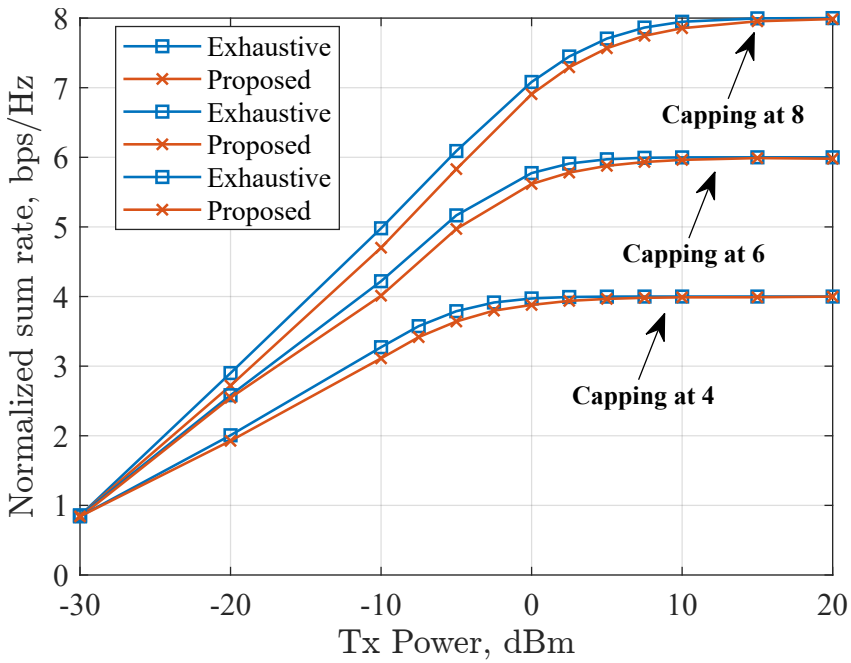


Figure 4.9: Performance tracking of the proposed scheme for different backhaul throttling levels for cache-disabled communications.

for different throttling settings. As expected, when the normalized backhaul capacity is configured at 4 bps/Hz, the achievable sum-rate gradually saturates towards the upper limit of 4 bps/Hz as the transmit power is increased. Similarly, when the capping level is set to 6 and 8 bps/Hz, the achievable sum-rate level is capped at 6 and 8 bps/Hz, respectively. In all three cases, it is clear that the proposed scheme can consistently support near-optimal sum-rate levels irrespective of the backhaul capping limit. This demonstrates the adaptability and effectiveness of the proposed solution.

Figure 4.10 demonstrates the importance of maintaining an optimized NOMA resource management operation even when proactive data caching is not needed for transmission scenarios enjoying unrestricted backhaul operation with no throttling effects. As stated, here we assume that the connected backhaul link is not imposing any throttling effects on the system's operation, and thus the achievable sum-rate level is solely determined by the attainable communication rate over the forward link between the transmitting base station and the activated NOMA users. As can be seen from the figure, the exhaustive search algorithm provides the highest possible sum-rate levels as expected. The proposed scheme, however, can support near-optimal speeds across the transmit power range, reaching about 99.75% of the exhaustive's performance at the 10 dBm point. Nonetheless, performance degradation of as much as 37% can be incurred if the operation is not optimized. This clearly demonstrates that even when the backhaul link capacity

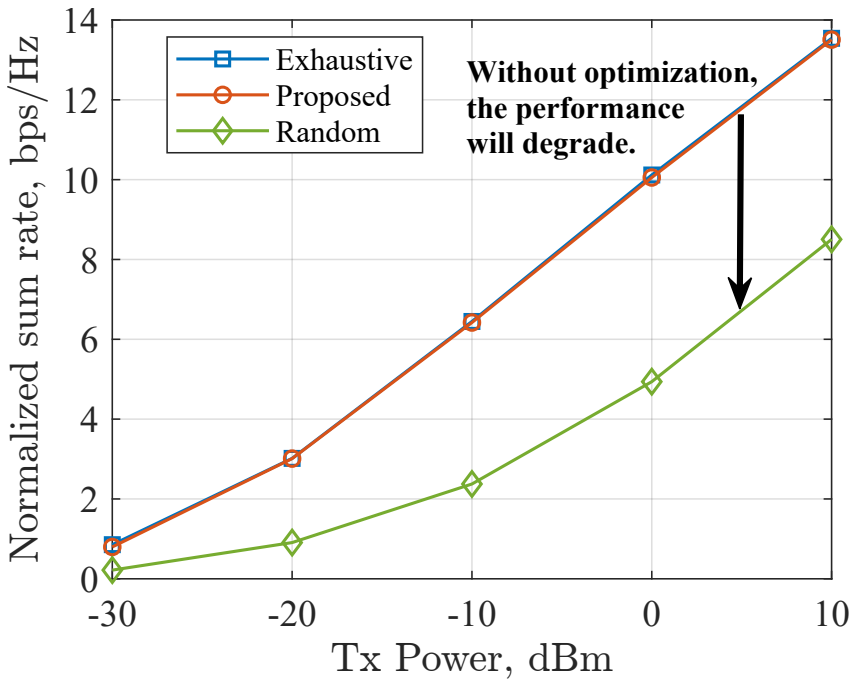


Figure 4.10: Performance tracking of the proposed scheme when the backhaul link is *not* throttled.

is not limited, the proposed NOMA user selection and power allocation optimization is still needed to maintain near-optimal communication rates.

Chapter 5

UAV-Assisted Mission-Critical 3D-NOMA Downlink

5.1 Introduction

Beyond 5G (B5G) and 6G cellular networks face design challenges due to their increased requirements on massive connectivity and communication speeds [122–128]. This can be particularly pressing in zones stricken by disasters where the primary base stations (BSs) infrastructure is momentarily out of commission owing to sustaining severe or mild impairment. In such scenarios, ad-hoc intervention based on dispatched unmanned aerial vehicles (UAVs) can allow for a quick and suitable remedy to maintain adequate coverage and provide high-speed reliable wireless connections to offload downlink data to appropriately activated receivers within the afflicted region [80, 81, 129–137]. The mobile UAV base stations (UAV-MBSs) can therefore play the pivotal role of rapidly dispatched mobile BSs within certain target regions.

Incorporating mmWaves into UAV-based transmissions can provide important advantages to the communication system. The huge bandwidth resources provided by mmWaves can help the UAV-mounted BS to support high-speed communications as well as flexible coverage [138, 139]. For example, the authors in [138] have studied UAV-mounted BSs to support dynamic rerouting for reconfigurable backhauls operating over mmWaves bands. In [139], the authors proposed a beamforming technique to support flexible coverage within target zones by exploiting mmWave-enabled UAV transmissions. Moreover, the availability of the LOS component in UAV-based systems is suitable for

mmWaves-based communications aiming to reach high gains [140, 141].

On the other hand, non-orthogonal multiple access (NOMA) downlink protocol is an efficient multiplexing candidate approach for the UAV-MBS to utilize to satisfy the connectivity and transfer speeds requirements set forth for B5G and 6G wireless systems. Contrary to conventional orthogonal multiple access, NOMA-based transmissions are preferred because they have been demonstrated to offer better overall performance through stacking the data of multiple receiving devices (RDs) using a unified resource block (RB) design, wherein jointly-multiplexed devices would enjoy a larger transmission bandwidth as well as more frequent scheduling [99, 100, 142–145]. Data frames conveying information of multiplexed receivers are sent over the unified RB at varying levels of transmission power to enable each device to successfully recover its own intended data by applying successive interference cancellation (SIC) to sequentially retrieve then remove the messages within the received NOMA stack until it extracts its intended message signal [43]. Energy-efficient planning of the dispatched UAV-MBS flying course throughout the entire communication period is imperative so that the UAV's battery use is optimized. In addition, it is of critical importance to optimize the continuous adaptation of various allocated power portions within a maximum allowable budget of available transmission power as well as the dynamic activation of the receiving devices to reap as much of the promised performance of NOMA-based operation as possible [105, 107]. Moreover, making proper choices regarding the selection of appropriate receiving devices to add to a certain NOMA message stack is important to attain boosted sum-rate levels [146]. In addition, each time the UAV moves position, power allocation and device paring need to be re-optimized, resulting in a surge in complexity and energy consumption. This is not acceptable especially when coupled with the UAV's limited battery life.

The aforementioned challenges have not been sufficiently addressed to the best of our knowledge. Recently, reinforcement learning (RL)-based methods have been attracting the attention of the research community due to their effectiveness and inherent flexibility in dealing with highly dynamic sequential decision problems. In this chapter, we present two proposed algorithms based on the powerful RL framework to address the joint issue of energy-efficient dynamic UAV-MBS path design, receiving device activation, and transmit power distribution for high-speed NOMA-UAV-based downlink wireless communications. In particular, multiarmed bandit (MAB) and double deep Q-network (DDQN) RL agents

are employed to leverage their highly adaptable nature to handle various dynamic and complex models. For the proposed MAB approach, we consider the two variants: minimax optimal stochastic strategy (MOSS) as well as upper confidence bound (UCB) for their simple yet effective deployment. For the DDQN-based algorithm, the RL agent training can be carried out offline where the agent engages in multiple interactions with the UAV-NOMA environment model before it is dispatched for operational deployment. The DDQN RL agent learns an effective deep neural network (DNN)-based strategy which it uses to determine the appropriate projections of the environment's subsequent states onto a series of decisions yielding high returns in the long-term. On the other hand, no DNNs are incorporated in the MAB-based approach which is deployed directly to make on-the-fly online decisions while aiming to attain adequate performance in terms of the achievable total data rate level through the dynamic selection of various allowable actions according to their varying levels of some appropriate fitness criteria to determine their effective utilities. The utility of making various decisions are updated continuously over the communication time horizon.

5.2 Related Work

Pairing NOMA receiving devices optimally for downlink transmissions generally requires a complete scan exhausting all possible groupings. However the computational burden of such a brute force approach is huge which deems the application of the optimal solver unrealistic from a practical standpoint [146]. Numerous strategies were devised for the appropriate grouping of receiving devices over unified NOMA RBs. Famous benchmarks include the random grouping algorithm (RGA) which samples the action space using a uniform random decisioning strategy and channel-state grouping (CSG) [102] wherein strong nodes (i.e., devices with boosted channel conditions) are grouped with weaker nodes (i.e., devices with attenuated channel conditions). The authors in [103] developed a technique leveraging unsupervised learning for node clustering wherein they developed an algorithm based on expectation maximization by harnessing spatial correlative patterns among different nodes to solve the device grouping problem.

Although optimized distribution of the power available for signal transmission over receivers multiplexed on a given RB can be attained for power-domain NOMA systems using the technique outlined in [104], optimized operation results can be accomplished

only by considering the joint problem of device grouping and power distribution, a taxing NP-hard problem [105]. Upon handling the problem in [105], the authors devised an allocation scheme that operates suboptimally by constructing a correlation structure for the downlink channels, and then deploying difference of convex optimization to distribute the available power. With a focus on reducing the amount of power consumption, the work in [106] leveraged the sparse nature associated with NOMA power distribution to formulate a convex-relaxed version of the power distribution and user grouping problem. However, the presented technique incurs high computational demand to solve the formulated string of problems. A relaxed version based on l_1-norm characterization is formed in [107] for the joint problem of power distribution and device grouping, wherein the authors applied a solution method based on compressive sensing.

Although the above body of research work handles the problem of NOMA user grouping and power distribution for a variety of conventional wireless communication scenarios, it fails to accommodate less common yet important mission-critical scenarios such as emergency-oriented communications wherein an effective and reliable transmission system which can be quickly deployed is needed. One possible use case, for example, is dispatching UAV-mounted mobile BSs to blackout sites where receiving devices are disconnected from the main servicing infrastructure. To incorporate this need, an emergency network UAV-aided framework is developed in [80] for operation in disaster zones. The scheduling and trajectory of UAV-MBSs are firstly designed to support wireless coverage to the receiving devices on the ground. Afterwards, to expand the UAV-MBS wireless service domain, the authors formed a ground-based multi-hop D2D system and studied the UAV-MBS transceiver design. However, the generic system presented in [80] does not account for NOMA resource management optimization. To address this point, the authors in [81] established a UAV-assisted framework for NOMA-based emergency communications. The proposed scheme started by establishing a UAV-active uplink line to collect information relevant to the IoT devices within the areas under emergency operation. Subsequently, to support coverage for IoT users, a joint power management and UAV dispatching scheme is proposed. However, downlink NOMA user grouping is not considered. To handle this issue, the work presented in [82] combines both NOMA power distribution and user grouping with UAV-MBS path design and optimizes the operation jointly with a sumrate-maximization objective in mind. However, energy-efficient

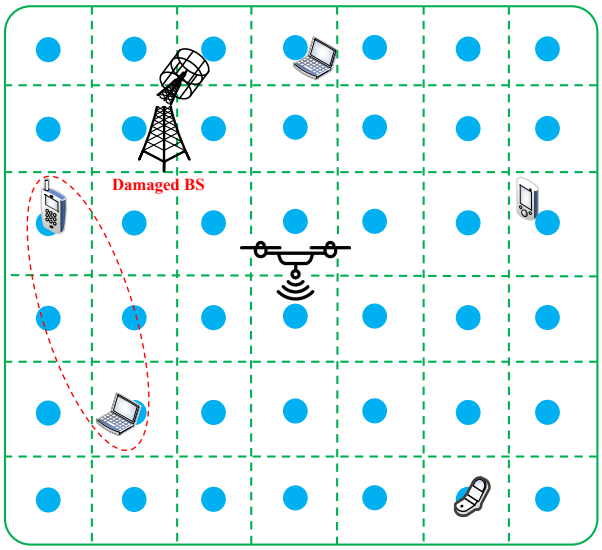


Figure 5.1: NOMA-UAV-MBS system model.

operation is not guaranteed since battery-aware design is not considered.

In this chapter, we consider energy-driven design of the joint problem of dynamic UAV-MBS path planning and downlink NOMA user grouping and power distribution where battery-constrained operation is taken into account and optimized routing is accomplished through the deployment of the proposed RL-based frameworks where the RL agents are aiming to maximize the total rate while operating in a battery-constrained mode for energy-efficient UAV-MBS deployment.

5.3 System Model

Consider a downlink UAV-MBS-based system operating via NOMA protocol to offload information data to a group of wireless receiving devices as illustrated in Fig. 5.1. Application scenarios for such a UAV-based communication system include deployment in emergency cases (e.g., environments affected by a natural disaster [80, 81]) where ground users are experiencing connection issues due to temporary damage of the nearby BS as depicted in Fig. 5.2. In this case, the serving UAV-MBS can offload emergency communications data to ground users via a multi-hop UAV relay chain or by utilizing HAPS link communications. Another practical use case is to establish temporary hotspot communication links for suburban or rural environments [147]. Initially, the transmitting UAV-MBS begins at some arbitrary position (e.g., center of the flying zone). The coverage area is split into multiple small regions of potential spots for hovering, with the blue circles marking the hovering positions' centers. The receiving devices are scattered

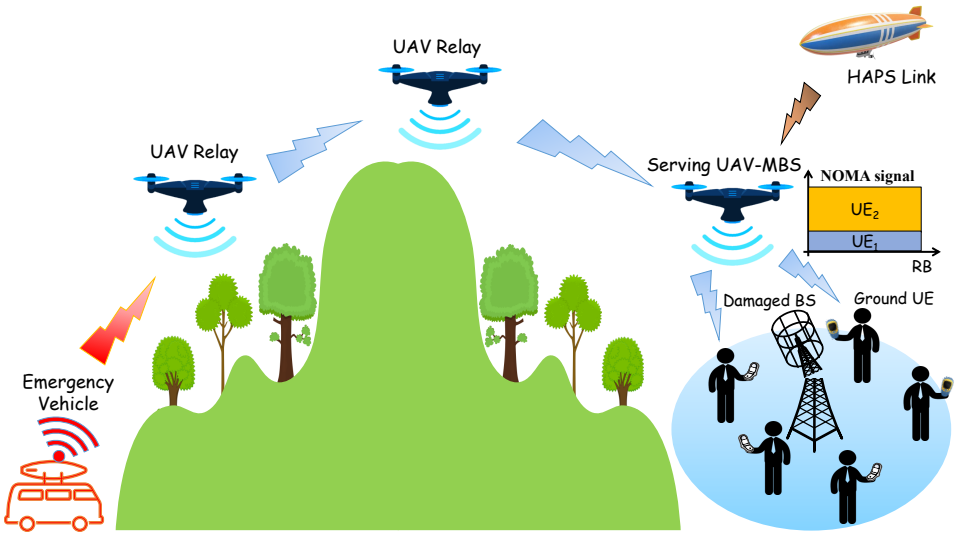


Figure 5.2: NOMA-UAV-MBS emergency environment [80, 81].

arbitrarily throughout the flying zone covered by the UAV-MBS. To optimize the rate of transferring data, the UAV-MBS transmitter must carefully select the receiving devices to activate and offload data to as well as allocate the available transmit power efficiently. In addition, enery-efficient dynamic course planning is important to position the UAV-MBS in an optimized manner across all the allowable hovering spots so as to accommodate and account for the evolving nature of the various wireless links while avoiding rapid depletion of the UAV's battery. In this work, we assume that the UAV-MBS is capable of predetermining its own location within the flying zone[1]. We also assume that the channel-state information (CSI) information is available at the UAV-MBS side. It is worth mentioning that the operation of the proposed agents does not require the UAV mobile BS to have explicit knowledge of the ground users' locations. CSI availability is sufficient for the proposed RL agents to operate. Alternatively, if the reward feedback information is readily available, the MAB agent can estimate the proper action and navigate the zone accordingly. Since downlink NOMA protocol [99] is utilized, the information offloaded to the activated receiving devices is sent over unified RBs where the corresponding total rate attained by the selected devices may be expressed as

$$R = \sum_{d=1}^{D} W \log_2 \left(1 + \frac{p_d g_d}{\sum_{i=1}^{d-1} p_i g_d + \sigma^2} \right), \tag{5.1}$$

[1]This is typically accomplished through a global navigation satellite system (GNSS). However, if the GNSS is unavailable or more precise location information is needed, then other techniques may be used (e.g., integrating ultra-wideband (UWB) technology with LiDAR-based range finders) [148].

where, without loss of generality, we assume a descending-order CSI gain information between the UAV-MBS and the receiving devices: $g_1 > g_2 > \cdots > g_D$. W represents the available transmission bandwidth, and σ^2 represents the variance of the zero-mean additive white Gaussian noise. p_d is the portion of the transmission power that the UAV-MBS allocates to the signal of the d-th selected receiving device. Under the assumption of a total number of K candidate receiving devices, we have $d \in \{1, 2, \ldots, D\}$, where $D < K$ represents the number of information streams offloaded via the downlink NOMA multiplexing protocol.

We assume that the UAV's battery has a finite capacity of χ energy units (EUs) and that the battery energy level decreases linearly as a function of the traveled distance according to

$$L(i) = \chi - \sum_{j=1}^{i} \eta \, Z(j) \,, \tag{5.2}$$

where $L(i)$ denotes the UAV's battery level at time step i and $Z(j)$ is the distance traveled at step j, $j = 1, 2, \ldots, i$. η is the energy expense per unit distance. Here, we mainly focus on tracking the energy consumed in moving the UAV around whereas hovering energy consumption is not included.

We assume that the communication channel's wireless link between the UAV-MBS transmitter and the d-th candidate receiving device is given by a Rician channel representation to model the presence of the LOS signal. Therefore, the channel link connecting receiving device d to the UAV-MBS may be written as

$$g_d = \sqrt{\frac{F_r}{F_r + 1}} \bar{g}_d + \sqrt{\frac{1}{F_r + 1}} \tilde{g}_d \,, \tag{5.3}$$

where $\bar{g}_d$ denotes the LOS deterministic component which is set to a typical value of 1 [149]. $\tilde{g}_d$ represents an NLOS random component that follows the Rayleigh distribution. F_r denotes the Rician channel parameter.

Dynamic spot selection for UAV-MBS adaptive hovering throughout the flying zone is of paramount importance to properly tune the effective wireless links connecting the receiving devices to the UAV-MBS in a way that maximizes the collective acquired rate without expending the UAV's battery inefficiently. By changing the UAV-MBS location, the wireless channel can then be controlled to combat the interference and thus

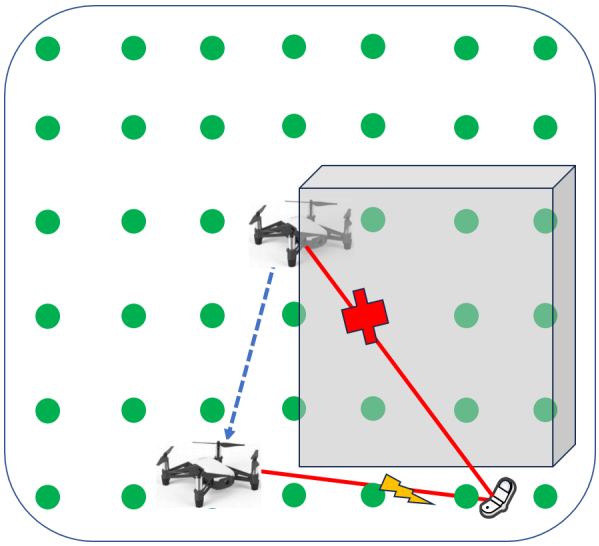


Figure 5.3: UAV Location update within environments with blockers.

boost the effective SINR level at the receiving devices, and consequently improve the sum-rate [140, 150, 151]. Adjusting the UAV hovering location within the service area is therefore important for the proposed scheme to mitigate the effects of interference among multiplexed NOMA users to maximize the achievable sum-rate level. Consider for example a two-user NOMA downlink transmission with SIC detection: the normalized sum-rate is $log_2(1 + \frac{p_1 g_1}{\sigma^2}) + log_2(1 + \frac{p_2 g_2}{p_1 g_2 + \sigma^2})$. Here, the channel gains, $g_d, d = 1, 2$, are time-varying and depend on the Tx-Rx separation distance. Therefore, by changing the UAV Tx location, the channel gains can be adjusted to maximize the sum-rate level. Moreover, changing the UAV's location can be useful for situations where blockers are present in the environment around stationary users. In this case, the UAV can change its location to achieve an LOS link to the ground user and provide better connectivity as shown in Fig. 5.3. In addition, careful activation of receiving devices as well as proper splitting of the UAV-MBS transmission power among the selected devices must be adjusted dynamically to provide and sustain high operational performance.

5.4 Proposed Algorithms

This section presents RL solution methods based on CS-MAB and DDQN to address the joint problem outlined earlier in section 5.3. Although both methods share the same underlying agent-environment interaction principle, they have distinctive features and operate on different basic concepts: DDQN agents rely on DNNs to represent their decision-making policy and can provide near-optimal performance if environment-related

information is collected adequately during offline training sessions. MAB agents, on the other hand, do not employ DNNs, and make on-the-fly decisions through direct online deployment where they adjust their decisions dynamically according to the rewards received during a series of successive interactions with the environment. Since no DNNs are present, MAB-based operation is generally less complex and simpler to implement than its DDQN-based counterpart. It may, however, provide less optimized performance than DDQN.

5.4.1 Deep RL-based Operation: Q network Method

Algorithm I outlines the proposed DDQN solution for the joint UAV-MBS path design and NOMA device activation and power allocation scheme. The fundamental operating premise is to develop a successful strategy to transfer sequential input environmental variables to highly rewarding decisions over the long run, which are reflected in the cumulative rates acquired by active receiving devices. The algorithm leverages the main Q-learning concept that maps a given action $\mathcal{A}$ to a fitness Q-value when taken in a state $\mathcal{L}$ by following some policy π according to

$$V_q^{\pi}(\mathcal{L},\mathcal{A}) = E[r_1 + \beta r_2 + \beta^2 r_3 + \ldots \,|\, \mathcal{A}_0 = \mathcal{A},\, \mathcal{L}_0 = \mathcal{L}]\,, \tag{5.4}$$

where $V_q^{\pi}(\mathcal{L},\mathcal{A})$ denotes the mean discounted summation of rewards acquired in a long-term sense. These rewards are earned when, starting at some arbitrary state $\mathcal{L}_0 = \mathcal{L}$, action $\mathcal{A}_0 = \mathcal{A}$ is applied then the subsequent state-action path is dictated in accordance with the policy π. r_1 denotes the immediately-acquired reward and $r_l \,\forall\, l > 1$ represents the rewards acquired subsequently during future states. $\beta \in [0,1]$ is a parameter to adjust the amount of discounting to apply to balance r_1 with future rewards.

A DDQN is used to provide $V_q(\mathcal{L},\mathcal{A};\Psi_i)$, a configurable parameteric implementation of the value function which can be tuned to generate a good approximation of the optimal function $V_q^*(\mathcal{L},\mathcal{A})$. A DNN is employed to store Ψ_i, the parameter set used to generate the optimal strategy. DDQN RL agents employ a separate set of parameters $\bar{\Psi}_i$ to generate Y_i which provides a target for the training of Ψ_i, the main strategy set for the i-th agent-environment interactive exchange. The learnable target is estimated as

$$Y_i = r_i + \beta \hat{V}_q(\mathcal{L}_{i+1}, \arg\max_{\mathcal{A}} V_q(\mathcal{L}_{i+1},\mathcal{A};\Psi_i);\bar{\Psi}_i)\,, \tag{5.5}$$

where the main DNN parameter set Ψ_i is reserved for decision making whereas the assessment of the corresponding fitness value is estimated through the target DNN, $\bar{\Psi}_i$. $\hat{V}_q(.)$ represents the response generated by the target DNN.

NOMA-UAV-MBS: State & Action Spaces, and Rewards

The state defining the environment of the NOMA-UAV-MBS system is formed for the i-th interactive step as

$$\mathcal{L}_i = \{g_1(i), g_2(i), \cdots, g_K(i), L(i)\}\ , \tag{5.6}$$

where $g_k(i)$ and $L(i)$ represent the channel gain of the k-th candidate and the UAV's battery level at time step i, respectively.

The action $\mathcal{A}_i$ applied by the proposed DDQN agent to state $\mathcal{L}_i$ is formed as

$$\begin{aligned}\mathcal{A}_i = \{\, & X(i), Y(i), s_1(i), s_2(i), \cdots, s_K(i), \\ & p_1(i), p_2(i), \cdots, p_K(i)\ \}\ ,\end{aligned} \tag{5.7}$$

where $X(i)$ and $Y(i)$ are the coordinates of the chosen UAV hovering spot at the i-th step whereas $s_k(i) = 1$ is a binary indicator denoting the selection of the k-th device and $p_k(i)$ represents its corresponding transmission power portion allocated by the UAV-MBS.

The immediate reward resulting from applying $\mathcal{A}_i$ in the i-th interaction step to the environment can be expressed as

$$r_i = \sum_{k=1}^{K} r_k(i)\,, \tag{5.8}$$

where the reward contribution of the k-th activated device is evaluated as

$$r_k(i) = \log_2\left(1 + \Gamma_k(i)\right), \tag{5.9}$$

with the ratio of desired signal power to the collective power of interference and noise (SINR) evaluated as

$$\Gamma_k(i) = \frac{s_k(i)p_k(i)g_k(i)}{\sum_{u=1}^{k-1} s_u(i)p_u(i)g_k(i) + \sigma^2}. \tag{5.10}$$

The β-discounted total reward accumulated throughout an interaction horizon I may

therefore be expressed as

$$R(I) = E[r_1 + \beta r_2 + \beta^2 r_3 + \cdots + \beta^{I-1} r_I]. \tag{5.11}$$

The proposed RL operation in Algorithm I is designed so that the DDQN agent interacts continuously for E episodes each comprising up to I interactions with the NOMA-UAV-MBS environment. The proposed agent aims to learn a successful strategy that sequentially projects the environment's states to a series of actions that maximize the total long-run reward in (5.11). At the start, two identical random instantiations of the parameter sets Ψ and $\bar{\Psi}$ are generated for the main and target DNN-based policies. The NOMA-UAV-MBS environment is then subjected to a uniformly distributed random action with probability α_s. The decision is made with probability $1 - \alpha_s$ depending on the existing DNN strategy, Ψ. Subsequently, the SIC-related QoS requirement for proper detection,

$$\left(p_d - \sum_{u=1}^{d-1} p_u \right) \frac{g_{d-1}}{\sigma^2} \geq \mu \,, \tag{5.12}$$

can be checked and, if necessary, enforced whenever the action is not conforming to the condition. p_d and μ respectively denote the power portion allocated for active receiving device d, $d = 2, 3, \ldots, D$, and a reliable detection threshold for the SIC operation. The sum-rate immediate reward as well as the next resultant NOMA-UAV-MBS system state can then be buffered into an experience-gathering memory unit $\mathcal{M}$ which collects important information relevant to the agent's ongoing interaction with the environment. The collected information helps the agent to form a concrete set of experiences which can be progressively fused and harnessed to update the agent's acting policy Ψ during each round of interaction. The buffering memory unit, $\mathcal{M}$, is propagated with training data objects taking the form of 4-tuple items each consisting of a possible state of the environment, an associated action performed during that state, a resultant subsequent state, and the reward collected from that interaction. A mini-batch of random data items is fetched from the experience memory unit for tuning and adjustment of the main policy DNN toward the agent's learnable target, Y_j. In *double* DQN operation, each training data item within the mini-batch is used to compute a corresponding learnable value over a two-step process: firstly the training item's next state information is passed through the main DNN, Ψ, to find the action with the associated highest Q-value. Secondly,

Algorithm I: Proposed DDQN Agent for Joint UAV-MBS Path Design and NOMA Device Activation and Power Allocation.

- **Set** P_t, χ, g, μ, E, η, I, δ, α_s, α_f, γ, α_d
- **Initialize** Ψ, $\bar{\Psi} = \Psi$, $d = 1$, $\alpha_1 = \alpha_s$, $\mathcal{M} = \text{NULL}$
- **While** $e \leq E$ run the following episode:
  - Set $i = 1$, $L(i) = \chi$
  - Initialize NOMA-UAV-MBS state $\mathcal{L}_i$
  - **While** $i \leq I$
    1. Draw a random sample s from a uniform distribution $U(0,1)$: **If** $s \leq \alpha_i \Rightarrow$ Pick decision $\mathcal{A}_i$ randomly. **Else**, $\mathcal{A}_i = \arg\max_{\mathcal{A}} V_q(\mathcal{L}_i, \mathcal{A}; \Psi)$.
    2. **If** $\mathcal{A}_i$ does not conform to the QoS requirement (5.12):
       * For $k \in \mathcal{V}_u$, $\mathcal{V}_u \equiv$ set of violating devices, enforce (5.12) on $\mathcal{A}_i$ by rectifying the allocated power portion for each active device in $\mathcal{V}_u$:
       $$p_k = \frac{\mu\sigma^2}{g_{k-1}} + \sum_{u=1}^{k-1} p_u$$
       * Normalize Tx power level of all active devices:
       $$p_d \leftarrow \frac{p_d}{\sum_{u=1}^{D} p_u} P_t$$
    3. Execute $\mathcal{A}_i$ in the environment, then monitor its resulting state $\mathcal{L}_{i+1}$ and the UAV's battery level $L(i)$, and acquire the generated reward r_i
    4. Append $(\mathcal{L}_i, \mathcal{A}_i, r_i, \mathcal{L}_{i+1})$, the experience gathered through interaction, to the memory unit $\mathcal{M}$
    5. Randomly pick an experience mini-batch $(\mathcal{L}_j, \mathcal{A}_j, r_j, \mathcal{L}_{j+1})$ from the memory $\mathcal{M}$:
       * Set $A^* = \arg\max_{\mathcal{A}} V_q(\mathcal{L}_{j+1}, \mathcal{A}; \Psi)$
       * Form the agent's training target as
       $$Y_j = r_j + \mathbf{1}[\mathcal{L}_j \neq \textit{Terminal}]\beta \hat{V}_q(\mathcal{L}_{j+1}, A^*; \bar{\Psi})$$
       * Take a single gradient descent step on
       $$\textstyle\sum_j (Y_j - V_q(\mathcal{L}_j, \mathcal{A}_j; \Psi))^2 \text{ w.r.t } \Psi$$
    6. **If** i *Mod* $\delta = 0$:Adjust $\bar{\Psi}$ softly using a smoothing parameter γ:
    $$\bar{\Psi} \leftarrow \gamma\Psi + (1-\gamma)\bar{\Psi},\ 0 < \gamma < 1$$
    7. **If** $\alpha_f < \alpha_i$:Reduce α_i further to approach the final level α_f through the annealing factor α_d:
    $$\alpha_{i+1} = \alpha_i(1-\alpha_d), 0 < \alpha_d < 1$$
    8. **If** $L(i) \leq (1-g)\chi$ (i.e., UAV's battery is depleted):
       * Penalize the reward for battery draining:
       $$r_i = r_i - \rho$$
       * Mark state $\mathcal{L}_i$ as *Terminal*, and set $\alpha_1 = \alpha_i$.
       * ***End Episode***
    9. Increment iteration index: $i \leftarrow i + 1$
  - **end While**
  - Configure probability of randomized decisioning for the upcoming episode:
  $$\alpha_1 = \alpha_I$$
  - Advance episode: $e \leftarrow e + 1$
- **end While**

where

- $\mathbf{1}[.]$ is the indicator function. E is the number of episodes and I is the maximum number of iterations within an episode.
- α_s is the initial probability of randomized decisioning whereas α_f represents the final probability of picking an action in a random fashion during advanced interactions.
- ρ is the penalty for battery draining and δ is a configurable controller to adjust the target network's update interval.

the the training item's immediate reward component is combined with the output of the secondary DNN, $\bar{\Psi}$, which corresponds to the action selected in the first step. Although the main DNN parameter set is updated at each interaction, the critic's network, $\bar{\Psi}$, is smoothly updated in a periodic fashion every δ iterations. To accomplish this soft update, a fractional smoothing factor, $\gamma \in (0, 1)$, is used to fuse the updated parameter set of the main DNN with the current parameter set of the target DNN.

To optimize the operation for battery-aware decisioning, at each interactive iteration, the agent's inspect the status of the UAV's battery to determine whether to prematurely halt the ongoing episode. If the UAV's battery is drained (i.e., $L(i) \leq (1-g)\chi$, g is the battery drain percentage), the agent's reward is discounted by a configurable penalty parameter, ρ. The agent then saves the current value of the probability of making its decision on a random basis, α_i, to be used as the starting value, α_1, for the next episode, then the current episode is abruptly terminated, thus giving the agent an incentive towards deciding in favor of actions that, in the long run, do not drain the UAV's battery rapidly. Otherwise, towards the end of the interaction iteration, if the current probability of executing a random decision is greater than a preconfigured minimum end value, α_f, then the probability is reduced through a controllable decay parameter α_d. This probability reduction mechanism is gradual and lasts for an amount of iterations controlled through α_d to allow the DDQN agent to build more confidence to follow its developing internal strategy, Ψ, more frequently while it absorbs more knowledge of the underlying characteristics of the environment it is interacting with. Once the probability reduction process halts, the agent subsequently maintains a fixed level of α_f for sampling decisions randomly throughout all remaining interactions.

5.4.2 Multiarmed bandit-based Solution

Algorithm II outlines the proposed CS-MAB solution for the joint UAV-MBS path design (deciding which coordinates to choose for subsequent hovering positions within the defined grid for UAV-MBS operation) and NOMA device activation and power allocation scheme. The algorithm's deployment can be implemented via either CS-MOSS or CS-UCB options. The UAV-MBS player starts initially at some arbitrary state of the environment and pulls a decision arm then loops successively over all the allowable decision arms in a number of interactive steps in order to form initial crude estimates for the

Algorithm II: Proposed Cost-Subsidized MAB Operation for Joint UAV-MBS Path Design and NOMA Device Activation and Power Allocation.

- **Set** P_t, λ, μ, χ, η, $\mathcal{H}$
- **Initialize** $i = 1$, NOMA-UAV-MBS state $\mathcal{L}_i$
- **While** $i \leq \mathcal{H}$
  - **If** $i \leq \mathcal{A}_{max}$
    1. Pull decision arm $\mathcal{A}_i$
    2. **If** $\mathcal{A}_i$ does not conform to the QoS requirement (5.12):
       * For $k \in \mathcal{V}_u$, $\mathcal{V}_u \equiv$ set of violating devices, enforce (5.12) on $\mathcal{A}_i$ by rectifying the allocated power portion for each active device in $\mathcal{V}_u$:
       $$p_k = \frac{\mu\sigma^2}{g_{k-1}} + \sum_{u=1}^{k-1} p_u$$
       * Normalize power level of all active devices:
       $$p_d \leftarrow \frac{p_d}{\sum_{u=1}^{D} p_u} P_t$$
    3. Set up a reward jar for current arm: $m_i = r_i$
    4. Initialize a pull counter for current arm: $n_i = 1$
    5. Evaluate the utility of pulling current arm:
    $$f_i = \begin{cases} m_i + \sqrt{2\log(i)}, & \textbf{CS-UCB} \\ m_i + \sqrt{\max(\log(i), 0)}, & \textbf{CS-MOSS} \end{cases} \tag{5.13}$$
  - **Otherwise**
    1. Set $k = \arg\max_j f_j$
    2. Form a candidate subset of decision arms:
    $$\Omega(i) = \{j : f_j \geq (1-\lambda) f_k\} \tag{5.14}$$
    3. Pull battery-aware arm $\mathcal{A}_{i^*}$: $$i^* = \underset{j \in \Omega(i)}{\arg\max}\, L(j) \tag{5.15}$$
    4. **If** $\mathcal{A}_{i^*}$ does not conform to (5.12):
       * For $k \in \mathcal{V}_u$, enforce (5.12) on $\mathcal{A}_{i^*}$ by rectifying the allocated power portion for each active device in $\mathcal{V}_u$:
       $$p_k = \frac{\mu\sigma^2}{g_{k-1}} + \sum_{u=1}^{k-1} p_u$$
       * Normalize power level of all active devices:
       $$p_d \leftarrow \frac{p_d}{\sum_{u=1}^{D} p_u} P_t$$
    5. Update reward jar of $\mathcal{A}_{i^*}$: $m_{i^*} \leftarrow m_{i^*} + r_i$
    6. Increment corresponding counter: $n_{i^*} \leftarrow n_{i^*} + 1$
    7. Update the utility corresponding to playing $\mathcal{A}_{i^*}$:
    $$f_{i^*} = \begin{cases} \frac{m_{i^*}}{n_{i^*}} + \sqrt{\frac{2\log(i)}{n_{i^*}}}, & \textbf{CS-UCB} \\ \frac{m_{i^*}}{n_{i^*}} + \sqrt{\frac{\max\left(\log\left(\frac{i}{n_{i^*}}\right), 0\right)}{n_{i^*}}}, & \textbf{CS-MOSS} \end{cases} \tag{5.16}$$
  - Move to new NOMA-UAV-MBS state, $\mathcal{L}_{i+1}$.
  - Increment iteration index: $i \leftarrow i + 1$
- **end While**

where

- The horizon, $\mathcal{H}$, is the total number of played arms.
- $\mathcal{A}_{max}$ denotes the number of arms available to the agent.
- λ is the cost-subsidizing control parameter.

various utilities of playing different decision arms in accordance with Eq. (5.13). During this initialization phase, a separate reward jar is dedicated to collect the achievements of each decision arm. Each jar's initial value is set to the immediate reward acquired by executing the corresponding decision arm in the environment. As for the DDQN case, Eq. (5.8) reflects how much reward is generated for a given arm play. Each decision arm $\mathcal{A}_i$ is a three-tuple action object containing the hovering position coordinates (X,Y) of the UAV-MBS, along with the UAV-MBS transmit power portions ($p_1, p_2, ...p_K$), and a set of active receiving devices as defined in Eq. (5.7). Moreover, a set of counters is initialized to record the frequencies of pulling various decision arms in later stages. During each interaction, and before executing the chosen decision arm on the NOMA-UAV-MBS system, if the chosen decision arm is not conforming to the data detection fidelity requirement in (5.12), the agent enforces the condition by boosting the power portions associated with the violating receiving devices and applying power normalization to maintain the feasibility of transmission power budget.

As soon as the initialization phase terminates, dynamic selection of decision arms is accomplished by harnessing the available information of various arm-pulling utilities as well as the UAV-MBS battery level, which are updated dynamically in every interactive iteration in accordance with Eqs. (5.16), and (5.15) and (5.2), respectively. In particular, a feasible subset of candidate receiving devices exceeding a configurable QoS threshold on utility is formed according to Eq. (5.14) where a recommended cost-subsidizing factor of $\lambda = 0.1$ is used [152]. Next, to control the battery energy consumption and facilitate for a battery-aware operation, the feasible decision arm resulting in the current highest battery level is played in the environment as dictated by Eq. (5.15). The corresponding generated reward is then added to the associated reward jar of the played arm. In addition, the corresponding counter of the played arm is incremented by one, and the associated utility value is updated. The state of the environment advances subsequently to a new state and another iteration of interaction begins. The CS-MAB algorithm keeps on interacting with the environment for a predefined horizon, $\mathcal{H}$.

Numerous hyper-parameters govern the process of training a DDQN RL agent and, generally speaking, high complexity is associated with the required computations. This concern may nonetheless be bypassed if computations are offloaded to a prior stage of offline training wherein the DDQN RL agent's skills are honed through interactive training

on the possibly-simulated environment. Afterwards, online operation is commenced and the experienced agent is deployed. The proposed DDQN agent (with A layers) has a deployment complexity of $\mathcal{O}(vAK)$ for K candidate devices within an area of v hovering positions. On the other hand, the deployment complexity of the CS-MAB agent is $\mathcal{O}(vK)$. By contrast, in optimal operation, UAV-MBS placement and the grouping of D devices for active operation over an allocated RB would incur $\mathcal{O}\left(v\binom{K}{D}\right)$, thereby demanding a far greater implementation cost.

5.5 Numerical Analysis

5.5.1 Simulation Environment

Table 5.1: Simulation settings: UAV-assisted mission-critical *mmWave* transmissions.

Scheme	Setting	Value
Common settings	*mmWave* Carrier frequency	60 GHz
	System bandwidth, W	100 MHz
	Channel exponent (pathloss), υ	2.1
	Standard deviation for Shadowing, ζ	4.4 dB
	Type of wireless channel	Rician
	Rician channel parameter, F_r	10 dB
	Noise spectral density	-174 dBm/Hz
	Grid zone size	100×100 m^2
	Grid zone spacing	10 m
	UAV-MBS battery capacity, χ	1 EU
	UAV-MBS energy parameter, η	10^{-4} EU
	UAV-MBS transmit power, P_t	20 dBm
Proposed DDQN solution	Episodes	100
	Episode interactive iterations	100
	DNN optimizer method	SGDM
	Learning rate	0.001
	Secondary DNN update period, δ	4
	Gradual-update softening parameter, γ	0.001
	Capacity of experience memory unit	5000
	Mini-batch training items	8
	Randomized initial action probability, α_s	1
	Randomized action probability end-level, α_f	0.01
	Probability decaying parameter, α_d	0.005
	Battery draining penalty, ρ	100

The settings of the simulated environment are given in Table 5.1. The table presents the default simulation values of the used settings for the evaluated scenarios. We assume

a data-offloading NOMA-UAV-MBS downlink system where 5 candidate receiving devices are scattered arbitrarily within a 100×100 m^2 2-D zone. Various grid sizes ranging from less than 100-by-100 m^2 and going well beyond 100-by-100 m^2 have been considered in the literature. For instance, the authors in [153] considered a grid size of 30-by-30 m^2 whereas the authors in [140] considered a 200-by-200 grid. Similarly, the authors in [81] considered a 200-by-200 emergency communications area with four deployed UAVs. In our work, we consider an in between 100-by-100 area which can be suitable for spots not covered yet by the main network's infrastructure (e.g., in remote areas) or for zones suffering network instability such as in areas hit by disasters thus temporarily rendering the main network out-of-service. It is worth mentioning that other grid sizes can be used depending on the application. All receiving devices employ antenna elements at a fixed 1 m height. Initially, and at a fixed 10 m height, the UAV-MBS begins at the (0,0) hovering point within the flying region. The carrier setting for the *mmWave* is configured to 60 GHz. The system bandwidth setting is 100 MHz. We also consider the close-in model (CI),

$$PL_{CI}^{dB}(f, Z) = PL_{FS}^{dB}(f, Z_0) + 10\upsilon \log_{10}\left(\frac{Z}{Z_0}\right) + \Upsilon_{CI}^{\zeta}, \tag{5.17}$$

for large-scale channel disturbances. The used channel exponent setting for pathloss is $\upsilon = 2.1$ (typical for line-of-sight propagation (LOS) of *mmWaves* in urban environments) [118]. The model simulates large-scale fluctuations which is combined with small-scale fading of Eq. (5.3) to account for overall variation of the wireless channel. The channel gain of the d-th candidate receiving device is thus $PL_{CI}^{dB} + 20\log_{10}|g_d|$ dB. The standard deviation, ζ, for CI model emulated shadowing, Υ_{CI}^{ζ}, is configured as 4.4 dB. PL_{FS}^{dB} is the nominal pathloss for free-space in dB. Since it delivers appropriate model accuracy and maintains parameter stability both for outdoor as well as indoor urban environments (including micro and macro variants) spanning a broad frequency spectrum within microwave and *mmWave* bands, a $Z_0 = 1$ m reference distance is utilized for typical CI models. A 1 m reference point might be crossing the boundaries of the near-field emitted by massive antennae arrays. However, the inaccuracy introduced by such small distance is mostly trivial from the perspective of practical wireless communication systems. [118, 119].

The UAV-MBS operates at a 20-dBm default transmission power level to offload the NOMA message relaying the data of active receiving devices. The simulated AWGN noise

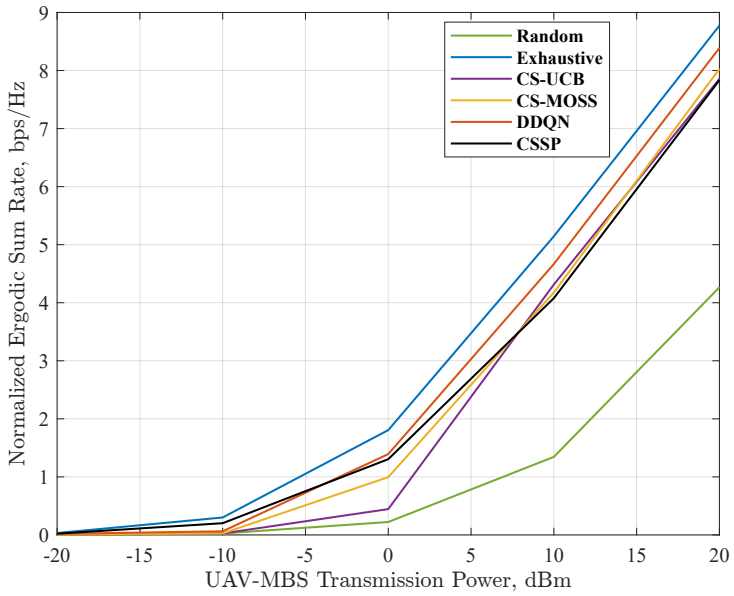


Figure 5.4: Evaluation of the performance in terms of the acquired sum-rate.

is generated using the typical -174 dBm/Hz for its PSD level.

The primary and secondary DNNs employed in the RL-based DDQN operation are built with a total of 5 convolutional layers (fully-connected type). Layers 1 through 4 have rectifying linear units (ReLUs) as non-linear activation functions: $A(x) = \max(0, x)$. The fifth layer (output layer) has the linear characteristic $A(x) = x$. The structure of the two sets of neurons configured for both DNNs is identical: 100 neurons within each of layer 1 and 5, whereas the remaining three layers in the middle consist of 95,90, and 85 neurons respectively. Moreover, each neuron in both DDNs is configured with an adjustable bias term.

DNN training for updating the main parameter set Ψ at each interactive iteration is accomplished by performing a single optimization step using stochastic gradient descent with momentum operation (SGDM optimizer) [154] towards an energy-efficient and highly rewarding mapping strategy from input environmental states to appropriate decisions that the agent can apply. The associated learning step size is configured as 0.001. The critic's secondary DNN set, $\bar{\Psi}$, is only periodically updated every fourth interactive step. The gradual update of $\bar{\Psi}$ is performed in a smooth fashion using a soft mixture of $\bar{\Psi}$ and Ψ with a 10^{-3} softening parameter. A limit of five thousand items is set on the size of the memory tank buffering experience data items collected by the DDQN agent over successive interactive steps.

The end policy for the trained DDQN agent is acquired by progressively updating the main DNN set Ψ, using an 8-item mini-batch (which is composed through random

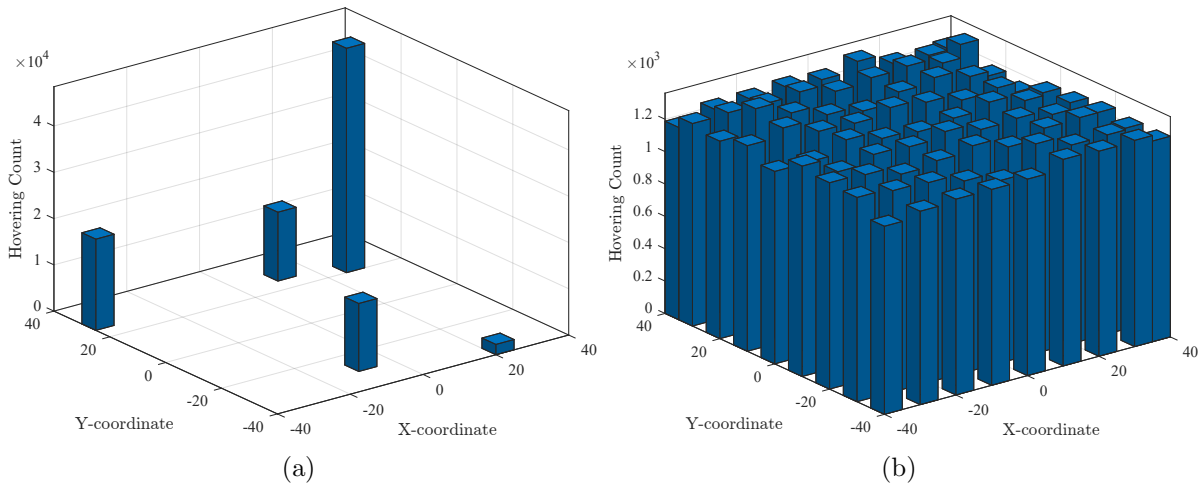


Figure 5.5: UAV hovering distribution for (a) Exhaustive search, and (b) Random selection strategy.

sampling from the previously stored experience items) during each update step. The DDQN agent's training starts with a 100% chance of choosing a random decision (with uniform selection distribution among the decision set). The agent gradually shifts away from this randomized decisioning by annealing its probability over successive interactive iterations until the chance of operating in a random fashion reaches a preset terminal level of 0.01. By setting the decaying parameter $\alpha_d = 5 \times 10^{-3}$, this probability decay process runs for slightly over 900 interactive iterations before it halts when the terminal level α_f is attained. The training phase comprises a 100 episodes in total where each one runs for up to a 100 interactive iterations.

5.5.2 Results and Discussion

In Fig. 5.4 we evaluate the performance of the proposed algorithms in terms of the normalized ergodic total rate with the total transmission power of the UAV-MBS ranging from -20 dBm to 20 dBm. Upper and lower performance benchmark references are respectively represented by the exhaustive and random solution strategies. In addition, conventional CSSP resource allocation reference baseline [102] is included with optimal UAV-MBS positioning. In CSSP, candidate receiving devices are arranged based on their channel state quality conditions wherein the devices experiencing larger gaps in their channel gains are paired. Clearly, the optimal exhaustive approach can support the highest total rate level in a consistent manner by virtue of exploring every available point within the action space before applying the most-rewarding alternative. On the other hand, random-based operation runs by choosing to execute some randomly sampled al-

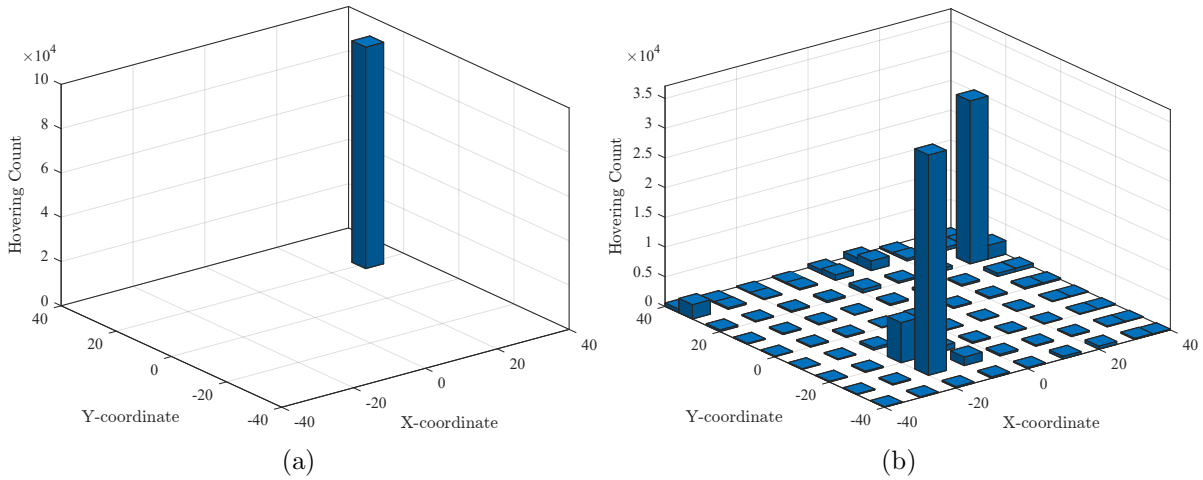
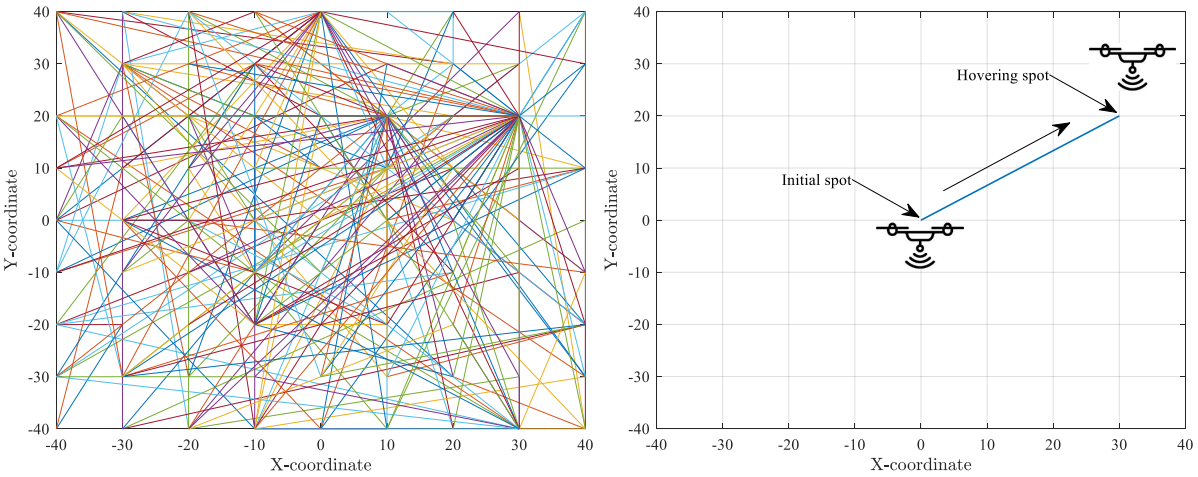


Figure 5.6: UAV hovering distribution for the proposed (a) DDQN RL agent, and (b) CS-MAB-UCB RL agent.

ternative from the available decision set, thereby resulting in a heavily non-optimized and weak performance which mainly functions as an indicator of achievable levels for the total rate. Coming on top of the proposed algorithms is the DDQN solution where it manages to remain within a close performance gap from the upper level provided by the exhaustive scan. Beginning with extremely low levels of UAV-MBS transmission power, the proposed algorithms produce performance results close to those achieved by a random approach, with CSSP taking the lead until around -4 dBm where the proposed DDQN first overtakes CSSP. As the UAV-MBS transmission power is increased, noticeable distinct performance gaps of the achievable sum-rate then begins to emerge until the DDQN agent rises to slightly over 77% of the exhaustive at 0 dBm as opposed to 72% for CSSP, with the CS-MOSS agent following next at about 55% whereas the CS-UCB agent falls behind around the 25% mark. The proposed CS-MAB agents then quickly rise to surpass CSSP starting around 8 dBm. All schemes then keep on rising until the DDQN agent manages to attain almost 96% of the sum-rate level achievable by an exhaustive scan at 20 dBm, whereas CS-MOSS rises to 91.5% and CS-UCB follows closely after at 89.5% which is immediately followed by CSSP at 89.3%.

Figure 5.5 illustrates the distribution of UAV hovering spot selection based on an exhaustive search for sum-rate maximization in part (a) and a random selection strategy in part (b). Figure 5.5 (a) shows the critical positions within the flying zone that the UAV-MBS visits for hovering during a series of 10^5 interactive steps while following an optimal solution policy as given by an exhaustive search covering the entirety of the action space. This intensive scanning of the grid filters the ineffective hovering spots while leaving intact



(a) (b)

Figure 5.7: Path traced by the UAV for DDQN-based operation, for (a) Training, and (b) Deployment.

the five positions represented by the blue bars due to their optimized positioning relative to device scatter pattern. These five critical hovering locations are chosen exclusively by the UAV-MBS to attain the transmission rates of Figs. 5.4 and 5.10. Despite the fact that all of the 5 locations are preferred and chosen by the optimal strategy, their associated visiting frequencies are not the same as indicated by the difference in their corresponding bar heights, and the UAV-MBS predominantly converges to the position (30,20) for hovering. Furthermore, in part (b) we can observe the non-optimized behavior of the random strategy where, as expected, all spots are selected with almost equal visiting frequencies.

Figure 5.6 illustrates in part (a) the distribution of UAV hovering spot selection based on a trained DDQN RL agent, whereas in part (b) the CS-UCB agent's selection strategy is presented. For the entire deployment stage, the trained DDQN agent decides to maintain a fixed position at the spot most frequently visited by the optimal scanning as indicated by the presence of a singular blue prism at the (30,20) grid point. On the other hand, the CS-UCB agent switches back and forth between the same spot and the spot at (-20,-30) while occasionally visiting other positions with less relative frequency as indicated by the blue bars of varying heights in part (b).

To get a complementary view of the DDQN operation, the trajectories traced by the UAV during both training and deployment are shown in Fig. 5.7. In part (a), each line in this figure represents a traced path segment connecting a departure point to a destination point where the UAV traverses the path connecting the two spots as it makes successive

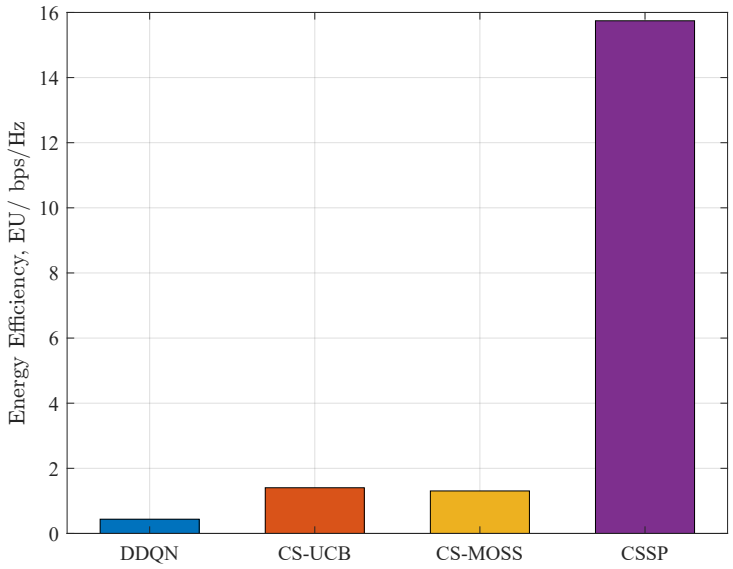
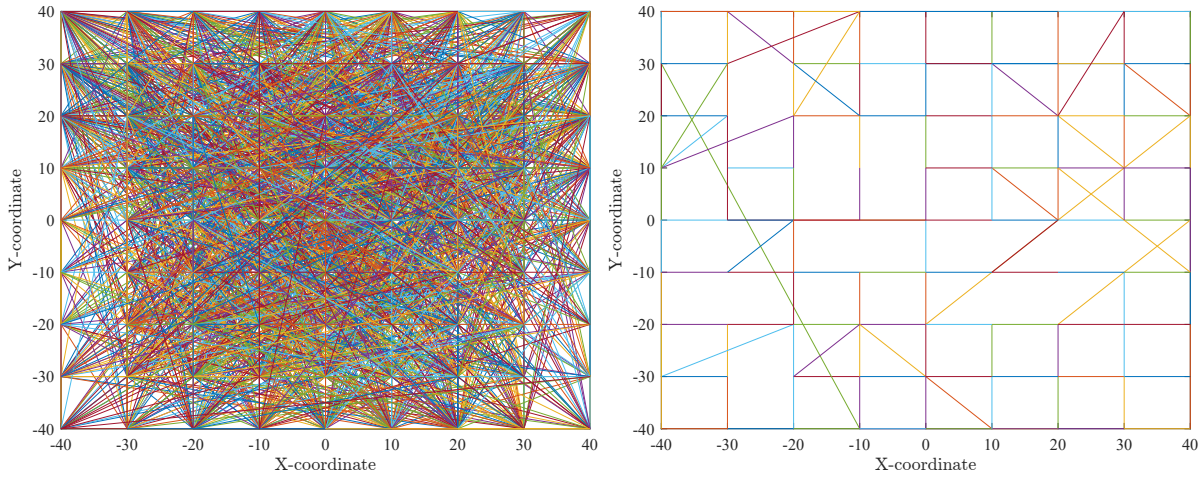


Figure 5.8: UAV movement energy efficiency of the proposed RL-based algorithms.

decisions while interacting with the environment. the agent learns by exploring various hovering positions to absorb the essential features of the environment prior to deployment. Noticeably, the agent goes through the three spots at (30,20), (10,20), and (0,40) more often than it routes through other spots on the grid. Upon deployment in part (b), the trained DDQN agent opts to route directly from the (0,0) starting point to the critical position found at (30,20) where it maintains position as discussed earlier in part (a) of Fig. 5.6.

The energy efficiency of the UAV movement is presented in Fig. 5.8 for CSSP and the proposed RL-based schemes where the Y-axis represents the total energy consumption level per normalized ergodic sum-rate. As shown, an energy efficiency level of about 0.44 EU per bps/Hz is achieved by the DDQN solution and is lower than those achieved by both CS-UCB and CS-MOSS solutions. The CS-MOSS solution comes second in line at around 1.31 EU per bps/Hz, thus requiring just a little below 200% higher energy to support the same total transfer speed provided when operating using the DDQN solution. The CS-UCB solution consumes about 1.4 EU per bps/Hz which sets it at around 220% and 7.5% behind the DDQN and CS-MOSS solutions, respectively. Lastly, to maintain adequate total rate levels, CSSP incurs a significant energy loss in comparison where it consumes over 11 times higher energy than CS-UCB. This demonstrates the effectiveness of the proposed RL-based approaches.

The trajectory traced by the UAV is shown in Fig. 5.9 for CS-MAB-based operation with and without battery optimization. In part (a), the subsidizing factor of the deployed



(a) (b)

Figure 5.9: Path traced by the UAV for MAB-based operation, for (a) No battery optimization: $\lambda = 0$ CS-UCB, and (b) With battery optimization: $\lambda = 0.1$ CS-UCB.

CS-UCB agent is configured to $\lambda = 0$ whereas in part (b) a recommended typical value of $\lambda = 0.1$ is used [152]. When the operation is not optimized for energy efficiency, the CS-UCB agent keeps rerouting through all the spots on the grid extremely intensively without any signs of filtering down the traced path. On the other hand, when battery optimization is turned on in part (b), the CS-UCB agent prunes the traversed path significantly and flies along a much smaller subset of path segments compared to the operation illustrated in part (a) when battery usage optimization is turned off. This clearly shows the effectiveness of cost-subsidizing operation when a MAB solution is deployed.

To verify the ability of our proposed solutions to support proportionally increasing total rate versus a variable transmit bandwidth range, Fig. 5.10 demonstrates the total achievable rate corresponding to $100 \sim 500$ MHz Tx bandwidths at 60-GHz NOMA *mmWave* carrier. When operating at a hundred MHz, all solutions result in sub-Giga data transfer speeds. All approaches then rapidly exceed 1 Gbps except for the random strategy which grows very slowly towards the 1-Gpbs level where it breaks it around the 500 MHz mark at which point the proposed methods manage to support rates beyond 2.5 Gbps. The DDQN agent comes on top of the proposed RL solutions where it achieves speeds as high as 2.92 Gbps while enjoying the full 500 MHz of system bandwidth. By contrast, CS-MOSS attains 2.81 Gbps whereas CS-UCB reaches 2.68 Gbps when utilizing the full bandwidth. CSSP trails behind and achieves 2.57 Gbps at the same 500-MHz point. Nonetheless, the optimal exhaustive scan can evidently support even

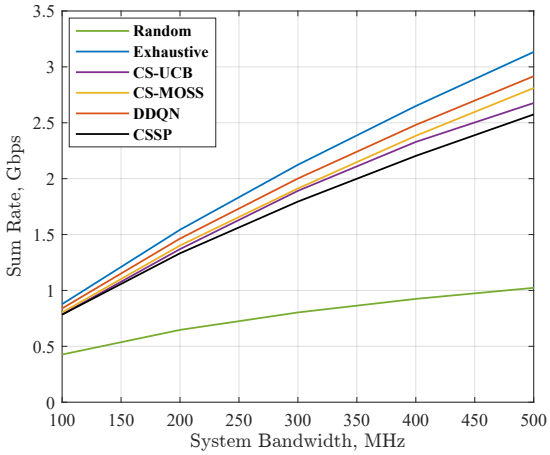


Figure 5.10: Evaluation of the overall system performance in terms of the sum-rate acquired at variable system bandwidth levels.

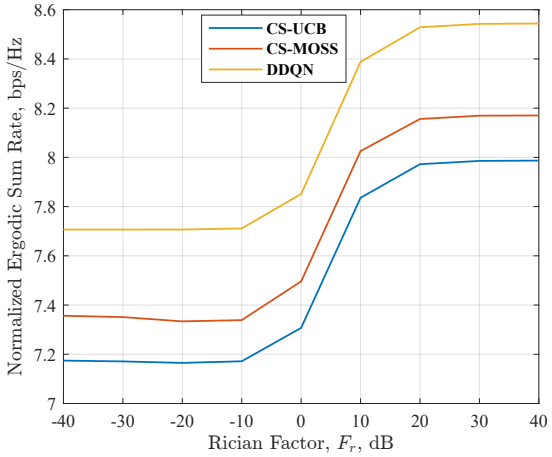


Figure 5.11: Tracking the maximum performance of the DDQN agent vs CS-MAB for varying levels of LOS component presence.

faster transmission rates going well beyond 3 Gbps by running through all permissible decisions.

The capability of the proposed algorithms to learn successful adaptations to changing line-of-sight circumstances within the wireless links of the used communication channels is validated in Fig. 5.11 where a comparative analysis of the DDQN agent versus cost-subsidized MAB methods is presented. The strength of the simulated LOS channel component is swept by adjusting the Rician parameter LOS control over the range $-40 \sim 40$ dB. Both DDQN and CS-MAB solutions show similar trends. When operating over non-LOS-dominated links (corresponding to $-40 \sim -10$ dB), the response exhibited by the DDQN agent is flat at around 7.7 bps/Hz. Similarly, CS-MOSS operates around 7.34 bps/Hz whereas CS-UCB provides 7.17 bps/Hz over the same region. Next, the total rates of both DDQN and CS-MAB solutions begin to accelerate as operation is shifted towards channels with more inherent presence of the line-of-sight component as illustrated in Fig. 5.11 for the $-10 \sim 10$ dB region. As shown for the region $10 \sim 40$ dB, the achieved levels of the total ergodic rate that can be supported by the proposed algorithms saturate eventually when operation is heavily geared toward LOS-dominated links where, at the end point of 40 dB, the DDQN solution provides 8.54 bps/Hz, whereas CS-MOSS supports 8.17 and CS-UCB comes next at 8 bps/Hz.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this thesis, we developed novel AI-based algorithms to address important issues in wireless communication systems to facilitate for more optimized technological solutions suitable for deployment in future 6G networks. The thesis focused on optimizing both PHY and MAC layers operation for 3D-NOMA networks with both terrestrial and non-terrestrial (UAV-based) communications modes considered. The overall objective of the thesis was to provide adequate AI-based solutions to challenging problems facing potential 6G candidate technologies where the developed solutions were specifically tailored for 3D-NOMA networks to ultimately address 6G target goals related to:

1. *Low-complexity PHY layer signal processing* to better accommodate 6G applications with stringent real-time demands (e.g., 3D holographic video streaming and haptic internet), with improved PHY layer transmission characteristics (e.g., enhanced OFDM PAPR reduction capability).

2. *Denser connectivity* where more devices can enjoy simultaneous access to the network without suffering performance degradation to improve the overall capacity of the servicing network (e.g., 3D-NOMA network employing power-domain user multiplexing to meet massive connectivity demands).

3. *Increased transmission speeds* where even higher data transfer rates can be supported over wireless links operating within the mmWave frequency range to satisfy the ever-growing need for faster connections for bandwidth-hungry 6G applications

which are expected to be more prevalent in the future (e.g., 16K conferencing, rich-content and immersive AR/VR applications).

4. *Robustness to emergencies and energy-oriented design* so that 6G networks including 3D-NOMA can adapt quickly and efficiently to disasters to maintain coverage support via mission-critical non-terrestrial UAV-mounted mobile base stations to provide reliable, fast wireless links to ground-based users.

In line with the above guiding targets, in this thesis, we proposed novel AI-based solutions to pave the way for the realization of the ultimate vision of 3D-NOMA networks for future 6G systems:

- For the PHY layer, in chapter 3, we proposed a novel neural network-based approach for PAPR reduction in OFDM signals with common conventional techniques for benchmark comparison. To surpass the limits of conventional solutions, we proposed a novel network-selection algorithm that scans non-linear maps in the neighborhood of conventional CF-based schemes to find the network with the lowest cubic metric reduction for certain QoS requirements. Comparative analysis in terms of the complexity required for processing OFDM signals showed significant improvement in favor of NN approaches which makes them more suitable for applications with tight low-latency real-time operation requirements. Our proposed SNN-CF scheme provided the highest performance in terms of OFDM cubic metric reduction while maintaining the $\mathcal{O}(N)$ complexity class of ANN. In addition, unlike ANN, the proposed scheme maintains much closer BER performance to conventional CF schemes at higher modulations. Moreover, the proposed scheme exhibits less OoB radiations than ANN making it more desirable for interference-sensitive transmission environments such as heterogeneous networks.

- For the MAC layer, in chapter 4, we proposed a cache-enabled RL-based power allocation and user selection scheme for opportunistic access to terrestrial downlink NOMA wireless systems. The proposed scheme learns to achieve near-optimal sum-rate levels through the continuous interaction with the NOMA environment. The trained RL agent managed to maintain consistent performance across a wide range of transmission scenarios including different cached-data availabilities, shadowing and Raleigh fading effects, and *mmWaves* transmissions. Although the training

algorithm may incur high computational complexity, the deployment complexity of a trained agent is not high. In addition, the training process can be accomplished by the agent-environment interaction in a simulated offline mode. Although other schemes such as CSP do not require convergence analysis, the proposed scheme manages to converge quickly within a few games to near-optimal policy. In addition, this is not a big concern since offloading the training process to an offline setting ensures that the trained agent will converge to its final policy before it is deployed for operation. Overall, the proposed RL-based solution provides adequate performance and exhibits high degree of adaptability to various transmission scenarios, thus making it suitable for further investigation for deployment in more advanced settings such as NOMA-based MIMO heterogeneous networks.

- In addition, non-terrestrial mission-critical UAV-based communications for emergency use cases was considered where operation optimization is accomplished to provide fast and energy-efficient UAV-assisted wireless downlink communication channels for ground-based users. To this end, in chapter 5, we have conducted an investigative study on the utilization of CS-MAB as well as DDQN agents as viable RL-based data-offloading solutions for emergency use cases deploying ready-to-dispatch UAV-MBSs for NOMA-based downlink transmissions. The DDQN agent's training was accomplished in an offline stage wherein the agent engages with the UAV-MBS-NOMA environment in a multi-iteration interactive mode prior to operational deployment. CS-MAB agents on the other hand were directly operated online as they do not utilize DNNs to require a training stage. Due to its tailored ability to resolve highly complex dynamic sequential decision problems, the proposed RL DDQN approach succeeded in supporting an energy-efficient near-optimal total rate level consistently in various battery-constrained transmission scenarios, whereas the proposed cost-subsidized MAB-based approaches followed closely after. Both proposed approaches have been tested via operation in *mmWave*-enabled propagation modes with varying dominance levels of the LOS Rician channel component. Both CS-MAB and DDQN solutions exhibited accelerated performance over links with strong LOS presence where they respectively supported as high an ergodic total rate level as 8.17 bps/Hz and 8.54 bps/Hz. We tackled the joint dynamic UAV-MBS trajectory design and NOMA transmit power splitting and receiving device

activation problem to adequately support ready-to-deploy energy-efficient solutions accommodating increased transfer speeds breaking beyond 2.5 Gbps which may prove critical importance for deployment in emergency cases with high-speed data-offloading demands as in the regions afflicted by disasters.

For PHY layer optimization, an improvement of 12% was obtained in the OFDM PAPR cubic metric measure with 99.9% CCDF fidelity. Moreover, 14.3% reduction in the associated computational cost was attained compared to conventional SCF for a 256-subcarrier OFDM system with QPSK modulation. On the other hand, for MAC layer NOMA resource management, an 18.4% faster communication speed was achieved by the proposed DDQN RL agent for cache-enabled NOMA downlink resource management over a capacity-limited backhaul link with a per-user cap of 8 bps/Hz when operating at a 38-GHz mmWave carrier using a 500 MHz of system bandwidth. In addition, significant UAV battery energy savings (reaching 91% of that required by conventional CSP) was accomplished for the dynamic path planning of mission-critical, non-terrestrial NOMA downlink transmissions using the proposed CS-UCB agent.

By solving critical PHY and MAC layer problems and boosting the performance in terms of various metrics beyond the limits of conventional approaches for both terrestrial and non-terrestrial communication scenarios, the proposed PHY and MAC layer-level AI-based schemes lay the foundational basis for an integrated and more comprehensive vision of the 3D-NOMA network in Fig. 1.4, bringing us one step closer towards more capable future 6G networks.

6.2 Future Work Suggestions

Future research directions may possibly cover areas including:

- Integrating novel AI-based PAPR reduction solutions within massive MIMO-based OFDM PHY layer operation.
- Developing novel system-level AI-based resource management solutions for intelligent reflecting surfaces (RIS)-aided multi-UAV communications with AI-driven guaranteed collision avoidance mechanisms.
- AI-aided designs for joint resource management and optimized operation of co-operative 3D-NOMA communication equipment harvesting and relaying wireless

energy.

- AI-driven coordination for HAPS nodes placement within a larger 3D-NOMA framework. Furthermore, Satellite-HAPS (and HAPS-UAV) link optimization is needed for the overall tuning of an encompassing 3D-NOMA 6G network.

LIST OF PUBLICATIONS

- *Journals*

  1. **Ahmad Gendia**, and Osamu Muta, **OFDM PAPR Reduction via Time-Domain Scattered Sampling and Hybrid Batch Training of Synchronous Neural Networks**, in *Electronics, 10, no. 14:1708,*
  *(**Published**).*
  2. **Ahmad Gendia**, Osamu Muta, and Ahmed Nasser, **Cache-enabled reinforcement learning scheme for power allocation and user selection in opportunistic downlink NOMA transmissions**, in *IEEJ Transactions on Electrical and Electronic Engineering, 17, no. 5, 722–731,2022,*
  *(**Published**).*

- *International conferences*

  1. **Ahmad Gendia**, Osamu Muta, Sherief Hashima, and Kohei Hatano, **UAV Positioning with Joint NOMA Power Allocation and Receiver Node Activation**, in *2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), 240–245, September 2022,*
  *(**Published**).*

REFERENCES

[1] D. Vaya and T. Hadpawat, "Internet of Everything (IoE): A New Era of IoT," in *ICCCE 2019: Proceedings of the 2nd International Conference on Communications and Cyber Physical Engineering*. Springer, 2020, pp. 1–6.

[2] Y. Liu, H.-N. Dai, Q. Wang, M. K. Shukla, and M. Imran, "Unmanned Aerial Vehicle for Internet of Everything: Opportunities and Challenges," *Computer communications*, vol. 155, pp. 66–83, 2020.

[3] V. C. Farias da Costa, L. Oliveira, and J. de Souza, "Internet of Everything (IoE) Taxonomies: A Survey and a Novel Knowledge-Based Taxonomy," *Sensors*, vol. 21, no. 2, p. 568, 2021.

[4] A. Shilpa, V. Muneeswaran, D. D. K. Rathinam, G. A. Santhiya, and J. Sherin, "Exploring the Benefits of Sensors in Internet of Everything (IoE)," in *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*. IEEE, 2019, pp. 510–514.

[5] A. Khanna and S. Kaur, "Internet of Things (IoT), Applications and Challenges: A Comprehensive Review," *Wireless Personal Communications*, vol. 114, pp. 1687–1762, 2020.

[6] U. S. Shanthamallu, A. Spanias, C. Tepedelenlioglu, and M. Stanley, "A Brief Survey of Machine Learning Methods and Their Sensor and IoT Applications," in *2017 8th International Conference on Information, Intelligence, Systems & Applications (IISA)*. IEEE, 2017, pp. 1–8.

[7] V. K. Quy, N. V. Hau, D. V. Anh, and L. A. Ngoc, "Smart Healthcare IoT Applications Based on Fog Computing: Architecture, Applications and Challenges," *Complex & Intelligent Systems*, vol. 8, no. 5, pp. 3805–3815, 2022.

[8] A. A. A. Sen and M. Yamin, "Advantages of Using Fog in IoT Applications," *International Journal of Information Technology*, vol. 13, pp. 829–837, 2021.

[9] Y. Chen, Q. Wang, H. Chen, X. Song, H. Tang, and M. Tian, "An Overview of Augmented Reality Technology," in *Journal of Physics: Conference Series*, vol. 1237, no. 2. IOP Publishing, 2019, p. 022082.

[10] L. F. de Souza Cardoso, F. C. M. Q. Mariano, and E. R. Zorzal, "A survey of industrial augmented reality," *Computers & Industrial Engineering*, vol. 139, p. 106159, 2020.

[11] T. Zhan, K. Yin, J. Xiong, Z. He, and S.-T. Wu, "Augmented Reality and Virtual Reality Displays: Perspectives and Challenges," *Iscience*, vol. 23, no. 8, p. 101397, 2020.

[12] P. A. Rauschnabel, B. J. Babin, M. C. tom Dieck, N. Krey, and T. Jung, "What Is Augmented Reality Marketing? Its Definition, Complexity, and Future," pp. 1140–1150, 2022.

[13] R. Aggarwal and A. Singhal, "Augmented reality and its effect on our life," in *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. IEEE, 2019, pp. 510–515.

[14] M. El Beheiry, S. Doutreligne, C. Caporal, C. Ostertag, M. Dahan, and J.-B. Masson, "Virtual Reality: Beyond Visualization," *Journal of molecular biology*, vol. 431, no. 7, pp. 1315–1321, 2019.

[15] W. R. Sherman and A. B. Craig, *Understanding Virtual Reality: Interface, Application, and Design*. Morgan Kaufmann, 2018.

[16] J. Radianti, T. A. Majchrzak, J. Fromm, and I. Wohlgenannt, "A Systematic Review of Immersive Virtual Reality Applications for Higher Education: Design Elements, Lessons Learned, and Research Agenda," *Computers & Education*, vol. 147, p. 103778, 2020.

[17] Y. Zhang, H. Liu, S.-C. Kang, and M. Al-Hussein, "Virtual Reality Applications for the Built Environment: Research Trends and Opportunities," *Automation in Construction*, vol. 118, p. 103311, 2020.

[18] M. Javaid and A. Haleem, "Virtual Reality Applications Toward Medical Field," *Clinical Epidemiology and Global Health*, vol. 8, no. 2, pp. 600–605, 2020.

[19] C. R. Ramachandiran, M. M. Chong, and P. Subramanian, "3D Hologram in Futuristic Classroom: A Review," *Periodicals of Engineering and Natural Sciences*, vol. 7, no. 2, pp. 580–586, 2019.

[20] L. Shi, B. Li, C. Kim, P. Kellnhofer, and W. Matusik, "Towards Real-Time Photorealistic 3D Holography with Deep Neural Networks," *Nature*, vol. 591, no. 7849, pp. 234–239, 2021.

[21] D. Blinder, A. Ahar, S. Bettens, T. Birnbaum, A. Symeonidou, H. Ottevaere, C. Schretter, and P. Schelkens, "Signal Processing Challenges for Digital Holographic Video Display Systems," *Signal Processing: Image Communication*, vol. 70, pp. 114–130, 2019.

[22] B. Lee, D. Kim, S. Lee, C. Chen, and B. Lee, "High-Contrast, Speckle-Free, True 3D Holography via Binary CGH Optimization," *Scientific reports*, vol. 12, no. 1, p. 2811, 2022.

[23] M. Alvarez-Mesa and C. C. Chi, "HEVC Software Media Player for Ultra-High-Quality Video: 8K and Beyond," in *Proceedings of the 25th International Display Workshops. Nagoya-Japan*, 2018, pp. 1437–1440.

[24] W. Zhang, F. Qian, B. Han, and P. Hui, "Deepvista: 16k Panoramic Cinema on Your Mobile Device," in *Proceedings of the Web Conference 2021*, 2021, pp. 2232–2244.

[25] C. Zheng, J. Yin, F. Wei, Y. Guan, Z. Guo, and X. Zhang, "STC: FoV Tracking Enabled High-Quality 16K VR Video Streaming on Mobile Platforms," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 2396–2410, 2021.

[26] T. H. Jung, H. Yoo, Y. Jin, C. E. Rhee, and C.-B. Chae, "Wireless VR/Haptic Open Platform for Multimodal Teleoperation," in *2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. IEEE, 2020, pp. 1–2.

[27] S. K. Sharma, I. Woungang, A. Anpalagan, and S. Chatzinotas, "Toward Tactile Internet in Beyond 5G Era: Recent Advances, Current Issues, and Future Directions," *Ieee Access*, vol. 8, pp. 56 948–56 991, 2020.

[28] N. Promwongsa, A. Ebrahimzadeh, D. Naboulsi, S. Kianpisheh, F. Belqasmi, R. Glitho, N. Crespi, and O. Alfandi, "A comprehensive survey of the tactile internet: State-of-the-art and research directions," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 1, pp. 472–523, 2020.

[29] Ericsson, "Ericsson Mobility Report," November 2022. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report/reports#emr2022report

[30] Radiocommunication sector of ITU, "Report ITU-R M.2370-0: IMT Traffic Estimates for the Years 2020 to 2030," *M Series: Mobile, radiodetermination, amateur and related satellite services*. [Online]. Available: https://www.itu.int/pub/R-REP-M.2370-2015

[31] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6G Wireless Systems: Vision, Requirements, Challenges, Insights, and Opportunities," *Proceedings of the IEEE*, vol. 109, no. 7, pp. 1166–1199, 2021.

[32] M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G Networks: Use Cases and Technologies," *IEEE Communications Magazine*, vol. 58, no. 3, pp. 55–61, 2020.

[33] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What Should 6G Be?" *Nature Electronics*, vol. 3, no. 1, pp. 20–29, 2020.

[34] 3GPP, Technical Specification Group Radio Access Network, "3GPP TS 38.211: NR; Physical Channels and Modulation," no. V17.3.0, 09 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3213

[35] Z. Pi and F. Khan, "An Introduction to Millimeter-Wave Mobile Broadband Systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, June 2011.

[36] W. Feng, Y. Wang, D. Lin, N. Ge, J. Lu, and S. Li, "When mmWave Communications Meet Network Densification: A Scalable Interference Coordination Perspective," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1459–1471, 2017.

[37] M. Mezzavilla, M. Zhang, M. Polese, R. Ford, S. Dutta, S. Rangan, and M. Zorzi, "End-to-end Simulation of 5G mmWave Networks," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 2237–2263, 2018.

[38] A. V. Lopez, A. Chervyakov, G. Chance, S. Verma, and Y. Tang, "Opportunities and Challenges of mmWave NR," *IEEE Wireless Communications*, vol. 26, no. 2, pp. 4–6, 2019.

[39] E. C. Strinati, S. Barbarossa, J. L. Gonzalez-Jimenez, D. Ktenas, N. Cassiau, L. Maret, and C. Dehos, "6G: The Next Frontier: From Holographic Messaging to Artificial Intelligence using Subterahertz and Visible Light Communication," *IEEE Vehicular Technology Magazine*, vol. 14, no. 3, pp. 42–50, 2019.

[40] V. Petrov, T. Kurner, and I. Hosako, "IEEE 802.15. 3d: First Standardization Efforts for Sub-Terahertz Band Communications Toward 6G," *IEEE Communications Magazine*, vol. 58, no. 11, pp. 28–33, 2020.

[41] N. Iswarya and L. Jayashree, "A Survey on Successive Interference Cancellation Schemes in Non-Orthogonal Multiple Access for Future Radio Access," *Wireless Personal Communications*, vol. 120, no. 2, pp. 1057–1078, 2021.

[42] T. Van Luong, N. Shlezinger, C. Xu, T. M. Hoang, Y. C. Eldar, and L. Hanzo, "Deep learning based successive interference cancellation for the non-orthogonal downlink," *IEEE transactions on vehicular technology*, vol. 71, no. 11, pp. 11 876–11 888, 2022.

[43] A. H. Gendia, M. Elsabrouty, and A. A. Emran, "Cooperative Multi-Relay Non-Orthogonal Multiple Access for Downlink Transmission in 5G Communication Systems," in *2017 Wireless Days*, 2017, pp. 89–94.

[44] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The Roadmap to 6G: AI Empowered Wireless Networks," *IEEE communications magazine*, vol. 57, no. 8, pp. 84–90, 2019.

[45] J. Wang, R. Li, J. Wang, Y.-q. Ge, Q.-f. Zhang, and W.-x. Shi, "Artificial Intelligence and Wireless Communications," *Frontiers of Information Technology & Electronic Engineering*, vol. 21, pp. 1413–1425, 2020.

[46] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep Learning for Massive MIMO CSI Feedback," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 748–751, 2018.

[47] M. Arnold, S. Dörner, S. Cammerer, S. Yan, J. Hoydis, and S. t. Brink, "Enabling FDD Massive MIMO through Deep Learning-Based Channel Prediction," *arXiv preprint arXiv:1901.03664*, 2019.

[48] Y. Huangfu, J. Wang, C. Xu, R. Li, Y. Ge, X. Wang, H. Zhang, and J. Wang, "Realistic Channel Models Pre-Training," in *2019 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2019, pp. 1–6.

[49] M. Sternad and D. Aronsson, "Channel Estimation and Prediction for Adaptive OFDM Downlinks [Vehicular Applications]," in *2003 IEEE 58th Vehicular Technology Conference. VTC 2003-Fall (IEEE Cat. No. 03CH37484)*, vol. 2. IEEE, 2003, pp. 1283–1287.

[50] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel State Information Prediction for 5G Wireless Communications: A Deep Learning Approach," *IEEE transactions on network science and engineering*, vol. 7, no. 1, pp. 227–236, 2018.

[51] Y. Yang, F. Gao, X. Ma, and S. Zhang, "Deep Learning-Based Channel Estimation for Doubly Selective Fading Channels," *IEEE Access*, vol. 7, pp. 36 579–36 589, 2019.

[52] C.-J. Chun, J.-M. Kang, and I.-M. Kim, "Deep Learning-Based Channel Estimation for Massive MIMO Systems," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1228–1231, 2019.

[53] N. Ginige, K. S. Manosha, N. Rajatheva, and M. Latva-aho, "Untrained DNN for Channel Estimation of RIS-Assisted Multi-User OFDM System with Hardware

Impairments," in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*. IEEE, 2021, pp. 561–566.

[54] X. Zheng and V. Lau, "Online DNN-based Channel Estimator for Massive MIMO Systems with Nonlinear Distortion," in *2021 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2021, pp. 1–6.

[55] J. Zhang, X. Ma, J. Qi, and S. Jin, "Designing Tensor-Train Deep Neural Networks for Time-Varying MIMO Channel Estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 15, no. 3, pp. 759–773, 2021.

[56] A. Naikoti and A. Chockalingam, "Low-Complexity Delay-Doppler Symbol DNN for OTFS Signal Detection," in *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*. IEEE, 2021, pp. 1–6.

[57] K. Singh, A. F. Makarim, H. Albinsaid, C.-P. Li, and Z. J. Haas, "Passive beamforming design and dnn-based signal detection in ris-assisted mimo systems with generalized spatial modulation," *IEEE Transactions on Vehicular Technology*, vol. 72, no. 2, pp. 1879–1892, 2022.

[58] H. Ye, G. Y. Li, and B.-H. Juang, "Power of Deep Learning for Channel Estimation and Signal Detection in OFDM Systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, 2017.

[59] T. Gruber, S. Cammerer, J. Hoydis, and S. Ten Brink, "On Deep Learning-Based Channel Decoding," in *2017 51st annual conference on information sciences and systems (CISS)*. IEEE, 2017, pp. 1–6.

[60] W. Lyu, Z. Zhang, C. Jiao, K. Qin, and H. Zhang, "Performance Evaluation of Channel Decoding with Deep Neural Networks," in *2018 IEEE International Conference on Communications (ICC)*. IEEE, 2018, pp. 1–6.

[61] Y. He, J. Zhang, S. Jin, C.-K. Wen, and G. Y. Li, "Model-Driven DNN Decoder for Turbo Codes: Design, Simulation, and Experimental Results," *IEEE Transactions on Communications*, vol. 68, no. 10, pp. 6127–6140, 2020.

[62] J. Armstrong, "Peak-to-Average Power Reduction for OFDM by Repeated Clipping and Frequency Domain Filtering," *Electronics Letters*, vol. 38, no. 5, pp. 246–247, Feb 2002.

[63] Y. . Wang and Z. . Luo, "Optimized Iterative Clipping and Filtering for PAPR Reduction of OFDM Signals," *IEEE Transactions on Communications*, vol. 59, no. 1, pp. 33–37, January 2011.

[64] S. Deng and M. Lin, "Recursive Clipping and Filtering With Bounded Distortion for PAPR Reduction," *IEEE Transactions on Communications*, vol. 55, no. 1, pp. 227–230, Jan 2007.

[65] Luqing Wang and C. Tellambura, "A Simplified Clipping and Filtering Technique for PAPR Reduction in OFDM Systems," *IEEE Signal Processing Letters*, vol. 12, no. 6, pp. 453–456, June 2005.

[66] R. J. Baxley, C. Zhao, and G. T. Zhou, "Constrained Clipping for Crest Factor Reduction in OFDM," *IEEE Transactions on Broadcasting*, vol. 52, no. 4, pp. 570–575, Dec 2006.

[67] K. Anoh, C. Tanriover, and B. Adebisi, "On the Optimization of Iterative Clipping and Filtering for PAPR Reduction in OFDM Systems," *IEEE Access*, vol. 5, pp. 12 004–12 013, 2017.

[68] X. Liu, X. Zhang, J. Xiong, F. Gu, and J. Wei, "An Enhanced Iterative Clipping and Filtering Method Using Time-Domain Kernel Matrix for PAPR Reduction in OFDM Systems," *IEEE Access*, vol. 7, pp. 59 466–59 476, 2019.

[69] I. Sohn and S. C. Kim, "Neural Network Based Simplified Clipping and Filtering Technique for PAPR Reduction of OFDM Signals," *IEEE Communications Letters*, vol. 19, no. 8, pp. 1438–1441, Aug 2015.

[70] M. Lin and Y. Zhao, "Artificial Intelligence-Empowered Resource Management for Future Wireless Communications: A Survey," *China Communications*, vol. 17, no. 3, pp. 58–77, 2020.

[71] Z. Yang and W. Bai, "Distributed Computation Offloading in Mobile Fog Computing: A Deep Neural Network Approach," *IEEE Communications Letters*, vol. 26, no. 3, pp. 696–700, 2021.

[72] E. Li, L. Zeng, Z. Zhou, and X. Chen, "Edge AI: On-Demand Accelerating Deep Neural Network Inference via Edge Computing," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 447–457, 2019.

[73] E. Kilcioglu, H. Mirghasemi, I. Stupia, and L. Vandendorpe, "An Energy-Efficient Fine-Grained Deep Neural Network Partitioning Scheme for Wireless Collaborative Fog Computing," *IEEE Access*, vol. 9, pp. 79 611–79 627, 2021.

[74] Z. Chen, S. Zhang, Z. Ma, S. Zhang, Z. Qian, M. Xiao, J. Wu, and S. Lu, "An Online Approach for DNN Model Caching and Processor Allocation in Edge Computing," in *2022 IEEE/ACM 30th International Symposium on Quality of Service (IWQoS)*. IEEE, 2022, pp. 1–10.

[75] M. Zaher, Ö. T. Demir, E. Björnson, and M. Petrova, "Distributed DNN Power Allocation in Cell-Free Massive MIMO," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2021, pp. 722–726.

[76] Y. Zhao, I. G. Niemegeers, and S. H. De Groot, "Power Allocation in Cell-Free Massive MIMO: A Deep Learning Method," *IEEE Access*, vol. 8, pp. 87 185–87 200, 2020.

[77] S. Xu, P. Liu, R. Wang, and S. S. Panwar, "Realtime Scheduling and Power Allocation using Deep Neural Networks," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2019, pp. 1–5.

[78] N. Yang, H. Zhang, K. Long, H.-Y. Hsieh, and J. Liu, "Deep Neural Network for Resource Management in NOMA Networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 876–886, 2019.

[79] H. Zhang, H. Zhang, K. Long, and G. K. Karagiannidis, "Deep learning based radio resource management in NOMA networks: User association, subchannel and power allocation," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 2406–2415, 2020.

[80] N. Zhao, W. Lu, M. Sheng, Y. Chen, J. Tang, F. R. Yu, and K.-K. Wong, "UAV-Assisted Emergency Networks in Disasters," *IEEE Wireless Communications*, vol. 26, no. 1, pp. 45–51, 2019.

[81] W. Feng, J. Tang, N. Zhao, Y. Fu, X. Zhang, K. Cumanan, and K.-K. Wong, "NOMA-based UAV-aided Networks for Emergency Communications," *China Communications*, vol. 17, no. 11, pp. 54–66, 2020.

[82] A. Gendia, O. Muta, S. Hashima, and K. Hatano, "UAV Positioning with Joint NOMA Power Allocation and Receiver Node Activation," in *2022 IEEE 33rd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2022, pp. 240–245.

[83] S. H. Muller and J. B. Huber, "OFDM with Reduced Peak-to-Average Power Ratio by Optimum Combination of Partial Transmit Sequences," *Electronics Letters*, vol. 33, no. 5, pp. 368–369, Feb 1997.

[84] J. Tellado and J. M. Cioffi, "Peak Power Reduction for Multicarrier Transmission," in *Proc. IEEE GLOBECOM. Conf*, 1999.

[85] K. Anoh, B. Adebisi, K. M. Rabie, and C. Tanriover, "Root-Based Nonlinear Companding Technique for Reducing PAPR of Precoded OFDM Signals," *IEEE Access*, vol. 6, pp. 4618–4629, 2018.

[86] H. Breiling, S. H. Muller-Weinfurtner, and J. B. Huber, "SLM Peak-Power Reduction without Explicit Side Information," *IEEE Communications Letters*, vol. 5, no. 6, pp. 239–241, June 2001.

[87] K. Yamashita, M. Ohta, and W. Jiang, "Reducing Peak-to-Average Power Ratio of Multicarrier Modulation by Hopfield Neural Network," *Electronics Letters*, vol. 38, no. 22, pp. 1370–1371, Oct 2002.

[88] H. J. J. and D. W. Tank, "Neural Computation of Decisions in Optimization Problems," *Biological Cybernetics*, vol. 52, no. 3, pp. 141–152, July 1985. [Online]. Available: https://doi.org/10.1007/BF00339943

[89] I. Sohn, "RBF Neural Network Based SLM Peak-to-Average Power Ratio Reduction in OFDM Systems," *ETRI journal*, vol. 29, no. 3, pp. 402–404, 2007.

[90] Y. Jabrane, V. P. G. Jiménez, A. G. Armada, B. A. E. Said, and A. A. Ouahman, "Reduction of Power Envelope Fluctuations in OFDM Signals by Using Neural Networks," *IEEE Communications Letters*, vol. 14, no. 7, pp. 599–601, July 2010.

[91] I. Sohn, "A Low Complexity PAPR Reduction Scheme for OFDM Systems via Neural Networks," *IEEE Communications Letters*, vol. 18, no. 2, pp. 225–228, February 2014.

[92] M. Kim, W. Lee, and D. Cho, "A Novel PAPR Reduction Scheme for OFDM System Based on Deep Learning," *IEEE Communications Letters*, vol. 22, no. 3, pp. 510–513, March 2018.

[93] T. R1-060023, "Cubic Metric in 3GPP-LTE," *3GPP TSGRAN WG1, Tech. Rep.*, Jan 2006.

[94] T. Kageyama, O. Muta, and H. Gacanin, "Performance Analysis of OFDM with Peak Cancellation Under EVM and ACLR Restrictions," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 6230–6241, 2020.

[95] D. W. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963. [Online]. Available: https://doi.org/10.1137/0111030

[96] C. Rapp, "Effects of HPA-Nonlinearity on a 4-DPSK/OFDM-Signal for a Digital Sound Broadcasting Signal," *ESA Special Publication*, vol. 332, pp. 179–184, 1991.

[97] Ericsson, "Ericsson Mobility Report," November 2020. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report/reports#emr2020report

[98] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[99] Y. Saito, A. Benjebbour, Y. Kishiyama, and T. Nakamura, "System-Level Performance Evaluation of Downlink Non-Orthogonal Multiple Access (NOMA)," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2013, pp. 611–615.

[100] ——, "System-Level Performance of Downlink Non-Orthogonal Multiple Access (NOMA) under Various Environments," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, 2015, pp. 1–5.

[101] S. Ali, E. Hossain, and D. I. Kim, "Non-Orthogonal Multiple Access (NOMA) for Downlink Multiuser MIMO Systems: User Clustering, Beamforming, and Power Allocation," *IEEE Access*, vol. 5, pp. 565–577, 2017.

[102] H. Zhang, D. Zhang, W. Meng, and C. Li, "User Pairing Algorithm with SIC in Non-Orthogonal Multiple Access System," in *2016 IEEE International Conference on Communications (ICC)*, 2016, pp. 1–6.

[103] J. Ren, Z. Wang, M. Xu, F. Fang, and Z. Ding, "An EM-Based User Clustering Method in Non-Orthogonal Multiple Access," *IEEE Transactions on Communications*, vol. 67, no. 12, pp. 8422–8434, 2019.

[104] Z. Yang, W. Xu, C. Pan, Y. Pan, and M. Chen, "On the Optimality of Power Allocation for NOMA Downlinks With Individual QoS Constraints," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1649–1652, 2017.

[105] S. Zhang, N. Zhang, G. Kang, and Z. Liu, "Energy and Spectrum Efficient Power Allocation with NOMA in Downlink HetNets," *Physical Communication*, vol. 31, pp. 121–132, 2018.

[106] Z. Yang, C. Pan, W. Xu, and M. Chen, "Compressive Sensing-Based User Clustering for Downlink NOMA Systems With Decoding Power," *IEEE Signal Processing Letters*, vol. 25, no. 5, pp. 660–664, 2018.

[107] A. Nasser, O. Muta, H. Gacanin, and M. Elsabrouty, "Joint User Pairing and Power Allocation With Compressive Sensing in NOMA Systems," *IEEE Wireless Communications Letters*, vol. 10, no. 1, pp. 151–155, 2021.

[108] W. Hao, O. Muta, and H. Gacanin, "Price-Based Resource Allocation in Massive MIMO H-CRANs With Limited Fronthaul Capacity," *IEEE Transactions on Wireless Communications*, vol. 17, no. 11, pp. 7691–7703, 2018.

[109] K. Long, P. Wang, W. Li, and D. Chen, "Spectrum Resource and Power Allocation With Adaptive Proportional Fair User Pairing for NOMA Systems," *IEEE Access*, vol. 7, pp. 80 043–80 057, 2019.

[110] S. Gurugopinath, Y. Al-Hammadi, P. C. Sofotasios, S. Muhaidat, and O. A. Dobre, "Non-Orthogonal Multiple Access with Wireless Caching for 5G-Enabled Vehicular Networks," *IEEE Network*, vol. 34, no. 5, pp. 127–133, 2020.

[111] L. Lei, T. X. Vu, L. Xiang, X. Zhang, S. Chatzinotas, and B. Ottersten, "Optimal Resource Allocation for NOMA-Enabled Cache Replacement and Content Delivery," in *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2019, pp. 1–6.

[112] K. N. Doan, M. Vaezi, W. Shin, H. V. Poor, H. Shin, and T. Q. S. Quek, "Power Allocation in Cache-Aided NOMA Systems: Optimization and Deep Reinforcement Learning Approaches," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 630–644, 2020.

[113] P. Yang, L. Li, W. Liang, H. Zhang, and Z. Ding, "Latency Optimization for Multi-user NOMA-MEC Offloading Using Reinforcement Learning," in *2019 28th Wireless and Optical Communications Conference (WOCC)*, 2019, pp. 1–5.

[114] A. Tatar, M. D. De Amorim, S. Fdida, and P. Antoniadis, "A Survey on Predicting the Popularity of Web Content," *Journal of Internet Services and Applications*, vol. 5, no. 1, pp. 1–20, 2014.

[115] M. S. Ali, H. Tabassum, and E. Hossain, "Dynamic User Clustering and Power Allocation for Uplink and Downlink Non-Orthogonal Multiple Access (NOMA) Systems," *IEEE Access*, vol. 4, pp. 6325–6343, 2016.

[116] V. Va and R. W. Heath, "Basic Relationship Between Channel Coherence Time and Beamwidth in Vehicular Channels," in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*. IEEE, 2015, pp. 1–5.

[117] S. Wang, K. Guan, D. He, G. Li, X. Lin, B. Ai, and Z. Zhong, "Doppler Shift and Coherence Time of 5G Vehicular Channels at 3.5 GHz," in *2018 IEEE International*

Symposium on Antennas and Propagation & USNC/URSI National Radio Science Meeting. IEEE, 2018, pp. 2005–2006.

[118] S. Sun, T. S. Rappaport, T. A. Thomas, A. Ghosh, H. C. Nguyen, I. Z. Kovács, I. Rodriguez, O. Koymen, and A. Partyka, "Investigation of Prediction Accuracy, Sensitivity, and Parameter Stability of Large-Scale Propagation Path Loss Models for 5G Wireless Communications," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 2843–2860, 2016.

[119] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband Millimeter-Wave Propagation Measurements and Channel Models for Future Wireless Communication System Design," *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3029–3056, 2015.

[120] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio and Y. LeCun, Eds., 2015. [Online]. Available: http://arxiv.org/abs/1412.6980

[121] T. Bai, R. Vaze, and R. W. Heath, "Using Random Shape Theory to Model Blockage in Random Cellular Networks," in *2012 International Conference on Signal Processing and Communications (SPCOM)*, 2012, pp. 1–5.

[122] W. Feng, N. Zhao, S. Ao, J. Tang, X. Zhang, Y. Fu, D. K. C. So, and K.-K. Wong, "Joint 3D Trajectory Design and Time Allocation for UAV-Enabled Wireless Power Transfer Networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 9265–9278, 2020.

[123] A. Dogra, R. K. Jha, and S. Jain, "A Survey on Beyond 5G Network With the Advent of 6G: Architecture and Emerging Technologies," *IEEE Access*, vol. 9, pp. 67 512–67 547, 2021.

[124] K. David and H. Berndt, "6G Vision and Requirements: Is There Any Need for Beyond 5G?" *IEEE Vehicular Technology Magazine*, vol. 13, no. 3, pp. 72–80, 2018.

[125] S. K. Sharma, I. Woungang, A. Anpalagan, and S. Chatzinotas, "Toward Tactile Internet in Beyond 5G Era: Recent Advances, Current Issues, and Future Directions," *IEEE Access*, vol. 8, pp. 56 948–56 991, 2020.

[126] M. Asad, S. Qaisar, and A. Basit, "Client-Centric Access Device Selection for Heterogeneous QoS Requirements in Beyond 5G IoT Networks," *IEEE Access*, vol. 8, pp. 219 820–219 836, 2020.

[127] R. Ali, Y. B. Zikria, A. K. Bashir, S. Garg, and H. S. Kim, "URLLC for 5G and Beyond: Requirements, Enabling Incumbent Technologies and Network Intelligence," *IEEE Access*, vol. 9, pp. 67 064–67 095, 2021.

[128] M. M. d. Silva and J. Guerreiro, "On the 5G and Beyond," *Applied Sciences*, vol. 10, no. 20, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/20/7091

[129] K. G. Panda, S. Das, D. Sen, and W. Arif, "Design and Deployment of UAV-Aided Post-Disaster Emergency Network," *IEEE Access*, vol. 7, pp. 102 985–102 999, 2019.

[130] M. Liu, J. Yang, and G. Gui, "DSF-NOMA: UAV-Assisted Emergency Communication Technology in a Heterogeneous Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 5508–5519, 2019.

[131] Z. Huang, C. Chen, and M. Pan, "Multiobjective UAV Path Planning for Emergency Information Collection and Transmission," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6993–7009, 2020.

[132] M. Y. Arafat and S. Moh, "Localization and Clustering Based on Swarm Intelligence in UAV Networks for Emergency Communications," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 8958–8976, 2019.

[133] G. Peng, Y. Xia, X. Zhang, and L. Bai, "UAV-Aided Networks for Emergency Communications in Areas with Unevenly Distributed Users," *Journal of Communications and Information Networks*, vol. 3, no. 4, pp. 23–32, 2018.

[134] T. Zhang, J. Lei, Y. Liu, C. Feng, and A. Nallanathan, "Trajectory Optimization for UAV Emergency Communication with Limited User Equipment Energy: A Safe-DQN Approach," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 3, pp. 1236–1247, 2021.

[135] W. Feng, J. Tang, Y. Yu, J. Song, N. Zhao, G. Chen, K.-K. Wong, and J. Chambers, "UAV-Enabled SWIPT in IoT Networks for Emergency Communications," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 140–147, 2020.

[136] S. K. Datta, J.-L. Dugelay, and C. Bonnet, "IoT Based UAV Platform for Emergency Services," in *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, 2018, pp. 144–147.

[137] P. Boccardo, F. Chiabrando, F. Dutto, F. G. Tonolo, and A. Lingua, "UAV Deployment Exercise for Mapping Purposes: Evaluation of Emergency Response Applications," *Sensors*, vol. 15, no. 7, pp. 15 717–15 737, 2015. [Online]. Available: https://www.mdpi.com/1424-8220/15/7/15717

[138] M. Gapeyenko, V. Petrov, D. Moltchanov, S. Andreev, N. Himayat, and Y. Koucheryavy, "Flexible and Reliable UAV-Assisted Backhaul Operation in 5G mmWave Cellular Networks," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 11, pp. 2486–2496, 2018.

[139] L. Zhu, J. Zhang, Z. Xiao, X. Cao, D. O. Wu, and X.-G. Xia, "3-D Beamforming for Flexible Coverage in Millimeter-Wave UAV Communications," *IEEE Wireless Communications Letters*, vol. 8, no. 3, pp. 837–840, 2019.

[140] Z. Xiao, H. Dong, L. Bai, D. O. Wu, and X.-G. Xia, "Unmanned Aerial Vehicle Base Station (UAV-BS) Deployment with Millimeter-Wave Beamforming," *IEEE Internet of Things Journal*, vol. 7, no. 2, pp. 1336–1349, 2019.

[141] Z. Xiao, P. Xia, and X.-G. Xia, "Enabling UAV Cellular with Millimeter-Wave Communication: Potentials and Approaches," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 66–73, 2016.

[142] Z. Chen, Z. Ding, X. Dai, and R. Zhang, "An Optimization Perspective of the Superiority of NOMA Compared to Conventional OMA," *IEEE Transactions on Signal Processing*, vol. 65, no. 19, pp. 5191–5202, 2017.

[143] M. Zeng, A. Yadav, O. A. Dobre, G. I. Tsiropoulos, and H. V. Poor, "Capacity Comparison Between MIMO-NOMA and MIMO-OMA with Multiple Users in a

Cluster," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2413–2424, 2017.

[144] ——, "On the Sum Rate of MIMO-NOMA and [MIMO-OMA systems," *IEEE Wireless Communications Letters*.

[145] Z. Wei, L. Yang, D. W. K. Ng, J. Yuan, and L. Hanzo, "On the Performance Gain of NOMA Over OMA in Uplink Communication Systems," *IEEE Transactions on Communications*, vol. 68, no. 1, pp. 536–568, 2020.

[146] A. Gendia, O. Muta, and A. Nasser, "Cache-Enabled Reinforcement Learning Scheme for Power Allocation and User Selection in Opportunistic Downlink NOMA Transmissions," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 17, no. 5, pp. 722–731, 2022.

[147] X. Mu, Y. Liu, L. Guo, J. Lin, and H. V. Poor, "Intelligent reflecting surface enhanced multi-UAV NOMA networks," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 10, pp. 3051–3066, 2021.

[148] F. Orjales, J. Losada-Pita, A. Paz-Lopez, and A. Deibe, "Towards Precise Positioning and Movement of UAVs for Near-Wall Tasks in GNSS-Denied Environments," *Sensors*, vol. 21, no. 6, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/6/2194

[149] X. Mu, Y. Liu, L. Guo, J. Lin, and H. V. Poor, "Intelligent Reflecting Surface Enhanced Multi-UAV NOMA Networks," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 10, pp. 3051–3066, 2021.

[150] L. Zhu, J. Zhang, Z. Xiao, and R. Schober, "Optimization of Multi-UAV-BS Aided Millimeter-Wave Massive MIMO Networks," in *GLOBECOM 2020-2020 IEEE Global Communications Conference*. IEEE, 2020, pp. 1–6.

[151] Z. Xiao, L. Zhu, Y. Liu, P. Yi, R. Zhang, X.-G. Xia, and R. Schober, "A Survey on Millimeter-Wave Beamforming Enabled UAV Communications and Networking," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 557–610, 2021.

[152] D. Sinha, K. A. Sankararaman, A. Kazerouni, and V. Avadhanula, "Multi-armed Bandits with Cost Subsidy," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2021, pp. 3016–3024.

[153] E. Koyuncu, M. Shabanighazikelayeh, and H. Seferoglu, "Deployment and Trajectory Optimization of UAVs: A Quantization Theory Approach," *IEEE Transactions on Wireless Communications*, vol. 17, no. 12, pp. 8531–8546, 2018.

[154] K. P. Murphy, *Probabilistic Machine Learning: An Introduction*. MIT Press, 2022. [Online]. Available: probml.ai