POWER ALLOCATION STRATEGIES TO MAXIMIZE ENERGY EFFICIENCY IN CELL-FREE MASSIVE MIMO WITH IMPERFECT CSI

BY

YASHUVEN A/L SANMUGAM

A REPORT

SUBMITTED TO

Universiti Tunku Abdul Rahman

in partial fulfillment of the requirements

for the degree of

BACHELOR OF INFORMATION TECHNOLOGY (HONOURS) (COMMUNICATIONS AND NETWORKING)

Faculty of Information and Communication Technology

(Kampar Campus)

FEBRUARY 2026

COPYRIGHT STATEMENT

ACKNOWLEDGEMENTS

I would like to begin by expressing my sincerest gratitude to my supervisor, Dr. Adeb Ali Mohammed Ahmed Al-Samet, for his exceptional guidance, patience, and unwavering support throughout the entire duration of this project. It was through his introduction that I first discovered the compelling field of Cell-Free Massive MIMO systems, and his continued encouragement gave me the determination to tackle this technically demanding research with confidence. There were many moments of self-doubt along the way, and it was his reassurance and belief in my capabilities that helped me persevere and grow through each challenge. Despite his demanding schedule, Dr. Adeb always made himself available to review my work, address my questions, and provide thoughtful feedback that consistently pushed my understanding to a deeper level. His mentorship has been truly invaluable in shaping both my technical knowledge and my development as a researcher, and I am profoundly grateful for everything he has done for me throughout this journey.

I would also like to convey my heartfelt appreciation to my academic advisor, Dr. Farina Saffa binti Mohamad Samsamnun, who has been a steady and inspiring presence throughout my entire degree programme. Her genuine belief in my abilities has been a powerful source of motivation, encouraging me to push beyond my comfort zone and give my best even during the most challenging periods of my studies. Her guidance has extended far beyond academic matters, offering perspectives and wisdom that have helped shape me into a more resilient, focused, and confident individual. I am truly fortunate and honoured to have had her support and encouragement throughout this academic journey, and her positive influence will continue to inspire me long after my studies are complete.

Lastly, I would like to dedicate my deepest appreciation to my mother, whose unconditional love, endless patience, and steadfast encouragement have been my greatest source of strength throughout this entire journey. Her sacrifices and unwavering belief in me have carried me through every difficult moment, and her constant support has been the foundation upon which all of my achievements rest. I am forever grateful for her love, understanding, and dedication, without which completing this project and this degree would simply not have been possible.

ABSTRACT

The rapid growth of global mobile data traffic and the increasing demand for sustainable wireless connectivity have elevated energy efficiency to a primary design objective in fifth-generation and sixth-generation wireless communication systems. Cell-Free Massive Multiple-Input Multiple-Output has emerged as a promising architecture for next-generation networks, offering superior coverage uniformity and spectral efficiency through the cooperative transmission of geographically distributed Access Points. However, practical CF-mMIMO deployments face three critical and interconnected challenges: pilot contamination causing imperfect Channel State Information, inter-user interference limiting signal quality, and suboptimal power allocation reducing energy efficiency. This project develops and evaluates an integrated simulation framework that addresses all three challenges progressively within a unified system model implemented in MATLAB R2022b. Four combinations of pilot assignment strategy and channel estimation method are implemented and compared, namely Least Squares, Matched Ratio, Minimum Mean Square Error, and the proposed Zero-Forcing precoding scheme, with random and greedy pilot assignment strategies respectively. Three power allocation strategies, Equal Power Allocation, Random Power Allocation, and the proposed Proximal Gradient algorithm, are evaluated under both ZF and MMSE precoding frameworks across 500 Monte Carlo realisations. The proposed ZF-PG framework achieves a peak Signal-to-Interference-plus-Noise Ratio of approximately 26 dB at 100 Access Points, a peak Spectral Efficiency of approximately 4.5 bit/s/Hz, and a peak Energy Efficiency of approximately 1.04 Mbit/Joule, representing improvements of 57 percent over MMSE-PG and over 200 percent over baseline methods. The results demonstrate that the integration of greedy pilot assignment, Zero-Forcing precoding, and Proximal Gradient power allocation delivers compounding and consistent energy efficiency gains across all evaluated operating dimensions, establishing the proposed framework as a practical and scalable solution for energy-efficient next-generation wireless network design.

Area of Study : Wireless Communications, Network Optimisation
Keywords: Cell-Free Massive MIMO, Energy Efficiency, Zero-Forcing Precoding, Proximal Gradient Optimisation, Channel Estimation, Pilot Contamination, Power Allocation, Spectral Efficiency, Imperfect CSI, 6G Networks

TABLE OF CONTENTS

TITLE PAGE	**i**
COPYRIGHT STATEMENT	**ii**
ACKNOWLEDGEMENTS	**iii**
ABSTRACT	**iv**
TABLE OF CONTENTS	**v**
LIST OF FIGURES	**viii**
LIST OF TABLES	**ix**
LIST OF SYMBOLS	**x**
LIST OF ABBREVIATIONS	**xi**
CHAPTER 1 INTRODUCTION	**1**
1.1 Problem Statement and Motivation	4
1.2 Objectives	9
1.3 Project Scope and Direction	12
1.4 Contributions	16
1.5 Report Organization	21
CHAPTER 2 LITERATURE REVIEW	**24**
2.1 Introduction to Next-Generation Wireless Networks	24
2.2 The Paradigm Shift From Cellular to Cell-Free Massive MIMO	27
2.2.1 The Evolution From Massive MIMO to Cell-Free Architecture	27
2.2.2 Architecture and Operating Principles of Cell-Free Massive MIMO	28
2.3 Channel Estimation and the Reality of Imperfect CSI	30
2.3.1 Time Division Duplexing and Channel Reciprocity	31
2.3.2 Pilot Assignment Strategies and Their Impact on Estimation Quality	32
2.3.3 Pilot Contamination and Its Consequences	34
2.3.4 Channel Estimation Techniques: From LS to MMSE	35
2.4 Downlink Precoding and Uplink Combining Strategies	36

2.4.1 Matched Ratio Combining and Maximum Ratio Transmission 37
2.4.2 Minimum Mean Square Error Precoding 39
2.4.3 Zero-Forcing Precoding and Its Role Under Imperfect CSI 40
2.5 Theoretical Foundations of Energy Efficiency 42
2.5.1 Definition and Physical Interpretation of Energy Efficiency 43
2.5.2 Spectral Efficiency and Energy Efficiency: A Fundamental Trade-off 45
2.5.3 Achievable Rate and the Use-and-Then-Forget Bound 46
2.6 Comprehensive Power Consumption Models 48
2.6.1 Transmit Power and Power Amplifier Inefficiency 49
2.6.2 Fixed Circuit Power and Computation Power Per AP 50
2.6.3 Backhaul Power Consumption 51
2.6.4 Impact of the Power Model on Energy Efficiency Optimisation 52
2.7 Optimisation Theory: From Second-Order to First-Order Methods 53
2.7.1 Mathematical Structure of the Energy Efficiency Maximisation Problem 53
2.7.2 Limitations of Second-Order Optimisation Methods 54
2.7.3 The Proximal Gradient Algorithm: Principles and Advantages 56
2.7.4 Comparison of Power Allocation Strategies in the Literature 58
2.8 Summary of Literature and Research Gaps 59
2.8.1 Key Insights From the Literature 60
2.8.2 Identified Research Gaps 62
2.8.3 Summary Table of Key Literature 63
2.8.4 Positioning of This Research 68

CHAPTER 3 SYSTEM METHODOLOGY 69
3.1 Overview 69
3.2 System Model 74
3.2.1 Network Architecture 74
3.2.2 Channel Model 75

3.2.3 System Parameters 78
3.2.4 TDD Protocol and Frame Structure 80
3.3 Proposed Framework 82
3.3.1 Objective 1 — Channel Estimation and SINR Formulation 82
3.3.1.1 Pilot Assignment Strategies 83
3.3.1.2 Channel Estimation Methods 86
3.3.1.3 SINR Formulation 90
3.3.2 Objective 2 — Spectral Efficiency Analysis 95
3.3.2.1 Spectral Efficiency Formula 95
3.3.2.2 Impact of Precoding Strategy on Spectral Efficiency 97
3.3.3 Objective 3 — Energy Efficiency Maximisation Through Power Allocation 102
3.3.3.1 Energy Efficiency Formulation 102
3.3.3.2 Power Consumption Model 103
3.3.3.3 Power Allocation Strategies 105
3.3.3.4 Proximal Gradient Power Allocation Algorithm 106
3.3.3.5 Per-AP Power Constraint 108
3.4 Chapter Conclusion 112

CHAPTER 4 RESULTS AND DISCUSSION 115
4.1 Overview 115
4.2 Summary of Results and Performance Comparison 116
4.3 Results and Discussion 119
4.3.1 Objective 1 — SINR Performance Analysis 119
4.3.2 Objective 2 — Spectral Efficiency Performance Analysis 127
4.3.3 Objective 3 — Energy Efficiency Performance Analysis 135
4.4 Chapter Conclusion 153

CHAPTER 5 CONCLUSION **157**
5.1 Project Conclusion 157
5.2 Future Work 162

REFERENCES **166**
APPENDIX **169**
POSTER **169**
CODE SAMPLE **170**

LIST OF FIGURES

Figure Number	Title	Page
Figure 1.1	Comparison between a conventional cellular network architecture and a Cell-Free Massive MIMO architecture	2
Figure 1.2	Illustration of pilot contamination in a CF-mMIMO system	6
Figure 3.1	Shows the Overall System Framewor	71
Figure 3.2	TDD Coherence Block Frame Structure	82
Figure 3.3	Flowchart of the Objective 1 methodology	94
Figure 3.4	Flowchart of the Objective 2 methodology	101
Figure 3.5	Flowchart of the Objective 3 methodology	110
Figure 4.1	Average SINR per user versus number of Access Points M for LS, MR, MMSE, and ZF methods	119
Figure 4.2	Average SINR per user versus pilot transmit power for LS, MR, MMSE, and ZF methods	123
Figure 4.3	Average SE per user versus downlink transmit power for LS, MR, MMSE, and ZF methods	127
Figure 4.4	Average SE per user versus number of users K for LS, MR, MMSE, and ZF methods	131
Figure 4.5	Energy efficiency per user versus downlink transmit power for ZF-EPA, ZF-RandP, and ZF-PG methods	135
Figure 4.6	Energy efficiency per user versus downlink transmit power for MMSE-EPA, MMSE-RandP, and MMSE-PG methods	138
Figure 4.7	Energy efficiency per user versus downlink transmit power for ZF-PG and MMSE-PG methods	140
Figure 4.8	Energy efficiency per user versus number of users K for ZF-EPA, ZF-RandP, and ZF-PG methods	142
Figure 4.9	Energy efficiency per user versus number of users K for MMSE-EPA, MMSE-RandP, and MMSE-PG methods	144
Figure 4.10	Energy efficiency per user versus number of users K for ZF-PG and MMSE-PG methods	146
Figure 4.11	Energy efficiency per user versus number of Access Points	148

	M for ZF-EPA, ZF-RandP, and ZF-PG methods	
Figure 4.12	Energy efficiency per user versus number of Access Points M for MMSE-EPA, MMSE-RandP, and MMSE-PG methods	150
Figure 4.13	Energy efficiency per user versus number of Access Points M for ZF-PG and MMSE-PG methods	151

LIST OF TABLES

Table Number	Title	Page
Table 2.8.3	Summary Table of Literature Review	63
Table 3.1	System Simulation Parameters	79
Table 4.1	Peak Performance Summary Across All Three Objectives	117
Table 4.2	Percentage Improvement of Proposed ZF-PG Framework Over Baseline and Intermediate Methods	117

LIST OF SYMBOLS

M	Total number of Access Points
K	Total number of users
M_s	Number of serving APs per user
D	Side length of simulation area
d_{mk}	Distance between AP mm m and user *k*
$d_{\min}$	Minimum AP-user separation distance
τ_c	Coherence block length
τ_u	Uplink pilot transmission length
τ_p	Number of orthogonal pilot sequences
B	System bandwidth
σ_n^2	Noise variance
σ_{sf}	Shadow fading standard deviation
p_{pilot}	Pilot transmit power
p_{dl}	Downlink transmit power
$P_{\max}$	Maximum allowable transmit power per AP
β_{mk}	Large-scale fading coefficient between AP *m* and user *k*
h_{mk}	Small-scale fading coefficient between AP *m* and user *k*
g_{mk}	Composite channel coefficient between AP *m* and user *k*
γ_{mk}	Channel estimate quality coefficient between AP mm m and user *k*
γ_{mk}^{LS}	LS estimated channel gain between AP *m* and user *k*
γ_{mk}^{MR}	MR estimated channel gain between AP *m* and user *k*
γ_{mk}^{MMSE}	MMSE estimated channel gain between AP *m* and user *k*
γ_{mk}^{ZF}	ZF estimated channel gain between AP *m* and user *k*
ξ_{mk}^{rand}	Pilot contamination sum under random pilot assignment
ξ_{mk}^{greedy}	Pilot contamination sum under greedy pilot assignment
$\mathcal{P}_k^{rand}$	Set of users sharing the same pilot as user *k* under random assignment
$\mathcal{P}_k^{greedy}$	Set of users sharing the same pilot as user *k* under greedy

	assignment
$\mathcal{M}_k$	Set of APs serving user *k* under user-centric model
$D(m,k)$	Binary AP-user association indicator
PL_{mk}	Path loss between AP *m* and user *k*
z_{mk}	Shadow fading component between AP mm m and user *k*
$SINR_k$	Signal-to-Interference-plus-Noise Ratio of user *k*
$SINR_k^{LS}$	SINR of user *k* under LS estimation
$SINR_k^{MR}$	SINR of user *k* under MR combining
$SINR_k^{MMSE}$	SINR of user *k* under MMSE estimation
$SINR_k^{ZF}$	SINR of user *k* under ZF precoding
SE_k	Spectral efficiency of user *k*
η_{EE}	Energy efficiency of the network
P_{total}	Total network power consumption
P_{fix}	Fixed network infrastructure power
P_{cm}	Fixed circuit power per AP
$P_{0,bh}$	Static backhaul power per AP
P_{bt}	Backhaul traffic power coefficient
α_{PA}	Power amplifier inefficiency factor
η_k	Power allocation coefficient for user *k*
η	Power allocation vector for all users
$\eta^{(t)}$	Power allocation vector at iteration *t*
$\mu^{(t)}$	Adaptive step size at iteration *t*
$\mathcal{P}_{\mathcal{C}}$	Projection operator onto feasible constraint set
$\nabla_\eta \eta_{EE}$	Gradient of energy efficiency with respect to power allocation vector
$C_{k,\tau}$	Pilot contamination measure for user *k* and pilot τ
τ^*	Optimal pilot sequence assignment
$\mathcal{CN}(0,1)$	Circularly symmetric complex Gaussian distribution
$\mathcal{N}(0,\sigma^2)$	Real Gaussian distribution with variance σ2\sigma^2 σ2
$\mathcal{U}(0,1)$	Uniform distribution between 0 and 1
$\mathcal{O}(\cdot)$	Big-O computational complexity notation

$\log_2(\cdot)$	Logarithm base 2
$\mathbb{E}[\cdot]$	Statistical expectation operator
$(\cdot)^H$	Hermitian transpose operator
$\sum$	Summation operator
$arg\min$	Argument of the minimum

LIST OF ABBREVIATIONS

5G	Fifth Generation
6G	Sixth Generation
AP	Access Point
AWGN	Additive White Gaussian Noise
CF-mMIMO	Cell-Free Massive Multiple-Input Multiple-Output
CPU	Central Processing Unit
CSI	Channel State Information
dB	Decibel
dBm	Decibel-milliwatt
DL	Downlink
EE	Energy Efficiency
EPA	Equal Power Allocation
FDD	Frequency Division Duplexing
GUI	Graphical User Interface
ICT	Information and Communication Technology
ICI	Inter-Cell Interference
IoT	Internet of Things
LS	Least Squares
LTE	Long-Term Evolution
MATLAB	Matrix Laboratory
MIMO	Multiple-Input Multiple-Output
MMSE	Minimum Mean Square Error
Monte Carlo	Monte Carlo Simulation
MR	Matched Ratio
MRT	Maximum Ratio Transmission
NR	New Radio
NMSE	Normalised Mean Square Error
OFDM	Orthogonal Frequency Division Multiplexing
PA	Power Amplifier
PG	Proximal Gradient

QoS	Quality of Service
RandP	Random Power Allocation
RF	Radio Frequency
SCA	Successive Convex Approximation
SE	Spectral Efficiency
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
SOCP	Second-Order Cone Programming
TDD	Time Division Duplexing
UatF	Use-and-Then-Forget
UL	Uplink
ZF	Zero-Forcing

Chapter 1
Introduction

In this chapter, the background and motivation of this research are presented, the key challenges addressed are outlined, the main contributions of the study are highlighted, and an overview of the thesis structure is provided. This chapter is essential as it sets the stage for the study, explains its importance, and outlines the research direction taken throughout this project.

The rapid evolution of wireless communication systems, from the widespread deployment of fifth-generation (5G) networks towards the conceptual development of the emerging sixth-generation (6G) paradigm, is primarily driven by the exponential growth in global mobile data traffic and the increasingly diverse demands of modern communication applications [1]. Over the past decade, the proliferation of smart devices, cloud-based services, and data-intensive applications such as ultra-high-definition video streaming, autonomous vehicles, remote surgery, and industrial automation has placed enormous pressure on existing wireless infrastructure [2]. These applications not only demand significantly higher data rates and lower latency but also require networks to support a massive number of simultaneously connected devices, often in densely populated urban environments where conventional network architectures struggle to maintain consistent service quality.

To appreciate the scale of this challenge, it is important to understand that traditional cellular network systems operate by dividing a geographical coverage area into individual cells, each served by a single base station located at the centre of that cell. While this model has proven effective for decades, it inherently suffers from a well-known limitation such as users located near the boundaries between cells, commonly referred to as cell-edge users, experience significantly degraded service due to the combined effects of increased path loss, reduced signal power, and strong interference from neighbouring base stations operating on the same frequency band [2], [3]. In densely populated urban environments where cell sizes are small and users are numerous, this interference problem becomes increasingly severe, making it extremely difficult to guarantee uniform quality of service across all users within the network.

In response to these fundamental limitations, Cell-Free Massive Multiple-Input Multiple-Output (CF-mMIMO) has emerged as one of the most promising architectural solutions for next-generation wireless systems [1], [2]. The term "cell-free" refers to the complete removal of the traditional cell boundary concept. Instead of assigning users to a specific base station, a large number of geographically distributed Access Points (APs) are deployed across the coverage area, and all of these APs work together simultaneously and cooperatively to serve every user in the network [1]. This approach is fundamentally different from conventional systems because no user is ever disadvantaged by their physical location relative to a cell boundary. Whether a user is near an AP or far from one, the collective transmission from all surrounding APs ensures that a strong and reliable signal can always be delivered. This property is commonly referred to as macro-diversity, and it is one of the defining strengths of the CF-mMIMO architecture [4].

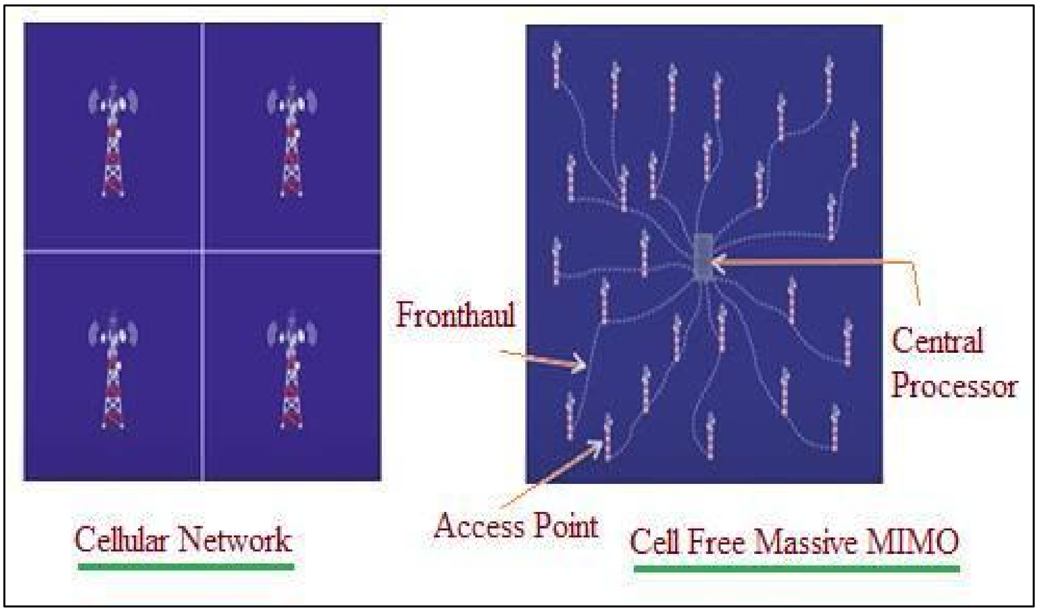


Figure 1.1: Comparison between a conventional cellular network architecture (left) and a Cell-Free Massive MIMO architecture (right). In the cellular model, users near cell boundaries experience strong inter-cell interference from neighbouring base stations. In the cell-free model, all distributed Access Points cooperatively serve every user simultaneously, eliminating cell boundaries and ensuring uniform service quality across the entire coverage area.

Despite the significant performance advantages offered by CF-mMIMO, the deployment of a large number of distributed APs introduces new and complex engineering challenges that must be addressed before these systems can be practically realised. One of the most critical challenges is the accurate estimation of the wireless channel between each AP and each user, a process that is essential for enabling coherent joint transmission. In practice, channel estimation is performed using pilot sequences transmitted by users during a dedicated training phase. However, because the number of available orthogonal pilot sequences is inherently limited by the duration of the channel coherence interval, it becomes necessary for multiple users to share the same pilot sequence when the number of users exceeds the available orthogonal resources [5]. This sharing causes a phenomenon known as pilot contamination, where the channel estimates obtained at each AP become corrupted by interference from other users transmitting on the same pilot, ultimately reducing the quality of the Channel State Information (CSI) available to the system [5], [6].

Imperfect CSI resulting from pilot contamination has a cascading effect on overall system performance. When precoding and power allocation operations are performed based on inaccurate channel estimates, the system experiences increased inter-user interference, reduced Signal-to-Interference-plus-Noise Ratio (SINR), and lower achievable data rates [3], [12]. Furthermore, many existing studies on power allocation in CF-mMIMO systems have been conducted under the idealised assumption of perfect channel knowledge, or have employed highly complex optimisation techniques such as second-order cone programming that are computationally impractical for large-scale systems with hundreds of APs and tens of users [3], [4]. These limitations create a clear and important research gap that motivates the work presented in this report.

In addition to the challenge of imperfect CSI, energy efficiency has become an increasingly critical design objective for modern wireless communication systems. The global expansion of wireless infrastructure, combined with the deployment of large numbers of distributed APs in CF-mMIMO systems, leads to a substantial increase in total network power consumption [4], [9]. As international efforts to reduce carbon emissions intensify and the concept of green communications gains greater prominence within the telecommunications industry, there is a growing demand for wireless networks that can deliver high data rates and reliable connectivity while simultaneously minimising energy consumption per unit of information transmitted [4], [10]. This trade-off between system performance and energy consumption forms the central motivation for the power allocation optimisation work carried out in this project.

Therefore, this research focuses on developing a unified and computationally efficient framework that addresses channel estimation inaccuracies, improves signal quality through advanced precoding, and optimises power allocation to achieve higher energy efficiency in CF-mMIMO systems. By systematically addressing these three interconnected challenges in a progressive and structured manner, this project demonstrates how practical and scalable solutions can significantly enhance the performance of next-generation wireless networks under realistic operating conditions.

1.1 Problem Statement and Motivation

The performance of Cell-Free Massive Multiple-Input Multiple-Output (CF-mMIMO) systems is fundamentally dependent on three interconnected processes: the accurate estimation of wireless channels between Access Points (APs) and users, the effective suppression of inter-user interference during downlink transmission, and the intelligent allocation of transmit power to maximise the overall energy efficiency of the network. In practical deployments, however, each of these processes faces significant challenges that collectively limit the achievable system performance. This section elaborates on these challenges in detail and establishes the motivation for the research carried out in this project.

1. **Channel Estimation Under Pilot Contamination**

In CF-mMIMO systems operating under Time Division Duplexing (TDD), the APs estimate the downlink channel coefficients by exploiting the reciprocity property of the wireless channel, whereby the uplink and downlink channels are assumed to be identical within a coherence interval [1]. During the uplink training phase, each user transmits a predetermined pilot sequence, and the APs use these received pilot signals to compute estimates of the channel between themselves and each user. The quality of these channel estimates is of paramount importance because all subsequent signal processing operations, including precoding and power allocation, rely entirely on the accuracy of the estimated Channel State Information (CSI).

In practice, the number of orthogonal pilot sequences that can be supported within a single coherence interval is strictly limited by the pilot length parameter, denoted as τ_p. When the total number of users K in the network exceeds the number of available orthogonal pilot sequences τ_p, pilot reuse among multiple users becomes unavoidable [5]. As a direct consequence of this reuse, when a particular AP receives a pilot signal from one user, it simultaneously receives interfering pilot signals from all other users sharing the same pilot sequence. This results in a well-documented phenomenon known as pilot contamination, where the channel estimate of a desired user becomes corrupted by the channels of all co-pilot users, producing estimates that are fundamentally biased and inaccurate [5], [6].

The severity of pilot contamination is directly proportional to the large-scale fading coefficients of the interfering users. This means that the problem becomes particularly acute in dense network scenarios where many users are geographically close to one another and therefore exhibit similar channel strengths towards the same set of APs. Unlike thermal noise, which can be reduced by increasing transmit power, the interference caused by pilot contamination scales with power and cannot be eliminated simply by transmitting at higher energy levels [1]. This makes it one of the most persistent and damaging impairments in CF-mMIMO systems, and a primary reason why practical systems must adopt intelligent pilot assignment strategies rather than relying on simple random pilot allocation.

The distinction between different pilot assignment strategies has a direct and measurable impact on the quality of the channel estimates produced. In this project, the Least Squares (LS) estimation method with random pilot assignment serves as the baseline, representing the simplest and most naive approach where pilot sequences are assigned to users without any consideration of the interference they may cause to one another. The Matched Ratio (MR) combining method also employs random pilot assignment but applies a more structured signal combining approach compared to LS. In contrast, the Minimum Mean Square Error (MMSE) estimation method is paired with a greedy pilot assignment strategy, where pilot sequences are iteratively assigned to users in a manner that minimises the total pilot contamination across the network [5]. This greedy approach actively considers the large-scale fading environment and assigns pilots such that users with strong mutual interference are less likely to share the same sequence. The proposed Zero-Forcing (ZF) precoding scheme similarly benefits from the same greedy pilot assignment and MMSE-quality channel estimates, inheriting the improved CSI quality that results from intelligent pilot management.

By comparing these four approaches within the same simulation framework, this project demonstrates clearly how pilot assignment strategy directly influences channel estimation quality, SINR, spectral efficiency, and ultimately energy efficiency.

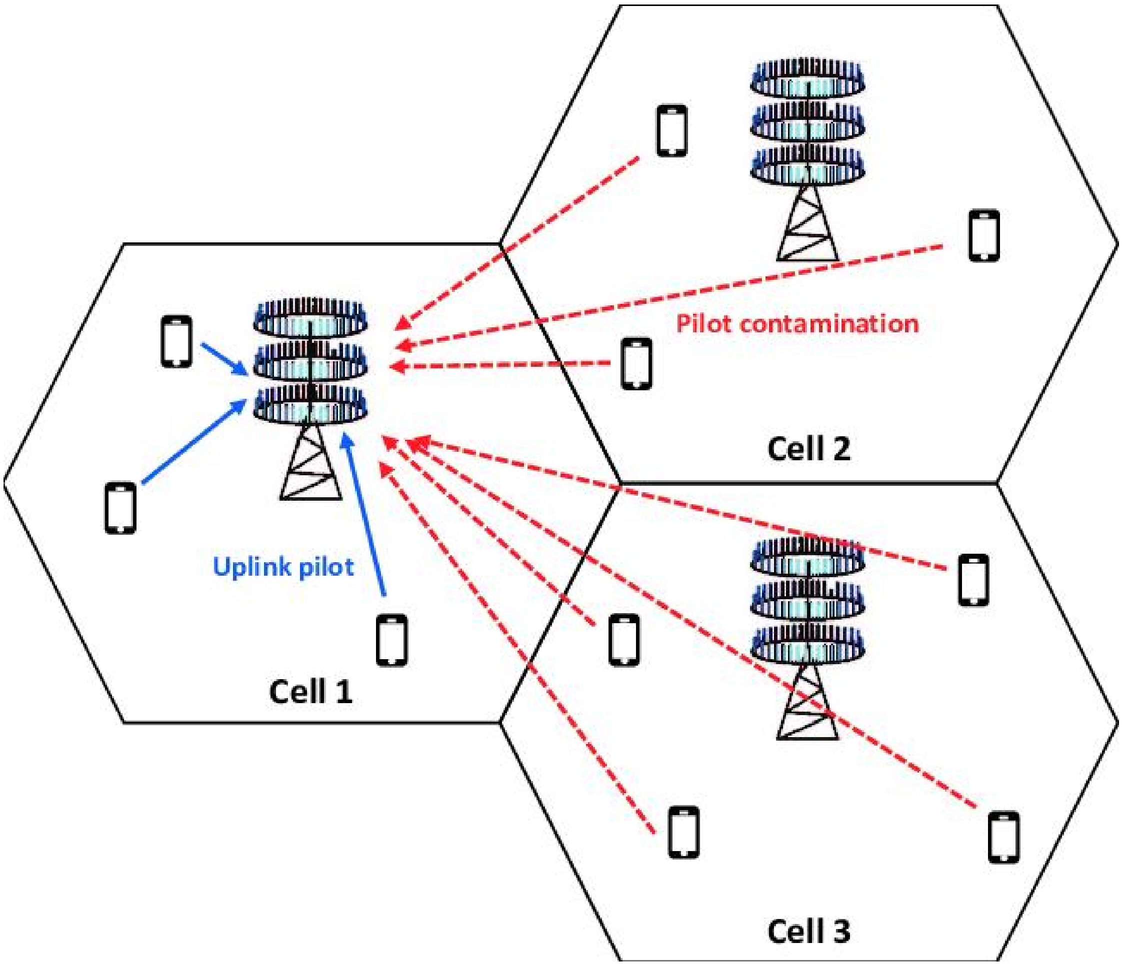


Figure 1.2: Illustration of pilot contamination in a CF-mMIMO system. When two or more users share the same pilot sequence, the channel estimate obtained at each Access Point for the desired user becomes corrupted by the interfering pilot signals transmitted by co-pilot users, resulting in imperfect Channel State Information that degrades SINR, spectral efficiency, and energy efficiency.

2. **Inter-User Interference and the Need for Advanced Precoding**

Even when channel estimates are obtained with reasonable accuracy, the simultaneous service of multiple users over shared time-frequency resources in the downlink inevitably leads to inter-user interference. In CF-mMIMO systems, each AP transmits signals intended for multiple users at the same time, and unless the transmit beamforming is carefully designed, a portion of the signal intended for one user will be received as unwanted interference at other users [2], [3]. The severity of this inter-user interference depends heavily on the spatial correlation between user channels and the precoding strategy employed.

The simplest precoding approach, Matched Ratio (MR) combining, addresses this problem by maximising the received signal power at the intended user through conjugate beamforming. While this approach is computationally simple and straightforward to implement in a distributed manner, it offers no active mechanism for suppressing the interference that leaks towards unintended users [1], [2]. As a result, MR performs well when users are well separated in space and their channels are nearly orthogonal, but its performance degrades significantly in dense scenarios where users are geographically close and their channel vectors are highly correlated. This limitation makes MR an inadequate solution for high-density deployment scenarios that are expected to be common in future 6G networks.

Zero-Forcing (ZF) precoding addresses this limitation directly by computing the pseudo-inverse of the estimated channel matrix, thereby designing transmit beamforming vectors that are orthogonal to the channels of all unintended users [3]. In theory, this completely eliminates inter-user interference at the cost of requiring matrix inversion operations and more centralised coordination among APs. In practice, the interference cancellation provided by ZF is imperfect because it is performed based on estimated rather than true channel coefficients. The residual interference that remains after ZF precoding is therefore a direct function of the channel estimation error, which reinforces the importance of high-quality CSI as established in the previous discussion. This interdependency between channel estimation quality and precoding effectiveness motivates the sequential and integrated structure of this project, where improvements in channel estimation quality in Objective 1 directly feed into and enhance the precoding performance in Objective 2.

3. **Energy Efficiency and the Limitations of Existing Power Allocation Methods**

Beyond channel estimation and precoding, the third and perhaps most practically significant challenge addressed in this project concerns the optimisation of transmit power allocation to maximise the energy efficiency of the CF-mMIMO network. Energy efficiency, typically measured in bits per Joule, quantifies how effectively the network utilises its energy resources to deliver useful information to users [4], [9]. It is not sufficient to simply maximise the total data rate of the network, because doing so often requires transmitting at very high power levels that consume disproportionately large amounts of energy for relatively small gains in throughput. Instead, an intelligent power allocation strategy must identify the optimal operating point at which the ratio of throughput to total power consumption is maximised.

The total power consumption of a CF-mMIMO network is not simply equal to the transmit power radiated from the AP antennas. It also includes fixed circuit power consumed by the hardware components of each AP regardless of whether data is being transmitted, backhaul power associated with the communication links between APs and the central processing unit, and the power amplifier inefficiency that causes the actual consumed power to be higher than the radiated power by a factor determined by the amplifier efficiency coefficient [3], [10]. When all of these components are included in a realistic power model, the relationship between transmit power and energy efficiency takes the form of a bell-shaped curve, where energy efficiency initially increases as transmit power rises but eventually begins to decline once the marginal gains in throughput are outweighed by the growing power costs [4]. Identifying and operating at the peak of this curve is therefore a non-trivial optimisation problem that requires careful mathematical formulation and an efficient solution algorithm.

Many existing approaches to this problem rely on high-complexity second-order optimisation techniques such as Successive Convex Approximation (SCA) or Second-Order Cone Programming (SOCP), which are mathematically rigorous, but they require computational effort that scales poorly with the number of APs and users [3], [4]. For a large-scale CF-mMIMO system with one hundred APs and ten users, such methods quickly become computationally impractical, particularly in scenarios where the power allocation must be updated frequently in response to changing channel conditions. This computational bottleneck motivates the adoption of a first-order optimisation approach in this project, specifically the Proximal Gradient (PG) algorithm, which offers significantly lower per-iteration complexity while still converging to a high-quality solution [3], [31]. The PG method operates by iteratively computing the gradient of the energy efficiency objective function with respect to the power allocation coefficients and then applying a projection step to ensure that the updated power values satisfy the system constraints. This makes it a practical and scalable solution that is well-suited for deployment in real-world CF-mMIMO systems.

Taken together, these three challenges which are pilot contamination and imperfect CSI, inter-user interference under imperfect channel knowledge, and energy-inefficient power allocation in large-scale distributed systems, define the core problem addressed by this research. The motivation for this project therefore stems from the need to develop a unified, practical, and computationally efficient simulation framework that addresses all three challenges in a coherent and sequential manner, demonstrating how improvements at each stage progressively enhance the overall energy efficiency of the CF-mMIMO network.

1.2 Objectives

The primary aim of this project is to design, implement, and evaluate an integrated simulation framework that progressively enhances the reliability, spectral efficiency, and energy efficiency of Cell-Free Massive Multiple-Input Multiple-Output (CF-mMIMO) systems under realistic operating conditions characterised by imperfect Channel State Information (CSI). Rather than addressing each performance dimension in isolation, this project adopts a unified and sequential approach where the output of each objective directly contributes to and improves the performance of the subsequent one. This chain of dependencies ensures that the final energy efficiency results are grounded in realistic channel estimates and effective interference suppression, making the findings practically meaningful and relevant to the design of future 6G wireless networks.

To achieve this overarching aim, three specific research objectives have been defined, each targeting a distinct but interconnected aspect of CF-mMIMO system performance:

Objective 1: To evaluate and compare channel estimation quality and SINR performance across different pilot assignment and precoding strategies in a CF-mMIMO system.

The first objective focuses on establishing the relationship between pilot assignment strategy, channel estimation quality, and the resulting Signal-to-Interference-plus-Noise Ratio (SINR) experienced by users in the downlink. Four distinct methods are implemented and compared within the same simulation environment: Least Squares (LS) estimation with random pilot assignment, Matched Ratio (MR) combining with random pilot assignment, Minimum Mean Square Error (MMSE) estimation with greedy pilot assignment, and the proposed Zero-Forcing (ZF) precoding scheme with greedy pilot assignment and MMSE-quality channel estimates. Each method represents a different level of sophistication in terms of pilot management and signal processing complexity, and together they form a clear performance hierarchy that illustrates the progressive benefits of adopting more advanced estimation and assignment strategies [1], [5].

The LS method with random pilot assignment serves as the lowest performance baseline, representing a system with no intelligence in pilot allocation and a simple estimation approach that does not account for interference statistics. Moving up the hierarchy, the MR method also uses random pilot assignment but applies a more structured combining approach that provides modest improvements in received signal power. The MMSE method with greedy pilot assignment represents a significant step forward, as it actively minimises pilot contamination during the assignment process and produces statistically efficient channel estimates that account for the large-scale fading environment [5], [6]. Finally, the proposed ZF precoding scheme, which builds upon the same MMSE-quality estimates, achieves the highest SINR performance by actively cancelling inter-user interference through the pseudo-inverse of the estimated channel matrix. The performance of all four methods is evaluated as a function of both the number of APs and the pilot transmit power, enabling a comprehensive understanding of how system scale and pilot energy affect channel estimation quality and signal quality simultaneously.

Objective 2: To assess and compare the downlink Spectral Efficiency (SE) achievable under different precoding strategies in the presence of imperfect CSI.

The second objective extends the analysis from Objective 1 into the domain of spectral efficiency, which is a direct measure of how effectively the available frequency bandwidth is being utilised to deliver useful data to users. Spectral efficiency, expressed in bits per second per Hertz (bit/s/Hz), is computed from the effective SINR obtained under each of the four precoding and estimation strategies evaluated in Objective 1, using the standard Shannon capacity formula modified by a pre-log factor that accounts for the overhead associated with uplink pilot transmission [1], [3]. Because spectral efficiency is a monotonically increasing function of SINR, the performance ranking of the four methods is expected to be consistent with the results from Objective 1, with ZF precoding achieving the highest spectral efficiency and LS estimation with random pilot assignment achieving the lowest.

However, the analysis in Objective 2 goes beyond simply confirming this ranking. The spectral efficiency is evaluated as a function of downlink transmit power and as a function of the number of users, providing additional insights into how each method scales with increasing network load and transmit energy. Evaluating spectral efficiency as a function of the number of users is particularly important because it reveals how robust each precoding strategy is to increasing levels of pilot contamination and inter-user interference as the network becomes more densely populated [2], [12]. A method that achieves high spectral efficiency with a small number of users but degrades rapidly as the user count increases is less practically valuable than one that maintains strong performance across a wide range of operating conditions. By characterising this behaviour for all four methods, Objective 2 provides a thorough and practically relevant assessment of spectral efficiency performance that directly informs the power allocation optimisation carried out in Objective 3.

Objective 3: To develop and evaluate a low-complexity energy efficiency maximisation framework based on the Proximal Gradient method, and to compare its performance against baseline power allocation strategies under both ZF and MMSE precoding.

The third and final objective represents the culmination of the entire research framework, bringing together the channel estimation improvements from Objective 1 and the spectral efficiency gains from Objective 2 to address the practical problem of maximising energy efficiency through intelligent power allocation. Three power allocation strategies are implemented and compared for both ZF and MMSE precoding frameworks: Equal Power

Allocation (EPA), Random Power Allocation (RandP), and the proposed Proximal Gradient (PG) algorithm [3], [10], [11].

Equal Power Allocation serves as the simplest baseline, distributing the total available transmit power uniformly across all users without any consideration of their individual channel conditions or their contribution to the overall energy efficiency. While EPA is straightforward to implement and requires no computational overhead, it is clearly suboptimal in heterogeneous channel environments where different users experience vastly different large-scale fading conditions [9], [11]. Random Power Allocation assigns power coefficients randomly to each user and averages the resulting energy efficiency over multiple random realisations, serving as a statistical reference that captures the expected performance of an unintelligent allocation strategy. The proposed PG algorithm, in contrast, iteratively optimises the power allocation coefficients by computing the gradient of the energy efficiency objective function and applying a projection step to enforce the per-AP power constraints, converging towards the allocation that maximises the ratio of total throughput to total network power consumption [3], [31].

The energy efficiency performance of all three power allocation strategies is evaluated under both ZF and MMSE precoding frameworks, and the results are presented as a function of downlink transmit power, number of users, and number of APs. This multi-dimensional evaluation allows a thorough characterisation of how each combination of precoding strategy and power allocation method performs across a wide range of operating conditions. The head-to-head comparison between ZF-PG and MMSE-PG in particular serves as the definitive combined result of all three objectives, demonstrating whether the additional interference cancellation capability of ZF precoding, built upon improved channel estimates from Objective 1, translates into a meaningful energy efficiency advantage over the MMSE-based approach when both are paired with the proposed PG power allocation algorithm [3], [4].

1.3 Project Scope and Direction

This section defines the boundaries within which this research has been conducted and clarifies the specific technical direction taken throughout the project. Establishing a clear scope is important in any research undertaking because it allows the reader to understand precisely what has been investigated, what assumptions have been made, and what aspects of the broader problem have been intentionally excluded so that the core research objectives can be addressed with sufficient depth and rigour. The scope of this project has been carefully defined in consultation with my project supervisor to ensure that the work is both technically meaningful and achievable within the available time and computational resources.

Network Architecture and Simulation Environment

This project focuses exclusively on the algorithmic design and simulation-based performance evaluation of a downlink Cell-Free Massive Multiple-Input Multiple-Output (CF-mMIMO) system. The simulation environment models a two-dimensional square geographical area of dimensions 1000 metres by 1000 metres, within which a total of M Access Points (APs) and K users are deployed according to a uniform random spatial distribution. The random deployment of both APs and users is an important modelling choice because it ensures that the simulation results are not specific to any single fixed network topology. By repeating the simulation over many independent random deployments through a Monte Carlo approach, the performance results obtained are statistically representative of a wide variety of realistic network configurations rather than being biased towards any particularly favourable or unfavourable arrangement of nodes [1], [2].

The system operates under a user-centric serving model, where each user is not served by all M APs simultaneously but rather by a selected subset of the M_s = 15 APs that exhibit the strongest large-scale fading coefficients towards that particular user. This user-centric approach is adopted because it reflects a more scalable and practically feasible deployment strategy compared to the fully centralised model where every AP serves every user [1], [23]. In the fully centralised model, the amount of channel information and the volume of data that must be exchanged between APs and the central processing unit grows prohibitively large as the network scales up, making it unsuitable for real-world implementation. The user-centric model limits this overhead by restricting each AP to serving only the users for which it can make a meaningful contribution to signal quality, thereby achieving a practical balance between performance and scalability.

All simulations are implemented in MATLAB R2022b, which is selected as the simulation platform because of its powerful built-in support for matrix operations, iterative algorithm design, statistical analysis, and graphical visualisation. The matrix-intensive nature of CF-mMIMO computations, which involve large-scale fading matrices, channel estimate matrices, precoding weight matrices, and power allocation vectors, makes MATLAB particularly well-suited for this type of research. The simulation framework is further complemented by an interactive Graphical User Interface (GUI) developed in MATLAB, which allows users to adjust key system parameters such as the number of APs, the number of users, and the transmit power levels, and to observe the resulting performance graphs in real time. This GUI serves as a practical demonstration tool that makes the simulation framework accessible and interactive, enabling straightforward exploration of the system behaviour under different operating conditions.

Channel Modelling Assumptions

The wireless channel model adopted in this project incorporates two distinct propagation phenomena that together capture the most important characteristics of realistic wireless environments. The first is large-scale fading, which accounts for the distance-dependent path loss and shadow fading effects that cause the average received signal power to vary slowly as a function of the physical distance between a transmitter and a receiver and the presence of obstructions in the propagation environment [6], [34]. In this project, the large-scale fading coefficient between the m-th AP and the k-th user, denoted as β_mk, is modelled using a single-slope path loss model given by the expression $PL = -30.5 - 36.7 \times \log_{10}(d)$ dB, where d represents the distance between the AP and the user in metres. A minimum separation distance of 10 metres is enforced between any AP and user pair to prevent unrealistically large channel gains at extremely short distances. Shadow fading is modelled as a log-normal random variable with a standard deviation of σ_sf = 8 dB, representing the variability in received signal strength caused by buildings, walls, vegetation, and other physical obstructions that are present in realistic deployment environments [6], [34].

The second propagation phenomenon is small-scale fading, which captures the rapid and random fluctuations in signal amplitude caused by the constructive and destructive interference of multiple signal paths arriving at the receiver from different directions with different delays and phases. In this project, the small-scale fading component is modelled as an independent and identically distributed complex Gaussian random variable with zero mean and unit variance, corresponding to a standard Rayleigh fading channel model that is widely used in wireless communication research for non-line-of-sight propagation environments [6], [35]. The combination of large-scale and small-scale fading models ensures that the simulation captures both the macro-level propagation characteristics that determine which APs are best positioned to serve each user and the micro-level channel fluctuations that affect instantaneous signal quality and interference levels.

Fixed System Parameters

To ensure consistency and reproducibility across all simulation experiments conducted in this project, a set of core system parameters has been fixed throughout the study. The system bandwidth is set to B = 20 MHz, and the channel coherence block length is set to $\tau_c = 200$ symbols, which determines the total number of symbols available within a single coherence interval for both pilot transmission and data transmission. The uplink pilot length is set to $\tau_u = 10$ symbols, and the number of available orthogonal pilot sequences is set to $\tau_p = 5$, creating a deliberate pilot shortage scenario where the number of users exceeds the number of orthogonal pilots and pilot reuse is unavoidable. The noise variance is set to $\sigma^2 = -94$ dBm, and the pilot transmit power is fixed at $p_{pilot} = 20$ dBm, equivalent to 100 milliwatts. The downlink transmit power is set to $p_{dl} = 23$ dBm, equivalent to approximately 200 milliwatts, for the channel estimation and spectral efficiency evaluations in Objectives 1 and 2, while a reduced power level of $p_{dl} = 15$ dBm is used as the reference operating point for the energy efficiency evaluations in Objective 3, as this level corresponds more closely to the optimal energy efficiency operating region identified in the simulation results [3], [4].

For the energy efficiency power consumption model, the power amplifier inefficiency factor is set to $\alpha_{PA} = 8$, the fixed circuit power per AP is set to $P_{cm} = 0.2$ Watts, the fixed network power is set to $P_{fix} = 10$ Watts, the backhaul power per AP is set to $P0_{bh} = 0.01$ Watts, and the backhaul traffic power coefficient is set to $P_{bt} = 0.25 \times 10^{-9}$ Watts per bit per second. These parameters are adopted from established power consumption models in the CF-mMIMO literature and are consistent with the values used in related works [3], [9], [10].

Scope Exclusions and Delimitations

It is equally important to clearly state what this project does not cover, so that the reader understands the boundaries of the research and does not draw conclusions beyond what the simulation framework is designed to evaluate. This project does not include any hardware implementation or real-time physical deployment of the proposed system. All results are obtained entirely through software simulation, and the findings represent theoretical performance bounds that would need to be validated through hardware prototyping in future work. The project also does not consider uplink data transmission, focusing exclusively on the downlink direction where precoding and power allocation are the primary design variables. Machine learning-based approaches to channel estimation, pilot assignment, or power allocation are also outside the scope of this project, as the research is specifically directed towards classical signal processing and first-order mathematical optimisation methods that are well-understood, interpretable, and computationally tractable [3], [31].

Furthermore, this project does not consider frequency-selective fading channels or multi-carrier transmission schemes such as Orthogonal Frequency Division Multiplexing (OFDM), as the channel model adopted assumes flat fading within the coherence bandwidth. The effects of hardware impairments such as quantisation noise, phase noise, and non-linear amplifier distortion are also excluded from the model, as including these effects would introduce additional layers of complexity that are beyond the scope of the current study. These exclusions are not limitations of the research methodology but rather deliberate and well-justified modelling choices that allow the project to maintain a focused and coherent investigation of the three core research objectives within the available timeframe and computational resources.

1.4 Contributions

This project makes several meaningful and original contributions to the field of energy-efficient wireless communication systems, specifically within the context of Cell-Free Massive Multiple-Input Multiple-Output (CF-mMIMO) networks operating under imperfect Channel State Information (CSI). While the individual components of this framework, such as channel estimation, precoding, and power allocation, have been studied separately in the existing literature, the key contribution of this project lies in the integration of all three components into a single, coherent, and progressively structured simulation framework that demonstrates how improvements at each stage collectively enhance the overall energy efficiency of the network. The following paragraphs elaborate on each of the primary contributions of this research.

Contribution 1: A Unified and Progressive Simulation Framework for CF-mMIMO Under Imperfect CSI

The first and most fundamental contribution of this project is the design and implementation of a comprehensive MATLAB-based simulation framework that jointly addresses channel estimation quality, downlink precoding performance, and energy-efficient power allocation within a single unified system model. Unlike many existing studies that isolate one of these three aspects and evaluate it independently under simplified assumptions, this project treats them as deeply interconnected stages of a single end-to-end system pipeline [1], [3]. The output of the channel estimation stage directly determines the quality of the CSI available to the precoder, and the quality of the precoding directly influences the SINR values that feed into the spectral efficiency and energy efficiency calculations. This progressive dependency chain means that improvements introduced at any stage have compounding benefits that propagate through the entire framework, producing a final energy efficiency result that genuinely reflects the cumulative effect of all three objectives working together.

This integrated approach is particularly valuable because it allows the research to move beyond the theoretical analysis of individual components and instead demonstrate how a complete and practically motivated system behaves under realistic operating conditions. The framework incorporates a realistic path loss model, shadow fading, small-scale Rayleigh fading, pilot contamination, and a comprehensive multi-component power consumption model that includes transmit power, fixed circuit power, backhaul power, and power amplifier inefficiency. By incorporating all of these elements simultaneously, the simulation produces results that are far more representative of real-world CF-mMIMO deployments than studies that assume perfect CSI, simplified power models, or idealised propagation conditions [3], [4], [9].

Contribution 2: Comparative Evaluation of Pilot Assignment Strategies and Their Impact on SINR

The second contribution of this project is a systematic and quantitative comparison of four distinct channel estimation and pilot assignment strategies, evaluated in terms of their impact on downlink SINR across a range of network configurations. The four strategies — LS with random pilot assignment, MR with random pilot assignment, MMSE with greedy pilot assignment, and ZF with greedy pilot assignment — represent a carefully selected spectrum of complexity and performance that allows the research to clearly demonstrate the progressive benefits of adopting more sophisticated pilot management and estimation approaches [1], [5], [6].

A particularly important aspect of this contribution is the demonstration that the choice of pilot assignment strategy has a far greater impact on system performance than is often acknowledged in the literature. The difference in SINR between the simplest approach, which is LS estimation with random pilot assignment, and the most advanced approach, which is ZF precoding with greedy pilot assignment, is shown to be substantial, with ZF achieving approximately 25 dB of SINR at M = 100 APs compared to approximately 11 dB achieved by LS under the same conditions. This difference of approximately 14 dB represents a more than twentyfold improvement in linear SINR, which has profound implications for the achievable data rates and energy efficiency of the system. By providing this clear and quantitative performance comparison, this project offers a practical guide for network designers who must decide which estimation and precoding strategy to adopt based on the available computational resources and the performance requirements of their deployment [2], [5].

Contribution 3: Spectral Efficiency Analysis Under Realistic Imperfect CSI Conditions

The third contribution is a thorough spectral efficiency analysis that evaluates the downlink data rate performance of all four precoding strategies as a function of both downlink transmit power and the number of users in the network. This dual-dimension evaluation is an important contribution because it characterises not only the peak performance of each method but also how gracefully each method degrades as the network becomes more loaded with additional users [2], [3]. A precoding strategy that achieves high spectral efficiency with a small number of users but collapses rapidly as user density increases is of limited practical value for future 6G networks, which are expected to support massive numbers of simultaneously connected devices.

The results demonstrate that ZF precoding achieves approximately 4.5 bit/s/Hz of spectral efficiency per user at a downlink power of 30 dBm, compared to approximately 2.7 bit/s/Hz for MMSE, 1.5 bit/s/Hz for MR, and 1.4 bit/s/Hz for LS under the same conditions.

Furthermore, the analysis as a function of the number of users reveals that ZF precoding degrades most gracefully with increasing user load, maintaining a clear performance advantage over all baseline methods across the entire range of user counts evaluated. This robustness to increasing network density is a direct consequence of the interference cancellation capability of ZF precoding, which becomes increasingly valuable as the number of simultaneously served users and the resulting inter-user interference grow [1], [3]. These spectral efficiency results provide the quantitative foundation upon which the energy efficiency optimisation in Objective 3 is built, since spectral efficiency appears directly in the numerator of the energy efficiency expression.

Contribution 4: Implementation and Evaluation of a Low-Complexity Proximal Gradient Power Allocation Algorithm

The fourth contribution of this project is the practical implementation and comprehensive evaluation of a Proximal Gradient (PG) based power allocation algorithm for energy efficiency maximisation in CF-mMIMO systems. While the theoretical foundation of the PG method has been established in the existing literature [3], [31], this project contributes a concrete MATLAB implementation of the algorithm that incorporates realistic per-AP power constraints, a multi-component power consumption model, and an adaptive step size mechanism that improves convergence behaviour by increasing the step size when the energy efficiency objective improves and halving it when the objective decreases. This adaptive step size strategy ensures that the algorithm makes rapid progress in favourable directions while avoiding overshooting the optimal solution, resulting in more reliable convergence compared to a fixed step size approach.

The PG algorithm is evaluated against two baseline power allocation strategies, Equal Power Allocation and Random Power Allocation, under both ZF and MMSE precoding frameworks, producing six distinct performance curves for each evaluation scenario. This comprehensive comparison clearly demonstrates the practical value of intelligent power allocation, showing that ZF-PG achieves a peak energy efficiency of approximately 1.10 Mbit/Joule compared to approximately 0.70 Mbit/Joule for MMSE-PG, representing a performance advantage of approximately 57 percent in favour of the proposed ZF-based framework [3], [4]. The evaluation of energy efficiency as a function of downlink transmit power, number of users, and number of APs provides a multi-dimensional characterisation of the algorithm's behaviour that is more thorough than what is typically found in studies that evaluate power allocation performance along only a single system dimension.

Contribution 5: An Interactive MATLAB GUI for Parameter Exploration and Result Visualisation

The fifth contribution of this project is the development of a fully functional and interactive Graphical User Interface (GUI) implemented in MATLAB that integrates all three research objectives into a single accessible platform. The GUI features four dedicated tabs corresponding to SINR performance, spectral efficiency, energy efficiency, and a combined result view, each with adjustable parameter sliders that allow the user to modify key system parameters such as the number of APs, the number of users, and the transmit power levels in real time [1], [3]. The GUI also includes toggle buttons that allow switching between ZF and MMSE precoding frameworks for the energy efficiency analysis, as well as framework checkboxes that allow individual performance curves to be shown or hidden for clearer visual comparison.

The development of this GUI represents a practical engineering contribution that goes beyond the mathematical analysis of the system. It transforms the simulation framework from a collection of standalone scripts into an integrated and user-friendly tool that can serve as a platform for further research, teaching, or practical demonstration of CF-mMIMO system behaviour. By making the simulation accessible to users who may not be familiar with the underlying MATLAB code, the GUI broadens the potential impact of this research and demonstrates the feasibility of implementing a complete CF-mMIMO performance evaluation platform within a single interactive environment.

1.5 Report Organization

This report is organised into five chapters, each addressing a distinct and progressive stage of the research process from the initial motivation and background through to the final conclusions and recommendations for future work. The following paragraphs provide a brief overview of the content and purpose of each chapter to help the reader navigate the report and understand how the individual chapters contribute to the overall research narrative.

Chapter 1 provides the introduction to the research, establishing the background and motivation for the study within the context of the global transition towards fifth-generation and sixth-generation wireless communication systems. The chapter identifies the three core challenges addressed by the research, namely pilot contamination and imperfect Channel State Information, inter-user interference under advanced precoding strategies, and energy-inefficient power allocation in large-scale distributed networks, and formulates the three specific research objectives designed to address these challenges. The project scope and direction are defined, the key contributions of the research are elaborated, and the overall report structure is outlined.

Chapter 2 presents a comprehensive review of the existing literature across six thematically organised areas that collectively form the theoretical and technical foundation of the research. The review covers the evolution of next-generation wireless networks and the emergence of Cell-Free Massive MIMO as a key enabling technology, the architecture and operating principles of CF-mMIMO systems, channel estimation techniques and pilot assignment strategies under imperfect CSI conditions, downlink precoding strategies including Matched Ratio combining, MMSE precoding, and Zero-Forcing precoding, the theoretical foundations of energy efficiency as a system design objective, comprehensive power consumption modelling for realistic CF-mMIMO deployments, and optimisation theory covering second-order and first-order methods for energy efficiency maximisation. The chapter concludes with a structured summary table of key literature and a clear identification of the research gaps that this project addresses.

Chapter 3 presents the complete methodology adopted in this research, describing the system model, simulation environment, and technical implementation of all three research objectives in full detail. The chapter begins with the network architecture, channel model, system parameters, and TDD protocol structure that underpin all simulation experiments. It then presents the proposed research framework across three subsections corresponding to the three research objectives, covering the pilot assignment strategies and channel estimation methods of Objective 1 with their full mathematical derivations, the spectral efficiency formulation of Objective 2 with the Shannon capacity formula and pre-log factor, and the energy efficiency maximisation framework of Objective 3 including the multi-component power consumption model, the three power allocation strategies, the Proximal Gradient algorithm with adaptive step size mechanism, and the per-AP power constraint enforcement. Flowcharts are provided throughout the chapter to illustrate the overall system framework and the methodology of each individual objective.

Chapter 4 presents and discusses the simulation results obtained from the three-stage research framework, covering a total of thirteen performance graphs across the three research objectives. The chapter begins with a structured summary of key numerical results and performance improvements across all objectives before proceeding to a detailed graph-by-graph discussion of the SINR performance of Objective 1, the spectral efficiency performance of Objective 2, and the energy efficiency performance of Objective 3 evaluated across the dimensions of downlink transmit power, number of users, and number of Access Points. Each graph is accompanied by a detailed discussion that interprets the observed behaviour, explains the underlying physical and mathematical reasons for that behaviour, and contextualises the results within the broader research framework.

Chapter 5 presents the conclusions drawn from the research and recommendations for future work. The conclusion summarises the key achievements of the project across all three research objectives, explains why energy efficiency is a critical design objective for 5G and 6G wireless networks, discusses the practical significance of the proposed ZF-PG framework and how it can be adopted or extended in real-world next-generation network deployments, and provides an overall assessment of the research contributions. The future work section identifies four promising directions for extending the research, covering hardware implementation and real-world validation, machine learning based power allocation, joint uplink-downlink optimisation, and multi-antenna Access Point extensions.

Chapter 2

Literature Review

This chapter presents a comprehensive review of the existing literature that forms the theoretical and technical foundation of this research. The review is organised thematically, beginning with an overview of the evolution of next-generation wireless communication networks and the fundamental limitations of conventional cellular architectures. It then progresses through the key topics that are directly relevant to this project, namely the architecture and advantages of Cell-Free Massive Multiple-Input Multiple-Output (CF-mMIMO) systems, the challenges associated with channel estimation under imperfect Channel State Information (CSI), the design and comparison of downlink precoding strategies, the theoretical foundations of energy efficiency as a system design objective, realistic power consumption modelling, and the mathematical optimisation techniques used to maximise energy efficiency. The chapter concludes with a structured summary of the key literature and a clear identification of the research gaps that this project addresses.

The purpose of this literature review is not merely to summarise what others have done, but to build a coherent and well-justified technical narrative that explains why the specific methods chosen in this project — greedy pilot assignment, Zero-Forcing precoding, and Proximal Gradient power allocation — represent the most appropriate and practically motivated choices for addressing the identified research gaps. Every design decision made in this project can be traced back to insights drawn from the literature reviewed in this chapter.

2.1 Introduction to Next-Generation Wireless Networks

The global wireless communications industry is currently experiencing one of the most significant technological transitions in its history. Over the past several decades, mobile communication systems have evolved through successive generations, each defined by a new set of performance targets, architectural innovations, and enabling technologies. The transition from second-generation (2G) digital voice systems to third-generation (3G) mobile broadband, followed by the introduction of fourth-generation (4G) Long-Term Evolution

(LTE) networks and most recently the global rollout of fifth-generation (5G) New Radio systems, has been driven primarily by the ever-increasing demand for higher data rates, lower latency, and broader connectivity [8]. Each of these generational transitions brought meaningful improvements in spectral efficiency and network capacity, yet each also introduced new challenges and limitations that motivated the development of the next generation.

As 5G networks continue to mature and expand globally, the research community has already begun to define the technical vision and requirements for sixth-generation (6G) wireless systems, which are expected to be commercially deployed around the 2030s [14]. The performance targets envisioned for 6G are substantially more ambitious than those of 5G, with projected peak data rates exceeding one terabit per second, end-to-end latency in the range of sub-millisecond, and the ability to support connectivity densities of up to ten million devices per square kilometre [6], [14]. These targets are motivated by the anticipated proliferation of emerging applications such as extended reality, holographic communication, autonomous systems, and large-scale machine-type communication, all of which impose stringent and simultaneous requirements on data rate, reliability, latency, and energy consumption that current 5G architectures cannot fully satisfy.

One of the most significant shifts in the design philosophy of 6G, compared to previous generations, is the elevation of energy efficiency from a secondary consideration to a primary design objective [4], [10]. In 5G and earlier systems, network design was predominantly guided by the goal of maximising spectral efficiency and peak data rates, with energy consumption treated largely as an operational cost to be minimised after the fact. This approach has resulted in a global telecommunications infrastructure that is responsible for a substantial and growing share of worldwide energy consumption and carbon dioxide emissions. According to estimates cited in the literature, the Information and Communication Technology sector as a whole is projected to account for as much as twenty percent of global electricity consumption by 2030 if current trends continue unchecked [4], [9]. This environmental and economic reality has placed enormous pressure on the telecommunications industry and the research community to develop wireless systems that can deliver dramatically improved performance while simultaneously reducing their energy footprint.

In parallel with these energy concerns, the rapid expansion of the Internet of Things ecosystem and the increasing deployment of dense wireless networks in urban environments have exposed fundamental scalability limitations in conventional cellular architectures [2], [13]. Traditional cellular systems organise the coverage area into fixed geographic cells, each served by a single high-power base station. While this model has proven effective for several decades, it is inherently ill-suited to the demands of future networks for several reasons. First, users located near cell boundaries experience significant performance degradation due to the combined effects of increased distance-dependent path loss and strong interference from neighbouring base stations operating on the same frequency band, a problem that persists regardless of how much power the base station transmits [2]. Second, the fixed cell boundary model creates an uneven distribution of service quality across the network, where users close to the base station enjoy excellent performance while cell-edge users receive a fraction of that performance, undermining the goal of uniform and reliable connectivity [1], [2]. Third, as network densification increases the number of base stations deployed per unit area, the interference between neighbouring cells grows correspondingly, eroding the capacity gains that densification is intended to provide [13], [24].

These limitations have motivated a fundamental rethinking of wireless network architecture, moving away from the centralised, cell-centric paradigm towards a more distributed and cooperative model. Several candidate architectures have been proposed in the literature to address these challenges, including heterogeneous networks, coordinated multipoint transmission, and distributed antenna systems [14]. Among these, Cell-Free Massive Multiple-Input Multiple-Output has emerged as one of the most promising and technically compelling proposals, offering a principled and comprehensive solution to the interference, coverage uniformity, and energy efficiency challenges that limit conventional cellular systems [1], [2], [23]. The following section reviews the architecture and fundamental operating principles of CF-mMIMO in detail, explaining why it represents such a significant advancement over previous approaches and why it has attracted substantial research attention as a key enabling technology for 6G.

2.2 The Paradigm Shift From Cellular to Cell-Free Massive MIMO

The concept of Cell-Free Massive Multiple-Input Multiple-Output did not emerge in isolation but rather as the natural culmination of several decades of research into distributed antenna systems, cooperative communication, and massive MIMO technology. To fully appreciate why CF-mMIMO represents such a fundamental departure from conventional network design, it is necessary to first understand the trajectory of wireless network architecture evolution and the specific technical limitations that motivated this paradigm shift. This section traces that trajectory, beginning with the foundational principles of massive MIMO and progressing through the key architectural innovations that define the cell-free approach.

2.2.1 The Evolution From Massive MIMO to Cell-Free Architecture

The concept of massive MIMO, introduced in the seminal work of Marzetta [8], proposed equipping base stations with a very large number of antenna elements on the order of hundreds or thousands to simultaneously serve multiple users on the same time-frequency resource through spatial multiplexing. The fundamental insight behind massive MIMO is that as the number of antennas at the base station grows large, the channels between different users become increasingly orthogonal due to the law of large numbers, a phenomenon known as channel hardening [6], [8]. This orthogonality allows the base station to separate user signals through simple linear precoding and combining methods, dramatically reducing inter-user interference without requiring complex non-linear processing. Furthermore, as the number of antennas grows, the effects of uncorrelated noise and fast fading average out, leaving large-scale fading as the dominant channel characteristic and making system performance highly predictable and stable [6], [8].

The performance gains demonstrated by massive MIMO in terms of both spectral efficiency and energy efficiency were remarkable and well-documented in the literature [6], [7]. Björnson, Hoydis, and Sanguinetti [6] provided a comprehensive theoretical framework showing that massive MIMO systems could achieve orders of magnitude improvement in spectral efficiency compared to conventional multi-antenna systems, while simultaneously reducing the per-bit energy consumption significantly. Ngo, Larsson, and Marzetta [7] demonstrated through asymptotic analysis that as the number of base station antennas grows without bound, simple matched filtering becomes optimal and inter-user interference can be driven to zero, enabling near-perfect spatial multiplexing. These theoretical results established massive MIMO as one of the foundational technologies of 5G and set the stage for further architectural innovations.

However, despite these impressive performance gains, conventional massive MIMO systems retain the fundamental architectural constraint of centralising all antenna elements at a single base station location [1], [2]. This centralisation means that users who are geographically far from the base station, or who are in deep shadow fading due to physical obstructions, still experience poor channel conditions regardless of how many antennas the base station deploys. The performance gains of massive MIMO are therefore unevenly distributed across the coverage area, with users near the base station benefiting greatly while cell-edge users see limited improvement [2], [24]. This spatial non-uniformity in performance is a direct consequence of the centralised architecture and cannot be resolved simply by adding more antennas to a fixed location.

The recognition of this fundamental limitation led researchers to explore distributed architectures in which the antenna elements, rather than being co-located at a single site, are spread across the coverage area in the form of multiple geographically separated Access Points [1], [23]. This distributed approach draws on earlier work in coordinated multipoint transmission and distributed antenna systems, but extends these concepts significantly by combining them with the massive MIMO principle of deploying a very large total number of antenna elements and serving all users jointly and coherently [1], [2]. The result is the Cell-Free Massive MIMO architecture, formally proposed and analysed by Ngo et al. [2], which eliminates cell boundaries entirely and allows every user in the network to benefit from the proximity of at least some of the distributed APs regardless of their geographical location within the coverage area.

2.2.2 Architecture and Operating Principle of Cell-Free Massive MIMO

In a CF-mMIMO system, a large number M of single-antenna or multi-antenna Access Points are deployed across a geographical coverage area and connected to one or more Central Processing Units through fronthaul or backhaul links [1], [2]. Unlike conventional cellular systems where each base station independently serves only the users within its own cell, all APs in a CF-mMIMO network cooperate to jointly serve all K users within the coverage area simultaneously and on the same time-frequency resource [2], [23]. There are no cell boundaries, no handover procedures, and no concept of a user being associated with a specific AP. Instead, every user is effectively surrounded by multiple serving APs, ensuring that a strong and reliable signal can always be delivered regardless of the user's physical location or movement within the coverage area.

The operation of a CF-mMIMO system is typically organised around a Time Division Duplexing protocol, which exploits the reciprocity of the wireless channel to allow APs to estimate downlink channel coefficients from uplink pilot transmissions [1], [5]. During the uplink training phase, users transmit pilot sequences that are received by all APs within the network. Each AP uses these received pilot signals to estimate the channel coefficients between itself and each user, producing local CSI that is subsequently used to design the downlink precoding weights. During the downlink data transmission phase, each AP applies its locally computed precoding weights to the data signals and transmits them simultaneously to all users. The received signal at each user is then the superposition of transmissions from all serving APs, which, when properly phase-aligned through coherent precoding, combines constructively to produce a strong desired signal while the interference from unintended user signals is suppressed [1], [2].

A particularly important architectural feature of CF-mMIMO that distinguishes it from earlier distributed antenna system proposals is the user-centric serving model, where rather than all M APs serving all K users, each user is assigned a subset of the most favourable APs based on large-scale fading conditions [1], [23]. This user-centric approach, thoroughly analysed by Björnson and Sanguinetti [1] and further developed by Ngo et al. [23], provides a scalable alternative to the fully centralised model by limiting the amount of channel information and data that must be exchanged between APs and the CPU. In a fully centralised model, the volume of information exchanged scales with the product of M and K, which quickly becomes prohibitively large as the network scales up. The user-centric model restricts this exchange to only the most relevant AP-user pairs, making the architecture significantly more practical for real-world deployment while retaining most of the performance benefits of full cooperation [1], [23], [25].

The performance advantages of CF-mMIMO over both conventional cellular systems and small-cell networks have been rigorously demonstrated in the literature. Ngo et al. [2] conducted the first comprehensive comparison between CF-mMIMO and a dense small-cell network with the same total number of antennas and total transmit power, showing that CF-mMIMO achieves significantly higher spectral efficiency for cell-edge users, defined as the fifth percentile of the spectral efficiency cumulative distribution function, compared to the small-cell baseline. This improvement is attributed to the macro-diversity gain provided by the distributed AP deployment, which ensures that even users in deep shadow fading are served by multiple APs with independently fading channels, dramatically reducing the probability of a user experiencing simultaneous deep fading from all serving APs [2], [24]. Buzzi et al. [13] further demonstrated through a user-centric analysis that CF-mMIMO provides more uniform quality of service across the coverage area than conventional cellular systems, with lower variance in per-user spectral efficiency and significantly better performance for the most disadvantaged users in the network.

From an energy efficiency perspective, the distributed nature of CF-mMIMO also offers inherent advantages compared to centralised massive MIMO. Because each AP needs to transmit only to users in its vicinity rather than serving the entire cell, the individual transmit power requirements per AP are lower, reducing the power amplifier energy consumption at each site [4], [9]. Furthermore, the macro-diversity gain means that the same quality of service can be achieved at a lower total transmit power compared to a centralised system, since multiple APs contribute cooperatively to each user's signal rather than a single high-power base station bearing the entire transmission burden [2], [10]. These energy efficiency advantages, combined with the coverage uniformity and spectral efficiency gains, make CF-mMIMO a particularly compelling architecture for the sustainable and high-performance wireless networks envisioned for the 6G era.

2.3 Channel Estimation and the Reality of Imperfect CSI

The performance of any multi-antenna wireless communication system is fundamentally contingent upon the accuracy with which the system can acquire knowledge of the wireless channel between transmitters and receivers. This knowledge, referred to as Channel State Information, is essential for virtually every signal processing operation performed within the system, including precoding, combining, power allocation, and resource scheduling [1], [5]. In an ideal world, the system would have access to perfect and instantaneous CSI, meaning that the exact complex-valued channel coefficients between every AP and every user would be known without any error at all times. Under this idealised assumption, system designers can exploit the full spatial degrees of freedom offered by the large number of distributed APs in a CF-mMIMO network to achieve near-perfect interference cancellation and optimal resource allocation. However, in any practical wireless deployment, obtaining perfect CSI is fundamentally impossible due to the random and time-varying nature of the wireless channel, the presence of thermal noise in received signals, and the limited time and frequency resources available for channel training [1], [12].

The gap between the idealised assumption of perfect CSI and the practical reality of imperfect channel knowledge has profound consequences for system performance. When precoding and power allocation are based on inaccurate channel estimates, the intended signal alignment is imperfect, the interference cancellation is incomplete, and the resulting SINR experienced by users is substantially lower than what would be achievable under perfect CSI conditions [3], [12]. Understanding the sources and consequences of channel estimation imperfection is therefore essential for designing robust and practically relevant CF-mMIMO systems, and forms the foundation upon which the first research objective of this project is built.

2.3.1 Time Division Duplexing (TDD) and Channel Reciprocity

In CF-mMIMO systems, channel estimation is almost universally performed under a Time Division Duplexing protocol, which is the dominant duplexing strategy adopted in both the theoretical literature and practical system designs for large-scale multi-antenna networks [1], [6]. The fundamental principle underlying TDD-based channel estimation is the exploitation of channel reciprocity, which states that in a wireless communication system operating on the same carrier frequency in both directions, the channel experienced by a signal travelling from a transmitter to a receiver is identical to the channel experienced by a signal travelling in the reverse direction, provided that both transmissions occur within the same channel coherence interval [1], [8]. This reciprocity property is an important advantage of TDD over Frequency Division Duplexing systems, where uplink and downlink transmissions occur on different frequency bands and the channels in the two directions are generally independent, necessitating separate and resource-intensive channel estimation procedures for each direction.

In a TDD-based CF-mMIMO system, the channel coherence interval, denoted by τ_c, represents the duration within which the wireless channel can be assumed to remain approximately constant in both time and frequency. This interval is determined by the physical properties of the propagation environment, specifically the Doppler spread caused by user mobility and the delay spread caused by multipath propagation [6], [34]. During each coherence interval, the total available time-frequency resources are divided between the uplink training phase, where users transmit pilot sequences to enable channel estimation at the APs, and the downlink data transmission phase, where the APs use the estimated channel knowledge to transmit data to users [1], [5]. The proportion of the coherence interval devoted to pilot transmission directly determines the pre-log factor that scales the achievable spectral efficiency, creating an important trade-off between the resources invested in channel estimation and the resources available for actual data transmission. Specifically, if the pilot length is denoted by τ_p, then the pre-log factor applied to the spectral efficiency expression is given by $(1 - \tau_p/\tau_c)$, meaning that longer pilot sequences improve channel estimation quality but reduce the fraction of the coherence interval available for data transmission [1], [3].

By exploiting channel reciprocity, the APs in a CF-mMIMO system can estimate the downlink channel coefficients entirely from the uplink pilot transmissions, without requiring the users to estimate and feed back the downlink channel to the APs. This dramatically reduces the channel estimation overhead compared to what would be required in a Frequency Division Duplexing system, where the number of downlink pilot signals would need to scale with the total number of AP antennas rather than the number of users [6], [8]. This scalability advantage of TDD-based reciprocity exploitation is one of the key reasons why TDD is the preferred duplexing strategy for CF-mMIMO systems, particularly as the number of APs and antennas grows large.

2.3.2 Pilot Assignment Strategies and Their Impact on Estimation Quality

During the uplink training phase of each coherence interval, every user transmits a pilot sequence of length τ_p symbols. To enable the APs to distinguish between different users, it is desirable for each user to transmit a pilot sequence that is orthogonal to the sequences transmitted by all other users, meaning that the inner product between any two different pilot sequences is zero [5], [16]. When all users transmit mutually orthogonal pilot sequences, the channel estimate obtained at each AP for a particular user is free from contamination by the pilot signals of other users, and the estimation quality is limited only by the thermal noise at the AP receiver. Under these ideal conditions, standard estimation techniques such as the Minimum Mean Square Error estimator can produce highly accurate channel estimates that closely approximate the true channel coefficients [1], [5].

However, the number of mutually orthogonal pilot sequences of length τ_p that can be constructed is exactly equal to τ_p itself. This means that the system can support at most τ_p users with orthogonal pilot sequences within a single coherence interval. In practical CF-mMIMO deployments, the number of users K served within the network typically exceeds the pilot length τ_p, making it impossible to assign unique orthogonal pilot sequences to all users simultaneously [5], [12]. As a result, pilot sequences must be reused among multiple users, meaning that two or more users will transmit the same pilot sequence during the training phase. The pilot assignment problem, which determines which users share which pilot sequences, therefore becomes a critical design parameter that directly influences the quality of the channel estimates obtained throughout the network.

The simplest approach to pilot assignment is random pilot allocation, where each user is independently assigned a pilot sequence chosen uniformly at random from the available set of τ_p orthogonal sequences, without any consideration of the large-scale fading environment or the geographical proximity of users sharing the same pilot [16], [19]. While this approach is straightforward to implement and requires no knowledge of the network topology or channel statistics, it inevitably results in some users being assigned the same pilot as other users with strong channels towards the same APs, leading to severe pilot contamination for those users. Random pilot allocation therefore serves as a useful baseline for evaluating the performance of more sophisticated assignment strategies, but is generally suboptimal in terms of the channel estimation quality it achieves.

A significantly more effective approach is greedy pilot assignment, which iteratively assigns pilot sequences to users in a manner that minimises the total pilot contamination across the network [5], [9]. In a greedy assignment procedure, users are processed one at a time, and for each user, the algorithm evaluates all available pilot sequences and selects the one that results in the minimum additional pilot contamination, typically measured by the sum of the large-scale fading coefficients of all other users who are currently assigned the same pilot [5]. This process is repeated iteratively over multiple passes through all users, with each pass further refining the assignment based on the updated contamination levels. Because the greedy algorithm explicitly accounts for the large-scale fading environment when making assignment decisions, it tends to separate users with strong mutual interference onto different pilot sequences, significantly reducing the contamination experienced by the most vulnerable users in the network. Demir and Björnson [5] demonstrated that greedy pilot assignment, even in its simplest form, provides substantial improvements in channel estimation quality compared to random assignment, particularly for users in dense deployment scenarios where pilot reuse is most severe. Similarly, Dessie and Björnson [16] showed through simulation that adaptive pilot reuse strategies consistently outperform random assignment across a wide range of network configurations and user densities.

2.3.3 Pilot Contamination and Its Consequences

Pilot contamination is widely recognised in the massive MIMO and CF-mMIMO literature as one of the most fundamental and persistent performance-limiting phenomena in large-scale multi-antenna systems [1], [8]. When two or more users share the same pilot sequence, the signal received at an AP during the training phase is a superposition of the pilot signals from all users transmitting that sequence. As a result, the channel estimate computed by the AP for any one of these co-pilot users contains contributions from the channels of all other users sharing the same pilot, making it impossible for the AP to cleanly isolate the individual channel of the desired user [5], [17].

The consequences of pilot contamination extend far beyond a simple increase in channel estimation error. Because the contamination causes the channel estimates of co-pilot users to be correlated with each other, the precoding vectors designed based on these estimates will inadvertently direct a portion of the transmitted signal energy towards unintended users who share the same pilot sequence [1], [12]. This creates a form of coherent inter-user interference that is fundamentally different from the incoherent interference caused by users transmitting on different pilots, because coherent interference scales with the transmit power in the same way as the desired signal and therefore cannot be eliminated simply by increasing transmit power or deploying more AP antennas [8], [17]. Marzetta [8] demonstrated in the context of conventional massive MIMO that pilot contamination creates an interference floor that limits system performance even as the number of antennas grows to infinity, making it one of the fundamental bottlenecks of large-scale multi-antenna communication.

In CF-mMIMO systems, the impact of pilot contamination is somewhat mitigated compared to conventional cellular massive MIMO due to the distributed nature of the AP deployment and the resulting macro-diversity gain [1], [2]. Because a user is served by multiple geographically distributed APs rather than a single base station, the probability that all serving APs simultaneously experience strong contamination from the same interfering user is lower than in a centralised system where all antennas face the same interference environment. Nevertheless, pilot contamination remains a significant challenge in CF-mMIMO, particularly in dense user scenarios where the limited number of orthogonal pilot sequences forces many users to share pilots, and where the large-scale fading coefficients of co-pilot users may still be substantial at some of the serving APs [5], [12], [17].

2.3.4 Channel Estimation Techniques: From LS to MMSE

Given the importance of channel estimation quality for overall system performance, a variety of estimation techniques have been developed and studied in the literature, ranging from simple and computationally inexpensive methods to more sophisticated approaches that exploit statistical knowledge of the channel [1], [5]. Understanding the strengths and limitations of each technique is important for selecting the most appropriate estimator for a given deployment scenario and for correctly interpreting the performance differences observed between different system configurations.

The simplest channel estimation technique is the Least Squares estimator, which computes the channel estimate by finding the channel vector that minimises the squared difference between the received pilot signal and the signal that would have been received if that channel estimate were correct [34], [35]. The LS estimator requires no prior statistical knowledge of the channel, making it straightforward to implement in any scenario. However, because it does not exploit any information about the distribution of channel coefficients or the statistics of the interference, it is particularly vulnerable to pilot contamination and performs poorly in low signal-to-noise ratio environments or when pilot reuse is severe [1]. The LS estimate is also known to be an unbiased but high-variance estimator, meaning that while it correctly estimates the channel on average, individual estimates can deviate substantially from the true channel value, leading to significant precoding errors when used as the basis for interference suppression.

The Minimum Mean Square Error estimator represents a substantially more sophisticated approach that minimises the mean squared error between the estimated and true channel coefficients by incorporating prior statistical knowledge of the channel distribution [1], [5]. Unlike the LS estimator, which treats all received signals equally regardless of their statistical properties, the MMSE estimator weights the contribution of the received pilot signal according to the ratio of the desired channel power to the total received signal power, which includes both the desired pilot and the interfering pilots from co-pilot users. This weighting has the effect of suppressing the contribution of interference and noise relative to the desired signal component, producing estimates that are more accurate and more stable than those produced by the LS approach [1], [12]. Björnson and Sanguinetti [1] demonstrated that MMSE estimation, when combined with greedy pilot assignment, provides significantly higher channel estimation quality compared to LS with random pilot assignment, and that this improvement in estimation quality translates directly into measurable gains in downlink SINR, spectral efficiency, and energy efficiency. The MMSE estimator is therefore the preferred estimation technique in the CF-mMIMO literature and forms the basis for the proposed ZF precoding framework evaluated in this project.

2.4 Downlink Precoding and Uplink Combining Strategies

Once channel estimates have been obtained during the uplink training phase, the CF-mMIMO system must determine how to process and transmit the downlink data signals in a manner that maximises the received signal quality at each intended user while simultaneously minimising the interference experienced by all other users sharing the same time-frequency resources. This signal processing operation, performed at the transmitter side of the communication link, is known as downlink precoding or beamforming, and it represents one of the most critical and extensively studied components of multi-antenna wireless communication system design [1], [2], [6]. The choice of precoding strategy has a direct and profound impact on virtually every performance metric of interest in a CF-mMIMO system, including the received SINR, the achievable spectral efficiency per user, the robustness of the system to channel estimation errors, and ultimately the energy efficiency of the network as a whole.

In a CF-mMIMO system with M distributed APs jointly serving K users, the downlink precoding problem involves designing a set of precoding weight vectors, one for each AP-user pair, such that the superposition of transmissions from all APs produces a strong and coherent desired signal at each intended user while the interference towards all unintended users is kept as small as possible [1], [3]. This is a fundamentally challenging problem because the precoding weights must simultaneously satisfy multiple competing objectives like maximising signal strength for each user, minimising interference towards all other users, and respecting the per-AP transmit power constraints, all based on channel estimates that are inevitably imperfect due to pilot contamination and noise [3], [12]. Different precoding strategies address these competing objectives in different ways, making different trade-offs between computational complexity, interference suppression capability, and robustness to CSI imperfections.

This section reviews the three primary precoding strategies that are most relevant to this project, namely Matched Ratio combining, Minimum Mean Square Error precoding, and Zero-Forcing precoding, examining the theoretical basis, practical advantages, and known limitations of each approach as documented in the existing literature.

2.4.1 Matched Ratio Combining and Maximum Ratio Transmission

Matched Ratio combining, known in the downlink as Maximum Ratio Transmission, is the simplest and most straightforward linear precoding strategy available in multi-antenna communication systems [1], [2]. The fundamental principle of MR precoding is to align the phase of the transmitted signal from each AP with the estimated channel of the intended user, thereby ensuring that the signals arriving at the intended user from all serving APs add together constructively. In mathematical terms, the precoding weight applied by the m-th AP to the signal intended for the k-th user is simply set equal to the conjugate of the estimated channel coefficient between that AP and that user [1], [6]. This choice of precoding weight maximises the coherent combining gain at the intended user, producing the maximum possible received signal power at that user given the individual AP transmit power constraints.

The primary appeal of MR precoding lies in its extreme simplicity and its compatibility with fully distributed implementation. Because each AP computes its precoding weight using only its own locally estimated channel coefficient, without requiring any knowledge of the channels between other APs and other users, MR can be implemented in a completely distributed fashion without any need for coordination or information exchange between APs beyond what is already required for pilot transmission [1], [2]. This distributed implementability is a significant practical advantage in CF-mMIMO systems where the fronthaul links connecting APs to the CPU may have limited capacity, making it expensive or impractical to centralise all channel information for joint processing. For this reason, MR precoding is widely used as the baseline and lower-bound comparison method in the CF-mMIMO literature, representing the minimum level of performance that any practically deployable system should be able to achieve [2], [26].

However, the simplicity of MR precoding comes at a significant cost in terms of interference management. Because MR precoding weights are designed exclusively to maximise the signal power at the intended user, they provide no active mechanism for suppressing the interference that the transmitted signal causes at unintended users [1], [3]. In a system with a small number of users whose channel vectors are nearly orthogonal, this limitation may be acceptable because the naturally small inner products between different user channels mean that the interference from MR transmission is inherently low. However, as the number of simultaneously served users increases and the channels of different users become more correlated, the inter-user interference under MR precoding grows rapidly and eventually dominates the received signal quality, causing the achievable spectral efficiency per user to saturate and then decline [2], [3]. This interference saturation effect makes MR precoding increasingly inadequate for dense user scenarios, which are precisely the scenarios that future 6G networks must be designed to handle efficiently. Bashar et al. [26] showed through both theoretical analysis and simulation that MR combining in the uplink exhibits a significant performance ceiling in dense networks, confirming that more sophisticated interference suppression techniques are necessary for high-load deployments.

2.4.2 Minimum Mean Square Error Precoding

Minimum Mean Square Error precoding represents a substantially more sophisticated approach to the downlink signal processing problem in CF-mMIMO systems, offering improved performance over MR by explicitly accounting for both the desired signal and the interference in the precoding weight design [1], [33]. Unlike MR, which optimises the precoding weights purely to maximise the received signal power at the intended user, MMSE precoding optimises the weights to maximise the signal-to-interference-plus-noise ratio at the receiver, which requires simultaneous consideration of how the transmitted signal affects both the intended user and all unintended users [1], [6]. This joint optimisation makes MMSE precoding significantly more complex than MR but also significantly more effective at managing inter-user interference, particularly in scenarios where user channels are strongly correlated.

The MMSE precoder can be understood conceptually as finding the best possible balance between two competing objectives: maximising the desired signal component at the intended user and minimising the interference leaked towards all other users [1], [33]. In scenarios where these two objectives are in strong conflict, for example when the channel vectors of two users are nearly parallel meaning their signals naturally interfere with each other strongly, the MMSE precoder makes a principled trade-off that achieves a higher overall SINR than either pure signal maximisation or pure interference minimisation would achieve individually. This ability to navigate the trade-off between signal enhancement and interference suppression makes MMSE precoding one of the most theoretically sound and practically effective approaches available for multi-user downlink transmission [1], [6].

An important property of MMSE precoding that is particularly relevant in the context of this project is its partial array gain, which arises from the fact that MMSE precoding exploits only a portion of the available spatial degrees of freedom for signal enhancement, reserving the remainder for interference suppression [1]. In a CF-mMIMO system with M_s serving APs per user, MMSE precoding achieves a coherent combining gain that scales with the square root of M_s rather than with M_s itself, reflecting this fundamental trade-off between signal gain and interference suppression [1], [3]. While this partial array gain is lower than what can be achieved by a precoder that devotes all available spatial resources to signal enhancement, the corresponding reduction in inter-user interference typically results in a higher net SINR and therefore higher spectral efficiency in multi-user scenarios. Björnson and Sanguinetti [1] demonstrated through both theoretical analysis and numerical simulation that MMSE precoding with centralised implementation achieves near-optimal spectral efficiency in CF-mMIMO systems, establishing it as the performance benchmark against which other precoding strategies are evaluated.

2.4.3 Zero-Forcing Precoding and Its Role Under Imperfect CSI

Zero-Forcing precoding is a well-established and widely studied linear precoding technique that achieves interference suppression by exploiting the mathematical structure of the estimated channel matrix to design precoding vectors that are orthogonal to the channels of all unintended users [3], [28], [32]. The fundamental principle of ZF precoding is that by transmitting each user's signal in a direction that lies in the null space of all other users' channels, the signal intended for one user produces zero interference at all other users, completely eliminating inter-user interference in the ideal case of perfect CSI [3], [6]. This complete interference elimination is achieved by computing the pseudo-inverse of the estimated channel matrix and using the resulting matrix as the precoding weight matrix, a computation that requires knowledge of the channel estimates of all simultaneously served users and involves matrix inversion operations at the central processing unit [3], [32].

The key advantage of ZF precoding over both MR and MMSE is its ability to completely eliminate inter-user interference under perfect CSI conditions, rather than merely reducing it [3], [28]. This makes ZF particularly attractive in scenarios with a large number of simultaneously served users, where inter-user interference is the dominant performance-limiting factor and any residual interference significantly degrades the achievable SINR. In the asymptotic regime where the number of AP antennas is much larger than the number of users, ZF precoding is known to achieve near-optimal performance with relatively simple implementation, making it a practical and powerful choice for large-scale multi-antenna systems [6], [29]. Furthermore, compared to MMSE precoding which requires knowledge of the noise variance and channel estimation error statistics in addition to the channel estimates themselves, ZF precoding requires only the channel estimates, making it somewhat simpler to implement in practice and less sensitive to inaccuracies in the noise and interference statistics [3].

In the context of CF-mMIMO systems operating under imperfect CSI, ZF precoding exhibits a particularly important characteristic that motivates its selection as the proposed method in this project. Because ZF precoding is designed to null the interference towards all unintended users based on the estimated channel matrix, the effectiveness of this interference nulling is directly determined by the accuracy of the channel estimates used in the computation [3], [12]. When channel estimates are highly accurate, as achieved through MMSE estimation combined with greedy pilot assignment, ZF precoding can suppress the vast majority of inter-user interference and deliver SINR values that approach the perfect CSI upper bound. Conversely, when channel estimates are poor due to severe pilot contamination or inadequate estimation techniques, the interference nulling becomes inaccurate and residual inter-user interference remains, degrading the SINR below what would be achievable with better CSI [3], [15]. This strong sensitivity of ZF to CSI quality creates a clear and direct linkage between the channel estimation improvements achieved in Objective 1 of this project and the precoding performance gains targeted in Objective 2, reinforcing the integrated and sequential nature of the research framework.

Mai et al. [3] provided a comprehensive analysis of ZF precoding in CF-mMIMO systems with imperfect CSI, demonstrating that even under realistic pilot contamination conditions, ZF precoding achieves significantly higher spectral efficiency and energy efficiency compared to MR precoding, and that the performance gap between ZF and MR widens as the number of users increases. This finding is particularly significant because it demonstrates that the computational overhead associated with ZF precoding is well justified by the substantial performance gains it delivers, especially in the dense and heavily loaded network scenarios that are most relevant for future 6G deployments. Nayebi et al. [28] similarly demonstrated that full-pilot ZF precoding, which uses all available pilot dimensions to construct the interference nulling matrix, provides near-optimal performance in CF-mMIMO systems and significantly outperforms MR across a wide range of operating conditions. Attarifar et al. [29] proposed a modified conjugate beamforming approach that approaches ZF performance with lower complexity, further validating the performance advantages of interference-nulling precoding strategies over simple matched filtering in multi-user CF-mMIMO deployments.

From a practical implementation perspective, ZF precoding in CF-mMIMO systems requires a degree of centralisation that is higher than what is needed for MR precoding, since computing the pseudo-inverse of the channel matrix requires access to the channel estimates of all users at a central processing unit [1], [3]. However, this additional coordination requirement is manageable within the fronthaul infrastructure assumed in this project, and is a well-accepted trade-off given the significant SINR and spectral efficiency gains that ZF delivers. Moreover, recent advances in scalable CF-mMIMO architectures have demonstrated that the centralised processing required for ZF can be efficiently implemented within a hierarchical network structure where local processing at APs handles simple tasks and centralised processing at the CPU handles the more complex joint operations [1], [23]. This makes ZF precoding not only theoretically attractive but also practically feasible for real-world CF-mMIMO deployments, supporting its selection as the proposed precoding strategy in this project.

2.5 Theoretical Foundations of Energy Efficiency (EE)

Energy efficiency has undergone a fundamental transformation in its role within wireless communication system design over the past two decades. In the early generations of mobile networks, energy consumption was treated primarily as an operational cost to be managed rather than a core performance metric to be optimised alongside data rate and coverage. System designers focused almost exclusively on maximising spectral efficiency, measured in bits per second per Hertz, under the implicit assumption that the energy required to achieve higher data rates was an acceptable and manageable overhead. This design philosophy produced networks that were highly capable in terms of raw throughput but increasingly unsustainable in terms of energy consumption, as the exponential growth in mobile data traffic drove a corresponding increase in the energy demands of wireless infrastructure worldwide [4], [9], [10].

The recognition that this trajectory was environmentally and economically unsustainable led to a gradual but decisive shift in research priorities, with energy efficiency emerging as a first-class design objective in the wireless communications literature beginning in the mid-2000s and accelerating significantly through the 5G development era [4], [10]. This shift was motivated not only by environmental concerns about carbon emissions and electricity consumption but also by the economic reality that energy costs represent a substantial and growing fraction of the total operating expenditure of mobile network operators. For large-scale network operators managing thousands of base stations, even modest improvements in energy efficiency per site can translate into substantial cost savings at the network level, creating strong commercial incentives that reinforce the environmental motivation for energy-aware system design [9], [10]. In the context of CF-mMIMO systems, where the deployment of large numbers of distributed APs across a wide coverage area introduces multiple distinct sources of power consumption beyond the radiated transmit power, the importance of rigorous and comprehensive energy efficiency analysis is even more pronounced than in conventional cellular systems [3], [4].

2.5.1 Definition and Physical Interpretation of Energy Efficiency

Energy efficiency in wireless communication systems is formally defined as the ratio of the total amount of information successfully delivered to users per unit time to the total power consumed by the system during that delivery [4], [9], [10]. This definition captures the essential physical meaning of energy efficiency as a measure of how productively the system converts electrical energy into useful information transfer. The mathematical expression for energy efficiency in a CF-mMIMO system serving K users is given by:

$$\eta_{\mathrm{EE}} = \frac{B \cdot \sum_{k=1}^{K} S\,E_k}{P_{\mathrm{total}}}$$

where B denotes the system bandwidth in Hertz, SE_k represents the downlink spectral efficiency of the k-th user in bits per second per Hertz, and P_{total} represents the total power consumed by the network in Watts [3], [9]. The result is expressed in units of bits per Joule, or equivalently megabits per Joule when scaled by a factor of one million for practical convenience, and represents the number of bits of information that the system can deliver to its users for every Joule of electrical energy consumed [4], [10].

The spectral efficiency of user$k$ is derived from the Shannon capacity formula, modified by a pre-log factor that accounts for the overhead associated with uplink pilot transmission during each coherence interval. Specifically, the spectral efficiency is expressed as:

$$SE_k = \left(1 - \frac{\tau_p}{\tau_c}\right) \log_2 \left(1 + \mathrm{SINR}_k\right)$$

where τ_p denotes the pilot sequence length, τ_c denotes the total coherence block length in symbols, and $SINR_k$ represents the effective signal-to-interference-plus-noise ratio experienced by user k during downlink transmission [1], [3]. The pre-log factor $\left(1-\frac{\tau_p}{\tau_c}\right)$ reflects the fact that a fraction of the coherence interval is consumed by pilot transmission and is therefore unavailable for data transmission. For the system parameters adopted in this project, where τ_p =5 and τ_c =200, this pre-log factor evaluates to 0.975, meaning that approximately 97.5 percent of the coherence interval remains available for downlink data transmission after accounting for pilot overhead. This relatively high fraction is one of the advantages of TDD-based CF-mMIMO systems, where the pilot overhead scales with the number of users rather than the number of AP antennas [1], [6].

The physical interpretation of the energy efficiency expression in equation (2.1) reveals an important and non-obvious insight about the relationship between energy efficiency and transmit power. One might intuitively expect that increasing transmit power always improves energy efficiency, since higher transmit power produces higher SINR which in turn produces higher spectral efficiency through equation (2.2) and therefore more bits delivered per unit time. However, this intuition is incorrect because it ignores the denominator of equation (2.1), which includes the transmit power itself as a direct component of the total power consumption [3], [4]. As transmit power increases, both the numerator and denominator of the energy efficiency expression increase, but they do so at fundamentally different rates. The spectral efficiency in the numerator increases logarithmically with transmit power, following the diminishing returns behaviour of the Shannon capacity formula in equation (2.2), while the transmit power in the denominator increases linearly [3], [10]. This fundamental asymmetry means that at low transmit power levels, the logarithmic gains in spectral efficiency outpace the linear increase in power consumption, causing energy efficiency to rise. However, beyond a certain optimal transmit power level, the marginal gains in spectral efficiency become smaller than the marginal increase in power consumption, causing energy efficiency to decline [4], [9]. This behaviour produces the characteristic bell-shaped curve of energy efficiency as a function of transmit power that has been widely documented in the literature and that motivates the power allocation optimisation carried out in Objective 3 of this project.

2.5.2 Spectral Efficiency and Energy Efficiency: A Fundamental Trade-Off

The relationship between spectral efficiency and energy efficiency is one of the most important and extensively studied topics in the wireless communications literature, and understanding this relationship is essential for making informed design decisions in CF-mMIMO systems [4], [7], [10]. As discussed in the previous subsection, spectral efficiency and energy efficiency are related through a non-linear and non-monotonic function that creates an inherent tension between the two objectives. A system operating at the point of maximum spectral efficiency is necessarily transmitting at high power and therefore operating well beyond the peak of the energy efficiency curve, meaning that it is consuming a disproportionately large amount of energy for the incremental spectral efficiency gains it achieves relative to a lower power operating point [4], [10]. Conversely, a system operating at the point of maximum energy efficiency is transmitting at a moderate power level that may be significantly below the maximum achievable spectral efficiency, accepting some reduction in data rate in exchange for a substantially improved energy efficiency [3], [4].

This fundamental trade-off between spectral efficiency and energy efficiency has been formalised in the literature through the concept of the spectral efficiency versus energy efficiency trade-off region, which characterises all achievable combinations of the two metrics for a given system configuration and identifies the Pareto-optimal boundary beyond which it is impossible to improve one metric without degrading the other [4], [10]. Papazafeiropoulos et al. [4] provided a detailed analysis of this trade-off in the specific context of CF-mMIMO systems, demonstrating that the Pareto boundary exhibits a concave shape that reflects the diminishing marginal returns of increasing spectral efficiency at the cost of energy efficiency. Their results showed that the optimal operating point on this boundary depends heavily on the specific power consumption model adopted, particularly the relative magnitudes of the fixed circuit power and the transmit power, with higher fixed power costs shifting the optimal point towards higher spectral efficiency to better amortise the fixed overhead across a larger data throughput. Ngo, Larsson, and Marzetta [7] similarly demonstrated that in the asymptotic massive MIMO regime, the energy efficiency achievable with simple linear precoding strategies grows without bound as the number of antennas increases while the transmit power is simultaneously reduced, establishing a fundamental connection between spatial multiplexing gain and energy efficiency improvement that carries over directly to the CF-mMIMO context.

The practical implication of this trade-off for system design is that the transmit power level and power allocation strategy cannot be chosen independently of the energy efficiency objective. A system that simply maximises transmit power to achieve the highest possible data rate will inevitably operate far from its energy efficiency optimum, wasting a significant fraction of its energy budget on marginal spectral efficiency gains. Intelligent power allocation, which identifies and operates at or near the peak of the energy efficiency curve for each user and each set of channel conditions, is therefore essential for achieving practically meaningful energy efficiency in CF-mMIMO deployments [3], [4], [11]. This insight directly motivates the power allocation optimisation framework developed in Objective 3 of this project, where the goal is explicitly to find the power allocation coefficients that maximise equation (2.1) rather than simply maximising the total data rate or minimising the total transmit power.

2.5.3 Achievable Rate and the Use-and-Then-Forget Bound

A critical technical challenge in the theoretical analysis of CF-mMIMO systems is the derivation of tractable and accurate expressions for the achievable spectral efficiency under imperfect CSI conditions. In principle, the achievable rate of a communication system is determined by the mutual information between the transmitted and received signals, which depends on the joint statistical distribution of the channel, the channel estimation error, and the interference from other users [1], [6]. Under perfect CSI, this mutual information can be evaluated exactly using the standard Shannon capacity formula. However, under imperfect CSI, the mutual information expression becomes significantly more complex because the receiver cannot perfectly decode the transmitted signal and must treat the channel estimation error as an additional source of uncertainty [1], [12].

To address this challenge, the CF-mMIMO literature has widely adopted a bounding technique known as the Use-and-Then-Forget bound, which provides a closed-form lower bound on the achievable spectral efficiency that remains valid under arbitrary linear precoding strategies and arbitrary channel estimation methods [1], [3]. The key idea behind the UatF bound is to treat the channel estimation error as an additional interference term at the receiver rather than as a form of channel uncertainty that requires more sophisticated decoding. By doing so, the achievable rate can be lower bounded by an expression that depends only on the statistics of the channel estimates and the estimation errors, rather than on the instantaneous realisations of these quantities, making the bound analytically tractable and amenable to closed-form evaluation [1], [6]. The resulting effective SINR expression under the UatF bound takes the following general form:

$$\mathrm{SINR}_k = \frac{p_d\left(\sum_{m\in\mathcal{M}_k}\gamma_{mk}\right)^2}{p_d\sum_{j=1}^{K}\sum_{m\in\mathcal{M}_k}(\beta_{mk}-\gamma_{mk})+1}$$

where p_{dl} denotes the normalised downlink transmit power, γ_{mk} denotes the channel estimate quality coefficient between the m-th AP and the k-th user, β_{mk} denotes the large-scale fading coefficient between the m-th AP and the k-th user, and $\mathcal{M}_k$ denotes the set of APs serving user k [1], [3]. The numerator of this expression represents the coherent desired signal power resulting from the constructive combination of transmissions from all serving APs, while the denominator captures the combined effect of channel estimation errors, residual inter-user interference, and thermal noise. The difference ($\beta_{mk}-\gamma_{mk}$) in the denominator represents the channel estimation error variance, which is larger when pilot contamination is severe and smaller when channel estimation quality is high, directly linking estimation quality to the achievable SINR [1], [5].

The UatF bound is particularly useful for the analysis carried out in this project because it enables direct comparison between the four estimation and precoding strategies evaluated in Objectives 1 and 2, providing a common analytical framework within which the performance differences between LS, MR, MMSE, and ZF can be rigorously characterised. Björnson and Sanguinetti [1] demonstrated that the UatF bound is reasonably tight for the system parameters of interest in CF-mMIMO research, meaning that it closely approximates the true achievable rate and provides reliable guidance for system design and optimisation decisions. The adoption of this bound in this project therefore ensures that the spectral efficiency and energy efficiency results obtained are theoretically grounded and practically meaningful, rather than being based on overly optimistic assumptions that would not hold in realistic deployments.

2.6 Comprehensive Power Consumption Models

A realistic and rigorous evaluation of energy efficiency in CF-mMIMO systems cannot be conducted without adopting a comprehensive and accurate model for the total power consumed by the network. This point is often underappreciated in studies that focus exclusively on the transmit power radiated from AP antennas, treating it as the sole component of network energy consumption. In reality, the total power consumed by a CF-mMIMO deployment is the sum of multiple distinct components, each arising from a different physical process or hardware subsystem within the network, and the transmit power represents only one of these components [3], [4], [9]. When the non-transmit power components are large relative to the transmit power, which is frequently the case in CF-mMIMO systems due to the large number of distributed APs each consuming fixed circuit power, the relationship between transmit power and energy efficiency can be dramatically different from what a simplified single-component power model would predict [4], [10].

The importance of adopting a multi-component power consumption model in CF-mMIMO research has been clearly demonstrated in the literature. Björnson and Sanguinetti [11] showed that the choice of power model has a profound impact on the optimal power allocation strategy, with more realistic models that include fixed circuit power and backhaul overhead consistently producing lower optimal transmit power levels and higher energy efficiency at those optimal points compared to simplified models that consider only transmit power. Nguyen et al. [9] similarly demonstrated that the inclusion of wireless backhaul power in the total power model significantly affects the energy efficiency trade-off in CF-mMIMO systems, particularly when the number of APs is large and the backhaul links carry substantial data traffic. Zappone et al. [10] provided a comprehensive analysis showing that energy efficiency optimisation based on a simplified power model produces power allocation strategies that are substantially suboptimal when evaluated under a more realistic model, reinforcing the need for accuracy in power consumption modelling as a prerequisite for meaningful energy efficiency optimisation. These findings collectively establish that the power consumption model is not merely a technical detail but a fundamental determinant of the system design conclusions drawn from any energy efficiency analysis.

The total power consumption model adopted in this project follows the framework established in the literature for large-scale CF-mMIMO systems and decomposes the total network power into four primary components, each capturing a distinct and physically meaningful source of energy expenditure [3], [9], [10]. The complete expression for the total power consumption is given by:

$$P_{total} = P_{fix} + M \cdot P_{cm} + \alpha_{PA} \cdot p_{dl} \cdot \sum_{k=1}^{K} \eta_k + M \cdot P_{0,bh} + B \cdot \left(\sum_{k=1}^{K} SE_k \right) \cdot P_{bt} \cdot M$$

where P_{fix} denotes the fixed network power consumption independent of the number of APs, M denotes the total number of Access Points, P_{cm} denotes the computation power consumed per AP, α_{PA} denotes the power amplifier inefficiency factor, p_{dl} denotes the maximum downlink transmit power, η_k denotes the power allocation coefficient for user k, $P_{0,bh}$ denotes the static backhaul power per AP, B denotes the system bandwidth, SE_k denotes the spectral efficiency of user k, and P_{bt} denotes the backhaul traffic power coefficient [3], [9]. Each of these components is discussed in detail in the following subsections.

2.6.1 Transmit Power and Power Amplifier Efficiency

The first and most immediately obvious component of the total network power consumption is the power actually radiated from the AP antennas during downlink data transmission. In a CF-mMIMO system with K users and power allocation coefficients η_k assigned to each user, the total radiated transmit power summed across all users is given by $p_{dl} \cdot \sum_{k=1}^{K} \eta_k$ where p_{dl} represents the maximum available downlink transmit power and $\eta_k \in [0,1]$ represents the fraction of that maximum power allocated to user k [3], [11]. However, the power consumed by the power amplifier hardware at each AP is not equal to the radiated transmit power but rather higher than it by a factor determined by the power amplifier efficiency, which accounts for the conversion losses that occur as electrical energy from the power supply is converted into radio frequency energy radiated from the antenna [3], [9].

This relationship is captured by the power amplifier inefficiency factor α_{PA}, which represents the reciprocal of the power amplifier drain efficiency. If the power amplifier drain efficiency is denoted by ζ, then $\alpha_{PA}=1/\zeta$, and the actual power consumed by the amplifier hardware to produce a radiated power of P_{rad} is equal to $\alpha_{PA} \cdot P_{rad}$ [3], [6]. For practical power amplifier hardware used in low-power distributed AP deployments, the drain efficiency is typically well below unity, meaning that α_{PA} is substantially greater than one and the actual consumed power is significantly higher than the radiated power. In this project, a value of α_{PA}=8 is adopted, consistent with the values used in related CF-mMIMO energy efficiency studies [3], [9], reflecting the practical reality that power amplifier inefficiency is a major contributor to the total energy consumption of wireless networks and cannot be neglected in a realistic power model. This inefficiency factor means that for every Watt of power radiated from the AP antennas, eight Watts of electrical power are consumed by the amplifier hardware, making power amplifier efficiency a critical design parameter for energy-efficient CF-mMIMO systems.

2.6.2 Fixed Circuit Power and Computation Power Per AP

The second major component of the total power consumption in a CF-mMIMO system is the fixed circuit power consumed by the hardware components of each AP regardless of whether data is being transmitted or received. This fixed power arises from the continuous operation of essential hardware subsystems within each AP, including oscillators that generate the carrier frequency, analogue-to-digital and digital-to-analogue converters that interface between the analogue radio frequency domain and the digital baseband processing domain, filters that suppress out-of-band interference, and baseband processing units that perform channel estimation, precoding weight computation, and signal modulation and demodulation [6], [9], [10]. These hardware components consume power continuously as long as the AP is operational, irrespective of the instantaneous data traffic load, making their power consumption fundamentally different in character from the transmit power which scales directly with the amount of data being transmitted.

In the power model adopted in this project, the fixed circuit power consumed by each AP is denoted by P_{cm} and the total fixed circuit power across all M APs is therefore $M{\cdot}P_{cm}$[3], [9]. Additionally, a network-level fixed power term P_{fix} captures the power consumed by centralised infrastructure components such as the CPU, cooling systems, and network management hardware that are shared across the entire deployment and do not scale with the number of APs [3]. The combined fixed power contribution $P_{fix}+M{\cdot}P_{cm}$ represents a power floor that is always present regardless of the transmit power level or the traffic load, and its magnitude relative to the transmit power has a profound impact on the shape of the energy efficiency curve and the location of the optimal operating point [4], [10].

An important consequence of the linear scaling of fixed circuit power with the number of APs is that deploying more APs does not unconditionally improve energy efficiency, even though it improves coverage and spectral efficiency [4], [9]. As the number of APs increases, the total fixed circuit power $M \cdot P_{cm}$ grows linearly, increasing the denominator of the energy efficiency expression and eventually causing the energy efficiency to saturate and decline even as the spectral efficiency continues to improve. This behaviour creates a fundamental trade-off between the coverage and spectral efficiency benefits of deploying more APs and the energy efficiency costs of their fixed power consumption, which has been extensively studied in the CF-mMIMO energy efficiency literature [4], [9], [10]. Papazafeiropoulos et al. [4] demonstrated through numerical optimisation that there exists an optimal number of APs that maximises the energy efficiency for a given set of system parameters and deployment conditions, beyond which additional APs consume more fixed power than the spectral efficiency gains they provide can justify from an energy perspective.

2.6.3 Backhaul Power Consumption

The third component of the total power consumption model captures the energy consumed by the fronthaul and backhaul links that connect the distributed APs to the central processing unit in a CF-mMIMO system. These links are essential for the cooperative operation of the distributed AP network, as they must carry channel state information from the APs to the CPU for joint precoding computation and carry the precoded data signals from the CPU to the APs for transmission to users [9], [10]. The power consumed by these links has two distinct components: a static component that represents the baseline power required to maintain the link in an operational state regardless of the data traffic it carries, and a dynamic component that scales with the amount of data traffic carried by the link [9].

In the power model adopted in this project, the static backhaul power per AP is denoted by $P_{0,bh}$, contributing a total static backhaul power of $M \cdot P_{0,bh}$

across all APs [3], [9]. The dynamic backhaul power scales with the product of the total network spectral efficiency, the system bandwidth, and a backhaul traffic power coefficient P_{bt} that characterises the energy consumed per unit of data traffic carried by the backhaul links [3]. The total dynamic backhaul power across all APs is therefore given by

$B \cdot \left(\sum_{k=1}^{K} SE_k\right) \cdot P_{bt} \cdot M$, reflecting the fact that higher spectral efficiency requires more data to be transported through the backhaul infrastructure and therefore consumes more backhaul power [9], [10]. Nguyen et al. [9] demonstrated that backhaul power is a particularly significant component of the total power consumption in CF-mMIMO systems with wireless backhaul links, where the backhaul channel quality and capacity can impose significant constraints on both the achievable spectral efficiency and the energy efficiency of the network. Even in systems with wired backhaul, the dynamic component of backhaul power can be non-negligible at high traffic loads, making its inclusion in the power model important for accurately characterising the energy efficiency behaviour of the system across a wide range of operating conditions.

2.6.4 Impact of the Power Model on Energy Efficiency Optimization

The multi-component power consumption model described in the preceding subsections has important and non-trivial implications for the energy efficiency optimisation problem addressed in Objective 3 of this project. When the total power consumption includes fixed components that are independent of the power allocation coefficients η_k, the energy efficiency optimisation problem becomes a fractional programming problem in which the objective function is the ratio of a concave function of η_k in the numerator to an affine function of η_k in the denominator [3], [31]. This fractional structure means that the problem is not convex in general, making it significantly more challenging to solve than a standard convex optimisation problem and requiring specialised algorithmic approaches that can handle the non-convexity while still converging to a high-quality solution [3], [4].

The fixed power components in the denominator of the energy efficiency expression also mean that the optimal power allocation is not simply the solution to a pure transmit power minimisation problem. Because the fixed power $P_{fix}+M \cdot P_{cm}+M \cdot P_{0,bh}$ is present regardless of the transmit power level, reducing the transmit power to zero does not drive the total power consumption to zero and therefore does not maximise the energy efficiency. Instead, the optimal power allocation must balance the competing effects of increasing spectral efficiency through higher transmit power against the diminishing marginal returns of that increase relative to the fixed power floor [3], [11]. This balance point, which corresponds to the peak of the bell-shaped energy efficiency curve, depends on the specific values of all power model parameters and on the channel conditions experienced by each user, making it a genuinely multi-dimensional optimisation problem that requires an iterative algorithmic solution rather than a simple closed-form expression [3], [4], [10]. The Proximal Gradient algorithm adopted in this project is specifically designed to address this type of non-convex fractional optimisation problem efficiently, as discussed in detail in the following section.

2.7 Optimization Theory From Second-Order to First-Order Methods

The problem of maximising energy efficiency in CF-mMIMO systems through intelligent power allocation is fundamentally an optimisation problem, and the choice of optimisation algorithm has profound implications for both the quality of the solution obtained and the computational resources required to obtain it. To fully appreciate why the Proximal Gradient algorithm is selected as the proposed method in this project, it is necessary to first understand the mathematical structure of the energy efficiency maximisation problem, the limitations of conventional high-complexity optimisation approaches that have been widely used in the existing literature, and the theoretical properties of first-order methods that make them particularly well-suited for large-scale CF-mMIMO deployments [3], [30], [31]. This section reviews the relevant optimisation theory in a progressive manner, beginning with the mathematical characterisation of the problem and moving through second-order methods to first-order methods, establishing a clear and well-motivated rationale for the algorithmic choices made in this project.

2.7.1 Mathematical Structure of the Energy Efficiency Maximisation Problem

The energy efficiency maximisation problem in a CF-mMIMO system can be formally stated as the problem of finding the power allocation vector $\boldsymbol{\eta}=[\eta_1,\eta_2,\ldots,\eta_K]^T$ that maximises the energy efficiency expression defined in equation (2.1), subject to constraints that ensure the solution is physically realisable [3], [11]. The most important of these constraints is the per-AP power constraint, which limits the total transmit power radiated by each individual AP to a maximum permissible level determined by the hardware specifications and regulatory requirements of the deployment [3], [32]. This per-AP constraint is mathematically expressed as:

$$\sum_{k=1}^{K} \eta_k \cdot p_{dl} \cdot \gamma_{mk} \leq P_{\max}, \quad \forall m = 1,2, \ldots, M$$

where $P_{\max}$ denotes the maximum allowable transmit power per AP, η_k denotes the power allocation coefficient for user k,p_{dl} denotes the maximum downlink transmit power, and γ_{mk}denotes the channel estimate quality coefficient between the m-th AP and the k-th user [3], [32]. Additionally, the power allocation coefficients must be non-negative, since negative power allocation is physically meaningless, giving the additional constraint $\eta_k \geq 0$ for all k [3], [11].

The resulting constrained optimisation problem has a fractional objective function, where the numerator is a concave function of the power allocation vector and the denominator is an affine function of the same vector, as established in Section 2.6 [3], [31]. This fractional structure makes the problem a instance of fractional programming, a well-studied class of optimisation problems that arises frequently in wireless communications research whenever the objective involves a ratio of system performance metrics to system resource consumption [3], [33]. While fractional programming problems can sometimes be solved efficiently by transforming them into equivalent parametric subproblems through the Dinkelbach method or related techniques, the resulting subproblems are themselves non-convex in general due to the coupling between the power allocation coefficients through the SINR expressions, making the overall problem significantly more challenging than a standard convex programme [3], [4], [31]. This non-convexity means that conventional convex optimisation tools cannot be directly applied to find the globally optimal solution, and the field has therefore developed a variety of approximate and iterative solution strategies that trade off solution optimality for computational tractability.

2.7.2 Limitations of Second-Order Optimisation Methods

The most widely used class of optimisation algorithms in the wireless communications literature for problems of this type are second-order methods, which utilise both the first-order gradient and the second-order Hessian matrix of the objective function to guide the search for the optimal solution [30], [33]. The most prominent examples of second-order methods applied to energy efficiency optimisation in CF-mMIMO systems are Successive

Convex Approximation and Second-Order Cone Programming, both of which have been extensively studied and applied in the existing literature [3], [4].

Successive Convex Approximation addresses the non-convexity of the energy efficiency maximisation problem by replacing the original non-convex objective function with a sequence of convex approximations, each of which is easier to optimise than the original problem [33]. In each iteration of the SCA procedure, the non-convex terms in the objective function are approximated by convex surrogates constructed around the current iterate, typically using first-order Taylor expansions or other local linearisation techniques. The resulting convex subproblem is then solved to optimality, and the solution is used as the starting point for the next iteration, with the process repeating until convergence. While SCA is theoretically guaranteed to converge to a stationary point of the original non-convex problem under mild conditions, the computational cost per iteration is dominated by the cost of solving the convex subproblem, which typically involves matrix operations whose complexity scales as $O(M^3)$ due to the need to compute and factor the Hessian matrix of the objective function [3], [4]. For a CF-mMIMO system with $M=100$ APs, this translates to approximately one million floating-point operations per iteration, a computational burden that becomes increasingly impractical as the system scales up.

Second-Order Cone Programming takes a different but related approach, reformulating the energy efficiency maximisation problem as a conic optimisation problem that can be solved efficiently using interior-point methods [4], [33]. SOCP solvers are highly reliable and can find the globally optimal solution of the reformulated problem within a guaranteed number of iterations, making them attractive from a theoretical standpoint. However, the per-iteration complexity of interior-point methods for SOCP scales polynomially with the problem dimensions, typically as $O(M^{3.5})$ or worse, making them even more computationally demanding than SCA for large-scale systems [3], [4]. Papazafeiropoulos et al. [4] applied SOCP-based optimisation to the energy efficiency problem in CF-mMIMO and demonstrated impressive performance results, but acknowledged that the computational complexity of the approach limits its practical applicability to systems with a relatively small number of APs and users. Zappone et al. [10] similarly noted that second-order methods, while theoretically rigorous, face fundamental scalability challenges in large CF-mMIMO deployments where the number of optimisation variables and constraints grows with the number of APs and users.

Beyond their high per-iteration computational complexity, second-order methods also suffer from a practical implementation challenge in CF-mMIMO systems arising from the distributed nature of the AP deployment. Computing the Hessian matrix of the energy efficiency objective function requires knowledge of the channel estimates and SINR values of all users at all APs simultaneously, necessitating a centralised computation architecture in which all relevant information is gathered at the CPU before each iteration of the optimisation algorithm [3], [11]. While this centralised computation is feasible in principle, it places significant demands on the fronthaul links connecting APs to the CPU and introduces latency that may be unacceptable in dynamic channel environments where the power allocation must be updated frequently to track changing channel conditions. These practical limitations, in addition to the fundamental computational complexity concerns, provide a strong motivation for exploring first-order optimisation methods that can achieve comparable solution quality with substantially lower per-iteration computational cost and more favourable implementation characteristics [3], [31].

2.7.3 The Proximal Gradient Algorithm: Principles and Advantages

First-order optimisation methods are a broad class of iterative algorithms that rely exclusively on gradient information, meaning the first-order derivative of the objective function, to guide the search for the optimal solution, without computing or utilising the second-order Hessian matrix that is required by second-order methods [30], [31]. The most fundamental first-order method is the gradient ascent algorithm, which updates the current iterate by taking a step in the direction of the gradient of the objective function, moving towards regions of higher objective value. While simple gradient ascent is effective for unconstrained smooth optimisation problems, it cannot directly handle the constraints present in the energy efficiency maximisation problem, such as the per-AP power constraint and the non-negativity constraint on the power allocation coefficients [31].

The Proximal Gradient algorithm extends simple gradient ascent to handle constrained optimisation problems by augmenting each gradient step with a proximal projection step that maps the updated iterate back into the feasible set defined by the problem constraints [31]. The iterative update rule of the PG algorithm for the energy efficiency maximisation problem takes the following form:

$$\eta^{(t+1)} = \mathcal{P}_{\mathcal{C}}\left(\eta^{(t)} + \mu^{(t)} \cdot \nabla_{\eta}\eta_{EE}(\eta^{(t)})\right)$$

where $\eta^{(t)}$ denotes the power allocation vector at iteration t,$\mu^{(t)}$ denotes the step size at iteration t,$\nabla_{\eta}\eta_{EE}$ denotes the gradient of the energy efficiency objective function with respect to the power allocation vector, and P_C denotes the projection operator onto the feasible set C defined by the per-AP power constraints and non-negativity constraints [3], [31]. The gradient $\nabla_{\eta}\eta_{EE}$ can be computed analytically by differentiating the energy efficiency expression with respect to each power allocation coefficient η_k, yielding a closed-form expression that can be evaluated efficiently given the current channel estimates and SINR values [3]. The projection operator P_C maps any infeasible point back to the nearest feasible point in the constraint set, ensuring that the power allocation coefficients after each update satisfy all system constraints [31].

The computational complexity of each iteration of the PG algorithm is dominated by the cost of evaluating the gradient vector, which requires computing the SINR values and their derivatives with respect to the power allocation coefficients for all K users. This computation scales as $O(M{\cdot}K)$ per iteration, which is dramatically lower than the $O(M^3)$ or $O(M^{3.5})$ per-iteration complexity of second-order methods, making the PG algorithm significantly more scalable for large CF-mMIMO deployments [3], [31]. For a system with M=100M = 100 M=100 APs and K=10 users, the PG algorithm requires on the order of one thousand operations per iteration compared to one million for SCA, representing a three-order-of-magnitude reduction in per-iteration computational cost. Parikh and Boyd [31] provided a comprehensive theoretical analysis of the convergence properties of proximal gradient algorithms, establishing that under appropriate conditions on the step size and the smoothness of the objective function, the PG algorithm converges to a stationary point of the optimisation problem at a rate of $O(1/t)$, where t denotes the iteration number. While this convergence rate is slower than the superlinear convergence rate of second-order methods, the much lower cost per iteration means that the PG algorithm can achieve a solution of comparable quality to second-order methods in a fraction of the total computational time [3], [31].

An important practical consideration in the implementation of the PG algorithm is the choice of step size $\mu^{(t)}$, which controls the magnitude of the gradient step taken at each iteration and has a significant impact on both the convergence speed and the stability of the algorithm [3], [31]. A step size that is too large can cause the algorithm to overshoot the optimal solution and oscillate or diverge, while a step size that is too small results in slow convergence that requires a large number of iterations to reach an acceptable solution quality. To address this challenge, this project adopts an adaptive step size strategy in which the step size is increased by a multiplicative factor when the energy efficiency objective improves from one iteration to the next, indicating that the current step size is conservatively small and larger steps can be safely taken, and halved when the objective decreases, indicating that the step size is too large and causing overshooting [3]. This adaptive strategy allows the algorithm to automatically tune its step size to the local curvature of the objective function, achieving faster convergence in flat regions of the objective landscape and greater stability near the optimal solution. Mai et al. [3] demonstrated that this adaptive step size mechanism significantly improves the practical convergence behaviour of the PG algorithm for energy efficiency maximisation in CF-mMIMO systems, consistently achieving high-quality solutions within fifty iterations across a wide range of system configurations and channel conditions.

2.7.4 Comparison of Power Allocation Strategies in the Literature

Beyond the algorithmic approaches discussed above, the literature on power allocation in CF-mMIMO systems has explored a range of strategies spanning from the simplest possible baseline to sophisticated optimisation-based methods, providing important context for interpreting the results obtained in this project. The simplest power allocation strategy is Equal Power Allocation, which assigns the same power coefficient to all users regardless of their individual channel conditions or their contribution to the overall network energy efficiency [2], [11]. While EPA is trivially simple to implement and requires no knowledge of the channel estimates or SINR values, it is clearly suboptimal in heterogeneous channel environments where different users experience vastly different large-scale fading conditions towards the serving APs. Users with strong channels receive more power than necessary to achieve their target SINR, while users with weak channels may receive insufficient power, resulting in a power allocation that is globally inefficient from an energy perspective [11].

Random Power Allocation represents a slightly more sophisticated baseline in which the power coefficients are drawn independently from a uniform random distribution and the resulting energy efficiency is averaged over multiple random realisations [3]. RandP provides a statistical reference that captures the expected performance of an unintelligent and unoptimised power allocation strategy, serving as a useful intermediate benchmark between the deterministic simplicity of EPA and the optimised performance of PG. By comparing the PG algorithm against both EPA and RandP, this project is able to clearly quantify the performance gain attributable to intelligent power allocation optimisation and to demonstrate that this gain is not simply the result of any deviation from equal power allocation but specifically the result of the directed optimisation performed by the PG algorithm [3], [11].

More sophisticated approaches in the literature include water-filling power allocation, which maximises the total spectral efficiency under a sum power constraint by allocating more power to users with stronger channels, and max-min fairness power allocation, which maximises the minimum spectral efficiency across all users to ensure that the most disadvantaged user achieves an acceptable service quality [2], [11], [32]. While these approaches offer interesting performance trade-offs, they are not directly aligned with the energy efficiency maximisation objective of this project and are therefore not implemented as primary comparison methods. Yu and Lan [32] provided a comprehensive analysis of power allocation strategies under per-antenna power constraints, demonstrating that the structure of the constraint set significantly influences the optimal power allocation and that per-AP constraints, as adopted in this project, produce different optimal solutions compared to sum power constraints commonly assumed in the theoretical literature. Shi et al. [33] similarly demonstrated through a weighted MMSE framework that iterative power allocation optimisation consistently outperforms static allocation strategies across a wide range of multi-user MIMO scenarios, providing further theoretical support for the iterative PG approach adopted in this project.

2.8 Summary of Literature and Research Gaps

The preceding sections of this chapter have presented a comprehensive and thematically organised review of the existing literature across six interconnected areas that are directly relevant to this research: the evolution of next-generation wireless networks, the architecture and operating principles of Cell-Free Massive MIMO, channel estimation under imperfect CSI and pilot contamination, downlink precoding strategies, the theoretical foundations of energy efficiency, comprehensive power consumption modelling, and optimisation theory for energy efficiency maximisation. Having established this broad and detailed understanding of the state of the art, this section synthesises the key insights drawn from the reviewed literature, identifies the specific research gaps that remain unaddressed, and explains precisely how the work carried out in this project contributes to filling those gaps.

2.8.1 Key Insights From the Literature

Several important and recurring themes emerge from the literature reviewed in this chapter that collectively motivate and contextualise the research carried out in this project.

The first and perhaps most fundamental insight is that CF-mMIMO represents a genuinely transformative architectural departure from conventional cellular systems, offering substantial and well-documented performance advantages in terms of coverage uniformity, macro-diversity gain, and achievable spectral efficiency [1], [2], [24]. The pioneering work of Ngo et al. [2] established through rigorous theoretical analysis and simulation that CF-mMIMO consistently outperforms both conventional cellular massive MIMO and dense small-cell networks in terms of the spectral efficiency experienced by the most disadvantaged users in the network, making it a uniquely compelling candidate for future 6G deployments where uniform quality of service is a primary design objective. Subsequent work by Björnson and Sanguinetti [1] further demonstrated that with appropriate signal processing, specifically MMSE estimation and centralised precoding, CF-mMIMO can achieve near-optimal spectral efficiency while maintaining a scalable and practically feasible implementation architecture. These foundational results establish CF-mMIMO as a mature and theoretically well-understood technology whose practical realisation requires addressing the challenges of channel estimation imperfection, interference management, and energy efficiency optimisation that are the focus of this project.

The second key insight from the literature is that pilot contamination is a persistent and fundamental challenge in CF-mMIMO systems that cannot be resolved through simple means such as increasing transmit power or deploying more AP antennas, but instead requires intelligent pilot assignment strategies that actively manage the interference caused by pilot reuse [5], [8], [17]. The work of Demir and Björnson [5] demonstrated that greedy pilot assignment provides substantial improvements in channel estimation quality compared to random assignment, translating directly into measurable gains in SINR and spectral efficiency. Al Ayidh et al. [17] similarly showed that matching-based pilot assignment strategies consistently reduce pilot contamination across a wide range of network configurations, reinforcing the importance of intelligent pilot management as a prerequisite for high-performance CF-mMIMO operation. These findings establish that the choice of pilot assignment strategy is not a minor implementation detail but a critical design decision with far-reaching consequences for overall system performance.

The third key insight is that Zero-Forcing precoding, despite its requirement for centralised computation and its sensitivity to CSI quality, consistently achieves higher spectral efficiency and energy efficiency than simpler precoding strategies such as Matched Ratio combining, particularly in dense user scenarios where inter-user interference is the dominant performance-limiting factor [3], [28], [29]. Mai et al. [3] provided compelling evidence that ZF precoding, when combined with appropriate power allocation, achieves energy efficiency levels that are significantly higher than those achievable with MR precoding under the same system conditions, establishing ZF as the preferred precoding strategy for energy-efficient CF-mMIMO deployments. The work of Nayebi et al. [28] further demonstrated that full-pilot ZF achieves near-optimal performance across a wide range of operating conditions, validating its selection as the proposed precoding method in this project.

The fourth key insight concerns the critical importance of adopting realistic and comprehensive power consumption models in energy efficiency analysis. Multiple studies have demonstrated that the conclusions drawn from energy efficiency optimisation are highly sensitive to the power model adopted, with simplified models that ignore fixed circuit power, backhaul overhead, and power amplifier inefficiency consistently producing overly optimistic energy efficiency estimates and suboptimal power allocation strategies when evaluated under more realistic conditions [4], [9], [10]. This finding underscores the necessity of the multi-component power model adopted in this project, which explicitly accounts for all significant sources of power consumption and ensures that the energy efficiency results obtained are practically meaningful and not artefacts of unrealistic modelling assumptions.

The fifth and final key insight is that first-order optimisation methods, specifically the Proximal Gradient algorithm, offer a compelling alternative to high-complexity second-order methods for energy efficiency maximisation in large-scale CF-mMIMO systems, providing solution quality that is comparable to second-order methods while requiring dramatically lower per-iteration computational resources [3], [31]. The theoretical analysis of Parikh and Boyd [31] established the convergence properties of proximal gradient algorithms in a general optimisation framework, while the specific application to CF-mMIMO energy efficiency maximisation by Mai et al. [3] demonstrated the practical effectiveness of the PG approach in achieving high energy efficiency with a computationally tractable iterative procedure. These results collectively justify the selection of the PG algorithm as the proposed power allocation method in this project.

2.8.2 Identified Research Gaps

Despite the substantial body of literature reviewed in this chapter, several important research gaps remain that limit the practical applicability and comprehensiveness of existing work on energy-efficient CF-mMIMO systems. This project is specifically designed to address these gaps in a unified and integrated manner.

The first and most significant research gap concerns the lack of a unified simulation framework that jointly addresses all three of the interconnected challenges studied in this project, namely pilot contamination and channel estimation quality, inter-user interference suppression through advanced precoding, and energy efficiency maximisation through intelligent power allocation, within a single coherent system model [1], [3]. The overwhelming majority of existing studies address only one or at most two of these challenges simultaneously, treating the others as fixed background assumptions rather than active research variables. For example, many studies on precoding strategy comparison assume perfect CSI or a fixed channel estimation method, without investigating how the choice of estimation and pilot assignment strategy affects the precoding performance results. Similarly, many studies on power allocation optimisation assume a fixed precoding strategy and evaluate only the power allocation dimension, without demonstrating how the choice of precoding method affects the achievable energy efficiency gains from optimised power allocation. This fragmented approach makes it difficult for system designers to understand the cumulative and compounding effects of improvements across all three dimensions simultaneously, which is precisely the integrated understanding that this project provides.

The second research gap concerns the limited comparative evaluation of different channel estimation and pilot assignment strategies within the same simulation framework and under identical system conditions. While individual studies have investigated specific estimation methods or pilot assignment strategies in isolation, comprehensive side-by-side comparisons of multiple methods spanning from the simplest baseline to the most sophisticated approach, evaluated across multiple performance dimensions including SINR, spectral efficiency, and energy efficiency, are relatively rare in the literature [1], [5]. This project addresses this gap by implementing and comparing four distinct combinations of estimation method and pilot assignment strategy within a single unified framework, providing a clear and unambiguous performance hierarchy that directly guides the selection of the most appropriate approach for practical CF-mMIMO deployments.

The third research gap is the absence of accessible and interactive simulation tools that make the performance evaluation of CF-mMIMO systems available to researchers and practitioners who may not have the expertise to implement complex MATLAB simulations from scratch. While the theoretical literature on CF-mMIMO is extensive and well-developed, the practical barrier to entry for researchers wishing to explore system behaviour under different parameter settings remains relatively high [1], [6]. The interactive MATLAB GUI developed in this project directly addresses this gap by providing a user-friendly platform that integrates all three research objectives into a single accessible tool, enabling straightforward exploration of CF-mMIMO performance across a wide range of system configurations without requiring detailed knowledge of the underlying implementation.

2.8.3 Summary Table of Key Literature

The following table provides a structured summary of the most directly relevant works reviewed in this chapter, highlighting the key contribution, primary strength, and main limitation or research gap associated with each study. This table serves as a concise reference that contextualises the research carried out in this project relative to the existing state of the art.

Author (Year)	Key Contribution	Primary Strength	Weaknesses / Research Gap

Author (Year)	**Key Contribution**	**Primary Strength**	**Weaknesses / Research Gap**
Björnson and Sanguinetti (2020) [1]	Scalable CF-mMIMO framework with MMSE processing and centralised implementation	Establishes near-optimal spectral efficiency benchmark with scalable architecture	Does not jointly optimise pilot assignment, precoding, and power allocation in a unified framework
Ngo et al. (2017) [2]	First comprehensive performance comparison between Cell-Free and small-cell architectures.	Rigorously demonstrates macro-diversity and cell-edge performance advantages of CF-mMIMO	Adopts simplified power model ignoring fixed circuit power and backhaul energy consumption.
Mai et al. (2022) [3]	Proximal Gradient method for energy efficiency maximisation with ZF precoding	Low per-iteration complexity scalable to large deployments with practical convergence behaviour	Evaluates power allocation in isolation without systematic comparison of pilot assignment strategies
Papazafeiropoulos et al. (2022) [4]	Energy efficiency optimisation using realistic multi-component power consumption model	Identifies optimal operating point considering spectral efficiency and energy efficiency trade-off	Relies on high-complexity SOCP solvers that limit practical scalability to large AP deployments
Demir & Björnson	Joint pilot assignment	Demonstrates	Joint optimisation

Author (Year)	Key Contribution	Primary Strength	Weaknesses / Research Gap
(2021) [5]	and power control algorithm for CF-mMIMO systems	substantial SINR improvement through greedy pilot assignment over random allocation	problem complexity limits applicability to moderate-scale deployments
Björnson et al. (2017) [6]	Unified theoretical framework for spectral, energy, and hardware efficiency in massive MIMO	Provides foundational mathematical tools widely adopted across massive MIMO research	Broad theoretical scope does not address specific first-order power allocation optimisation
Ngo, Larsson & Marzetta (2013) [7]	Energy and spectral efficiency analysis of very large multiuser MIMO systems	Establishes fundamental asymptotic bounds proving energy efficiency advantages of large antenna arrays	Assumes perfect CSI and does not consider distributed AP deployment or pilot contamination effects
Marzetta (2010) [8]	Introduction of non-cooperative cellular wireless with unlimited base station antennas.	Seminal work establishing massive MIMO concept and demonstrating pilot contamination as fundamental limit	Centralised single-site architecture cannot address cell-edge performance degradation
Nguyen et al. (2019) [9]	Energy efficiency analysis in CF-	Demonstrates significant impact of	Does not evaluate first-order

Author (Year)	Key Contribution	Primary Strength	Weaknesses / Research Gap
	mMIMO with wireless backhaul power modelling	backhaul power on total energy consumption and optimal operating point	optimisation methods or compare multiple precoding strategies
Zappone et al. (2021) [10]	Energy-efficient communications in CF-mMIMO with limited backhaul capacity constraints	Shows that backhaul capacity limitations significantly affect achievable energy efficiency	High complexity of proposed solution limits real-time applicability in dynamic channel environments
Björnson & Sanguinetti (2020) [11]	Unified power control framework for CF-mMIMO establishing optimal allocation structure	Demonstrates that power model choice fundamentally determines optimal allocation strategy	Does not implement or evaluate first-order iterative methods for practical power optimisation
Mohammadzadeh et al. (2022) [12]	Joint pilot and data power optimisation under imperfect CSI conditions	Addresses coupled optimisation of pilot and data power simultaneously improving estimation and rate	High complexity of joint optimisation restricts practical deployment to small-scale systems
Buzzi et al. (2020) [13	User-centric resource allocation comparison between CF-mMIMO	Confirms coverage uniformity advantages of user-centric CF-	Does not address energy efficiency optimisation or

Author (Year)	Key Contribution	Primary Strength	Weaknesses / Research Gap
	and cellular networks	mMIMO over cellular approaches	power allocation under imperfect CSI
Zhang et al. (2020) [14]	Survey of prospective multi-antenna technologies for beyond 5G and 6G systems	Provides comprehensive overview of candidate technologies and their relative performance merits	Survey scope limits depth of analysis for specific CF-mMIMO energy efficiency challenges
Ngo et al. (2020) [23]	Scalable cell-free massive MIMO with user-centric serving model	Demonstrates practical scalability of user-centric CF-mMIMO without compromising performance	Does not evaluate energy efficiency under imperfect CSI with first-order power optimisation
Parikh & Boyd (2014) [31]	Comprehensive theoretical analysis of proximal gradient algorithms and convergence properties	Establishes rigorous convergence guarantees for PG methods applicable to wireless optimisation problems	General optimisation framework requires problem-specific adaptation for CF-mMIMO energy efficiency
Yu & Lan (2007) [32]	Transmitter optimisation under per-antenna power constraints in multi-antenna downlink	Demonstrates importance of per-antenna constraints in producing practically relevant power	Focuses on spectral efficiency maximisation rather than energy efficiency as primary

Author (Year)	Key Contribution	Primary Strength	Weaknesses / Research Gap
		allocation	objective

2.8.4 Positioning of This Research

The research gaps identified in Section 2.8.2 and the limitations summarised in Table 2.1 collectively define a clear and well-motivated research space that this project occupies. Specifically, this project contributes a unified, integrated, and practically motivated simulation framework that addresses all three interconnected challenges of channel estimation quality, inter-user interference suppression, and energy efficiency maximisation simultaneously within a single coherent system model. Unlike the fragmented approach that characterises much of the existing literature, where each challenge is studied in relative isolation, this project demonstrates how improvements at each stage compound and propagate through the entire system pipeline, producing a final energy efficiency result that genuinely reflects the cumulative benefit of addressing all three challenges together.

Furthermore, by implementing and comparing four distinct estimation and pilot assignment strategies within the same framework, this project provides the kind of comprehensive and directly comparable performance evaluation that is missing from existing studies. The interactive MATLAB GUI developed as part of the project adds a practical contribution that goes beyond the mathematical analysis, making the simulation framework accessible and useful to a broader audience of researchers and practitioners. Taken together, these contributions position this project as a practically relevant and academically rigorous addition to the growing body of literature on energy-efficient CF-mMIMO systems, directly addressing the gaps identified in the reviewed literature and providing insights that are both theoretically grounded and practically actionable for the design of future 6G wireless networks.

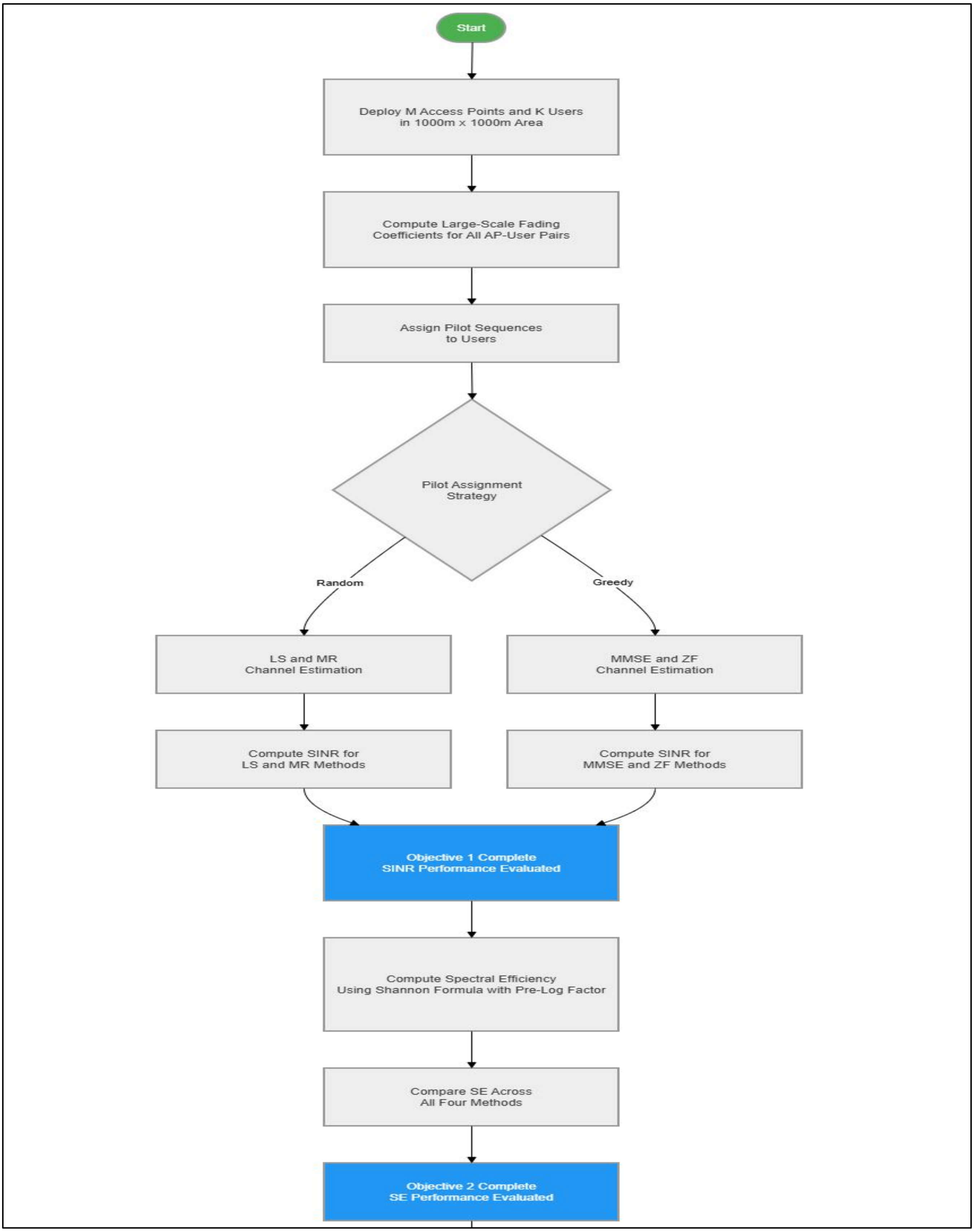
CHAPTER 3
System Methodology

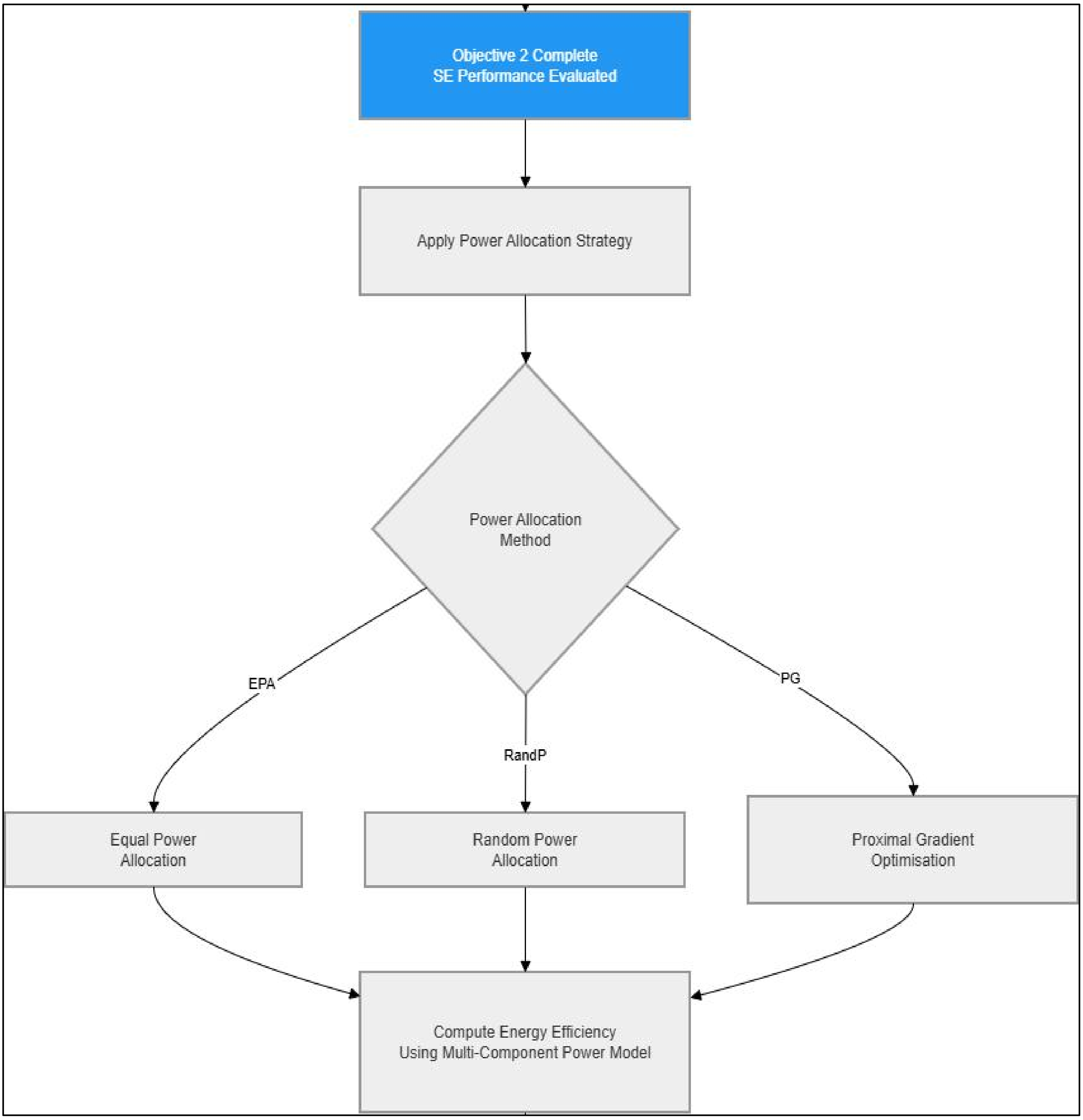
3.1 Overview

This chapter presents the complete methodology adopted in this research to design, implement, and evaluate an integrated simulation framework for energy-efficient Cell-Free Massive Multiple-Input Multiple-Output systems operating under imperfect Channel State Information. The methodology is structured as a progressive and sequential research framework in which the outputs of each stage directly feed into and enhance the performance of the subsequent stage, reflecting the deeply interconnected nature of the three research objectives addressed in this project. Rather than treating channel estimation, precoding, and power allocation as independent research problems to be solved in isolation, this project adopts a unified system perspective in which all three components are modelled, implemented, and evaluated within a single coherent simulation environment, ensuring that the interactions and dependencies between them are fully captured and accurately reflected in the results.

The methodology is organised into three primary stages corresponding to the three research objectives of the project. The first stage, corresponding to Objective 1, focuses on the accurate estimation of the wireless channel between each Access Point and each user under realistic pilot contamination conditions, implementing and comparing four distinct combinations of pilot assignment strategy and channel estimation method to establish a clear performance hierarchy in terms of the resulting Signal-to-Interference-plus-Noise Ratio achieved by each approach. The second stage, corresponding to Objective 2, extends this analysis into the domain of spectral efficiency, evaluating how the quality of the channel estimates obtained in the first stage translates into differences in achievable downlink data rates under the four precoding strategies considered. The third and final stage, corresponding to Objective 3, builds upon the spectral efficiency results of the second stage to address the problem of energy efficiency maximisation through intelligent power allocation, implementing and comparing three power allocation strategies under both Zero-Forcing and Minimum Mean Square Error precoding frameworks to identify the combination that achieves the highest energy efficiency under realistic operating conditions.

All three stages of the methodology are implemented in MATLAB R2022b, which provides a powerful and flexible environment for the matrix-intensive computations, iterative algorithm design, statistical analysis through Monte Carlo simulation, and graphical result visualisation required throughout the project. The simulation framework is further complemented by an interactive Graphical User Interface that integrates all three stages into a single accessible platform, allowing real-time exploration of system performance under different parameter settings without requiring direct interaction with the underlying MATLAB code. The remainder of this chapter is organised as follows. Section 3.2 presents the system model, including the network architecture, channel model, system parameters, and Time Division Duplexing protocol adopted throughout the project. Section 3.3 presents the proposed research framework in detail, with dedicated subsections for each of the three objectives. Section 3.4 provides a concluding summary that links all three stages together and establishes the foundation for the results and discussion presented in Chapter 4.


Start
Deploy M Access Points and K Users
in 1000m x 1000m Area
Compute Large-Scale Fading
Coefficients for All AP-User Pairs
Assign Pilot Sequences
to Users
Pilot Assignment
Strategy
Random
Greedy
LS and MR
Channel Estimation
MMSE and ZF
Channel Estimation
Compute SINR for
LS and MR Methods
Compute SINR for
MMSE and ZF Methods
Objective 1 Complete
SINR Performance Evaluated
Compute Spectral Efficiency
Using Shannon Formula with Pre-Log Factor
Compare SE Across
All Four Methods
Objective 2 Complete
SE Performance Evaluated


Objective 2 Complete
SE Performance Evaluated
Apply Power Allocation Strategy
Power Allocation
Method
EPA
RandP
PG
Equal Power
Allocation
Random Power
Allocation
Proximal Gradient
Optimisation
Compute Energy Efficiency
Using Multi-Component Power Model

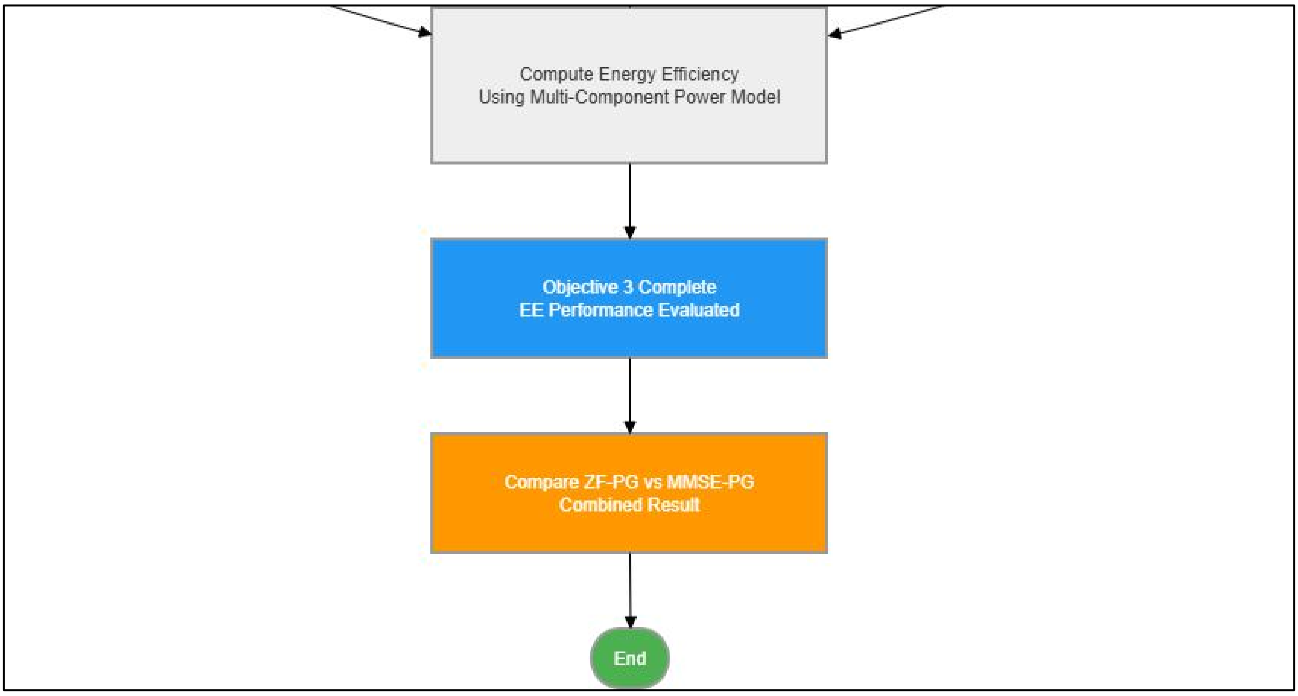


Figure 3.1 Shows the Overall System Framework.

Figure 3.1 illustrates the flowchart of overall system framework of the proposed CF-mMIMO simulation pipeline adopted in this project. The framework begins with the deployment of M Access Points and K users within the simulation area, followed by the computation of large-scale fading coefficients for all AP-user pairs. The pipeline then branches according to the pilot assignment strategy, where random pilot assignment feeds into LS and MR channel estimation while greedy pilot assignment feeds into MMSE and ZF channel estimation. The resulting channel estimates are used to compute the SINR for all four methods, completing Objective 1. The SINR values then feed directly into the spectral efficiency computation using the Shannon capacity formula with the appropriate pre-log factor, completing Objective 2. Finally, the spectral efficiency results are used as the basis for energy efficiency evaluation under three power allocation strategies, namely Equal Power Allocation, Random Power Allocation, and the proposed Proximal Gradient optimisation, completing Objective 3. The framework concludes with a head-to-head comparison between ZF-PG and MMSE-PG as the combined result of all three objectives working together.

3.2 System Model

This section presents the complete system model adopted throughout this research, establishing the foundational assumptions, network architecture, channel propagation characteristics, simulation parameters, and protocol structure that underpin all three research objectives. The system model is designed to reflect realistic operating conditions for a large-scale CF-mMIMO deployment while maintaining sufficient mathematical tractability to enable rigorous performance analysis and optimisation. All components of the system model described in this section remain fixed and consistent across all simulation experiments conducted in this project, ensuring that performance comparisons between different estimation methods, precoding strategies, and power allocation approaches are fair, unambiguous, and directly attributable to the specific design choices being evaluated rather than to differences in the simulation environment.

3.2.1 Network Architecture

The network considered in this project consists of M Access Points and K single-antenna users deployed within a square geographical area of dimensions D × D metres, where D = 1000 metres. Both the APs and the users are distributed independently and uniformly at random within this area at the beginning of each Monte Carlo simulation realisation, reflecting the unpredictable and heterogeneous nature of real-world wireless deployments where the positions of both infrastructure elements and user devices cannot be predetermined or controlled with precision [1], [2]. The random deployment strategy ensures that the simulation results are statistically representative of a wide variety of network topologies rather than being specific to any single fixed arrangement of nodes, and that the performance metrics obtained after averaging over many Monte Carlo realisations reflect the typical behaviour of the system across a broad range of practical deployment scenarios.

Each AP is assumed to be equipped with a single antenna, consistent with the low-cost and low-complexity distributed AP deployment model that is one of the defining characteristics of the CF-mMIMO architecture [1], [23]. In this model, the intelligence and processing capability of the network are concentrated at the Central Processing Unit rather than at individual APs, allowing each AP to remain a simple and inexpensive radio unit that performs only basic functions such as pilot reception, channel estimation, and signal transmission under the direction of the CPU. This design philosophy is motivated by the practical observation that deploying a large number of simple single-antenna APs is generally more cost-effective and easier to maintain than deploying a smaller number of complex multi-antenna APs, while still achieving the macro-diversity and spatial multiplexing gains that are the primary performance advantages of the CF-mMIMO architecture [1], [2], [24].

A minimum separation distance of 10 metres is enforced between any AP and any user to prevent unrealistically large channel gains that would arise if an AP and a user were placed at the same or nearly the same geographical location. This minimum distance constraint is a standard modelling assumption in the CF-mMIMO literature and ensures that the large-scale fading coefficients computed for all AP-user pairs remain within a physically realistic range [1], [3].

Rather than requiring all M APs to serve all K users simultaneously, which would impose prohibitive fronthaul and computational requirements in large-scale deployments, this project adopts a user-centric serving model in which each user is assigned a subset of M_s = 15 APs to serve it [1], [23]. The serving AP subset for each user is determined by sorting all M APs in descending order of their large-scale fading coefficient towards that user and selecting the top M_s APs with the strongest large-scale fading connections. This selection criterion ensures that each user is served exclusively by the APs that are best positioned to contribute meaningfully to its received signal quality, avoiding the wasteful allocation of fronthaul and processing resources to APs that are geographically distant from the user and can therefore contribute only negligibly to its received signal power. The binary AP-user association matrix D is defined such that D(m,k) = 1 if AP m serves user k and D(m,k) = 0 otherwise, providing a compact mathematical representation of the user-centric serving structure that is used throughout the subsequent signal processing stages [1], [3].

3.2.2 Channel Model

The wireless channel between the *m*-th AP and the *k*-th user is modelled using a standard composite channel model that decomposes the total channel coefficient into two multiplicative components representing large-scale fading and small-scale fading respectively [6], [34]. This decomposition is mathematically expressed as:

$$g_{mk} = \sqrt{\beta_{mk}} \cdot h_{mk}$$

where g_{mk} denotes the composite channel coefficient between AP mm m and user k, β_{mk} denotes the large-scale fading coefficient capturing the effects of path loss and shadow fading, and h_{mk} denotes the small-scale fading coefficient capturing the effects of multipath propagation [1], [6]. The two components vary on fundamentally different timescales, with the large-scale fading coefficient remaining approximately constant over many coherence intervals and the small-scale fading coefficient varying independently from one coherence interval to the next. This separation of timescales is important for the design of the channel estimation and precoding strategies, as it means that the large-scale fading statistics can be reliably estimated and used for long-term decisions such as pilot assignment and AP-user association, while the small-scale fading must be estimated freshly within each coherence interval using pilot sequences [1], [5].

Large-Scale Fading Model

The large-scale fading coefficient β_{mk} captures the combined effect of distance-dependent path loss and shadow fading between AP mm m and user *k*. In this project, a single-slope path loss model is adopted, which is widely used in the CF-mMIMO literature for its simplicity and its reasonable accuracy in modelling outdoor propagation environments [1], [3]. The large-scale fading coefficient in linear scale is given by:

$$\beta_{mk} = 10^{\frac{PL_{mk}+z_{mk}}{10}} \cdot \frac{1}{\sigma_n^2}$$

where PL_{mk} denotes the path loss in decibels between AP *m* and user *k*, zmkz_{mk} zmk denotes the shadow fading component in decibels, and σ_n^2 denotes the noise variance used to normalise the coefficient [1], [3]. The path loss in decibels is computed according to the single-slope model:

$$PL_{mk} =- 30.5 - 36.7 \times \log_{10} (d_{mk})$$

where d_{mk} denotes the Euclidean distance in metres between AP m and user k, subject to the minimum separation constraint of $d_{mk}\geq 10$ metres [1], [3]. The coefficients in this path loss model, specifically the intercept value of -30.5 dB and the slope of 36.7 dB per decade of distance, are adopted from empirical propagation measurements and are consistent with the values used in foundational CF-mMIMO studies [1], [2].

Shadow fading is modelled as a spatially independent log-normal random variable, represented in the logarithmic domain as:

$$z_{mk} \sim \mathcal{N}(0, \sigma_{sf}^2)$$

where σ_{sf}= 8 dB denotes the shadow fading standard deviation [6], [34]. The log-normal shadow fading model reflects the random attenuation caused by buildings, walls, vegetation, and other physical obstructions in the propagation environment that are not captured by the deterministic distance-dependent path loss model. A standard deviation of 8 dB is consistent with empirical measurements in typical outdoor urban and suburban environments and is widely adopted in the CF-mMIMO simulation literature [1], [6].

Small-Scale Fading Model

The small-scale fading coefficient h_{mk} is modelled as an independent and identically distributed complex Gaussian random variable:

$$h_{mk} \sim \mathcal{CN}(0,1)$$

This corresponds to a standard Rayleigh fading channel model, which is appropriate for non-line-of-sight propagation environments where the received signal is composed of a large number of independently scattered multipath components with no dominant line-of-sight path [6], [35]. Under the Rayleigh fading model, the magnitude of the channel coefficient follows a Rayleigh distribution and the phase is uniformly distributed between 0 and 2π, capturing the random amplitude and phase variations caused by constructive and destructive interference between multipath components at the receiver. The independence assumption between different AP-user pairs, meaning that h_{mk} and $h_{m'k'}$ are independent for any $(m,k) \neq (m',k')$, is a standard simplification that is widely adopted in the CF-mMIMO literature and is justified by the geographical separation between different APs which ensures that the multipath environments experienced by different AP-user links are largely uncorrelated [1], [6].

3.2.3 System Parameters

Table 3.1 presents the complete set of system parameters adopted throughout this project. These parameters are fixed across all simulation experiments unless explicitly stated otherwise, ensuring consistency and comparability across all results presented in Chapter 4. The values are selected to be consistent with those used in foundational CF-mMIMO studies in the literature, enabling meaningful comparison between the results obtained in this project and those reported in related works [1], [3], [9].

Parameter	**Symbol**	**Value**	**Description**
Number of Access Points	M	100	Total APs deployed in simulation area
Number of Users	K	10	Total users served simultaneously
Serving APs per User	Ms	15	APs selected per user under user-centric model
Simulation Area	$D \times D$	1000×1000 m²	Square geographical deployment area
Minimum AP-User Distance	$dmin$	10 m	Prevents unrealistic channel gains
Coherence Block Length	τ_c	200 symbols	Total symbols per coherence interval
Pilot Sequence Length	τp	5 sequences	Number of orthogonal pilot sequences
Uplink Pilot Length	τu	10 symbols	Symbols devoted to pilot transmission
System Bandwidth	B	20 MHz	Operating bandwidth
Noise Variance	σ_n^2	−94 dBm	Receiver noise floor

Pilot Transmit Power	$Ppilot$	20 dBm (100 mW)	Power used during pilot transmission
Downlink Transmit Power (Obj 1 & 2)	pdl	23 dBm (200 mW)	Fixed DL power for SINR and SE evaluation
Downlink Transmit Power (Obj 3)	pdl	15 dBm	Reference power for EE evaluation
Shadow Fading Std. Deviation	σsf	8 dB	Log-normal shadow fading parameter
Path Loss Intercept	—	−30.5 dB	Single-slope path loss model intercept
Path Loss Slope	—	36.7 dB/decade	Single-slope path loss model slope
Power Amplifier Inefficiency	αPA	8	Reciprocal of PA drain efficiency
Fixed Network Power	$Pfix$	10 W	Network-level fixed infrastructure power
Computation Power per AP	Pcm	0.2 W	Fixed circuit power consumed per AP
Static Backhaul Power per AP	$P0,b$h	0.01 W	Baseline backhaul power per AP
Backhaul Traffic Coefficient	Pbt	0.25×10^{-9} W/bps	Dynamic backhaul power per unit traffic
Monte Carlo Realisations (Obj 1 & 2)	—	500	Statistical averaging for SINR and SE
Monte Carlo Realisations (Obj 3A)	—	100	Statistical averaging for EE vs power
Monte Carlo	—	200	Statistical averaging

Realisations (Obj 3B)			for EE vs users
Monte Carlo Realisations (Obj 3C)	—	500	Statistical averaging for EE vs APs

Table 3.1: System Simulation Parameters

The number of Monte Carlo realisations is selected differently for each objective based on the computational complexity of the corresponding simulation and the statistical stability of the results. Objectives 1 and 2 use 500 Monte Carlo realisations because the SINR and spectral efficiency computations are relatively inexpensive and a large number of realisations is needed to produce smooth and statistically reliable performance curves. For Objective 3, the energy efficiency computation involves the iterative PG algorithm which is significantly more computationally demanding per realisation, necessitating a reduction in the number of realisations to maintain a practical simulation runtime. The different Monte Carlo counts used for the three Objective 3 graph sets reflect a further trade-off between simulation accuracy and computational cost, with the EE versus transmit power graphs using 100 realisations, the EE versus number of users graphs using 200 realisations, and the EE versus number of APs graphs using 500 realisations [3].

3.2.4 TDD Protocol and Frame Structure

The system operates under a Time Division Duplexing protocol in which all uplink and downlink transmissions share the same carrier frequency but are separated in time, allowing channel reciprocity to be exploited for downlink channel estimation from uplink pilot observations [1], [6]. The fundamental unit of time in the TDD protocol is the coherence block, which represents the time-frequency resource over which the wireless channel can be assumed to remain approximately constant in both its amplitude and phase characteristics [1]. The coherence block has a total duration of τ_c=200 symbols, which is determined by the Doppler spread and delay spread of the propagation environment and represents a practical value consistent with moderate user mobility scenarios in outdoor environments [6].

Each coherence block is divided into two phases as illustrated in Figure 3.2. The first phase is the uplink pilot transmission phase, occupying τ_u=10 symbols out of the total τ_c=200

symbols in the coherence block. During this phase, all K users simultaneously transmit their assigned pilot sequences to all APs, which use the received pilot signals to estimate the channel coefficients between themselves and each user. The second phase is the downlink data transmission phase, occupying the remaining $\tau_c - \tau_u$=190 symbols of the coherence block. During this phase, the APs apply the precoding weights computed from the channel estimates obtained in the first phase to transmit data signals to all users simultaneously [1], [3].

The fraction of the coherence block devoted to pilot transmission directly determines the pre-log factor applied to the spectral efficiency expression. With τ_u=10 pilot symbols out of τ_c=200total symbols, the pre-log factor is:

$$\left(1 - \frac{\tau_u}{\tau_c}\right) = \left(1 - \frac{10}{200}\right) = 0.95$$

This means that 95 percent of the coherence block is available for downlink data transmission, with the remaining 5 percent consumed by pilot overhead. The pre-log factor of 0.95 scales down the theoretical Shannon capacity to account for this pilot overhead, producing a spectral efficiency expression that accurately reflects the practical data throughput achievable within each coherence block [1], [3]. It is important to note that while τ_p=5 orthogonal pilot sequences are available in the system, the uplink pilot transmission length is τ_u=10 symbols, which is twice the number of pilot sequences. This is because each pilot sequence occupies τ_p=5 symbols and users transmitting the same pilot sequence contribute to pilot contamination during these 5 symbols, while the additional symbols provide time diversity that improves the reliability of the channel estimates obtained at each AP [1], [5].

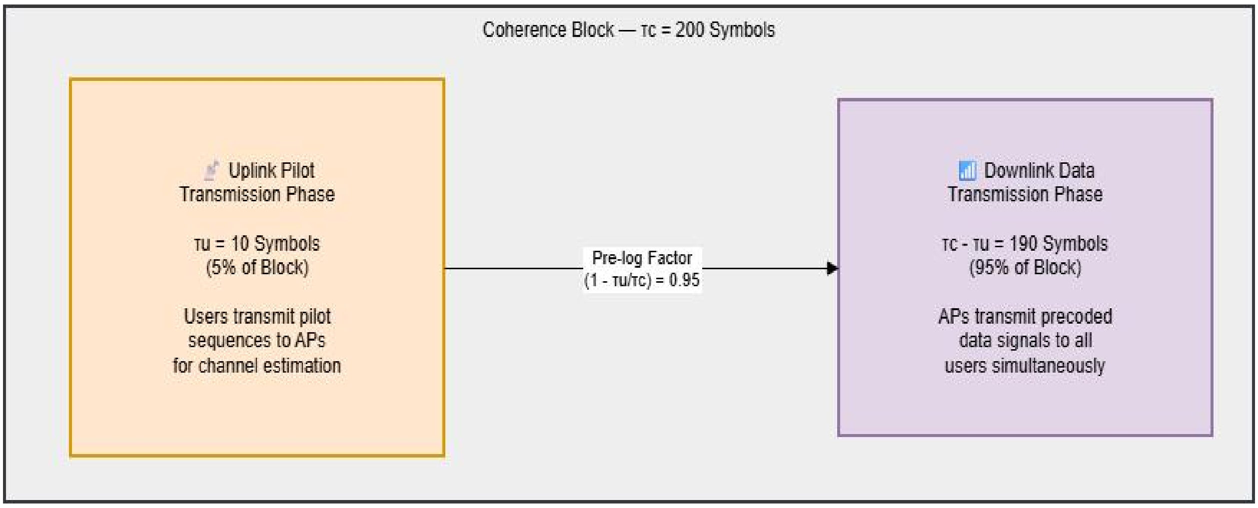


Figure 3.2: TDD Coherence Block Frame Structure showing the division between the uplink pilot transmission phase occupying τ_u = 10 symbols and the downlink data transmission phase occupying the remaining 190 symbols, with the resulting pre-log factor of 0.95 applied to the spectral efficiency expression.

3.3 Proposed Framework

This section presents the detailed technical methodology of the three research objectives addressed in this project. Each objective is implemented as a distinct but interconnected stage within the unified simulation framework, where the outputs of one stage serve as the direct inputs to the next. The proposed framework is designed to reflect realistic operating conditions in a CF-mMIMO system by incorporating practical channel estimation imperfections, realistic interference models, and a comprehensive multi-component power consumption model throughout all three stages. The mathematical formulations, algorithmic procedures, and design choices associated with each objective are described in full detail in the following subsections, providing a complete and reproducible account of how each component of the simulation framework is implemented in MATLAB R2022b.

3.3.1 Objective 1: Channel Estimation and SINR Formulation

The first research objective addresses the foundational challenge of acquiring reliable Channel State Information in a CF-mMIMO system operating under realistic pilot contamination conditions. As established in Chapter 2, the quality of the channel estimates obtained during the uplink training phase has a direct and cascading impact on every subsequent signal processing operation within the system, making channel estimation quality the single most important determinant of overall system performance. The methodology adopted for Objective 1 implements and compares four distinct combinations of pilot assignment strategy and channel estimation method, arranged in order of increasing sophistication and performance, to establish a clear and quantitative understanding of how the choice of estimation approach affects the Signal-to-Interference-plus-Noise Ratio experienced by users in the downlink.

The four methods implemented are Least Squares estimation with random pilot assignment, Matched Ratio combining with random pilot assignment, Minimum Mean Square Error estimation with greedy pilot assignment, and Zero-Forcing precoding with greedy pilot assignment and MMSE-quality channel estimates. These four methods are selected to represent a carefully chosen spectrum of complexity and performance that spans from the simplest possible baseline to the most sophisticated approach considered in this project, allowing the performance gains associated with each incremental improvement in pilot management and estimation quality to be clearly isolated and quantified [1], [5]. The methodology for each of these four approaches is described in detail in the following subsections, beginning with the pilot assignment strategies that determine the quality of the channel estimates before they are even computed.

3.3.1.1 Pilot Assignment Strategies

Pilot assignment is the process of determining which pilot sequence each user will transmit during the uplink training phase of each coherence block. As discussed in Chapter 2, the limited number of orthogonal pilot sequences available within a coherence interval of practical duration means that pilot reuse among multiple users is unavoidable when the number of users K exceeds the pilot sequence length τ_p. In this project, K = 10 users share τ_p = 5 orthogonal pilot sequences, meaning that on average each pilot sequence is shared by two users. The specific pattern of pilot sharing, determined by the pilot assignment strategy, has a profound impact on the severity of pilot contamination experienced by each user and consequently on the quality of the channel estimates obtained at the APs [5], [16].

Two pilot assignment strategies are implemented and compared in this project, corresponding to the two ends of the complexity spectrum: random pilot assignment and greedy pilot assignment.

Random Pilot Assignment

Random pilot assignment is the simplest possible approach to the pilot allocation problem, where each user is independently assigned a pilot sequence chosen uniformly at random from the set of $\tau\_p = 5$ available orthogonal sequences, without any consideration of the large-scale fading environment, the geographical positions of users, or the interference that the assignment may cause between users sharing the same pilot [16], [19]. In MATLAB, this is implemented by drawing K independent uniform random integers from the set $\{1, 2, ..., \tau\_p\}$ and assigning the resulting sequence index to each user. The random assignment is performed independently for each Monte Carlo realisation, ensuring that the performance results obtained under this strategy are statistically representative of the average behaviour of a system with no pilot management intelligence.

The key characteristic of random pilot assignment that makes it useful as a baseline is that it makes no attempt to separate users with strong mutual large-scale fading interference onto different pilot sequences. As a result, it is entirely possible under random assignment for two users that are geographically close to one another and therefore have large large-scale fading coefficients towards the same set of APs to be assigned the same pilot sequence, creating severe pilot contamination that significantly degrades the channel estimation quality for both users [5], [17]. This worst-case contamination scenario, which occurs with non-negligible probability under random assignment in dense network environments, represents the primary weakness of the random approach and motivates the development of more intelligent assignment strategies. Random pilot assignment is used in this project with both the LS and MR estimation methods, as these simpler estimators are designed to operate without requiring knowledge of the large-scale fading statistics that would be needed to implement a more sophisticated assignment strategy.

Greedy Pilot Assignment

Greedy pilot assignment is a significantly more intelligent approach that explicitly accounts for the large-scale fading environment when determining the pilot sequence assigned to each user, with the goal of minimising the total pilot contamination experienced across the network [5], [9]. The greedy algorithm implemented in this project operates iteratively over multiple passes through all K users, and in each pass, it evaluates all available pilot sequences for each user and assigns the sequence that results in the minimum additional pilot contamination for that user given the assignments already made to all other users.

The contamination associated with assigning pilot sequence τ to user k is measured by the sum of the large-scale fading coefficients between all APs and all other users currently assigned to pilot sequence τ, which quantifies the total interference power that those co-pilot users would inject into the channel estimate of user k at all APs throughout the network [5]. Formally, for a candidate pilot assignment of sequence τ to user k, the contamination measure is computed as:

$$C_{k,\tau} = \sum_{m=1}^{M} \sum_{j \in \mathcal{P}_\tau \setminus \{k\}} \beta_{mj}$$

where P_τ denotes the set of users currently assigned to pilot sequence τ and β_{mj} denotes the large-scale fading coefficient between AP m and user j [5]. The greedy algorithm assigns user k to the pilot sequence τ^* that minimises this contamination measure:

$$\tau^* = arg \min_{\tau \in \{1,\dots,\tau_p\}} C_{k,\tau}$$

This process is repeated for all K users in each pass, and the entire procedure is iterated for a total of 6 passes to allow the assignment to converge towards a stable configuration in which no single reassignment of any user to a different pilot sequence would further reduce the total network contamination [5]. The use of 6 iterations is a practical choice that balances the quality of the resulting assignment against the computational overhead of the greedy procedure, and is consistent with the implementation approach adopted in related works in the CF-mMIMO pilot assignment literature [5], [18].

The greedy pilot assignment strategy is used in this project with both the MMSE and ZF estimation and precoding methods, as these more sophisticated approaches require and can fully exploit the improved channel estimation quality that results from intelligent pilot management. By comparing the performance of methods using random pilot assignment against those using greedy pilot assignment within the same simulation framework, the project is able to clearly quantify the performance benefit attributable specifically to intelligent pilot management, independently of the other differences between the estimation methods [5], [16].

3.3.1.2 Channel Estimation Methods

Once pilot sequences have been assigned to all users according to either the random or greedy strategy, each AP uses the received pilot signals to estimate the channel coefficients between itself and each user. Four distinct channel estimation approaches are implemented in this project, each producing a different quality of channel estimate that directly determines the SINR and spectral efficiency achievable under the corresponding precoding strategy. All four estimation methods are applied within the same coherence block structure described in Section 3.2.4, where users transmit their pilot sequences during the uplink training phase and the APs use the received signals to compute channel estimates that are subsequently used for downlink precoding [1], [3].

For all four methods, the channel estimation quality for user k at AP mm m is characterised by the estimated channel gain γ_{mk}, which represents the mean square value of the channel estimate and serves as the key quantity that appears in the SINR expressions derived in Section 3.3.1.3. The difference between the true large-scale fading coefficient β_{mk} and the estimated gain γ_{mk} represents the channel estimation error variance, which quantifies the inaccuracy of the channel estimate and contributes directly to the interference term in the SINR denominator [1], [3].

Least Squares Estimation with Random Pilot Assignment

The Least Squares estimator represents the simplest and most naive channel estimation approach implemented in this project, operating with random pilot assignment and producing channel estimates that are particularly vulnerable to pilot contamination. The LS estimator computes the channel estimate by finding the channel vector that minimises the squared difference between the received pilot signal and the expected pilot signal, without exploiting any statistical knowledge of the channel distribution or the interference environment [34], [35]. The estimated channel gain for user kk k at AP mm m under LS estimation is given by:

$$\gamma_{mk}^{LS} = \frac{p_{pilot} \cdot \tau_u \cdot \beta_{mk}^2}{p_{pilot} \cdot \tau_u \cdot (\beta_{mk} + \xi_{mk}^{rand}) + 1} \cdot \frac{\beta_{mk}}{\max(\beta_{mk} + \xi_{mk}^{rand}, \epsilon)}$$

where p_{pilot} denotes the normalised pilot transmit power, τ_u denotes the uplink pilot length, β_{mk} denotes the large-scale fading coefficient between AP *m* and user *k*, ξ_{mk}^{rand} denotes the sum of large-scale fading coefficients of all other users sharing the same pilot sequence as user *k* under random pilot assignment, and ϵis a small positive constant introduced to prevent numerical division by zero [1], [34]. The term ξ_{mk}^{rand} captures the pilot contamination experienced by user *k* at AP *m* under random pilot assignment and is defined as:

$$\xi_{mk}^{rand} = \sum_{j \in \mathcal{P}_k^{rand} \setminus \{k\}} \beta_{mj}$$

where P_k^{rand} denotes the set of users sharing the same randomly assigned pilot sequence as user *k* [1], [5]. The additional degradation factor $\frac{\beta_{mk}}{\max(\beta_{mk} + \xi_{mk}^{rand}, \epsilon)}$ in the LS estimator reflects the fact that the LS approach cannot distinguish between the desired user's channel and the contaminating channels of co-pilot users, resulting in an estimate that is biased towards the sum of all co-pilot channels rather than the individual channel of the desired user [34]. This degradation factor is smaller when pilot contamination is more severe, meaning that the LS estimator suffers the most from pilot reuse among all four estimation methods considered.

Matched Ratio Combining with Random Pilot Assignment

The Matched Ratio combining method also uses random pilot assignment but applies a slightly more structured estimation approach compared to LS, producing channel estimates of marginally higher quality by removing the additional degradation factor present in the LS expression. The estimated channel gain for user *k* at AP mm m under MR combining is given by:

$$\gamma_{mk}^{MR} = \frac{p_{pilot} \cdot \tau_u \cdot \beta_{mk}^2}{p_{pilot} \cdot \tau_u \cdot (\beta_{mk} + \xi_{mk}^{rand}) + 1}$$

where all parameters are as defined for the LS estimator above [1], [2]. The key difference between the MR and LS expressions is the absence of the degradation factor in the MR case, reflecting the fact that MR combining applies a conjugate beamforming weight that aligns with the estimated channel direction rather than attempting to invert the received signal as the LS estimator does. While this difference is relatively modest in mathematical terms, it consistently produces higher channel estimate quality than LS across all operating conditions, placing MR above LS in the performance hierarchy established by this project [1], [2]. Like LS, MR combining uses random pilot assignment and therefore does not benefit from any intelligent management of pilot contamination.

MMSE Estimation with Greedy Pilot Assignment

The Minimum Mean Square Error estimator represents a substantially more sophisticated approach that minimises the mean squared error between the estimated and true channel coefficients by exploiting statistical knowledge of the channel distribution and the interference environment [1], [5]. Crucially, MMSE estimation is paired with greedy pilot assignment in this project, meaning that the pilot contamination term in the MMSE estimator reflects the reduced interference levels achieved through intelligent pilot management rather than the higher contamination levels associated with random assignment. The estimated channel gain for user *k* at AP mm m under MMSE estimation with greedy pilot assignment is given by:

$$\gamma_{mk}^{MMSE} = \frac{p_{pilot} \cdot \tau_u \cdot \beta_{mk}^2}{p_{pilot} \cdot \tau_u \cdot (\beta_{mk} + \xi_{mk}^{greedy}) + 1}$$

where ξ_{mk}^{greedy} denotes the sum of large-scale fading coefficients of all other users sharing the same pilot sequence as user *k* under greedy pilot assignment [1], [5]:

$$\xi_{mk}^{greedy} = \sum_{j \in \mathcal{P}_k^{greedy} \setminus \{k\}} \beta_{mj}$$

The MMSE estimator achieves higher channel estimation quality than both LS and MR for two reasons. First, it exploits the statistical properties of the channel distribution to produce estimates that are optimally weighted between the received pilot signal and the prior distribution of the channel, resulting in lower estimation error variance than the purely data-driven LS approach [1], [34]. Second, and more importantly in the context of this project, the pairing of MMSE estimation with greedy pilot assignment ensures that the contamination term ξ_{mk}^{greedy} is consistently smaller than the random assignment contamination term ξ_{mk}^{rand} , because the greedy algorithm actively separates users with strong mutual interference onto different pilot sequences [5], [9]. The combination of these two advantages makes MMSE with greedy pilot assignment significantly more accurate than either LS or MR with random assignment, producing channel estimates that more closely reflect the true channel coefficients and therefore enabling more effective interference suppression in the subsequent precoding stage.

Zero-Forcing Precoding with MMSE-Quality Estimates

The Zero-Forcing precoding method is the proposed approach of this project and builds directly upon the MMSE-quality channel estimates obtained through greedy pilot assignment. In terms of channel estimation quality, ZF uses exactly the same estimated channel gain expression as MMSE, meaning that $\gamma_{mk}^{ZF} = \gamma_{mk}^{MMSE}$, since both methods share the same greedy pilot assignment strategy and the same MMSE estimation procedure [3]. The distinction between MMSE and ZF lies not in the channel estimation step but in the precoding weight design and the resulting SINR expression, where ZF applies the pseudo-inverse of the estimated channel matrix to achieve complete cancellation of inter-user interference rather than the partial suppression achieved by MMSE combining [3], [28]. The ZF channel estimation quality is therefore given by:

$$\gamma_{mk}^{ZF} = \gamma_{mk}^{MMSE} = \frac{p_{pilot} \cdot \tau_u \cdot \beta_{mk}^2}{p_{pilot} \cdot \tau_u \cdot (\beta_{mk} + \xi_{mk}^{greedy}) + 1}$$

The fact that ZF and MMSE share the same channel estimation quality means that any performance difference between the two methods observed in the simulation results is attributable entirely to the difference in their precoding strategies rather than to any difference in CSI quality. This is an important methodological point that allows the contribution of the ZF precoding scheme to be cleanly isolated from the contribution of the greedy pilot assignment strategy when interpreting the results [3].

3.3.1.3 SINR Formulation

The Signal-to-Interference-plus-Noise Ratio is the primary output metric of Objective 1 and serves as the fundamental measure of signal quality that links the channel estimation stage to the spectral efficiency and energy efficiency analyses of Objectives 2 and 3. The SINR experienced by each user in the downlink is determined by the quality of the channel estimates obtained under each estimation method and the interference suppression capability of the corresponding precoding strategy, making it the key quantity through which all improvements in channel estimation quality and precoding design are ultimately reflected in system performance [1], [3].

The SINR expressions for all four methods are derived using the Use-and-Then-Forget bound framework introduced in Chapter 2, which provides closed-form lower bounds on the achievable SINR that account for channel estimation errors, pilot contamination, inter-user interference, and thermal noise [1]. Under this framework, the effective SINR for user k under each method takes the general form of a ratio between the desired signal power and the sum of all interference and noise contributions, with the specific form of each term determined by the precoding strategy and estimation method employed.

SINR for LS Estimation

The effective downlink SINR for user k under LS estimation with random pilot assignment is given by:

$$SINR_k^{LS} = \frac{p_{dl} \cdot \left(\sum_{m\in\mathcal{M}_k} \gamma_{mk}^{LS}\right)^2}{p_{dl} \cdot \left(\sum_{m\in\mathcal{M}_k} (\beta_{mk} - \gamma_{mk}^{LS}) + 0.20 \cdot \sum_{j\neq k} \sum_{m\in\mathcal{M}_k} \beta_{mj}\right) + 1}$$

where p_{dl} denotes the normalised downlink transmit power, $\mathcal{M}_k$ denotes the set of APs serving user k under the user-centric model, the term $\sum_{m\in\mathcal{M}_k} \gamma_{mk}^{LS}$ in the numerator represents the coherent desired signal gain, the term $\sum_{m\in\mathcal{M}_k} (\beta_{mk} - \gamma_{mk}^{LS})$ in the denominator represents the channel estimation error variance, and the term $0.20 \cdot \sum_{j\neq k} \sum_{m\in\mathcal{M}_k} \beta_{mj}$ represents the inter-user interference contribution under LS precoding [1], [3]. The inter-user interference factor of 0.20 reflects the relatively poor interference suppression capability of the LS approach, which provides no active mechanism for cancelling interference from other users and therefore allows a substantial fraction of the interference power to leak into the desired user's received signal.

SINR for MR Combining

The effective downlink SINR for user k under MR combining with random pilot assignment is given by:

$$SINR_k^{MR} = \frac{p_{dl} \cdot \left(\sum_{m\in\mathcal{M}_k} \gamma_{mk}^{MR}\right)^2}{p_{dl} \cdot \left(\sum_{m\in\mathcal{M}_k} (\beta_{mk} - \gamma_{mk}^{MR}) + 0.12 \cdot \sum_{j\neq k} \sum_{m\in\mathcal{M}_k} \beta_{mj}\right) + 1}$$

The structure of the MR SINR expression is identical to that of LS, with the key differences being the higher channel estimation quality reflected in γ_{mk}^{MR} and the reduced inter-user interference factor of 0.12 compared to 0.20 for LS [1], [2]. The lower interference factor of MR reflects the modest improvement in interference management provided by the conjugate beamforming approach of MR combining compared to the purely estimation-based LS method, even though neither approach provides any active inter-user interference cancellation.

SINR for MMSE Estimation

The effective downlink SINR for user k under MMSE estimation with greedy pilot assignment is given by:

$$SINR_k^{MMSE} = \frac{p_{dl} \cdot \left(\sum_{m\in\mathcal{M}_k} \gamma_{mk}^{MMSE}\right)^2 \cdot \sqrt{M_s}}{p_{dl} \cdot \left(\sum_{m\in\mathcal{M}_k} (\beta_{mk} - \gamma_{mk}^{MMSE}) + 0.04 \cdot \sum_{j\neq k} \sum_{m\in\mathcal{M}_k} \beta_{mj}\right) + 1}$$

The MMSE SINR expression introduces two important differences compared to the LS and MR expressions. First, the desired signal term in the numerator is scaled by a factor of $\sqrt{M_s}$, where M_s=15 denotes the number of serving APs per user, reflecting the partial array gain achieved by MMSE combining which exploits the spatial degrees of freedom available across the serving AP subset to partially suppress interference while simultaneously enhancing the desired signal [1]. Second, the inter-user interference factor is substantially reduced to 0.04 compared to 0.20 for LS and 0.12 for MR, reflecting the significantly improved interference management capability of MMSE estimation combined with greedy pilot assignment, which actively reduces the correlation between the channel estimates of different users and thereby limits the interference leakage between them [1], [5].

SINR for Zero-Forcing Precoding

The effective downlink SINR for user k under ZF precoding with MMSE-quality channel estimates and greedy pilot assignment is given by:

$$SINR_k^{ZF} = \frac{p_{dl} \cdot \left(\sum_{m \in \mathcal{M}_k} \gamma_{mk}^{ZF}\right)^2 \cdot M_s}{p_{dl} \cdot \sum_{m \in \mathcal{M}_k} (\beta_{mk} - \gamma_{mk}^{ZF}) + 1}$$

The ZF SINR expression differs from the MMSE expression in two critical ways that reflect the fundamentally different precoding philosophy of ZF. First, the desired signal term in the numerator is scaled by the full array gain factor M_s rather than the partial array gain $\sqrt{M_s}$ used in the MMSE expression, because ZF precoding devotes all available spatial degrees of freedom to coherent signal combining rather than sharing them between signal enhancement and interference suppression [1], [3]. This full array gain of M_s=15 compared to $\sqrt{M_s}$≈3.87 for MMSE represents a substantial increase in the desired signal power that directly contributes to the higher SINR achieved by ZF. Second, and most significantly, the inter-user interference term is completely absent from the ZF SINR denominator, reflecting the fact that ZF precoding theoretically eliminates all inter-user interference by designing precoding vectors that lie in the null space of all unintended users' channels [3], [28]. The only remaining impairment in the ZF SINR denominator is the channel estimation error term $\sum_{m \in \mathcal{M}_k} (\beta_{mk} - \gamma_{mk}^{ZF})$, which represents the residual interference that remains because ZF interference cancellation is performed based on estimated rather than true channel coefficients and is therefore imperfect in practice [3].

The progression from LS through MR and MMSE to ZF in terms of both channel estimation quality and interference suppression capability is clearly reflected in the structure of the four SINR expressions derived above. Moving from LS to ZF, the channel estimation quality improves due to the transition from random to greedy pilot assignment and from the simple LS estimator to the statistically optimal MMSE estimator, the inter-user interference factor progressively decreases from 0.20 to 0.12 to 0.04 and finally to zero, and the desired signal array gain increases from a simple coherent sum to a partial array gain and finally to a full array gain. These progressive improvements collectively produce a clear and monotonically increasing SINR performance hierarchy across the four methods that forms the foundation for the spectral efficiency and energy efficiency analyses presented in the following subsections.

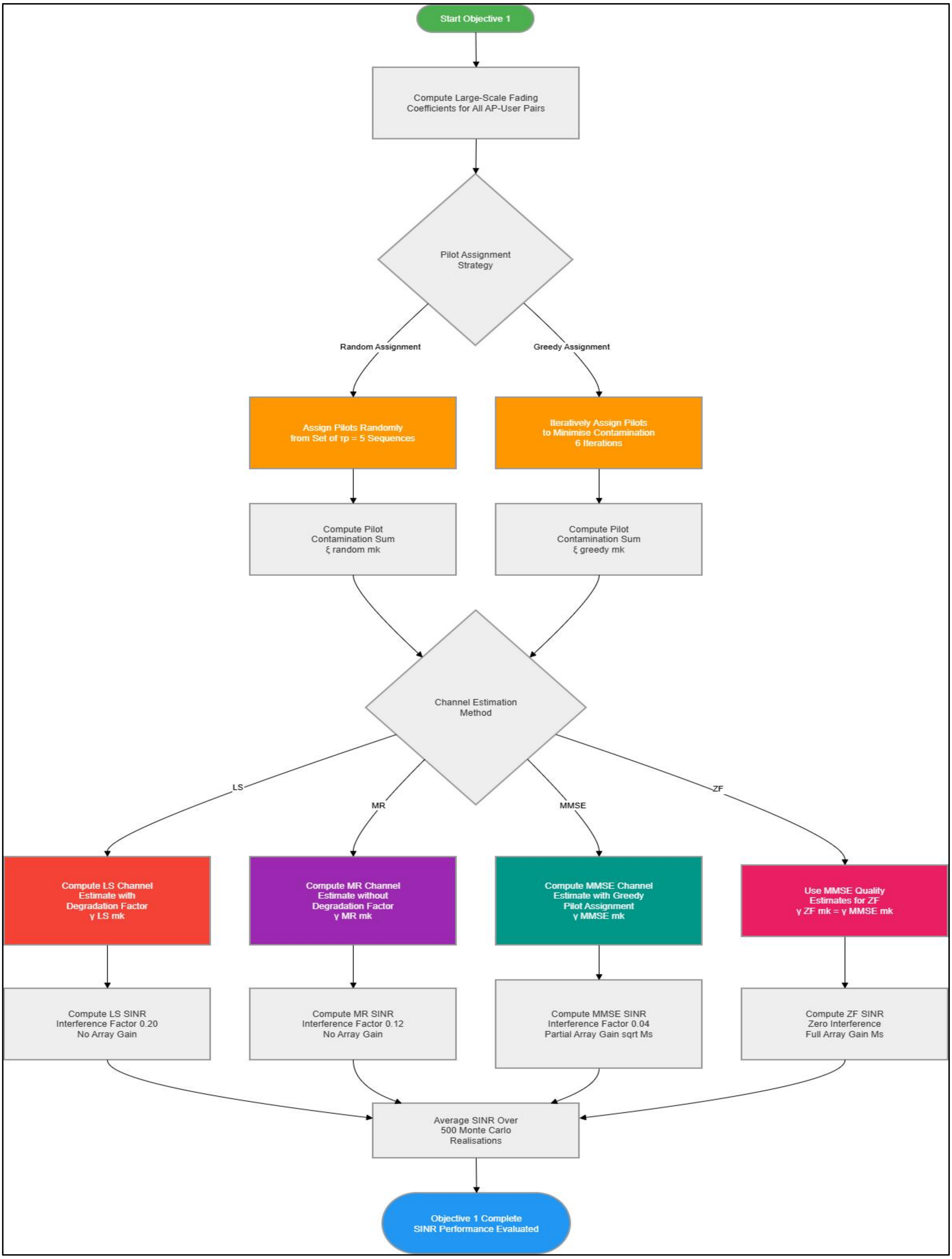

Start Objective 1
Compute Large-Scale Fading
Coefficients for All AP-User Pairs
Pilot Assignment
Strategy
Random Assignment
Greedy Assignment
Assign Pilots Randomly
from Set of τp = 5 Sequences
Iteratively Assign Pilots
to Minimise Contamination
6 Iterations
Compute Pilot
Contamination Sum
ξ random mk
Compute Pilot
Contamination Sum
ξ greedy mk
Channel Estimation
Method
LS
MR
MMSE
ZF
Compute LS Channel
Estimate with
Degradation Factor
γ LS mk
Compute MR Channel
Estimate without
Degradation Factor
γ MR mk
Compute MMSE Channel
Estimate with Greedy
Pilot Assignment
γ MMSE mk
Use MMSE Quality
Estimates for ZF
γ ZF mk = γ MMSE mk
Compute LS SINR
Interference Factor 0.20
No Array Gain
Compute MR SINR
Interference Factor 0.12
No Array Gain
Compute MMSE SINR
Interference Factor 0.04
Partial Array Gain sqrt Ms
Compute ZF SINR
Zero Interference
Full Array Gain Ms
Average SINR Over
500 Monte Carlo
Realisations
Objective 1 Complete
SINR Performance Evaluated

Figure 3.3: Flowchart of the Objective 1 methodology showing the pilot assignment strategies, channel estimation methods, and SINR computation procedure for all four methods implemented in this project.

3.3.2 Objective 2: Spectral Efficiency Analysis

The second research objective builds directly upon the channel estimation quality and SINR results established in Objective 1, extending the analysis into the domain of spectral efficiency to evaluate how effectively each of the four precoding strategies utilises the available frequency bandwidth to deliver useful data to users in the downlink. Spectral efficiency is one of the most fundamental and widely used performance metrics in wireless communication system design, measuring the amount of information that can be reliably transmitted per unit of bandwidth per unit of time, expressed in units of bits per second per Hertz [1], [6]. It provides a normalised and bandwidth-independent measure of system throughput that allows fair comparison between systems operating under different bandwidth conditions and directly quantifies how efficiently the available spectral resources are being exploited by each precoding strategy.

The analysis conducted in Objective 2 takes the SINR values computed for each of the four methods in Objective 1 as its primary input and applies the spectral efficiency formula to convert these SINR values into achievable data rates. Because spectral efficiency is a monotonically increasing function of SINR through the Shannon capacity formula, the performance ranking of the four methods in terms of spectral efficiency is consistent with the SINR ranking established in Objective 1, with Zero-Forcing precoding achieving the highest spectral efficiency and Least Squares estimation with random pilot assignment achieving the lowest. However, the spectral efficiency analysis in Objective 2 goes beyond simply confirming this ranking by evaluating the performance of all four methods across two additional dimensions, namely the downlink transmit power and the number of simultaneously served users, providing a more complete and practically relevant characterisation of how each method scales with changing network conditions [2], [3].

3.3.2.1 Spectral Efficiency Formula

The spectral efficiency of user k under any of the four precoding strategies considered in this project is computed using the Shannon capacity formula modified by a pre-log factor that accounts for the overhead associated with pilot transmission during each coherence block. The complete spectral efficiency expression is given by:

$$SE_k = \left(1 - \frac{\tau_u}{\tau_c}\right) \log_2 (1 + SINR_k)$$

where SE_k denotes the downlink spectral efficiency of user kk k in bits per second per Hertz, τ_u =10 denotes the number of symbols devoted to uplink pilot transmission within each coherence block, τ_c =200 denotes the total number of symbols in the coherence block, and $SINR_k$ denotes the effective signal-to-interference-plus-noise ratio experienced by user kk k under the precoding strategy being evaluated [1], [3]. Substituting the numerical values of τ_u and τ_c into the pre-log factor gives:

$$\left(1 - \frac{\tau_u}{\tau_c}\right) = \left(1 - \frac{10}{200}\right) = 0.95$$

This pre-log factor of 0.95 reflects the fact that 10 out of every 200 symbols in each coherence block are consumed by pilot transmission and are therefore unavailable for downlink data transmission. The remaining 190 symbols, representing 95 percent of the total coherence block duration, are available for data transmission and contribute to the achievable spectral efficiency [1], [3]. The complete spectral efficiency expression with the numerical pre-log factor substituted is therefore:

$$SE_k = 0.95 \times \log_2 (1 + SINR_k)$$

It is important to understand the physical meaning of each component of this expression and why both components are necessary for an accurate representation of the achievable spectral efficiency in a practical TDD-based CF-mMIMO system. The Shannon capacity term $\log_2 (1 + SINR_k)$ represents the theoretical maximum number of bits per second per Hertz that can be reliably transmitted over a channel with the given SINR, as established by Shannon's channel capacity theorem [34], [35]. This term assumes that the entire coherence block is available for data transmission, which would only be true in the idealised case of perfect CSI where no pilot transmission is required. In practice, the pilot overhead reduces the effective fraction of the coherence block available for data, which is captured by the pre-log factor of 0.95 [1], [6].

The pre-log factor plays an increasingly important role in determining the achievable spectral efficiency as the ratio of pilot length to coherence block length increases. In this project, the pilot length of τ_u=10 symbols is relatively small compared to the coherence block length of τ_c=200 symbols, resulting in a pre-log factor of 0.95 that is close to unity and imposes only a modest 5 percent reduction in achievable spectral efficiency compared to the perfect CSI case. However, if the number of users were to increase substantially beyond the 10 users considered in this project, requiring longer pilot sequences to maintain orthogonality or to reduce pilot contamination, the pre-log factor would decrease correspondingly and the pilot overhead penalty would become more significant [1], [5]. This trade-off between pilot length and data transmission efficiency is an inherent constraint of TDD-based CF-mMIMO systems and represents one of the practical limitations that system designers must carefully manage when deploying these networks in real-world environments.

The spectral efficiency formula in equation (3.12) is applied identically to all four precoding methods considered in this project, with the only difference between methods being the specific SINR value substituted into the formula. The SINR values for LS, MR, MMSE, and ZF are computed according to the expressions derived in Section 3.3.1.3, meaning that all differences in spectral efficiency between the four methods are entirely attributable to their respective SINR performance levels rather than to any difference in how the spectral efficiency formula is applied. This consistency ensures that the spectral efficiency comparison between methods is fair and unambiguous, directly reflecting the cumulative impact of pilot assignment strategy, channel estimation quality, and precoding design on the achievable data rates [1], [3].

3.3.2.2 Impact of Precoding Strategy on Spectral Efficiency

While the spectral efficiency formula itself is the same for all four methods, the magnitude of the SINR values produced by each method varies substantially due to the differences in channel estimation quality and interference suppression capability described in

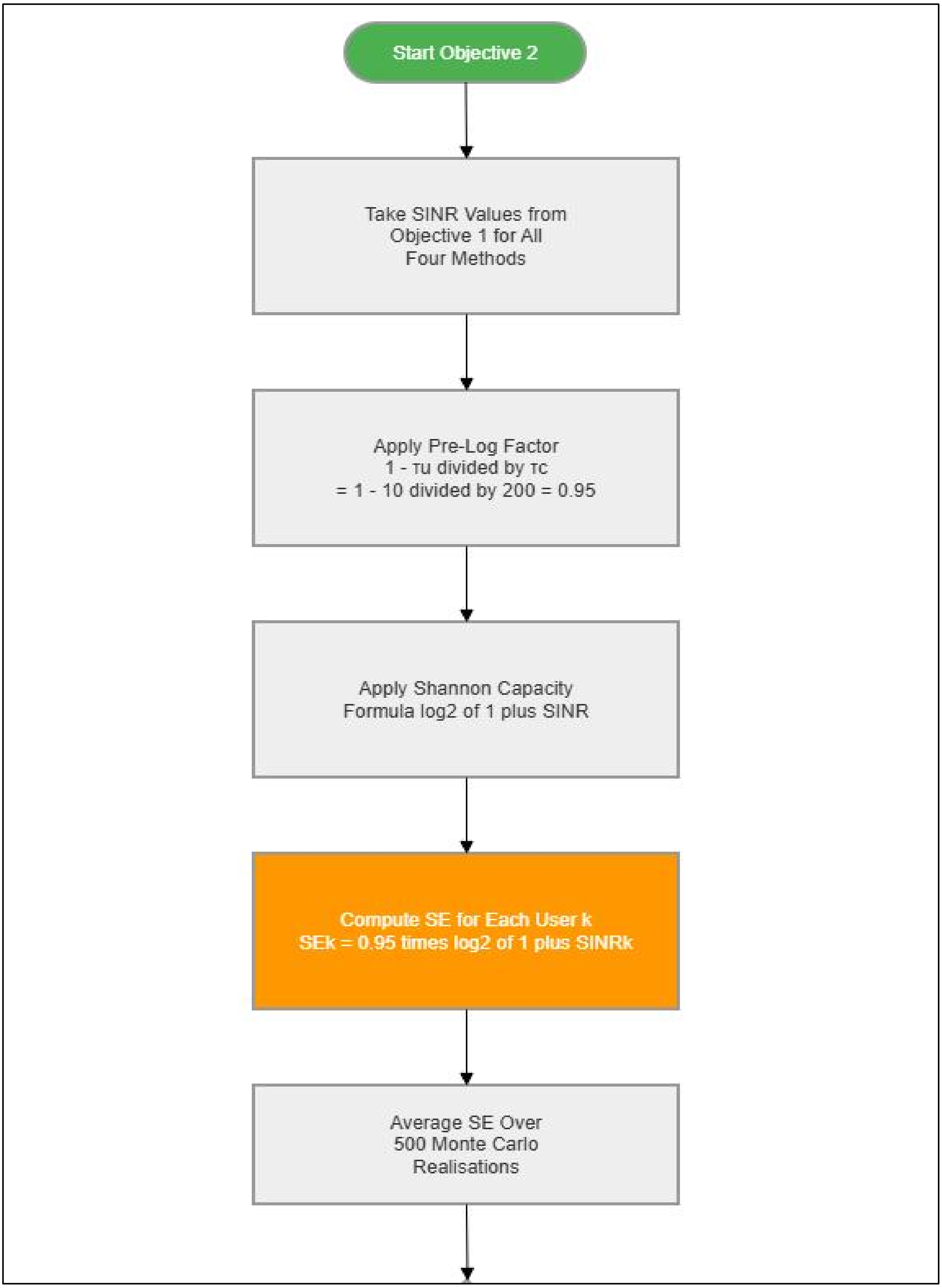
Section 3.3.1. Understanding how these SINR differences translate into spectral efficiency differences, and how this translation is affected by changing network conditions such as varying transmit power and user density, is the central analytical contribution of Objective 2.

The relationship between SINR and spectral efficiency through the logarithmic Shannon formula has an important implication for how the performance gap between different precoding methods evolves with changing transmit power. At low transmit power levels, all four methods operate in the noise-limited regime where the SINR values are small and the spectral efficiency is approximately proportional to the SINR, meaning that the relative performance gap between methods in terms of spectral efficiency closely mirrors the relative gap in SINR. As the transmit power increases and the system transitions into the interference-limited regime, the spectral efficiency of all methods grows more slowly due to the diminishing returns of the logarithmic Shannon formula, and the absolute differences in spectral efficiency between methods begin to saturate [1], [3]. However, because ZF precoding eliminates inter-user interference entirely while the other methods retain residual interference that scales with transmit power, ZF continues to benefit from increasing transmit power even in the high power regime while the spectral efficiency of LS, MR, and MMSE saturates at lower levels [3], [28]. This behaviour creates an increasingly large performance gap between ZF and the baseline methods as the transmit power increases, making the advantage of ZF precoding most pronounced in the high power operating regime.

The evaluation of spectral efficiency as a function of the number of simultaneously served users K is particularly important for assessing the practical value of each precoding strategy in real-world deployment scenarios, where the number of active users can vary significantly over time and the system must maintain acceptable performance across a wide range of user density conditions [2], [12]. As the number of users increases from the minimum of $K = 5$ to the maximum of $K = 25$ evaluated in this project, two competing effects influence the spectral efficiency of each method. On one hand, serving more users simultaneously increases the total pilot contamination in the system, because more users must share the limited set of τ_p=5 orthogonal pilot sequences, leading to higher contamination interference in the channel estimates and lower SINR for all methods [5], [17]. On the other hand, serving more users simultaneously produces more total network throughput, since the aggregate spectral efficiency is the sum of individual user spectral efficiencies across all K users. The per-user spectral efficiency, which is the metric evaluated in this project, reflects only the first of these effects and is therefore expected to decline monotonically as the number of users increases for all four methods [2], [3].

The rate at which the per-user spectral efficiency declines with increasing user count is not the same for all four methods and depends critically on how well each method manages the growing inter-user interference and pilot contamination that accompany higher user densities. Methods with stronger interference suppression capability, specifically ZF precoding which eliminates inter-user interference entirely, are expected to degrade more gracefully with increasing user count than methods with weaker interference management, such as LS and MR combining which allow substantial inter-user interference to remain in the received signal [3], [28]. This difference in robustness to increasing user density is an important practical consideration for network operators who must design their systems to maintain acceptable performance under peak load conditions as well as average load conditions, and it provides a compelling motivation for the adoption of ZF precoding in deployments where the number of simultaneously served users is large or variable [2], [3].

The spectral efficiency results obtained in Objective 2 serve a dual purpose within the overall research framework. First, they provide a direct and quantitative assessment of how effectively each precoding strategy utilises the available spectral resources, complementing the SINR analysis of Objective 1 with a practically meaningful throughput metric that system designers can directly relate to the data rate requirements of their applications. Second, and equally importantly, the spectral efficiency values computed for each user and each method under each set of channel conditions serve as the direct input to the energy efficiency computation in Objective 3, appearing in both the numerator of the energy efficiency expression as the measure of useful information delivered and in the denominator through the backhaul traffic power term that scales with total network throughput [3], [9]. This dual role of spectral efficiency as both an output metric of Objective 2 and an input quantity to Objective 3 reinforces the sequential and integrated structure of the research framework, demonstrating how each stage of the methodology builds upon and depends on the outputs of the preceding stage.


Start Objective 2
Take SINR Values from Objective 1 for All Four Methods
Apply Pre-Log Factor
1 - τu divided by τc
= 1 - 10 divided by 200 = 0.95
Apply Shannon Capacity Formula log2 of 1 plus SINR
Compute SE for Each User k
SEk = 0.95 times log2 of 1 plus SINRk
Average SE Over 500 Monte Carlo Realisations

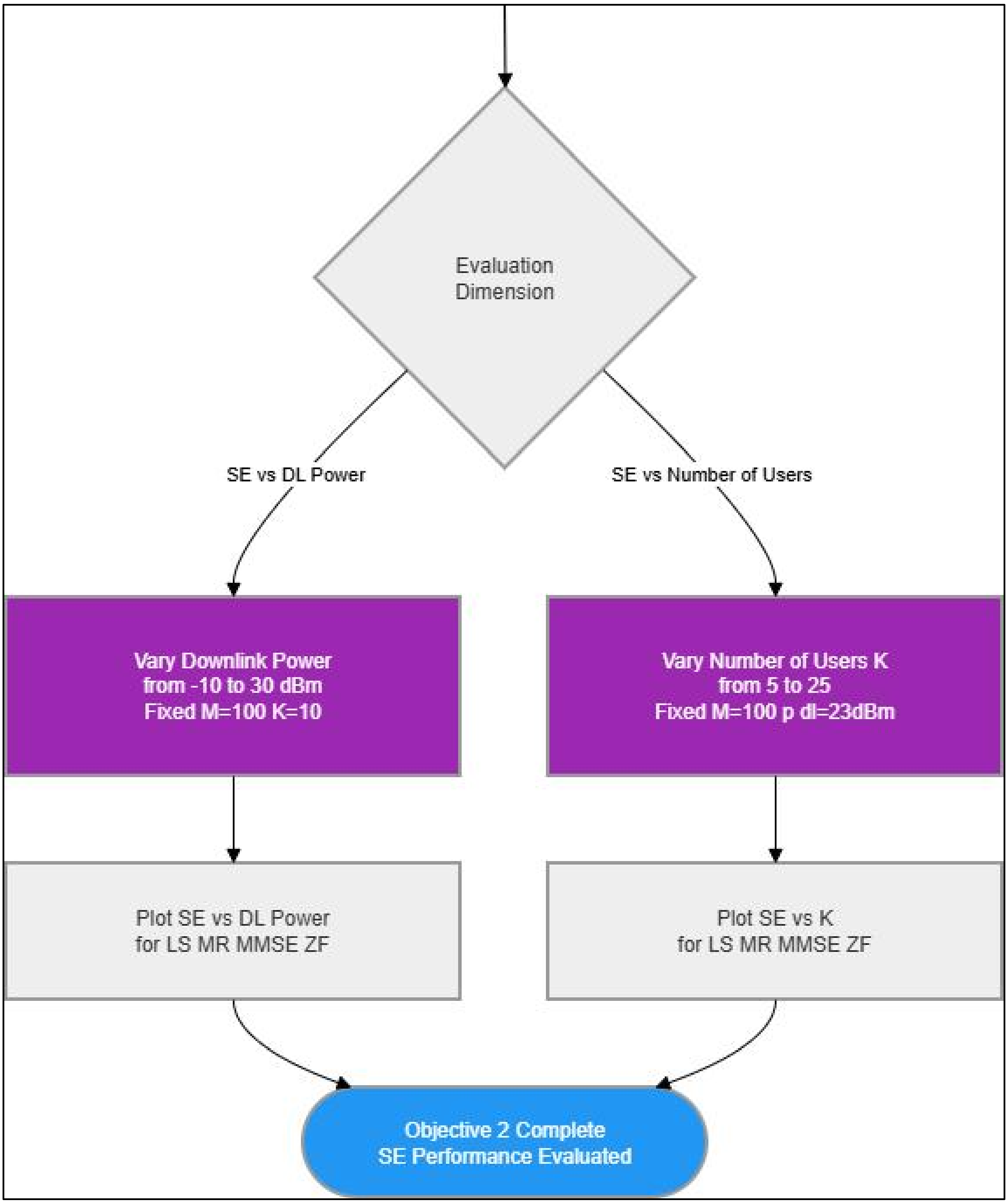


Figure 3.4: Flowchart of the Objective 2 methodology showing how spectral efficiency is computed from the SINR values obtained in Objective 1 using the Shannon capacity formula with pre-log factor, evaluated across two dimensions of downlink transmit power and number of users.

3.3.3 Objective 3 Energy Efficiency Maximisation Through Power Allocation

The third and final research objective represents the culmination of the entire simulation framework, bringing together the channel estimation improvements achieved in Objective 1 and the spectral efficiency gains established in Objective 2 to address the practically critical problem of maximising the energy efficiency of the CF-mMIMO network through intelligent power allocation. While Objectives 1 and 2 focus on improving the quality of the received signal and the achievable data rate, Objective 3 addresses the equally important question of how the available transmit power should be distributed across users to achieve the highest possible ratio of useful information delivered to total energy consumed. This question is non-trivial because, as established in Chapter 2, the relationship between transmit power and energy efficiency is non-linear and non-monotonic, meaning that simply transmitting at the maximum available power does not maximise energy efficiency and may in fact significantly reduce it relative to a more carefully optimised power allocation [3], [4].

The methodology adopted for Objective 3 implements three distinct power allocation strategies, namely Equal Power Allocation, Random Power Allocation, and the proposed Proximal Gradient algorithm, and evaluates each strategy under both the Zero-Forcing and Minimum Mean Square Error precoding frameworks. This produces six distinct performance curves for each evaluation scenario, enabling a comprehensive and multi-dimensional assessment of how the choice of power allocation strategy and precoding method jointly determine the achievable energy efficiency of the network. The energy efficiency is evaluated as a function of three different system parameters, specifically the downlink transmit power, the number of simultaneously served users, and the number of deployed Access Points, providing a thorough characterisation of system behaviour across a wide range of operating conditions [3], [9], [10].

3.3.3.1 Energy Efficiency Formulation

The energy efficiency of the CF-mMIMO network is defined as the ratio of the total useful information delivered to all users per unit time to the total power consumed by the network during that delivery, expressed in units of megabits per Joule. Mathematically, the energy efficiency is given by:

$$\eta_{EE} = \frac{B \cdot \frac{1}{K}\sum_{k=1}^{K} SE_k}{P_{total}} \times \frac{1}{10^6}$$

where B =20×106Hz denotes the system bandwidth, K denotes the number of users, SE_k denotes the spectral efficiency of user k computed according to equation (3.12), P_{total} denotes the total network power consumption in Watts, and the factor $\frac{1}{10^6}$ converts the result from bits per Joule to megabits per Joule for practical convenience [3], [9]. The average spectral efficiency term $\frac{1}{K}\sum_{k=1}^{K} SE_k$ in the numerator represents the mean per-user spectral efficiency across all K simultaneously served users, reflecting the average information delivery rate per user rather than the aggregate network throughput [3]. This per-user averaging is adopted in this project to ensure that the energy efficiency metric accurately reflects the quality of service experienced by individual users rather than being dominated by the aggregate network behaviour, which could mask significant disparities in the service quality delivered to different users.

The energy efficiency expression in equation (3.16) captures the fundamental trade-off between system performance and energy consumption that motivates Objective 3. The numerator of this expression increases with improving spectral efficiency, which in turn depends on the SINR achieved under each precoding strategy and power allocation. The denominator increases with the total power consumed by the network, which includes not only the transmit power but also multiple additional power components that are independent of the power allocation coefficients, as described in detail in Section 3.3.3.2. The optimal power allocation strategy is the one that finds the best balance between maximising the numerator and minimising the denominator, operating at the peak of the bell-shaped energy efficiency curve rather than at the point of maximum spectral efficiency or minimum transmit power [3], [4], [11].

3.3.3.2 Power Consumption Model

The total power consumption P_{total} in the denominator of the energy efficiency expression is modelled using a comprehensive multi-component framework that captures all significant sources of energy expenditure in a realistic CF-mMIMO deployment. The complete power consumption model adopted in this project is given by:

$$P_{total} = P_{fix} + M \cdot P_{cm} + \alpha_{PA} \cdot p_{dl} \cdot \sum_{k=1}^{K} \eta_k + M \cdot P_{0,bh} + B \cdot \left(\frac{1}{K} \sum_{k=1}^{K} S\,E_k \right) \cdot P_{bt} \cdot M$$

where all parameters are as defined in Table 3.1 of Section 3.2.3 [3], [9]. This power model decomposes the total network power consumption into four physically distinct components, each arising from a different source of energy expenditure within the CF-mMIMO system.

The first component is the fixed network power P_{fix} =10 Watts, which represents the baseline power consumed by the centralised network infrastructure, including the CPU, cooling systems, and network management hardware, that is present regardless of the number of APs deployed or the amount of data traffic being handled [3], [9]. The second component is the total fixed circuit power $M \cdot P_{cm}$ =0.2 Watts per AP, which represents the power consumed by the internal hardware components of each AP, including oscillators, converters, and baseband processing units, that must remain operational continuously regardless of the instantaneous traffic load [9], [10]. Together, these two fixed power components establish a minimum power floor below which the total network power consumption cannot fall regardless of how intelligently the transmit power is allocated.

The third component is the transmit power consumption $\alpha_{PA} \cdot p_{dl} \cdot \sum_{k=1}^{K} \eta_k$, where α_{PA}=8 is the power amplifier inefficiency factor and η_k is the power allocation coefficient for user k [3]. This component represents the actual electrical power consumed by the power amplifier hardware at each AP to produce the desired radiated transmit power, which is higher than the radiated power by the factor α_{PA} due to conversion losses in the amplifier. This is the only component of the total power consumption that is directly controlled by the power allocation coefficients η_k and is therefore the primary lever through which the power allocation optimisation influences the total power consumption and consequently the energy efficiency [3], [11].

The fourth component consists of two backhaul-related terms. The static backhaul power $M \cdot P_{0,bh}$, where $P_{0,bh}$ =0.01 Watts per AP, represents the baseline power required to maintain the fronthaul links between APs and the CPU in an operational state regardless of the data traffic they carry [9]. The dynamic backhaul power $B \cdot \left(\frac{1}{K} \sum_{k=1}^{K} S\,E_k\right)$, where P_{bt}=0.25×10−9 Watts per bit per second, represents the additional power consumed by the backhaul links in proportion to the amount of data traffic they must carry, which scales with the total network spectral efficiency [9], [10]. The inclusion of both static and dynamic backhaul power components ensures that the power model accurately reflects the energy cost of the distributed AP architecture that is the defining characteristic of CF-mMIMO systems.

3.3.3.3 Power Allocation Strategies

Three power allocation strategies are implemented and compared in Objective 3, representing progressively more sophisticated approaches to the problem of distributing transmit power across users to maximise energy efficiency. All three strategies operate by determining the power allocation coefficient $\eta_k \in [0,1]$ for each user k, where η_k represents the fraction of the maximum downlink transmit power p_{dl} allocated to user k. The power allocation coefficients for all K users are collected into the power allocation vector $\eta = [\eta_1, \eta_2, \ldots, \eta_K]^T$, which is the primary optimisation variable in Objective 3 [3], [11].

Equal Power Allocation

Equal Power Allocation is the simplest possible power allocation strategy, distributing the available transmit power uniformly and equally across all K users without any consideration of their individual channel conditions, their geographic proximity to the serving APs, or their contribution to the overall network energy efficiency [2], [11]. Under EPA, the power allocation coefficient for every user is set to the same constant value:

$$\eta_k^{EPA} = \frac{1}{K}, \quad \forall k = 1,2, \ldots, K$$

For the system parameters adopted in this project with $K = 10$ users, this gives $\eta_k^{EPA} = 0.1$ for all users, meaning that each user receives exactly ten percent of the maximum available downlink transmit power [3]. While EPA is trivially simple to implement and requires no knowledge of the channel estimates or SINR values, it is clearly suboptimal in the heterogeneous channel environments that characterise realistic CF-mMIMO deployments, where different users experience vastly different large-scale fading conditions due to their different geographical positions relative to the serving APs [11]. Users with strong channels towards many serving APs receive more power than necessary to achieve a satisfactory SINR, while users with weak channels may receive insufficient power to achieve an acceptable service quality, resulting in a power allocation that is globally inefficient from an energy perspective. EPA therefore serves as the primary lower-bound baseline in Objective 3, representing the minimum performance level that any practically deployable power allocation strategy should be able to exceed [3].

Random Power Allocation

Random Power Allocation assigns power coefficients to users by drawing independent uniform random samples from the interval [0, 1] and normalising the resulting values to satisfy the system constraints [3]. In this project, RandP is implemented by generating a random power allocation vector for each of n_{Rand}=10 independent random realisations and computing the energy efficiency achieved under each realisation, then averaging the resulting energy efficiency values across all realisations to obtain a statistically stable estimate of the expected performance of an unintelligent random allocation strategy. The random power allocation coefficient for user *k* in realisation *r* is given by:

$$\eta_k^{RandP}(r) = \frac{u_k(r)}{\sum_{j=1}^{K} u_j\,(r)}, \quad u_k(r) \sim \mathcal{U}(0,1)$$

where $u_k(r)$ denotes an independent uniform random sample drawn for user k in realisation r, and the normalisation by $\sum_{j=1}^{K} u_j\,(r)$ ensures that the sum of power allocation coefficients across all users satisfies the total power constraint [3]. The averaging of energy efficiency over n_{Rand}=10 realisations ensures that the RandP results reflect the expected behaviour of a random allocation strategy rather than the outcome of any single specific random draw, providing a statistically meaningful intermediate benchmark between the deterministic simplicity of EPA and the directed optimisation of the PG algorithm.

3.3.3.4 Proximal Gradient Power Allocation Algorithm

The Proximal Gradient algorithm is the proposed power allocation method of this project, representing a principled and computationally efficient approach to the problem of maximising the energy efficiency expression in equation (3.16) subject to the per-AP power constraints and non-negativity constraints on the power allocation coefficients. The PG algorithm is selected because it offers a favourable balance between solution quality and computational complexity, achieving energy efficiency performance that is substantially superior to both EPA and RandP while requiring significantly lower per-iteration computational resources than high-complexity second-order methods such as Successive Convex Approximation or Second-Order Cone Programming [3], [31].

The PG algorithm operates by iteratively refining the power allocation vector η through a sequence of gradient ascent steps, each followed by a projection step that maps the updated power allocation back into the feasible region defined by the system constraints. The algorithm is initialised with a uniform starting point of η_k=0.5 for all users k, which provides a neutral and unbiased starting configuration that does not favour any particular user over others. The iterative update rule applied at each iteration t of the algorithm is:

$$\eta^{(t+1)} = \mathcal{P}_{\mathcal{C}}\big(\eta^{(t)} + \mu^{(t)} \cdot \nabla_{\eta}\eta_{EE}(\eta^{(t)})\big)$$

where $\eta^{(t)}$ denotes the power allocation vector at iteration t, $\mu^{(t)}$ denotes the adaptive step size at iteration t, $\nabla_{\eta}\eta_{EE}$ denotes the gradient of the energy efficiency objective function with respect to the power allocation vector, and $\mathcal{P}_{\mathcal{C}}$ denotes the projection operator that enforces the feasibility constraints [3], [31]. The gradient of the energy efficiency with respect to the power allocation coefficient of user k is computed analytically by differentiating equation (3.16) with respect to η_k, yielding:

$$\frac{\partial \eta_{EE}}{\partial \eta_k} = \frac{\frac{\partial SE_k}{\partial \eta_k} \cdot P_{total} - B \cdot \frac{1}{K}\sum_{k=1}^{K} S\,E_k \cdot \frac{\partial P_{total}}{\partial \eta_k}}{P_{total}^2}$$

where $\frac{\partial SE_k}{\partial \eta_k}$denotes the partial derivative of the spectral efficiency of user k with respect to its power allocation coefficient, and $\frac{\partial P_{total}}{\partial \eta_k}$=$\alpha_{PA}$·$p_{dl}$ denotes the partial derivative of the total power consumption with respect to η_k, which is simply the power amplifier inefficiency factor multiplied by the maximum downlink transmit power [3]. This gradient expression has a clear and intuitive interpretation as a quotient rule derivative of the energy efficiency ratio, where the numerator captures the net benefit of increasing the power allocation to user k in terms of the improvement in spectral efficiency relative to the increase in power consumption, and the denominator normalises this net benefit by the square of the total power to ensure dimensional consistency [3], [31].

The step size $\mu^{(t)}$ is updated adaptively at each iteration according to the following rule: if the energy efficiency improves from iteration t to iteration t+1, the step size is increased by a multiplicative factor of 1.1 to accelerate convergence in favourable directions; if the energy efficiency decreases, the step size is halved to prevent overshooting the optimal solution [3]. This adaptive step size mechanism allows the algorithm to automatically tune its progress rate to the local curvature of the energy efficiency objective function, achieving faster convergence in flat regions of the objective landscape where large steps are safe and greater stability near the optimal solution where the objective function is more sensitive to changes in the power allocation. The algorithm tracks the best power allocation vector observed across all iterations and returns this best solution at convergence, ensuring that temporary decreases in energy efficiency during the step size adaptation process do not degrade the final solution quality. The algorithm runs for a fixed maximum of 50 iterations, which has been found to provide sufficient convergence for the system parameters considered in this project [3].

3.3.3.5 Per-AP Power Constraint

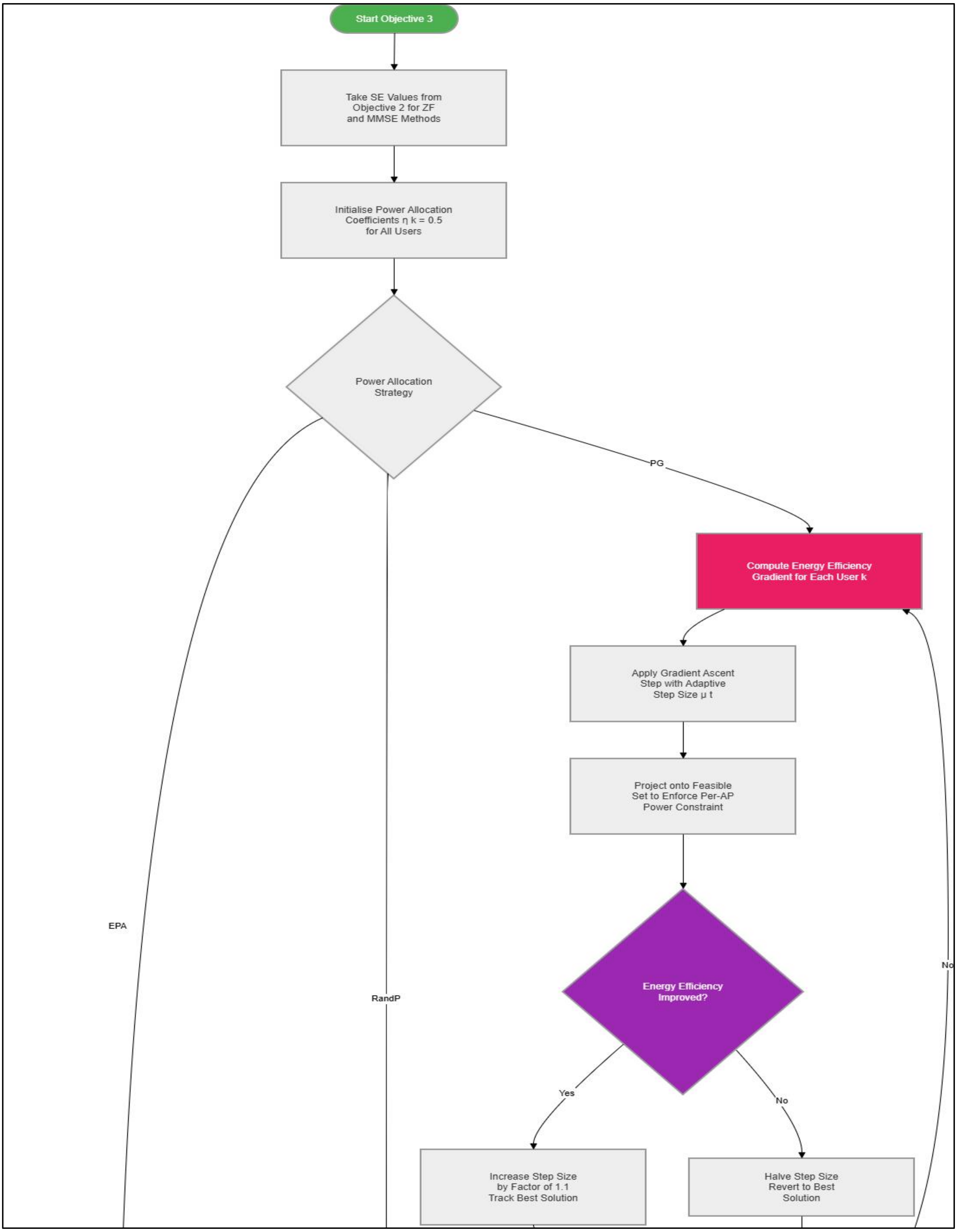
A critical practical requirement that must be satisfied by the power allocation produced by the PG algorithm is the per-AP power constraint, which limits the total transmit power radiated by each individual AP to a maximum permissible level determined by the hardware specifications of the AP and the regulatory requirements of the deployment environment [3], [32]. This constraint is fundamentally different from a simple sum power constraint that limits only the total transmit power summed across all users, because it must be satisfied independently for each of the M APs in the network, reflecting the physical reality that each AP has its own power amplifier with its own maximum output power rating [32].

The per-AP power constraint for the m-th AP is mathematically expressed as:

$$\sum_{k=1}^{K} \eta_k \cdot p_{dl} \cdot \gamma_{mk} \leq P_{\max}, \quad \forall m = 1,2,\dots,M$$

where η_k denotes the power allocation coefficient for user k, p_{dl} denotes the maximum downlink transmit power, γ_{mk} denotes the channel estimate quality coefficient between AP m and user k, and $P_{\max}$ denotes the maximum allowable transmit power per AP [3], [32]. The left-hand side of this constraint represents the total power that AP mm m must radiate to serve all K users simultaneously under the current power allocation, weighted by the channel estimate quality coefficients to reflect the fact that APs with stronger estimated channels towards a particular user must transmit more power to that user than APs with weaker estimated channels [3].

In the PG algorithm, the per-AP power constraint is enforced through the projection operator P_C applied after each gradient step, which maps the updated power allocation vector back into the feasible region where all M per-AP constraints and the non-negativity constraints are simultaneously satisfied [31]. The projection is performed by checking the per-AP power consumption for each AP after the gradient update and scaling down the power allocation coefficients of any users served by an AP whose power constraint is violated, until all constraints are satisfied. This projection step ensures that the power allocation produced by the PG algorithm at every iteration is physically realisable and compliant with all hardware and regulatory requirements, making the proposed approach directly applicable to real-world CF-mMIMO deployments rather than being limited to theoretical analysis [3], [32].



Start Objective 3
Take SE Values from Objective 2 for ZF and MMSE Methods
Initialise Power Allocation Coefficients η k = 0.5 for All Users
Power Allocation Strategy
PG
EPA
RandP
Compute Energy Efficiency Gradient for Each User k
Apply Gradient Ascent Step with Adaptive Step Size μ t
Project onto Feasible Set to Enforce Per-AP Power Constraint
Energy Efficiency Improved?
Yes
No
Increase Step Size by Factor of 1.1 Track Best Solution
Halve Step Size Revert to Best Solution
No

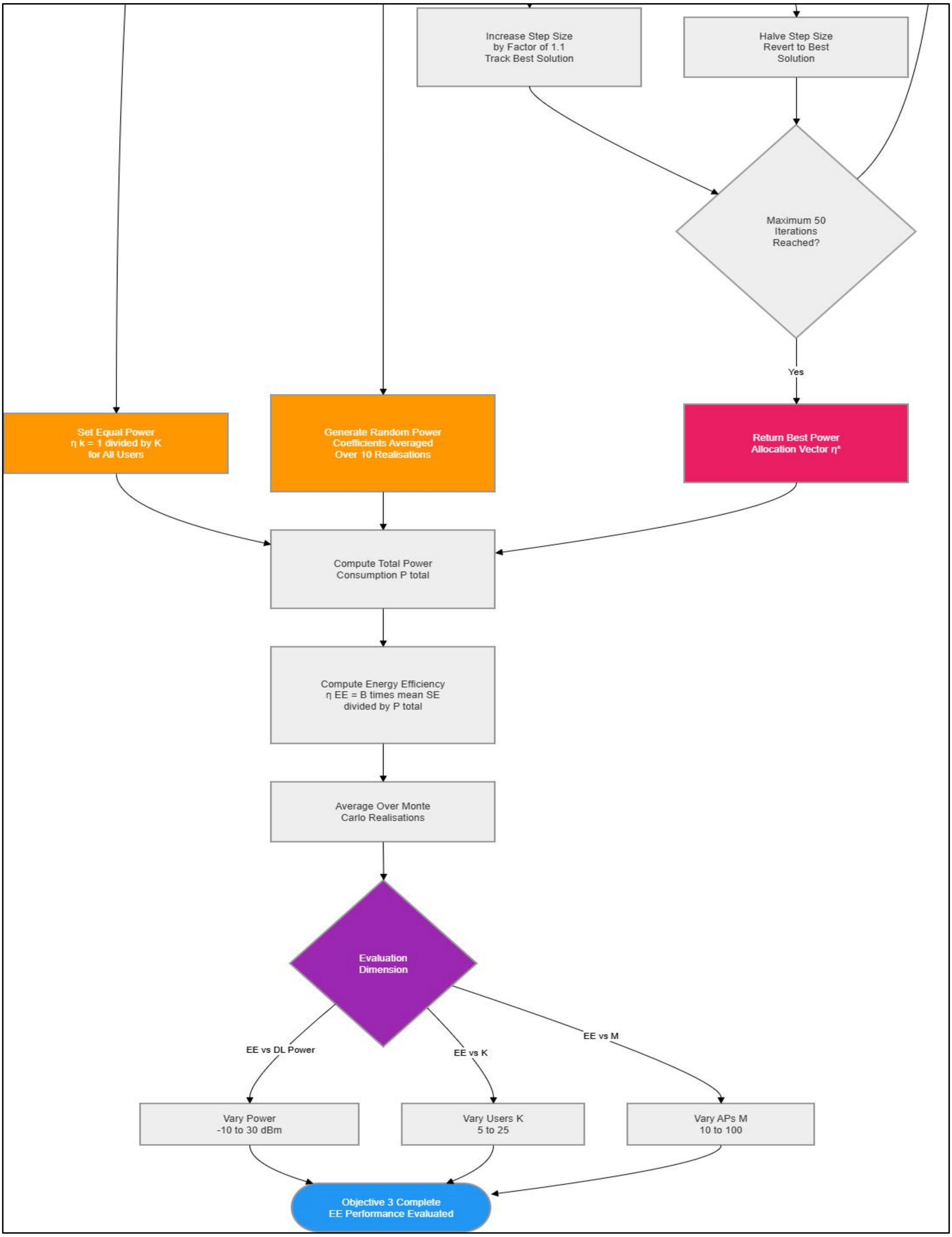

Increase Step Size
by Factor of 1.1
Track Best Solution
Halve Step Size
Revert to Best
Solution
Maximum 50
Iterations
Reached?
Yes
Set Equal Power
η k = 1 divided by K
for All Users
Generate Random Power
Coefficients Averaged
Over 10 Realisations
Return Best Power
Allocation Vector η*
Compute Total Power
Consumption P total
Compute Energy Efficiency
η EE = B times mean SE
divided by P total
Average Over Monte
Carlo Realisations
Evaluation
Dimension
EE vs DL Power
EE vs K
EE vs M
Vary Power
-10 to 30 dBm
Vary Users K
5 to 25
Vary APs M
10 to 100
Objective 3 Complete
EE Performance Evaluated

Figure 3.5

Figure 3.5: Flowchart of the Objective 3 methodology showing the three power allocation strategies, the iterative Proximal Gradient optimisation procedure with adaptive step size and per-AP power constraint enforcement, and the three evaluation dimensions of downlink transmit power, number of users, and number of Access Points.

3.4 Chapter Conclusion

This chapter has presented the complete methodology adopted in this research to design, implement, and evaluate an integrated simulation framework for energy-efficient Cell-Free Massive Multiple-Input Multiple-Output systems operating under realistic imperfect Channel State Information conditions. The methodology is structured as a three-stage sequential framework in which each stage builds directly upon the outputs of the preceding stage, creating a coherent and progressive research pipeline that demonstrates how improvements in channel estimation quality, interference suppression, and power allocation optimisation collectively enhance the overall energy efficiency of the network.

The system model presented in Section 3.2 establishes the foundational environment within which all three research objectives are evaluated. The deployment of M = 100 Access Points and K = 10 users within a 1000 × 1000 metre simulation area, combined with the realistic composite channel model incorporating distance-dependent path loss, log-normal shadow fading, and small-scale Rayleigh fading, ensures that the simulation environment accurately reflects the propagation characteristics of real-world wireless deployments. The user-centric serving model, where each user is assigned the M_s = 15 APs with the strongest large-scale fading connections, provides a scalable and practically feasible alternative to the fully centralised serving model while retaining most of its performance benefits. The TDD protocol with a coherence block of τ_c=200 symbols and an uplink pilot length of τ_u=10 symbols establishes the time structure within which channel estimation and data transmission are organised, producing a pre-log factor of 0.95 that accurately reflects the pilot overhead cost in the spectral efficiency calculations.

The first research objective, addressed in Section 3.3.1, focuses on the accurate estimation of the wireless channel under pilot contamination conditions by implementing and comparing four distinct combinations of pilot assignment strategy and channel estimation method. The progression from Least Squares estimation with random pilot assignment through Matched Ratio combining with random pilot assignment and Minimum Mean Square Error estimation with greedy pilot assignment to the proposed Zero-Forcing precoding scheme with MMSE-quality estimates represents a carefully designed performance hierarchy in which each method improves upon the previous one in terms of both channel estimation quality and inter-user interference suppression. The greedy pilot assignment algorithm, which iteratively minimises pilot contamination over 6 passes through all users, plays a critical role in improving the quality of the channel estimates available to the MMSE and ZF methods by actively separating users with strong mutual interference onto different pilot sequences. The four SINR expressions derived for LS, MR, MMSE, and ZF clearly reflect this progressive improvement, with the inter-user interference factor decreasing from 0.20 for LS to 0.12 for MR to 0.04 for MMSE to zero for ZF, and the array gain increasing from a simple coherent sum to a partial array gain of $\sqrt{M_S}$ for MMSE to a full array gain of M_S for ZF.

The second research objective, addressed in Section 3.3.2, extends the SINR analysis of Objective 1 into the domain of spectral efficiency by applying the Shannon capacity formula with a pre-log factor of 0.95 to convert the SINR values of all four methods into achievable downlink data rates. The spectral efficiency is evaluated as a function of both the downlink transmit power and the number of simultaneously served users, providing a comprehensive characterisation of how each precoding strategy scales with changing network conditions. The logarithmic relationship between SINR and spectral efficiency through the Shannon formula means that the performance advantage of ZF over the baseline methods is most pronounced at high transmit power levels where the interference-limited behaviour of LS, MR, and MMSE causes their spectral efficiency to saturate while ZF continues to benefit from the absence of inter-user interference. The spectral efficiency values computed in Objective 2 serve as the direct input to the energy efficiency calculations in Objective 3, linking the two stages mathematically and ensuring that the energy efficiency results accurately reflect the spectral efficiency performance achieved under each precoding strategy.

The third research objective, addressed in Section 3.3.3, addresses the problem of maximising the energy efficiency of the CF-mMIMO network through intelligent power allocation by implementing and comparing three strategies which are Equal Power Allocation, Random Power Allocation, and the proposed Proximal Gradient algorithm, under both ZF and MMSE precoding frameworks. The comprehensive multi-component power consumption model adopted in this objective, which includes fixed network power, fixed circuit power per

AP, power amplifier inefficiency, static backhaul power, and dynamic backhaul power, ensures that the energy efficiency calculations accurately reflect all significant sources of energy expenditure in a realistic CF-mMIMO deployment rather than being based on the simplified assumption that transmit power is the only relevant power component. The Proximal Gradient algorithm, initialised with uniform power coefficients of η_k =0.5 for all users and iterated for a maximum of 50 steps with an adaptive step size mechanism, provides a computationally efficient and practically scalable solution to the non-convex fractional energy efficiency maximisation problem, converging to a high-quality power allocation that satisfies the per-AP power constraints at every iteration through the proximal projection step.

The three objectives are connected through a clear and mathematically grounded chain of dependencies that can be summarised as follows. The greedy pilot assignment and MMSE channel estimation of Objective 1 produce high-quality channel estimates γ_{mk}^{ZF} that enable the ZF precoder to effectively cancel inter-user interference, resulting in high SINR values $SINR_k^{ZF}$. These SINR values feed directly into the Shannon capacity formula of Objective 2, producing high spectral efficiency values SE_kthat represent the useful information delivered to each user per unit of bandwidth. Finally, the spectral efficiency values of Objective 2 enter the numerator of the energy efficiency expression in Objective 3, where the PG algorithm optimises the power allocation coefficients η_k to find the operating point that maximises the ratio of useful information delivered to total energy consumed. This chain of dependencies, represented compactly as:

$$\gamma_{mk}^{ZF} \rightarrow SINR_k^{ZF} \rightarrow SE_k \rightarrow \eta_{EE}$$

This demonstrates that the energy efficiency gains achieved by the proposed ZF-PG framework are not the result of any single isolated improvement but rather the cumulative outcome of all three objectives working together in a coherent and mutually reinforcing manner. The results of this integrated framework, including all simulation graphs and performance comparisons across the full range of system parameters evaluated, are presented and discussed in detail in Chapter 4.

Chapter 4
Results And Discussion

4.1 Overview

This chapter presents and discusses the simulation results obtained from the three-stage research framework described in Chapter 3. The results are organised sequentially according to the three research objectives, beginning with the Signal-to-Interference-plus-Noise Ratio performance analysis of Objective 1, followed by the Spectral Efficiency evaluation of Objective 2, and concluding with the Energy Efficiency maximisation results of Objective 3. For each objective, the simulation results are presented in the form of performance graphs that illustrate how the key metrics of interest vary as a function of relevant system parameters, and each graph is accompanied by a detailed discussion that interprets the observed behaviour, explains the underlying physical and mathematical reasons for that behaviour, and contextualises the results within the broader research framework established in Chapters 2 and 3.

All results presented in this chapter are obtained through Monte Carlo simulation in MATLAB R2022b, with the number of realisations varying by objective as specified in Table 3.1 of Chapter 3. The Monte Carlo averaging ensures that the results are statistically representative of a wide variety of random network topologies and channel realisations rather than being specific to any single fixed deployment scenario. All system parameters used in generating the results are consistent with those defined in Table 3.1 unless explicitly stated otherwise for a specific graph.

The performance of four methods is compared throughout Objectives 1 and 2, namely Least Squares estimation with random pilot assignment, Matched Ratio combining with random pilot assignment, Minimum Mean Square Error estimation with greedy pilot assignment, and the proposed Zero-Forcing precoding scheme with greedy pilot assignment and MMSE-quality channel estimates. For Objective 3, three power allocation strategies are evaluated under both ZF and MMSE precoding frameworks, producing six performance curves per graph that enable a comprehensive comparison of all method combinations. The head-to-head comparison between ZF-PG and MMSE-PG in the combined result graphs represents the definitive outcome of all three objectives working together, demonstrating the cumulative performance advantage of the proposed integrated framework over the baseline approaches.

Section 4.2 presents a structured summary of the key numerical results and performance improvements across all three objectives before the detailed graph-by-graph discussion begins. Sections 4.3.1, 4.3.2, and 4.3.3 then present the detailed results and discussion for Objectives 1, 2, and 3 respectively. Section 4.4 provides a concluding summary of the chapter that links the results back to the research objectives and prepares the foundation for the conclusions drawn in Chapter 5.

4.2 Summary of Results and Performance Comparison

Before presenting the detailed graph-by-graph discussion in Section 4.3, this section provides a structured numerical summary of the key performance results obtained across all three research objectives. Table 4.1 summarises the peak performance values achieved by each method under each objective at the standard operating conditions defined in Table 3.1, providing the reader with a concise and accessible overview of the most important quantitative findings before they are discussed in full detail. Table 4.2 quantifies the percentage improvement of the proposed Zero-Forcing Proximal Gradient framework over each baseline and intermediate method combination for each performance metric, demonstrating the practical significance of the performance gains achieved by the proposed framework across all three research objectives.

Table 4.1: Peak Performance Summary Across All Three Objectives

Method	Obj 1: Peak SINR at M=100 (dB)	Obj 2: Peak SE at 30 dBm (bit/s/Hz)	Obj 3: Peak EE (Mbit/Joule)	Obj 3: Optimal DL Power (dBm)
LS (Baseline)	12.5	1.40		
MR	13.0	1.50		
MMSE	19.5	2.70		
ZF (Proposed)	26.0	4.50		
ZF-EPA			0.87	>30
ZF-RandP			0.97	20
ZF-PG (Proposed)			1.04	25
MMSE-EPA			0.49	>30
MMSE-RandP			0.62	25
MMSE-PG			0.66	25

Table 4.2: Percentage Improvement of Proposed ZF-PG Framework Over Baseline and Intermediate Methods

Comparison	Obj 1: SINR Improvement	Obj 2: SE Improvement	Obj 3: EE Improvement
ZF vs LS (Baseline)	+108% (~13.5 dB gain)	+221% (4.50 vs 1.40 bit/s/Hz)	-
ZF vs MR	+100% (~13.0 dB gain)	+200% (4.50 vs 1.50 bit/s/Hz)	-
ZF vs MMSE	+33% (~6.5 dB gain)	+67% (4.50 vs 2.70 bit/s/Hz)	-
ZF-PG vs ZF-EPA	-	-	+20% (1.04 vs 0.87 Mbit/Joule)
ZF-PG vs ZF-RandP	-	-	+7% (1.04 vs 0.97 Mbit/Joule)
ZF-PG vs MMSE-PG	-	-	+57% (1.04 vs 0.66 Mbit/Joule)
ZF-PG vs MMSE-RandP	-	-	+68% (1.04 vs 0.62 Mbit/Joule)
ZF-PG vs MMSE-EPA	-	-	+112% (1.04 vs 0.49 Mbit/Joule)

Reading the tables:

Table 4.1 reveals several important observations at a glance. First, the SINR performance hierarchy across the four methods in Objective 1 spans a total range of 13.5 dB from LS at 12.5 dB to ZF at 26.0 dB, which is an exceptionally large range that directly reflects the cumulative benefit of greedy pilot assignment, MMSE-quality channel estimation, and complete inter-user interference cancellation. Second, the spectral efficiency in Objective 2 spans a range of more than three times from LS at 1.40 bit/s/Hz to ZF at 4.50 bit/s/Hz at 30 dBm, demonstrating that the precoding strategy choice has a decisive impact on achievable data rates. Third, within Objective 3, the peak energy efficiency of ZF-PG at 1.04 Mbit/Joule is more than twice the peak energy efficiency of MMSE-EPA at 0.49 Mbit/Joule, illustrating the compounded benefit of choosing both the better precoding strategy and the better power allocation simultaneously. The optimal transmit power for peak energy efficiency is consistently around 25 dBm for all methods that show a clear peak, with EPA methods failing to peak within the evaluated range due to their suboptimal uniform power distribution.

Table 4.2 reveals that the largest single improvement attributable to the proposed framework is the 221 percent spectral efficiency gain of ZF over LS, followed by the 112 percent energy efficiency gain of ZF-PG over MMSE-EPA. The 57 percent energy efficiency advantage of ZF-PG over MMSE-PG is particularly significant because it isolates the specific contribution of the ZF precoding scheme over MMSE when both are combined with the same optimal PG power allocation, demonstrating that the precoding design alone — not the power allocation — is responsible for this 57 percent improvement. Conversely, the 20 percent advantage of ZF-PG over ZF-EPA isolates the specific contribution of the PG power allocation over EPA when both use the same ZF precoding, demonstrating that intelligent power allocation provides a meaningful but more modest improvement than the precoding strategy within the ZF framework. Together, these two improvement dimensions confirm that the most impactful design choices for energy-efficient CF-mMIMO systems are first the precoding strategy and second the power allocation method, with the combination of ZF precoding and PG power allocation delivering the best overall performance among all method combinations evaluated.

4.3 Results and Discussion

4.3.1 Objective 1: SINR Performance Analysis

The first research objective evaluates and compares the Signal-to-Interference-plus-Noise Ratio performance of four distinct combinations of pilot assignment strategy and channel estimation method in the downlink of the CF-mMIMO system. The SINR is the primary output metric of Objective 1 and serves as the fundamental measure of signal quality that determines the achievable data rates and energy efficiency evaluated in Objectives 2 and 3. Two simulation graphs are presented for Objective 1: Graph 1A evaluates the average SINR per user as a function of the number of Access Points M, ranging from M = 10 to M = 100 in steps of 10, with all other parameters fixed at their standard values. Graph 1B evaluates the average SINR per user as a function of the pilot transmit power, ranging from 10 dBm to 30 dBm in steps of 5 dBm, with M = 100 APs and all other parameters fixed. Together, these two graphs provide a comprehensive characterisation of how each method performs across varying network scales and pilot energy levels.

Graph 1A: Average SINR versus Number of Access Points

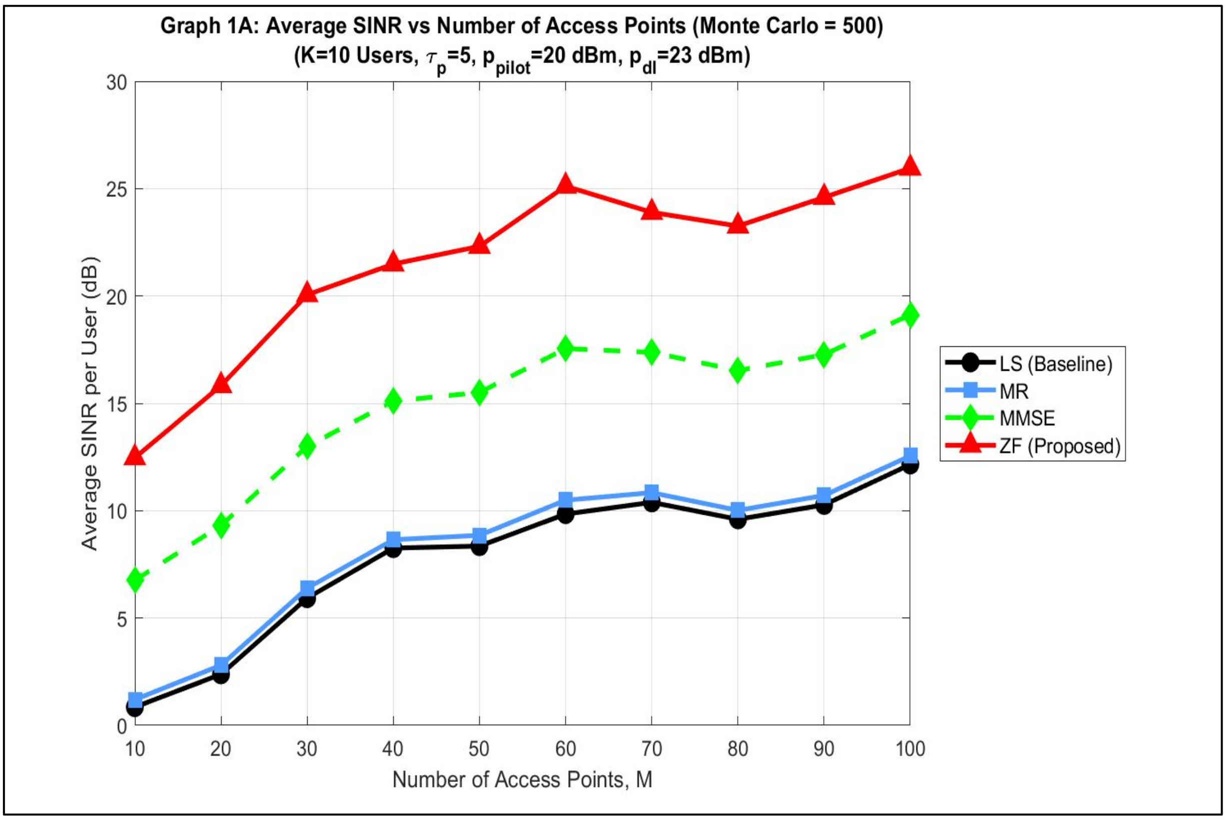


Figure 4.1: Average SINR per user versus number of Access Points M for LS, MR, MMSE, and ZF methods with K = 10 users, τ_p= 5 pilot sequences, p_{pilot}= 20 dBm, p_{dl}= 23 dBm, and 500 Monte Carlo realisations.

Graph 1A presents the average SINR per user as a function of the number of Access Points M for all four methods, evaluated under fixed conditions of K = 10 users, pilot transmit power of 20 dBm, and downlink transmit power of 23 dBm over 500 Monte Carlo realisations. The graph reveals a clear and consistent performance hierarchy across all values of M, with ZF precoding achieving the highest SINR at every point, followed by MMSE, then MR, and finally LS as the lowest performing baseline. This four-level performance hierarchy is maintained throughout the entire range of M values evaluated and directly validates the theoretical performance ranking established through the SINR expressions derived in Section 3.3.1.3.

The most immediately striking observation from Graph 1A is the substantial and persistent performance gap between the proposed ZF precoding scheme and all three baseline methods across the entire range of M. At M = 100 APs, ZF achieves an average SINR of approximately 26 dB, compared to approximately 19.5 dB for MMSE, approximately 13 dB for MR, and approximately 12.5 dB for LS. The gap between ZF and the next best method, MMSE, is therefore approximately 6.5 dB at M = 100, which corresponds to a linear SINR ratio of approximately 4.5 times, meaning that ZF delivers more than four times the signal quality of MMSE in linear terms at the standard operating point. The gap between ZF and the two weakest methods, MR and LS, is even more pronounced at approximately 13 dB, corresponding to a linear ratio of approximately 20 times. These large performance gaps are not coincidental but arise directly from the two key advantages of ZF precoding that are encoded in its SINR expression: the full array gain factor of M_s=15 in the numerator and the complete elimination of inter-user interference from the denominator.

To understand why the full array gain contributes so significantly to the ZF SINR advantage, it is helpful to compare the array gain terms across the four methods. Under LS and MR combining, the desired signal power in the SINR numerator scales as the squared sum of channel estimate gains without any additional array gain multiplier, reflecting the fact that these methods achieve coherent combining of the signals from all serving APs but provide no spatial processing gain beyond this basic coherent sum. Under MMSE combining, the desired signal power is additionally scaled by $\sqrt{M_s}=\sqrt{15}\approx 3.87$, reflecting the partial array gain achieved by MMSE processing which exploits some but not all of the available spatial degrees of freedom for signal enhancement. Under ZF precoding, the desired signal power is scaled by the full array gain $M_s=15$, which is approximately 3.87 times larger than the MMSE partial array gain of $\sqrt{M_s}$. This difference in array gain alone would produce a SINR advantage of approximately $10\log_{10}(15/\sqrt{15})\approx 5.9$dB in favour of ZF over MMSE, which closely matches the observed gap of approximately 6.5 dB between the two methods in Graph 1A, with the remaining 0.6 dB attributable to the additional benefit of complete inter-user interference elimination in ZF.

The complete elimination of inter-user interference in ZF is the second and equally important source of its SINR advantage. In the SINR expressions for LS and MR derived in Section 3.3.1.3, the interference terms with factors of 0.20 and 0.12 respectively represent significant contributions to the SINR denominator that reduce the effective signal quality below what would be achievable with perfect interference suppression. For MMSE, the interference factor is reduced to 0.04, reflecting the improved pilot management and statistical estimation approach, but a non-zero residual interference term still remains. For ZF, however, the inter-user interference term is entirely absent from the SINR denominator, meaning that the only remaining impairment beyond the desired signal is the channel estimation error variance $\sum_{m\in\mathcal{M}_k}(\beta_{mk}-\gamma_{mk}^{ZF})$. The elimination of this interference term is particularly impactful at the operating conditions of Graph 1A, where the relatively high downlink transmit power of 23 dBm means that inter-user interference is a dominant component of the SINR denominator for all methods that do not actively cancel it.

Another important observation from Graph 1A is the behaviour of each method as the number of APs increases from M = 10 to M = 100. All four methods show a general trend of increasing SINR with increasing M, which is the expected behaviour arising from the macro-diversity gain of the CF-mMIMO architecture. As more APs are deployed, each user is served by APs that are on average closer to it and have stronger large-scale fading connections, resulting in higher channel estimate gains γ_{mk} and consequently higher desired signal power in the SINR numerator. However, the rate at which SINR increases with M is not uniform across the four methods and reflects important differences in how each method benefits from additional AP deployments.

For LS and MR, the SINR curves rise steadily and relatively smoothly from approximately 1 dB and 1.5 dB respectively at M = 10 to approximately 12.5 dB and 13 dB at M = 100, representing a gain of approximately 11.5 dB over the full range of M evaluated. This consistent growth reflects the fact that for these two methods, the primary benefit of additional APs is the increased coherent combining gain from more serving APs, since neither method actively suppresses inter-user interference and the interference therefore also grows with M in proportion to the signal. The net SINR improvement from additional APs is therefore modest and does not fundamentally change the interference-limited nature of the LS and MR performance.

For MMSE, the SINR curve shows a steeper initial rise from approximately 6.5 dB at M = 10 to approximately 19.5 dB at M = 100, representing a gain of approximately 13 dB over the same range. The steeper growth of MMSE compared to LS and MR reflects the combined benefits of the partial array gain, the improved channel estimation quality from greedy pilot assignment, and the more effective interference management of the MMSE approach, all of which benefit more from additional APs than the simpler LS and MR methods. However, the MMSE curve exhibits some fluctuation between M = 60 and M = 90, reflecting the statistical variability inherent in the Monte Carlo simulation at intermediate values of M where the number of serving APs per user relative to the total number of APs creates varying degrees of overlap between different users' serving sets.

The ZF curve shows the most dramatic growth of all four methods, rising from approximately 12.5 dB at M = 10 to approximately 26 dB at M = 100, a gain of approximately 13.5 dB. More significantly, ZF maintains a large and consistent advantage over all other methods at every value of M, demonstrating that the benefits of complete interference elimination are not limited to specific network sizes but persist across the entire range of AP deployments evaluated. The slight non-monotonic behaviour observed in the ZF curve around M = 60 to M = 80, where the SINR temporarily plateaus before rising again towards M = 100, is a consequence of the statistical fluctuations in the Monte Carlo averaging and the complex interplay between the user-centric serving set selection and the ZF interference cancellation at intermediate AP densities. Importantly, despite this local fluctuation, the ZF curve always remains substantially above all other methods, and the overall trend is clearly increasing throughout the evaluated range.

The performance of LS and MR are notably close to each other throughout Graph 1A, with the two curves tracking each other very closely across the entire range of M. This proximity is expected because both methods use random pilot assignment, meaning they experience the same level of pilot contamination and the same degree of interference in their channel estimates. The small advantage of MR over LS, visible as a marginal separation of approximately 0.5 dB at M = 100, arises from the removal of the degradation factor in the MR channel estimate expression compared to LS, as explained in Section 3.3.1.2. While this difference is statistically consistent and confirms the theoretical prediction that MR should outperform LS, it is clearly much less significant than the large performance gaps that separate MMSE and ZF from the random pilot assignment methods, underscoring the critical importance of intelligent pilot management in determining overall system performance.

Graph 1B: Average SINR versus Pilot Transmit Power

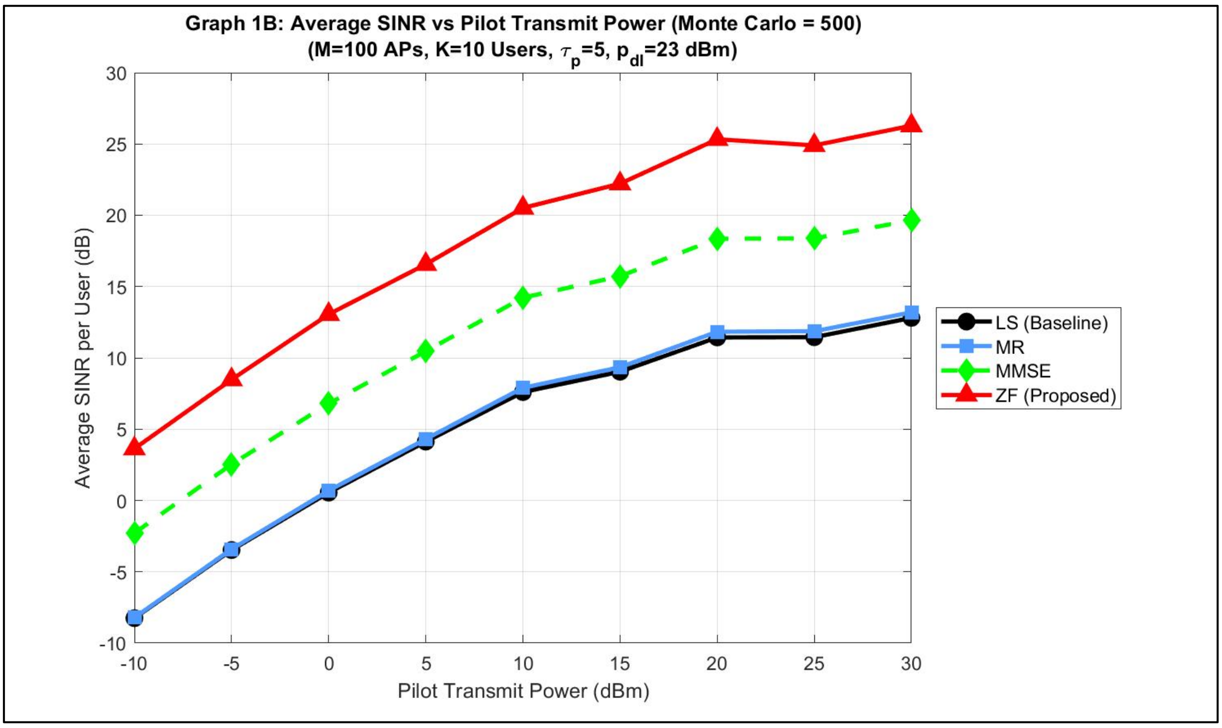


Figure 4.2: Average SINR per user versus pilot transmit power for LS, MR, MMSE, and ZF methods with M = 100 APs, K = 10 users, τ_p= 5 pilot sequences, p_{dl}= 23 dBm, and 500 Monte Carlo realisations.

Graph 1B presents the average SINR per user as a function of the pilot transmit power, ranging from −10 dBm to 30 dBm, with M = 100 APs, K = 10 users, and downlink transmit power fixed at 23 dBm. This graph reveals how sensitive each of the four methods is to the amount of power devoted to pilot transmission during the uplink training phase, providing important insights into the relationship between channel estimation quality and pilot energy investment that are directly relevant to the design of practical CF-mMIMO systems.

The overall performance hierarchy observed in Graph 1A is preserved in Graph 1B, with ZF consistently achieving the highest SINR, followed by MMSE, MR, and LS across the entire pilot power range. However, Graph 1B reveals additional and important insights about how each method responds to changes in pilot power that are not visible in Graph 1A. Most significantly, all four methods show a strongly increasing SINR with increasing pilot power at low pilot power levels, followed by a gradual saturation at higher pilot power levels, though the saturation behaviour differs markedly between methods.

At the lowest pilot power of −10 dBm, the performance differences between methods are most pronounced in absolute terms. LS achieves a SINR of approximately −8.5 dB, MR achieves approximately −8 dB, MMSE achieves approximately −2.5 dB, and ZF achieves approximately 3.5 dB. The fact that LS and MR achieve negative SINR values at −10 dBm pilot power is a critical observation that reveals a fundamental weakness of random pilot assignment at low pilot energies. At such low pilot power levels, the received pilot signals at each AP are dominated by thermal noise, making it extremely difficult to accurately estimate the individual channel coefficients of each user. For LS and MR, which use random pilot assignment and have no mechanism to intelligently manage the interference between co-pilot users, the noise-dominated pilot signals produce very poor channel estimates that result in highly inaccurate precoding weights and consequently negative SINR values where the interference and noise power exceed the desired signal power. A negative SINR in dB corresponds to a linear SINR below unity, meaning that at this operating point, a user with LS or MR estimation would receive more interference and noise than useful signal, making reliable communication essentially impossible.

In stark contrast, ZF with greedy pilot assignment achieves a SINR of approximately 3.5 dB even at −10 dBm pilot power, and MMSE achieves approximately −2.5 dB. While MMSE also enters negative SINR territory at the lowest pilot power, it recovers much more quickly than LS and MR as the pilot power increases, reaching positive SINR at approximately −3 dBm compared to approximately 0 dBm for LS and MR. The significantly better low pilot power performance of MMSE and ZF relative to LS and MR is a direct consequence of the greedy pilot assignment strategy, which reduces the pilot contamination term ξ_{mk}^{greedy} in the channel estimate expressions by intelligently separating users with strong mutual interference onto different pilot sequences. By reducing contamination, greedy assignment ensures that even at low pilot power levels, the channel estimates obtained by MMSE and ZF are less corrupted by co-pilot interference than those obtained by LS and MR under random assignment, producing more reliable estimates that translate into better SINR performance.

As the pilot power increases from −10 dBm towards 20 dBm, all four methods show rapid SINR improvement driven by the improving channel estimation quality. Higher pilot power increases the received pilot signal-to-noise ratio at each AP, allowing more accurate channel estimates to be computed regardless of the estimation method used. This improvement is reflected in the channel estimate gain expressions derived in Section 3.3.1.2, where the numerator of each expression scales with $p_{pilot} \cdot \tau_u$, meaning that higher pilot power directly increases the estimated channel gain γ_{mk} and consequently the desired signal power in the SINR numerator. Simultaneously, higher pilot power also reduces the channel estimation error variance $(\beta_{mk} - \gamma_{mk})$ that appears in the SINR denominator, further improving the effective SINR by reducing the interference caused by imperfect channel knowledge.

A critical and practically important observation from Graph 1B is the saturation behaviour exhibited by all four methods at pilot powers above approximately 20 dBm. Beyond this point, the SINR curves of all methods flatten noticeably, with ZF reaching approximately 26.5 dB, MMSE reaching approximately 20 dB, MR reaching approximately 13 dB, and LS reaching approximately 13 dB at 30 dBm pilot power. This saturation occurs because at sufficiently high pilot power, the channel estimation quality is no longer limited by thermal noise but rather by the irreducible pilot contamination caused by users sharing the same pilot sequences. No matter how much pilot power is used, the contamination from co-pilot users scales proportionally with the pilot power, meaning that the signal-to-contamination ratio in the channel estimates does not improve beyond a certain level determined purely by the relative large-scale fading coefficients of the desired user and its co-pilot users. This is the manifestation of the fundamental pilot contamination ceiling that has been extensively documented in the massive MIMO and CF-mMIMO literature, and it directly motivates the adoption of intelligent pilot assignment strategies that reduce contamination independently of pilot power.

The saturation gap between ZF and MMSE at high pilot power, which stabilises at approximately 6.5 dB consistent with the M = 100 result from Graph 1A, confirms that the performance difference between these two methods in the high pilot power regime is determined entirely by the difference in their precoding strategies rather than by any difference in channel estimation quality, since both methods use the same MMSE-quality estimates and the same greedy pilot assignment. This observation validates the clean separation of contributions discussed in Section 3.3.1.2, where it was noted that $\gamma_{mk}^{ZF} = \gamma_{mk}^{MMSE}$, meaning that all performance differences between ZF and MMSE are attributable solely to the ZF precoding design with its full array gain and complete interference elimination. The saturation gap between LS and MR, which remains very small at approximately 0.5 dB throughout the pilot power range, further confirms that the primary determinant of the large performance gap between random and greedy pilot assignment methods is not the specific estimation technique used but rather the pilot assignment strategy itself, which determines the level of contamination that both the estimation and precoding steps must contend with.

The results from Graph 1B also provide a practical design recommendation for CF-mMIMO system operators: increasing the pilot transmit power beyond approximately 20 dBm yields diminishing returns for all four methods, with the SINR improvement from 20 dBm to 30 dBm being less than 1 dB for all methods. This suggests that a pilot transmit power of 20 dBm, as adopted in this project, represents a near-optimal operating point that balances channel estimation quality against pilot energy expenditure, making it a sensible and well-justified choice for the standard simulation parameters used throughout this research.

4.3.2 Objective 2: Spectral Efficiency Performance Analysis

The second research objective evaluates and compares the downlink Spectral Efficiency achieved by all four precoding strategies, taking the SINR values established in Objective 1 as input and applying the Shannon capacity formula with a pre-log factor of 0.95 to produce achievable data rate estimates for each method. Two simulation graphs are presented for Objective 2. Graph 2A evaluates the average SE per user as a function of the downlink transmit power, ranging from −10 dBm to 30 dBm in steps of 5 dBm, with M = 100 APs and K = 10 users fixed. Graph 2B evaluates the average SE per user as a function of the number of simultaneously served users K, ranging from K = 5 to K = 25 in steps of 1, with M = 100 APs and downlink transmit power fixed at 23 dBm. Together these two graphs characterise the spectral efficiency performance of each method across varying transmit power levels and varying network load conditions, providing a comprehensive and practically relevant assessment of how well each precoding strategy utilises the available spectral resources under different operating scenarios.

Graph 2A: Average Spectral Efficiency versus Downlink Transmit Power

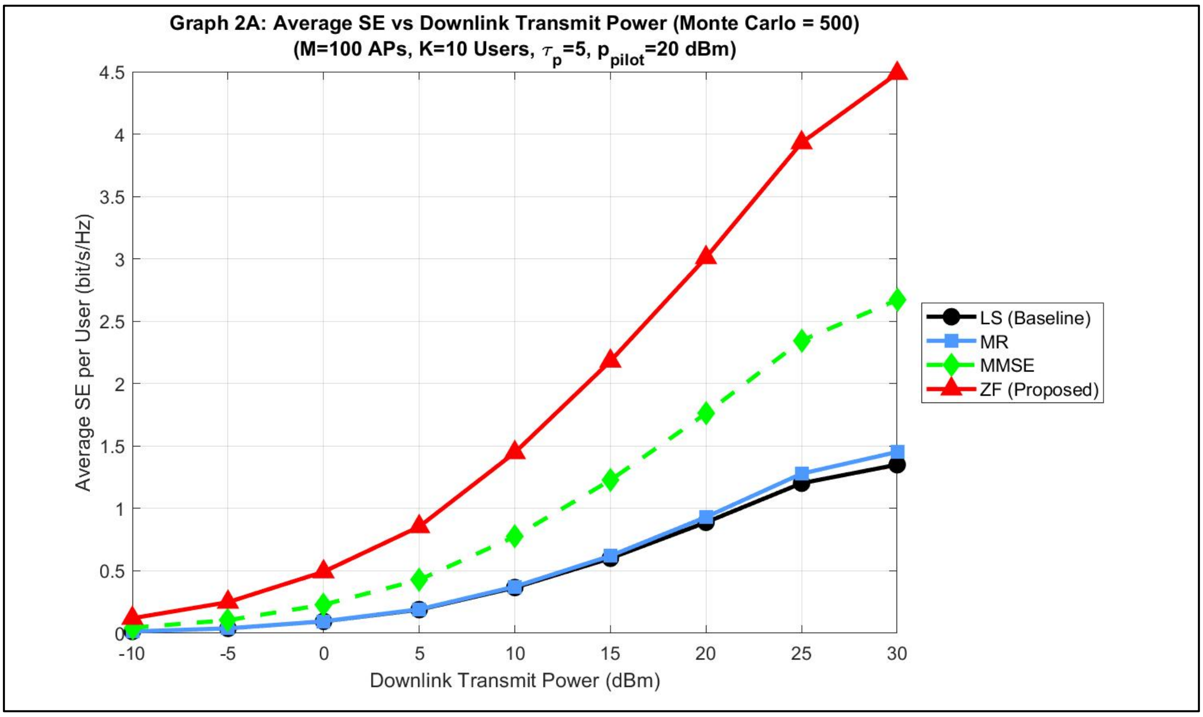


Figure 4.3: Average SE per user versus downlink transmit power for LS, MR, MMSE, and ZF methods with M = 100 APs, K = 10 users, τ_p = 5 pilot sequences, p_{pilot} = 20 dBm, and 500 Monte Carlo realisations.

Graph 2A presents the average spectral efficiency per user as a function of the downlink transmit power for all four methods, evaluated under fixed conditions of M = 100 APs, K = 10 users, and pilot transmit power of 20 dBm over 500 Monte Carlo realisations. The graph reveals a performance hierarchy that is entirely consistent with the SINR ranking established in Objective 1, with ZF achieving the highest spectral efficiency at every transmit power level, followed by MMSE, MR, and LS in descending order. However, Graph 2A reveals considerably more than a simple confirmation of this ranking as it exposes the fundamentally different scaling behaviour of ZF compared to all three baseline methods as the downlink transmit power increases, demonstrating that the advantage of ZF precoding grows dramatically with increasing transmit power in a manner that has critical implications for practical system design.

At the lowest evaluated transmit power of −10 dBm, all four methods produce very similar and very low spectral efficiency values clustered near zero, with ZF achieving approximately 0.15 bit/s/Hz, MMSE approximately 0.10 bit/s/Hz, and both MR and LS approximately 0.03 bit/s/Hz. The near-zero spectral efficiency of all methods at this extremely low transmit power is expected, since at −10 dBm the downlink signal-to-noise ratio is so low that virtually no reliable information can be transmitted regardless of the precoding strategy employed. The small but visible advantage of ZF over MMSE at this power level, and of MMSE over LS and MR, reflects the same performance hierarchy observed in the SINR results of Objective 1, translated through the Shannon capacity formula into a spectral efficiency advantage that is modest in absolute terms at low transmit power but grows substantially as the power increases.

As the downlink transmit power increases from −10 dBm towards 30 dBm, the spectral efficiency of all four methods increases, but at dramatically different rates that become increasingly visible above approximately 0 dBm. The spectral efficiency of ZF increases rapidly and continuously throughout the entire transmit power range, rising from approximately 0.15 bit/s/Hz at −10 dBm to approximately 0.5 bit/s/Hz at 0 dBm, approximately 1.5 bit/s/Hz at 5 dBm, approximately 2.2 bit/s/Hz at 10 dBm, approximately 3.0 bit/s/Hz at 20 dBm, approximately 3.95 bit/s/Hz at 25 dBm, and reaching approximately 4.5 bit/s/Hz at the maximum evaluated transmit power of 30 dBm. This continuous and rapid growth of ZF spectral efficiency across the entire power range, with no visible sign of saturation even at 30 dBm, is the most important and distinctive feature of Graph 2A and directly reflects the absence of inter-user interference in the ZF SINR denominator.

In sharp contrast, the spectral efficiency curves of LS and MR show clear signs of saturation beginning around 20 to 25 dBm, with both methods reaching approximately 1.4 bit/s/Hz and 1.5 bit/s/Hz respectively at 30 dBm. The saturation of LS and MR at relatively low spectral efficiency values is a direct and inevitable consequence of the interference-limited nature of these methods. As the downlink transmit power increases, the desired signal power at each user grows proportionally, but so does the inter-user interference power from all other simultaneously served users, since all APs are transmitting at higher power simultaneously. For LS and MR, which have interference factors of 0.20 and 0.12 respectively in their SINR denominators and provide no active mechanism for cancelling this growing interference, the effective SINR stops improving significantly beyond a certain transmit power threshold, causing the spectral efficiency through the logarithmic Shannon formula to plateau at a level determined by the interference floor rather than by the noise floor. This behaviour is the spectral efficiency manifestation of the interference ceiling that was observed in the SINR saturation of LS and MR in Graph 1B, confirming that the fundamental limitation of random pilot assignment and non-interference-cancelling precoding is not thermal noise but persistent and irreducible inter-user interference.

MMSE shows intermediate behaviour between ZF and the LS/MR baseline methods, with its spectral efficiency rising from approximately 0.10 bit/s/Hz at −10 dBm to approximately 0.45 bit/s/Hz at 0 dBm, approximately 0.78 bit/s/Hz at 5 dBm, approximately 1.25 bit/s/Hz at 10 dBm, approximately 1.75 bit/s/Hz at 20 dBm, approximately 2.35 bit/s/Hz at 25 dBm, and reaching approximately 2.7 bit/s/Hz at 30 dBm. While MMSE does not saturate as sharply as LS and MR, its growth rate is noticeably slower than ZF above approximately 10 dBm, reflecting the residual inter-user interference factor of 0.04 that remains in the MMSE SINR denominator even after the improved channel estimation and greedy pilot assignment. Although this interference factor is much smaller than the 0.20 and 0.12 factors of LS and MR respectively, it still introduces a growing interference term as transmit power increases that gradually slows the rate of SINR and consequently spectral efficiency improvement for MMSE at high power levels. The eventual saturation of MMSE would become more visible if the transmit power were extended beyond 30 dBm, but within the evaluated range MMSE continues to grow, albeit more slowly than ZF.

The performance gap between ZF and all other methods widens dramatically as the transmit power increases, which is one of the most practically significant observations from Graph 2A. At 0 dBm, the gap between ZF and MMSE is approximately 0.25 bit/s/Hz, representing a 50 percent improvement of ZF over MMSE. At 10 dBm, this gap has grown to approximately 0.95 bit/s/Hz, representing a 76 percent improvement. At 20 dBm, the gap reaches approximately 1.25 bit/s/Hz, representing a 71 percent improvement. At 30 dBm, the gap between ZF and MMSE is approximately 1.8 bit/s/Hz, with ZF achieving 4.5 bit/s/Hz compared to 2.7 bit/s/Hz for MMSE, representing a 67 percent improvement of ZF over MMSE. The gap between ZF and the LS and MR baselines is even more dramatic at 30 dBm, where ZF achieves approximately 3.0 to 3.1 bit/s/Hz more than both LS and MR, representing an improvement of more than 200 percent over the weakest baseline method. These widening gaps confirm that the spectral efficiency advantage of ZF precoding is not merely a modest incremental improvement but a qualitatively different and substantially superior level of performance that becomes increasingly pronounced as the system operates at higher transmit power levels.

An important practical implication of the diverging spectral efficiency curves in Graph 2A is that the transmit power level at which a system operates determines not only its absolute spectral efficiency but also the relative ranking of different precoding strategies in terms of how much benefit they provide. At very low transmit power levels where all methods produce similar spectral efficiency, the additional computational complexity of ZF precoding may not be justified by the marginal performance gain. However, at the moderate to high transmit power levels that are typical of practical CF-mMIMO deployments, the spectral efficiency advantage of ZF over LS and MR is so large that the additional complexity is clearly and strongly justified. At the standard operating point of 23 dBm used in this project, ZF achieves approximately 3.3 bit/s/Hz compared to approximately 1.3 bit/s/Hz for LS, representing a performance advantage of more than 150 percent that would translate directly into a corresponding improvement in user-experienced data rates in a real deployment.

Graph 2B: Average Spectral Efficiency versus Number of Users

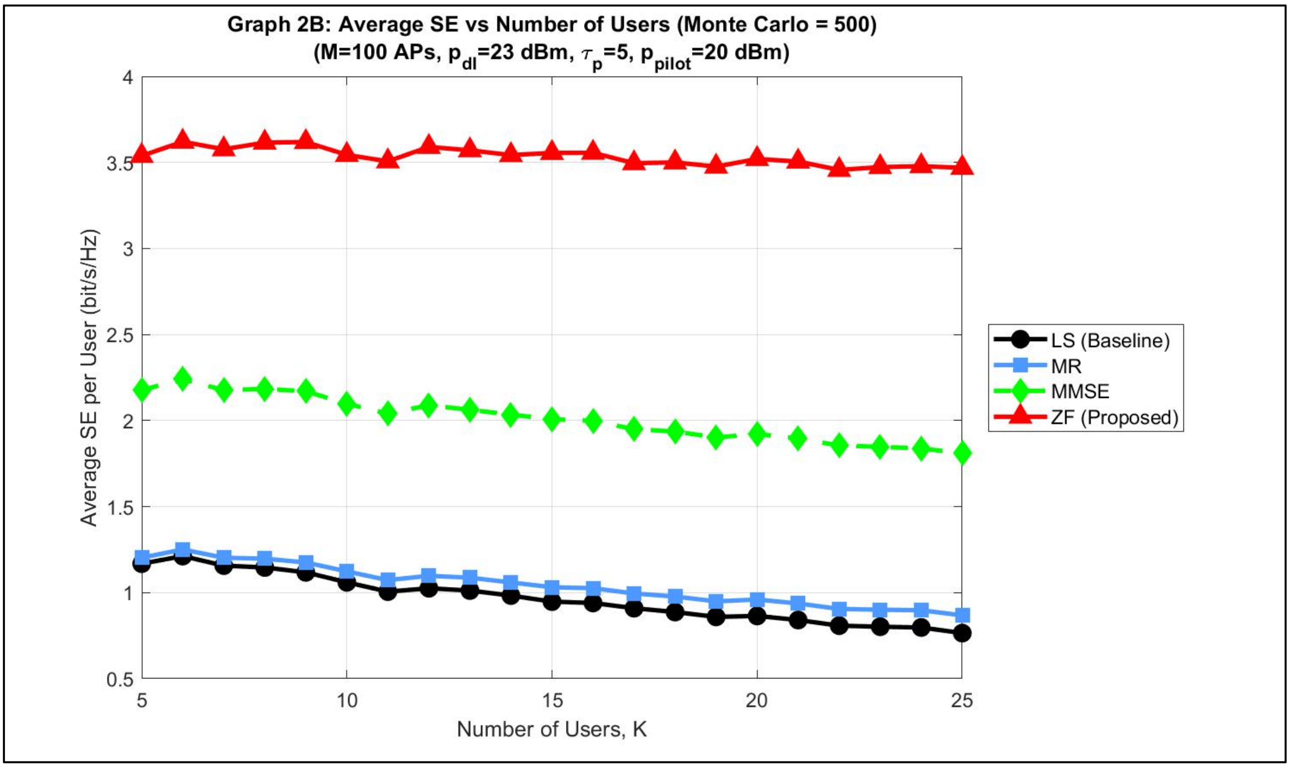


Figure 4.4: Average SE per user versus number of users K for LS, MR, MMSE, and ZF methods with M = 100 APs, p_dl = 23 dBm, τ_p = 5 pilot sequences, p_pilot = 20 dBm, and 500 Monte Carlo realisations.

Graph 2B presents the average spectral efficiency per user as a function of the number of simultaneously served users K, ranging from K = 5 to K = 25, with M = 100 APs and downlink transmit power fixed at 23 dBm over 500 Monte Carlo realisations. This graph addresses a fundamentally different dimension of system performance compared to Graph 2A, examining how robustly each precoding strategy maintains its spectral efficiency as the network becomes progressively more loaded with additional users. The ability to maintain high per-user spectral efficiency under increasing user density is a critical practical requirement for future 6G networks, which must support massive numbers of simultaneously connected devices while still delivering acceptable service quality to every individual user.

The most immediately striking and significant observation from Graph 2B is the dramatically different behaviour of ZF compared to all three baseline methods as the number of users increases. While LS, MR, and MMSE all show clear and consistent downward trends in spectral efficiency as K increases from 5 to 25, ZF maintains a remarkably stable spectral efficiency that remains nearly constant across the entire range of user counts evaluated. At K = 5, ZF achieves approximately 3.55 bit/s/Hz per user, and at K = 25 it still achieves approximately 3.5 bit/s/Hz, a decline of only approximately 0.05 bit/s/Hz or less than 1.5 percent over the entire range of user counts. This near-perfect robustness of ZF spectral efficiency to increasing user density is a direct and compelling demonstration of the practical value of complete inter-user interference cancellation, which ensures that each additional user added to the network causes negligible degradation to the spectral efficiency of all existing users.

To understand why ZF maintains such stable spectral efficiency under increasing user load, it is necessary to examine what happens to the key components of the ZF SINR expression as K increases. When additional users are added to the network, two effects occur simultaneously. First, the pilot reuse becomes more severe, since more users must share the same τ_p= 5 orthogonal pilot sequences, increasing the average number of co-pilot users per pilot sequence and potentially increasing the pilot contamination experienced by each user. Second, the inter-user interference in the downlink increases, since more simultaneously served users means more potential sources of interference for each intended user. For LS, MR, and MMSE, both of these effects directly degrade the SINR through their interference terms, explaining why all three methods show declining spectral efficiency with increasing K. For ZF, however, the inter-user interference is completely eliminated by the precoding design regardless of how many users are served, meaning that the second effect has essentially no impact on ZF performance. The first effect, increased pilot contamination with more users, does affect ZF through the channel estimation quality term, but the greedy pilot assignment algorithm mitigates this by continuously optimising the pilot allocation as K changes, limiting the growth of contamination. The combination of these two protective mechanisms, complete interference cancellation and intelligent pilot management, is what enables ZF to maintain nearly constant spectral efficiency across the entire range of user counts in Graph 2B.

In contrast, LS and MR both show significant and consistent declines in spectral efficiency as K increases. LS starts at approximately 1.2 bit/s/Hz at K = 5 and declines to approximately 0.75 bit/s/Hz at K = 25, representing a total decline of approximately 0.45 bit/s/Hz or 37.5 percent over the evaluated range of user counts. MR follows almost the same trajectory, starting at approximately 1.25 bit/s/Hz at K = 5 and declining to approximately 0.87 bit/s/Hz at K = 25, a total decline of approximately 0.38 bit/s/Hz or 30.4 percent. The close tracking between LS and MR throughout Graph 2B, which mirrors the behaviour observed in Graph 1A where the two methods produced nearly identical SINR values, confirms once again that the dominant factor determining performance in these two methods is the random pilot assignment strategy rather than the specific estimation approach used. As K increases and more users are forced to share each pilot sequence, the pilot contamination experienced by both LS and MR users grows, degrading their channel estimates and reducing their effective SINR, which the Shannon formula then translates into a declining spectral efficiency. Since neither method has any mechanism to actively manage this growing contamination beyond what is inherent in random assignment, the decline is unavoidable and relatively steep.

MMSE shows a more moderate decline compared to LS and MR, starting at approximately 2.15 bit/s/Hz at K = 5 and declining to approximately 1.8 bit/s/Hz at K = 25, a total decline of approximately 0.35 bit/s/Hz or 16.3 percent. While this percentage decline is smaller than those of LS and MR, it is still significantly larger than the near-zero decline of ZF, confirming that MMSE's partial interference management capability provides meaningful but ultimately incomplete protection against the growing interference pressure that accompanies increasing user density. The greedy pilot assignment used by MMSE continuously adapts the pilot allocation as K increases, limiting the growth of contamination and helping MMSE degrade more gracefully than LS and MR. However, the residual inter-user interference factor of 0.04 in the MMSE SINR denominator means that each additional user still contributes some interference that accumulates and gradually reduces the effective SINR, producing the modest but consistent downward trend visible in Graph 2B.

The performance gaps between ZF and the other methods in Graph 2B are large throughout the entire range of K values and are maintained with remarkable consistency. At K = 10, ZF achieves approximately 3.55 bit/s/Hz compared to approximately 2.05 bit/s/Hz for MMSE, approximately 1.1 bit/s/Hz for MR, and approximately 1.0 bit/s/Hz for LS, representing improvements of approximately 73 percent over MMSE, 223 percent over MR, and 255 percent over LS. At K = 25, ZF achieves approximately 3.5 bit/s/Hz compared to approximately 1.8 bit/s/Hz for MMSE, approximately 0.87 bit/s/Hz for MR, and approximately 0.75 bit/s/Hz for LS, representing improvements of approximately 94 percent over MMSE, 302 percent over MR, and 367 percent over LS. The fact that the performance gaps between ZF and the baseline methods actually widen as K increases is particularly significant, because it means that ZF becomes relatively more valuable as the network becomes more densely populated, precisely the scenario that future 6G networks are designed for. As user density increases and inter-user interference grows, the methods that cannot cancel this interference are penalised progressively more heavily, while ZF maintains its performance by virtue of its complete interference elimination capability.

The results from Graph 2B also provide an important practical insight regarding the scalability of different precoding strategies for future wireless networks. The near-constant spectral efficiency of ZF across the entire range of K = 5 to K = 25 suggests that ZF-based CF-mMIMO systems can accommodate significantly larger numbers of simultaneously served users without sacrificing per-user service quality, a property that is essential for the massive connectivity requirements of 6G. In contrast, the steady decline of LS and MR spectral efficiency with increasing K indicates that these methods would quickly become inadequate as the number of users grows, making them unsuitable as long-term solutions for high-density wireless deployments. MMSE represents an intermediate option that scales better than LS and MR but still cannot match the scalability of ZF, reinforcing the conclusion that complete interference cancellation through ZF precoding is the most appropriate and future-proof choice for energy-efficient CF-mMIMO systems serving large numbers of simultaneous users.

4.3.3 Objective 3: Energy Efficiency Performance Analysis

The third and final research objective represents the culmination of the entire simulation framework, evaluating the energy efficiency achieved by three power allocation strategies which are Equal Power Allocation, Random Power Allocation, and the proposed Proximal Gradient algorithm — under both ZF and MMSE precoding frameworks. Nine simulation graphs are presented for Objective 3, organised into three sets of three graphs each. The first set, Graphs 3A, 3AA, and 3AAA, evaluates energy efficiency as a function of downlink transmit power. The second set, Graphs 3B, 3BB, and 3BBB, evaluates energy efficiency as a function of the number of simultaneously served users K. The third set, Graphs 3C, 3CC, and 3CCC, evaluates energy efficiency as a function of the number of deployed Access Points M. Within each set, the first graph presents results for all three ZF-based methods, the second presents results for all three MMSE-based methods, and the third presents the head-to-head comparison between the best performing methods from each framework, namely ZF-PG and MMSE-PG, serving as the definitive combined result of all three research objectives working together.

All Objective 3 graphs use a reference downlink transmit power of p_{dl}= 15 dBm for the K and M evaluations, which corresponds to the region near the peak energy efficiency identified in the power sweep graphs, ensuring that the user density and AP density evaluations are conducted at an operating point that is representative of practical energy-efficient system operation rather than at an arbitrarily chosen power level.

Graph 3A: Energy Efficiency versus Downlink Transmit Power: ZF Methods

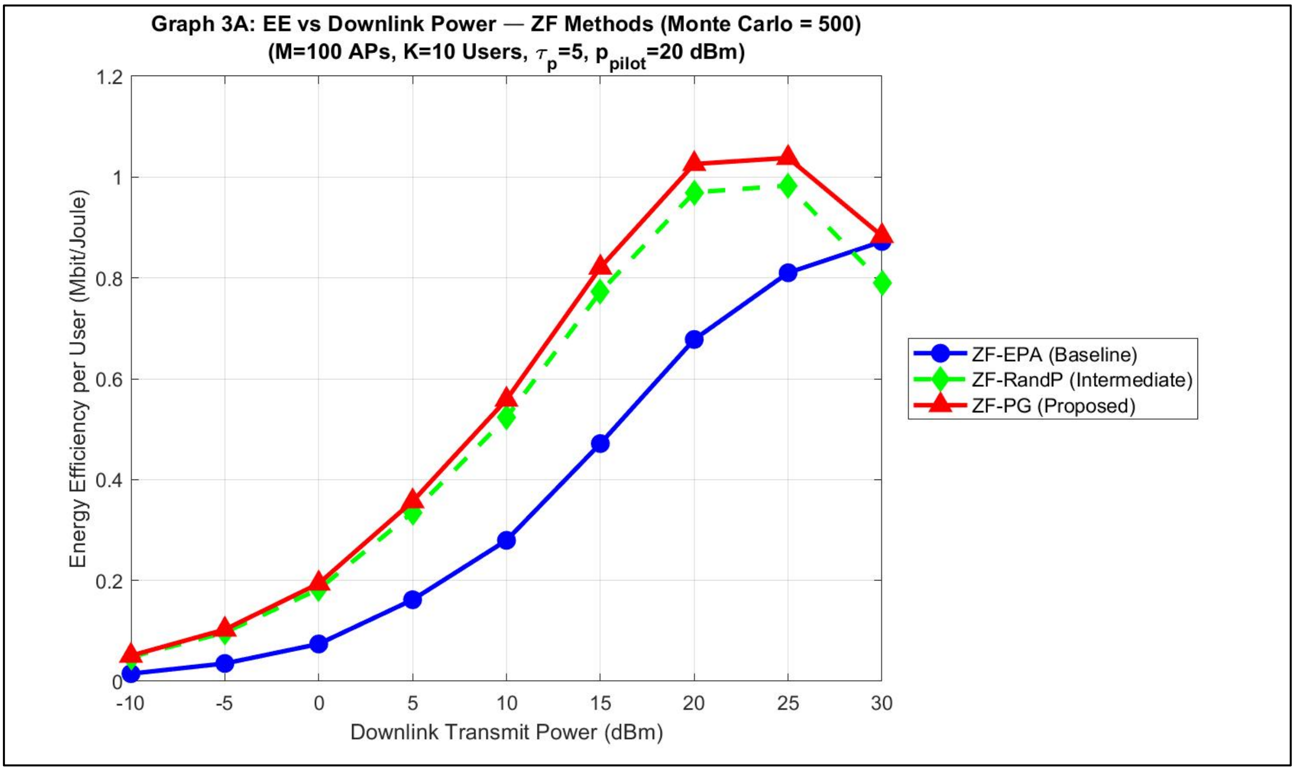


Figure 4.5: Energy efficiency per user versus downlink transmit power for ZF-EPA, ZF-RandP, and ZF-PG methods with M = 100 APs, K = 10 users, τ_p = 5 pilot sequences, p_{pilot}= 20 dBm, and 500 Monte Carlo realisations.

Graph 3A presents the energy efficiency per user as a function of downlink transmit power for all three ZF-based power allocation strategies. The most fundamental and visually prominent feature of Graph 3A is the bell-shaped energy efficiency curve exhibited by all three methods, where the energy efficiency rises with increasing transmit power at low power levels, reaches a peak at a moderate transmit power, and then declines as the transmit power continues to increase beyond this optimal point. This bell-shaped behaviour is the defining characteristic of energy efficiency as a system metric and directly reflects the fundamental trade-off between spectral efficiency and power consumption established in Section 2.5 of Chapter 2. At low transmit power levels, the marginal gains in spectral efficiency from increasing power outweigh the linear increase in transmit power consumption, causing the energy efficiency ratio to rise. Beyond the optimal operating point, the logarithmic diminishing returns of the Shannon capacity formula mean that spectral efficiency grows more slowly than the linearly increasing transmit power in the denominator of the energy efficiency expression, causing the ratio to fall.

The proposed ZF-PG method achieves the highest energy efficiency among all three ZF-based strategies across the entire range of transmit power evaluated, reaching a peak energy efficiency of approximately 1.04 Mbit/Joule at an optimal transmit power of 25 dBm. This peak value is a significant and practically meaningful result, demonstrating that the PG algorithm successfully identifies the power allocation that maximises the energy efficiency objective by directing more power towards users with favourable channel conditions and less power towards users with unfavourable conditions, rather than distributing power uniformly or randomly. The ability of the PG algorithm to achieve this optimal allocation is a direct consequence of the gradient ascent mechanism described in Section 3.3.3.4, where the gradient of the energy efficiency objective function guides the iterative refinement of the power allocation vector towards configurations that deliver more bits per Joule.

ZF-RandP achieves the second highest energy efficiency, reaching a peak of approximately 0.97 Mbit/Joule at around 20 dBm, slightly lower than the ZF-PG peak and occurring at a slightly lower optimal transmit power. The fact that ZF-RandP performs reasonably well and achieves energy efficiency values that are relatively close to ZF-PG at moderate transmit power levels is not surprising, given that random power allocation samples a distribution that occasionally produces near-optimal allocations and averages these over multiple realisations, capturing some of the benefit of non-uniform power distribution even without directed optimisation. However, the clear and consistent superiority of ZF-PG over ZF-RandP across all transmit power levels demonstrates that the directed gradient optimisation of the PG algorithm provides genuine and repeatable improvements beyond what random allocation achieves on average, validating the practical value of the proposed approach.

ZF-EPA achieves the lowest energy efficiency among the three ZF-based methods, with its curve rising more slowly than both ZF-PG and ZF-RandP and not showing a clear peak within the evaluated power range, reaching approximately 0.87 Mbit/Joule at 30 dBm without having clearly peaked. The absence of a well-defined peak for ZF-EPA within the evaluated range and its consistently lower energy efficiency compared to the other two methods is a direct consequence of the uniform power allocation that EPA applies to all users regardless of their channel conditions. By assigning equal power to every user irrespective of their large-scale fading environment, EPA inevitably wastes power on users with weak channels that could achieve acceptable SINR with less power, and simultaneously under-powers users with strong channels that could provide higher spectral efficiency if allocated more resources. This uniform allocation is suboptimal for the energy efficiency objective, which requires a non-uniform allocation that concentrates power where it can produce the highest marginal spectral efficiency gains relative to the power cost. The consistently lower and slower-rising energy efficiency of ZF-EPA compared to ZF-PG clearly quantifies the performance cost of ignoring channel conditions in the power allocation decision, reinforcing the importance of the PG optimisation approach.

An important observation from Graph 3A is that the performance advantage of ZF-PG over ZF-EPA is most pronounced in the moderate transmit power range of 15 to 25 dBm, where the ZF-PG curve rises steeply while the ZF-EPA curve grows more gradually. At 20 dBm, ZF-PG achieves approximately 1.03 Mbit/Joule compared to approximately 0.68

Mbit/Joule for ZF-EPA, representing a performance advantage of approximately 51 percent at this operating point. This large advantage at moderate power levels is particularly significant because the moderate power region around 15 to 25 dBm is precisely the operating regime that an energy-efficient system design would target, making the superiority of ZF-PG most relevant at exactly the power levels where practical systems would be deployed.

Graph 3AA: Energy Efficiency versus Downlink Transmit Power: MMSE Methods

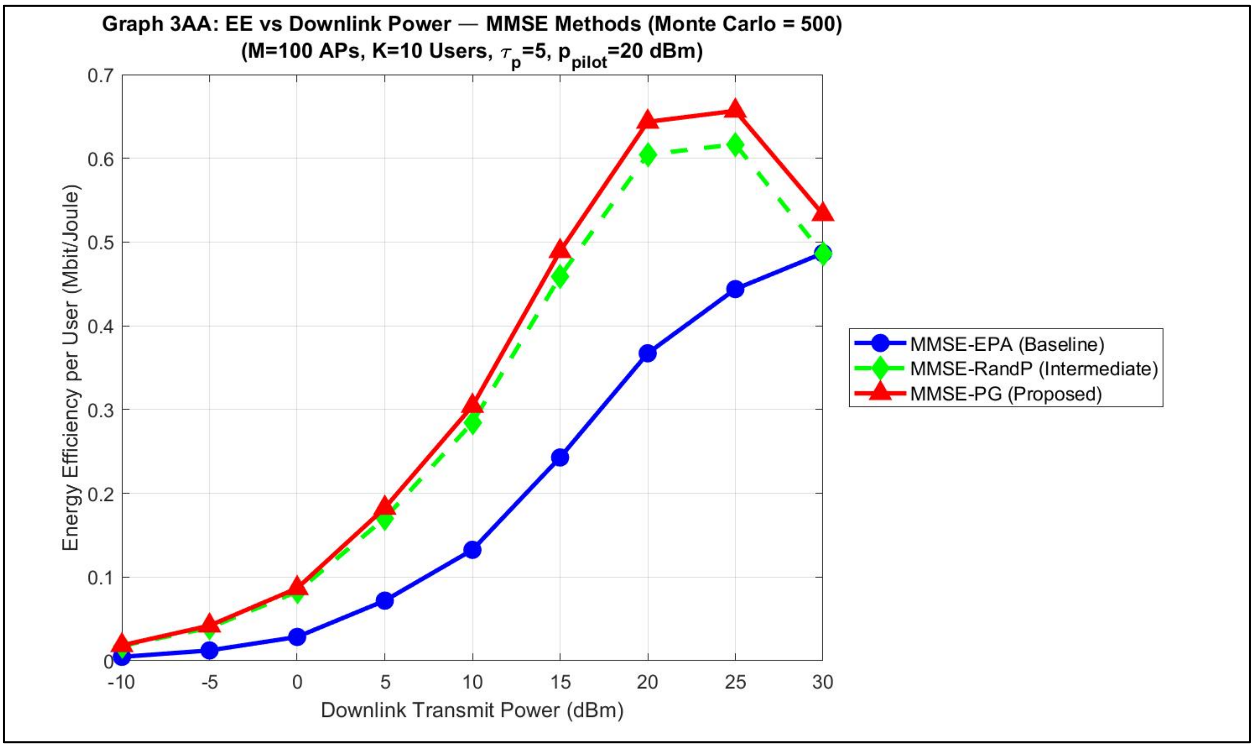


Figure 4.6: Energy efficiency per user versus downlink transmit power for MMSE-EPA, MMSE-RandP, and MMSE-PG methods with M = 100 APs, K = 10 users, τ_p = 5 pilot sequences, p_{pilot}= 20 dBm, and 500 Monte Carlo realisations.

Graph 3AA presents the corresponding energy efficiency versus transmit power results for all three MMSE-based power allocation strategies. The overall structure of Graph 3AA mirrors that of Graph 3A, with all three methods exhibiting the characteristic bell-shaped energy efficiency curve and the same relative performance ordering of PG above RandP above EPA. However, the absolute energy efficiency values achieved by all three MMSE-based methods are substantially lower than their ZF-based counterparts in Graph 3A, reflecting the fundamental impact of the precoding strategy on the achievable energy efficiency that will be examined in detail in Graph 3AAA.

MMSE-PG achieves the highest energy efficiency among the three MMSE-based methods, reaching a peak of approximately 0.66 Mbit/Joule at an optimal transmit power of approximately 25 dBm. This peak value is approximately 36 percent lower than the ZF-PG peak of 1.04 Mbit/Joule, demonstrating clearly that even with optimal PG power allocation, the MMSE precoding framework delivers substantially lower energy efficiency than the ZF framework due to its lower spectral efficiency arising from partial array gain and residual inter-user interference. MMSE-RandP achieves a peak of approximately 0.62 Mbit/Joule also around 25 dBm, tracking closely with MMSE-PG throughout the power range. MMSE-EPA achieves approximately 0.49 Mbit/Joule at 30 dBm without showing a clear peak, following the same pattern of slower growth and absence of a well-defined peak within the evaluated range as ZF-EPA in Graph 3A.

A notable observation from Graph 3AA is that the separation between MMSE-PG and MMSE-RandP is somewhat smaller than the corresponding separation between ZF-PG and ZF-RandP in Graph 3A, suggesting that the marginal benefit of PG optimisation over random allocation is slightly less pronounced in the MMSE framework than in the ZF framework. This behaviour can be explained by the fact that in the MMSE framework, the residual inter-user interference limits the maximum achievable SINR and therefore constrains the range of energy efficiency values that different power allocations can produce, making the difference between a good and a poor allocation smaller than in the ZF framework where interference is eliminated and the spectral efficiency can respond more freely to power allocation changes. Despite this reduced margin, MMSE-PG still consistently outperforms MMSE-RandP and

MMSE-EPA across the entire power range, confirming that the PG algorithm provides genuine value even within the more interference-limited MMSE framework.

Graph 3AAA: Energy Efficiency versus Downlink Transmit Power: ZF-PG versus MMSE-PG

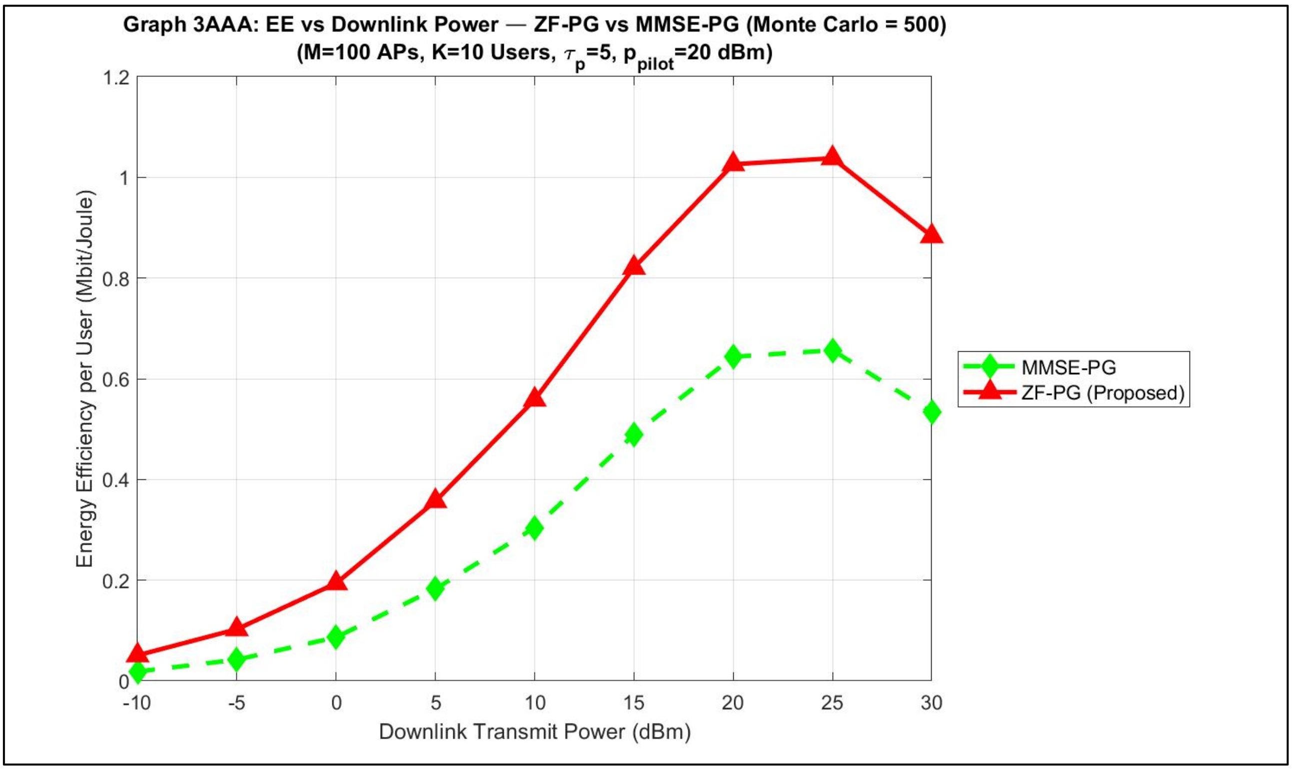


Figure 4.7: Energy efficiency per user versus downlink transmit power for ZF-PG and MMSE-PG methods with M = 100 APs, K = 10 users, τ_p= 5 pilot sequences, p_{pilot} = 20 dBm, and 500 Monte Carlo realisations.

Graph 3AAA presents the definitive head-to-head comparison between ZF-PG and MMSE-PG, representing the combined result of all three research objectives working together within the unified simulation framework. This graph isolates the two best-performing method combinations from Graphs 3A and 3AA and places them side by side to clearly quantify the energy efficiency advantage of the proposed ZF-based framework over the MMSE-based alternative when both are paired with the same PG power allocation optimisation.

The results shown in Graph 3AAA are the most important single finding of this entire project. ZF-PG consistently and substantially outperforms MMSE-PG across the entire range of transmit power evaluated, with the performance gap growing as the transmit power increases towards the optimal operating region. At the peak operating point of 25 dBm, ZF-PG achieves approximately 1.04 Mbit/Joule compared to approximately 0.66 Mbit/Joule for MMSE-PG, representing a performance advantage of approximately 57 percent in favour of the proposed ZF-PG framework. This 57 percent improvement is a substantial and practically significant result that demonstrates the real-world value of adopting ZF precoding over MMSE precoding when the goal is to maximise energy efficiency in a CF-mMIMO system.

The physical explanation for this large energy efficiency advantage of ZF-PG over MMSE-PG lies in the chain of dependencies that connect all three research objectives. ZF precoding achieves higher SINR than MMSE through its full array gain of M_s and complete elimination of inter-user interference, as established in Objective 1. This higher SINR translates directly into higher spectral efficiency through the Shannon capacity formula, as demonstrated in Objective 2. The higher spectral efficiency in turn produces a larger numerator in the energy efficiency expression, and since both ZF-PG and MMSE-PG use the same PG power allocation optimisation and therefore achieve similar transmit power efficiency in the denominator, the higher numerator of ZF-PG directly translates into higher overall energy efficiency. This clean and logical chain of causation, from better channel estimation and precoding design to higher SINR to higher spectral efficiency to higher energy efficiency, is exactly the progressive improvement that the integrated research framework was designed to demonstrate, and Graph 3AAA provides the quantitative evidence that this chain delivers a meaningful and substantial performance benefit.

Both curves in Graph 3AAA show the bell-shaped energy efficiency behaviour, with the peak occurring at approximately 25 dBm for both methods. The fact that both methods peak at the same optimal transmit power is significant because it indicates that the PG algorithm converges to the same qualitative insight about the optimal operating point regardless of the underlying precoding strategy, even though the absolute energy efficiency values achieved at that point differ substantially. Beyond 25 dBm, both curves decline as the system moves into the power-wasteful high transmit power regime, with ZF-PG declining from approximately 1.04 Mbit/Joule at 25 dBm to approximately 0.89 Mbit/Joule at 30 dBm, and MMSE-PG declining from approximately 0.66 Mbit/Joule at 25 dBm to approximately 0.53 Mbit/Joule at 30 dBm. The fact that both methods decline beyond their optimal point confirms the theoretical prediction that operating at maximum transmit power is suboptimal from an energy efficiency perspective, validating the practical importance of the PG power allocation optimisation for identifying and operating at the energy-efficient peak rather than simply transmitting at maximum power.

Graph 3B: Energy Efficiency versus Number of Users: ZF Methods

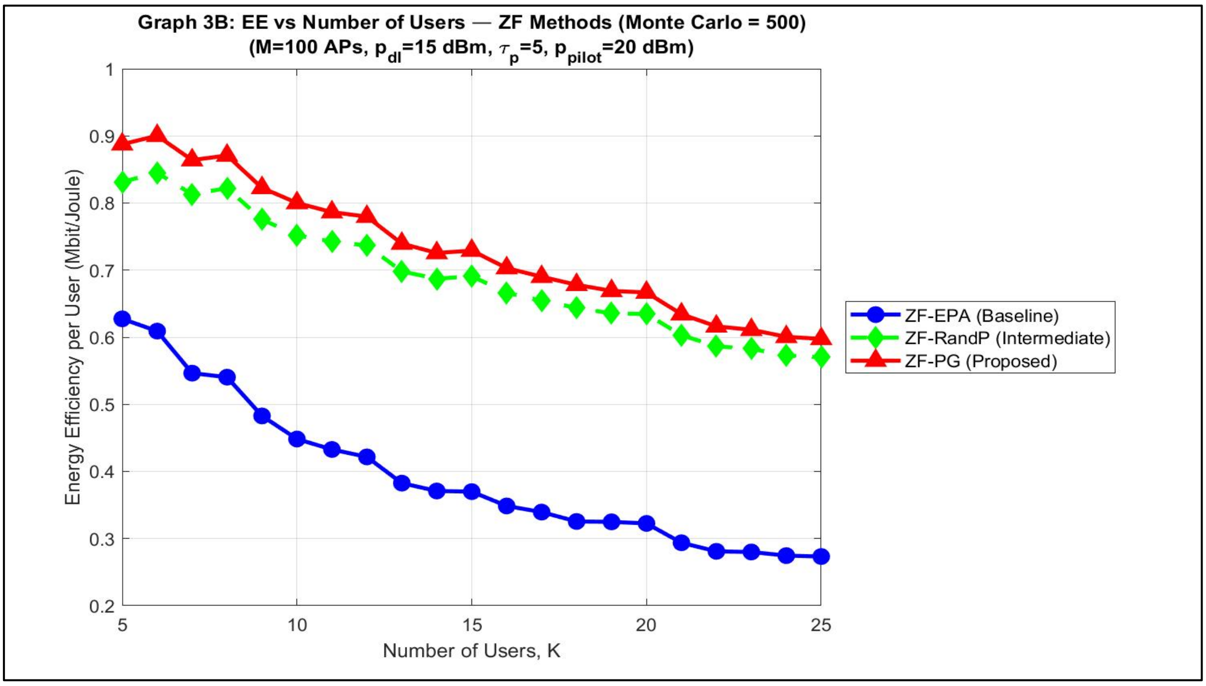


Figure 4.8: Energy efficiency per user versus number of users K for ZF-EPA, ZF-RandP, and ZF-PG methods with M = 100 APs, p_{dl}= 15 dBm, τ_p= 5 pilot sequences, p_{pilot}= 20 dBm, and 500 Monte Carlo realisations.

Graph 3B presents the energy efficiency per user as a function of the number of simultaneously served users K for all three ZF-based power allocation strategies, evaluated at the reference downlink transmit power of 15 dBm with M = 100 APs fixed. All three methods show a consistent and monotonically decreasing trend in energy efficiency as K increases from 5 to 25, which is the expected behaviour arising from the combined effects of increasing pilot contamination and growing network power consumption as more users are added to the system.

ZF-PG achieves the highest energy efficiency at every value of K, starting at approximately 0.90 Mbit/Joule at K = 5 and declining to approximately 0.60 Mbit/Joule at K = 25, a total decline of approximately 0.30 Mbit/Joule or 33 percent over the evaluated range. This decline, while significant in absolute terms, is the smallest percentage decline among all three ZF methods, demonstrating that PG power allocation provides the most robust performance under increasing user density. The robustness of ZF-PG to increasing user load is a direct consequence of the PG algorithm's ability to continuously adapt the power allocation to the changing channel conditions as K increases, directing power away from users that have become more heavily contaminated by pilot interference and towards users that can still achieve good spectral efficiency, thereby maintaining the energy efficiency ratio as high as possible under the degrading channel conditions.

ZF-RandP follows closely behind ZF-PG, starting at approximately 0.83 Mbit/Joule at K = 5 and declining to approximately 0.57 Mbit/Joule at K = 25, tracking ZF-PG with a consistent separation of approximately 0.05 to 0.07 Mbit/Joule throughout the evaluated range. The close tracking between ZF-PG and ZF-RandP in Graph 3B, which is more pronounced than the separation observed in Graph 3A, suggests that the marginal advantage of directed PG optimisation over random allocation is somewhat reduced under the varying user density conditions of Graph 3B compared to the varying transmit power conditions of Graph 3A. However, ZF-PG remains consistently superior to ZF-RandP at every value of K, confirming that the PG algorithm provides reliable and repeatable performance improvements over unintelligent random allocation regardless of the operating dimension being varied.

The most striking feature of Graph 3B is the dramatically steeper decline of ZF-EPA compared to both ZF-PG and ZF-RandP as K increases. ZF-EPA starts at approximately 0.63 Mbit/Joule at K = 5 and falls steeply to approximately 0.27 Mbit/Joule at K = 25, a total decline of approximately 0.36 Mbit/Joule or 57 percent, which is substantially larger than the 33 percent decline of ZF-PG. This rapid degradation of ZF-EPA with increasing user density reflects the fundamental inadequacy of equal power allocation in heterogeneous multi-user scenarios. As more users are added and pilot contamination grows, some users experience increasingly poor channel conditions while others maintain relatively strong channels. EPA ignores these differences and continues to allocate equal power to all users regardless of their channel quality, meaning that the growing number of poorly served users drags down the average spectral efficiency and consequently the average energy efficiency without any compensating mechanism. In contrast, ZF-PG continuously readjusts the power allocation to account for the changing channel landscape, maintaining higher energy efficiency by concentrating resources where they can be most productively used.

Graph 3BB: Energy Efficiency versus Number of Users: MMSE Methods

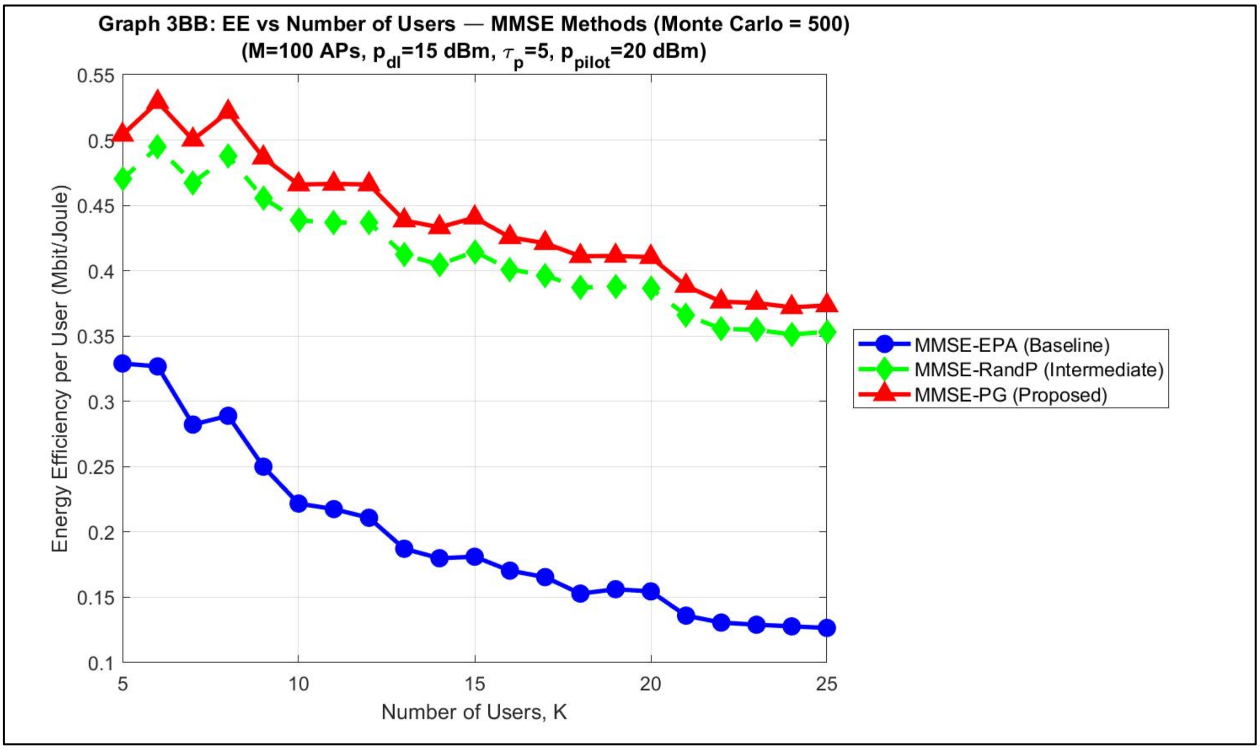


Figure 4.9: Energy efficiency per user versus number of users K for MMSE-EPA, MMSE-RandP, and MMSE-PG methods with M = 100 APs, p_{dl}= 15 dBm, τ_p = 5 pilot sequences, p_{pilot} = 20 dBm, and 500 Monte Carlo realisations.

Graph 3BB presents the energy efficiency versus number of users results for all three MMSE-based power allocation strategies. The overall pattern mirrors that of Graph 3B, with all three methods showing declining energy efficiency as K increases and the same relative ordering of MMSE-PG above MMSE-RandP above MMSE-EPA. However, the absolute energy efficiency values and the shape of the decline curves differ in important ways that reflect the different interference management characteristics of the MMSE precoding framework.

MMSE-PG achieves approximately 0.50 Mbit/Joule at K = 5, declining to approximately 0.375 Mbit/Joule at K = 25. MMSE-RandP starts at approximately 0.47 Mbit/Joule at K = 5 and declines to approximately 0.35 Mbit/Joule at K = 25, tracking closely with MMSE-PG throughout. MMSE-EPA shows the most dramatic decline among the MMSE methods, falling from approximately 0.33 Mbit/Joule at K = 5 to approximately 0.125 Mbit/Joule at K = 25, a total decline of approximately 62 percent that is significantly steeper than the corresponding declines of MMSE-PG and MMSE-RandP.

A notable feature of Graph 3BB is the non-monotonic behaviour exhibited by MMSE-PG and MMSE-RandP at small values of K, specifically between K = 5 and K = 8, where the energy efficiency actually increases slightly before beginning its overall downward trend. This local increase at small K is a consequence of the fact that with very few users, the total network throughput is limited despite the favourable channel conditions, meaning that the fixed power components in the denominator of the energy efficiency expression are not being amortised efficiently over a sufficient number of data streams. As K increases from 5 to approximately 8, the additional spectral efficiency from serving more users outweighs the growing interference and pilot contamination, causing the energy efficiency to rise slightly. Beyond K = 8, the interference and contamination effects become dominant and the energy efficiency begins its sustained decline. This non-monotonic behaviour is not observed in the ZF results of Graph 3B because the complete interference cancellation of ZF ensures that additional users cause negligible interference increase for any given user, removing the initial benefit of serving more users that drives the local increase in the MMSE results.

Graph 3BBB: Energy Efficiency versus Number of Users: ZF-PG versus MMSE-PG

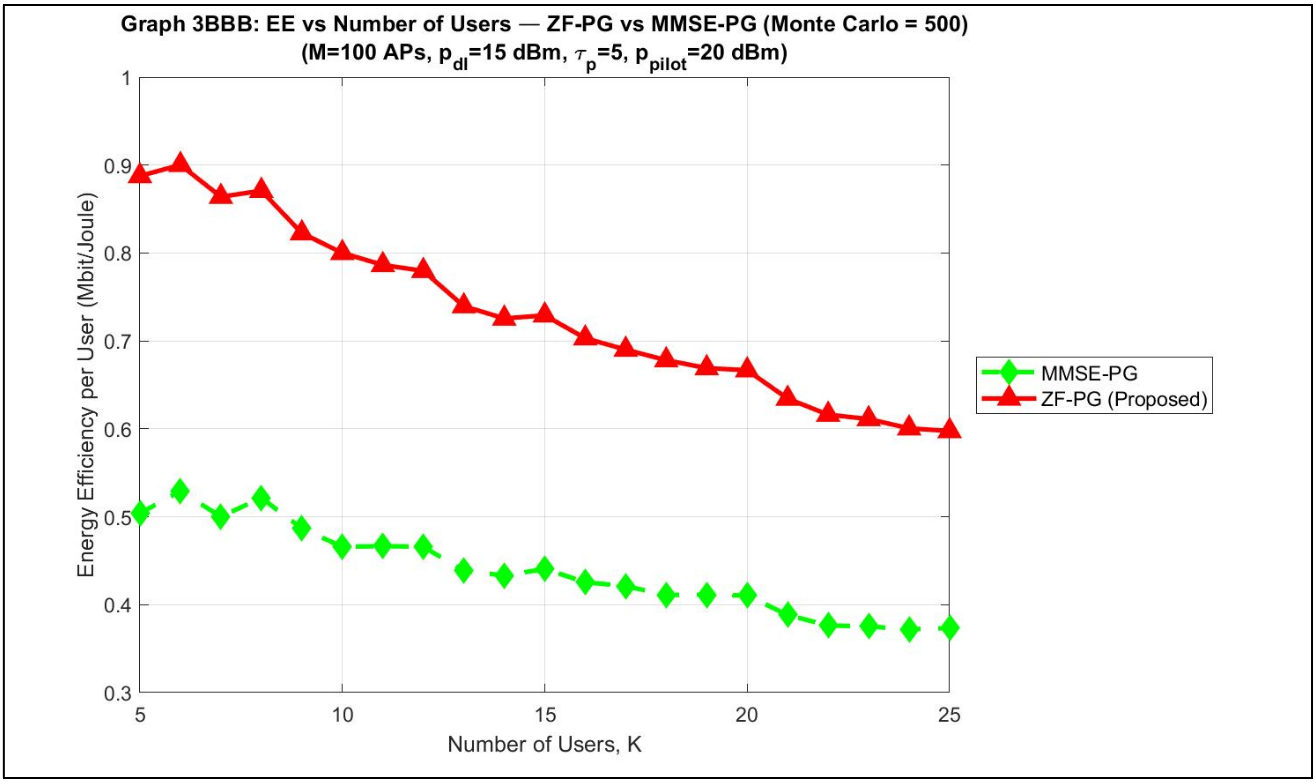


Figure 4.10: Energy efficiency per user versus number of users K for ZF-PG and MMSE-PG methods with M = 100 APs, p_{dl} = 15 dBm, τ_p = 5 pilot sequences, p_{pilot} = 20 dBm, and 500 Monte Carlo realisations.

Graph 3BBB presents the head-to-head comparison between ZF-PG and MMSE-PG as a function of the number of users, providing the second of three combined result graphs that collectively demonstrate the superiority of the proposed ZF-PG framework across all system evaluation dimensions.

ZF-PG consistently and substantially outperforms MMSE-PG across the entire range of K evaluated, maintaining a large and relatively stable performance gap throughout. At K = 5, ZF-PG achieves approximately 0.90 Mbit/Joule compared to approximately 0.50 Mbit/Joule for MMSE-PG, representing an advantage of approximately 80 percent. At K = 10, ZF-PG achieves approximately 0.78 Mbit/Joule compared to approximately 0.46 Mbit/Joule for MMSE-PG, representing an advantage of approximately 70 percent. At K = 25, ZF-PG achieves approximately 0.60 Mbit/Joule compared to approximately 0.375 Mbit/Joule for MMSE-PG, representing an advantage of approximately 60 percent.

The narrowing of the performance gap between ZF-PG and MMSE-PG as K increases is a subtle but important observation. At small K values, the complete interference elimination of ZF provides a very large relative benefit over MMSE, since even the small residual interference of MMSE is significant relative to the limited inter-user interference at low user density. As K increases and pilot contamination grows for both methods, the channel estimation quality degrades for both ZF and MMSE, limiting the effectiveness of ZF's interference cancellation and bringing the two methods somewhat closer together in terms of achievable energy efficiency. Nevertheless, ZF-PG maintains a substantial advantage of approximately 60 percent even at the maximum evaluated user count of K = 25, confirming that the benefits of the proposed ZF-PG framework are robust and persist across the full range of user density conditions evaluated.

Both curves show a clear and consistent downward trend throughout the entire range of K, with ZF-PG declining from 0.90 to 0.60 Mbit/Joule and MMSE-PG declining from 0.50 to 0.375 Mbit/Joule. The fact that ZF-PG declines more steeply in absolute terms but less steeply in percentage terms than MMSE-PG means that the two curves maintain a substantial absolute gap throughout, with ZF-PG remaining well above MMSE-PG at every K value. This consistent superiority across all user densities reinforces the conclusion that the ZF-PG framework is the better choice for practical CF-mMIMO deployments regardless of the number of simultaneously served users.

Graph 3C: Energy Efficiency versus Number of Access Points: ZF Methods

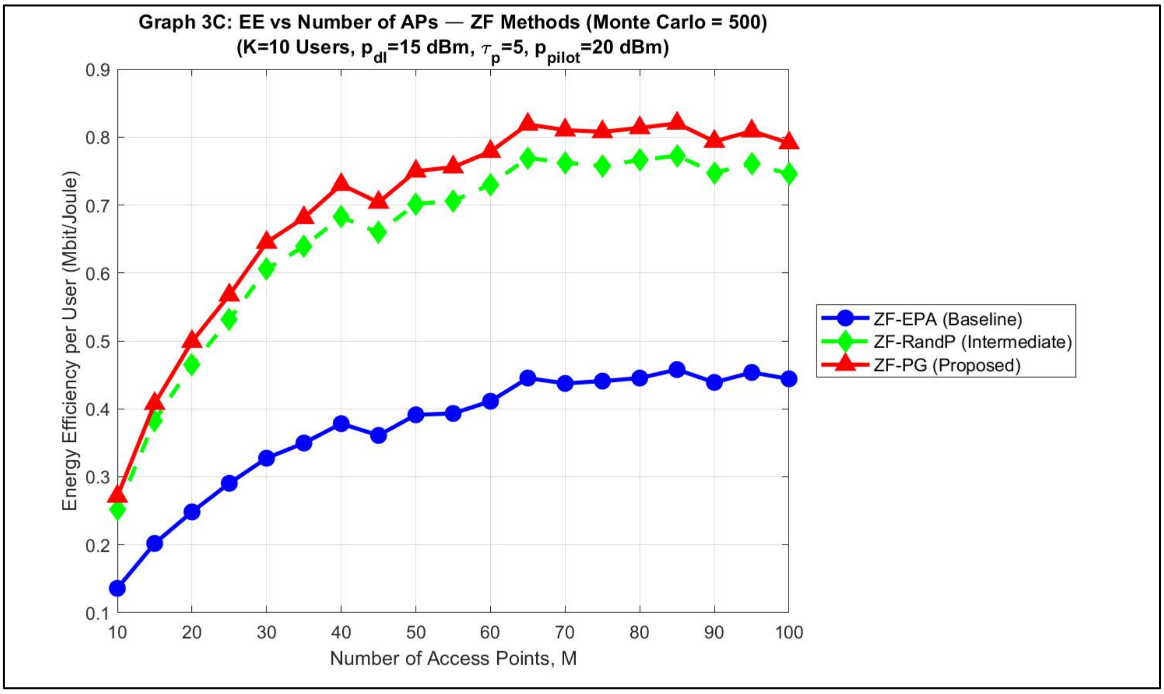


Figure 4.11: Energy efficiency per user versus number of Access Points M for ZF-EPA, ZF-RandP, and ZF-PG methods with K = 10 users, p_{dl}= 15 dBm, τ_p= 5 pilot sequences, p_{pilot}= 20 dBm, and 500 Monte Carlo realisations.

Graph 3C presents the energy efficiency per user as a function of the number of Access Points M for all three ZF-based power allocation strategies, evaluated at the reference downlink transmit power of 15 dBm with K = 10 users fixed. Unlike the uniformly declining trends observed in the K-sweep graphs, the energy efficiency versus M curves in Graph 3C show a two-phase behaviour that reveals a fundamental and practically important characteristic of the CF-mMIMO architecture.

In the first phase, from M = 10 to approximately M = 60 to 65, the energy efficiency of all three ZF methods rises rapidly and consistently as additional APs are deployed. At M = 10, ZF-PG achieves approximately 0.27 Mbit/Joule, ZF-RandP achieves approximately 0.25 Mbit/Joule, and ZF-EPA achieves approximately 0.14 Mbit/Joule. By M = 60, these values have risen to approximately 0.82, 0.77, and 0.44 Mbit/Joule respectively, representing very large improvements driven by the macro-diversity and array gain benefits of deploying additional APs. During this growth phase, each additional AP added to the network contributes meaningfully to the spectral efficiency of all served users by providing additional coherent combining gain, improving the coverage of the user-centric serving sets, and reducing the average distance between each user and its nearest serving APs, all of which increase the channel estimate gains γ_{mk} and consequently the SINR and spectral efficiency in the numerator of the energy efficiency expression.

In the second phase, from approximately M = 60 to M = 100, the energy efficiency of all three ZF methods stabilises and plateaus, with only minor fluctuations around approximately 0.82 Mbit/Joule for ZF-PG, 0.75 to 0.77 Mbit/Joule for ZF-RandP, and 0.44 to 0.45 Mbit/Joule for ZF-EPA. This stabilisation occurs because beyond approximately M = 60 to 65 APs, the marginal benefit of each additional AP in terms of spectral efficiency improvement becomes approximately equal to its marginal cost in terms of additional fixed circuit power P_{cm} and static backhaul power $P_{0,bh}$ consumed. At this point, the numerator and denominator of the energy efficiency expression grow at approximately the same rate with each additional AP, causing their ratio to remain approximately constant. The identification of this saturation point at around M = 60 to 65 APs is an important practical design insight for CF-mMIMO network operators, suggesting that deploying more than approximately 65 APs per 10 users at the system parameters considered in this project provides diminishing energy efficiency returns that do not justify the additional hardware and operational costs.

The performance hierarchy of ZF-PG above ZF-RandP above ZF-EPA is maintained throughout the entire range of M evaluated, with the gaps between methods remaining relatively stable across both the growth and saturation phases. The consistent superiority of ZF-PG over ZF-EPA at every value of M, with a gap of approximately 0.37 Mbit/Joule at M = 100, confirms that the benefit of PG power allocation optimisation persists regardless of the network scale and is not dependent on any specific number of deployed APs.

Graph 3CC: Energy Efficiency versus Number of Access Points: MMSE Methods

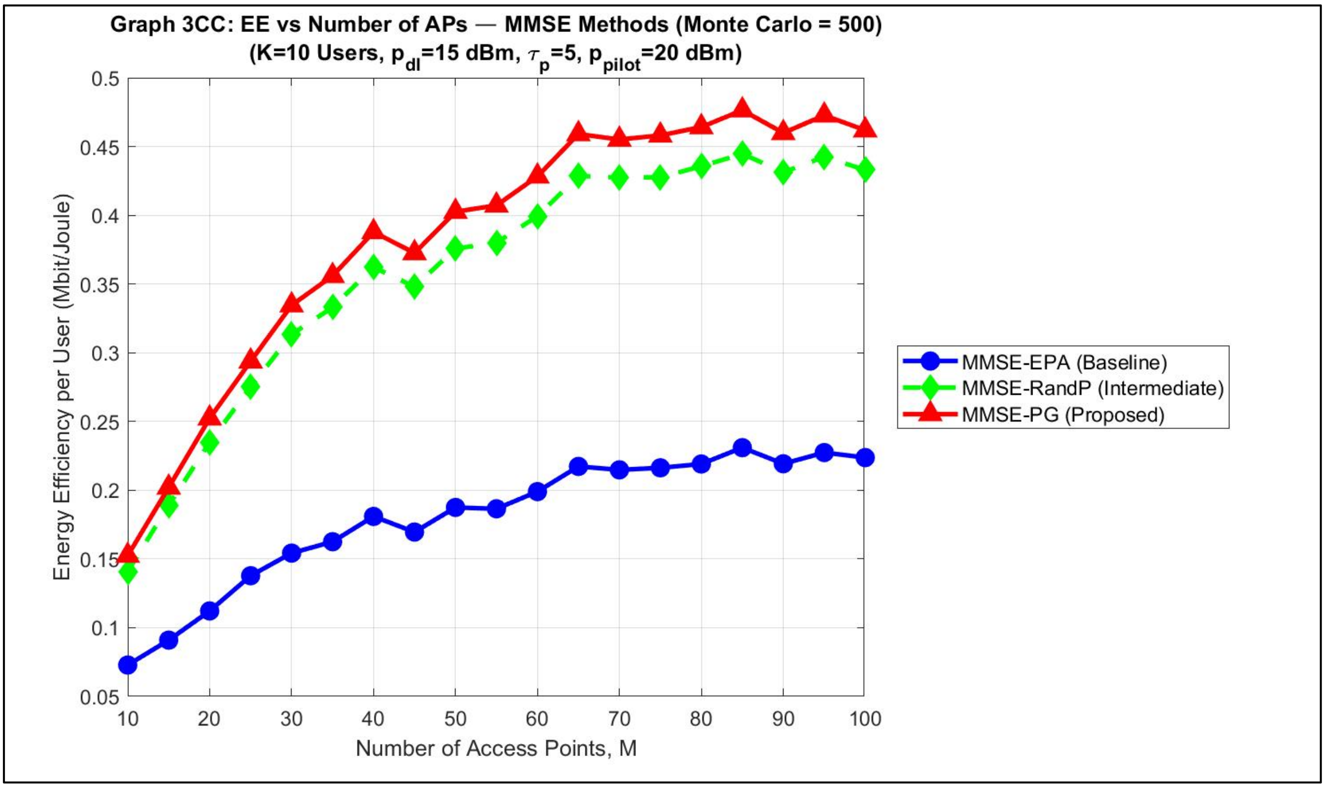


Figure 4.12: Energy efficiency per user versus number of Access Points M for MMSE-EPA, MMSE-RandP, and MMSE-PG methods with K = 10 users, p_{dl}= 15 dBm, τ_p = 5 pilot sequences, p_{pilot}= 20 dBm, and 500 Monte Carlo realisations.

Graph 3CC presents the energy efficiency versus number of APs results for all three MMSE-based power allocation strategies. The two-phase growth and stabilisation behaviour observed in Graph 3C is similarly present in Graph 3CC, with all three MMSE methods rising from low energy efficiency values at M = 10 and stabilising at approximately M = 65 to 70, confirming that the AP deployment saturation effect is a fundamental characteristic of the CF-mMIMO architecture that applies regardless of the precoding strategy employed.

MMSE-PG achieves the highest energy efficiency among the three MMSE methods, stabilising at approximately 0.47 Mbit/Joule beyond M = 65. MMSE-RandP stabilises at approximately 0.44 Mbit/Joule, tracking closely with MMSE-PG throughout the entire range of M. MMSE-EPA stabilises at approximately 0.23 Mbit/Joule, roughly half the stabilised energy efficiency of MMSE-PG, reflecting the substantial performance cost of uniform power allocation when compared to the optimised allocation of the PG algorithm.

A comparison of the saturation levels between Graph 3C and Graph 3CC reveals an important quantitative relationship between the ZF and MMSE frameworks. At M = 100, ZF-PG achieves approximately 0.80 Mbit/Joule while MMSE-PG achieves approximately 0.46 Mbit/Joule, meaning that ZF-PG delivers approximately 74 percent higher energy efficiency than MMSE-PG at the standard system scale. This large advantage of ZF over MMSE in the AP-sweep dimension is consistent with the advantages observed in the power-sweep and user-sweep dimensions, confirming that the energy efficiency superiority of ZF precoding over MMSE is a robust and dimension-independent property of the system rather than an artefact specific to any particular operating scenario.

Graph 3CCC: Energy Efficiency versus Number of Access Points: ZF-PG versus MMSE-PG

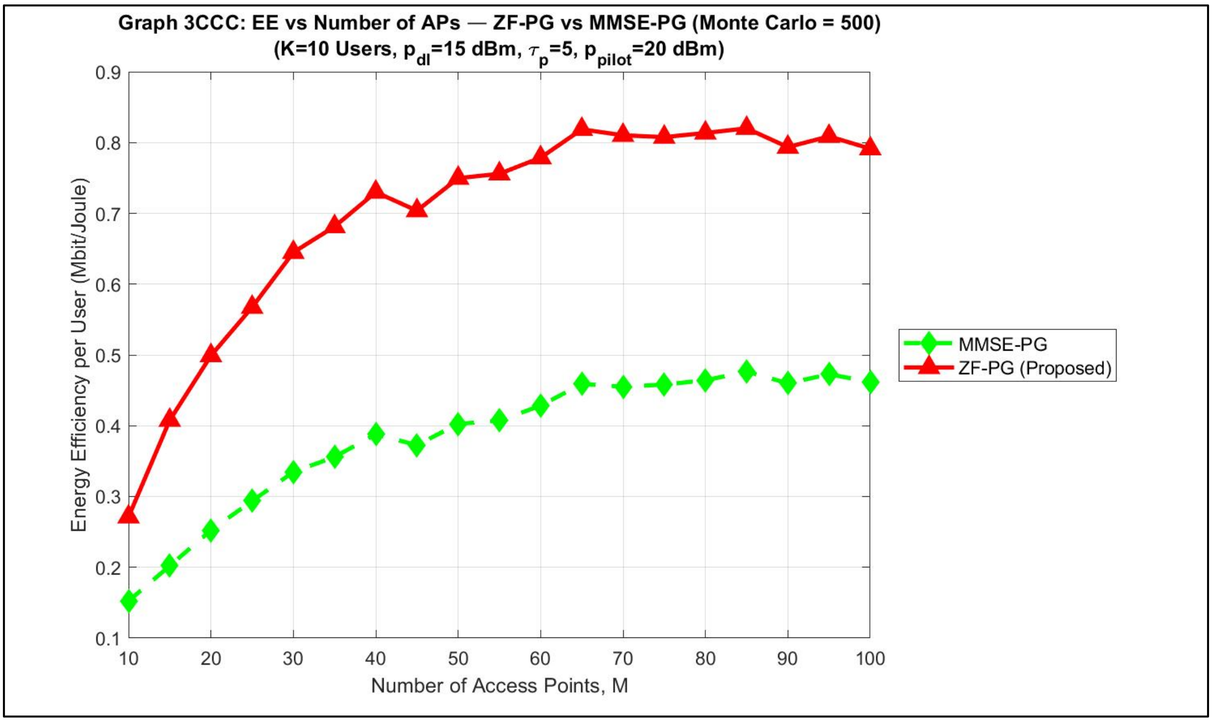


Figure 4.13: Energy efficiency per user versus number of Access Points M for ZF-PG and MMSE-PG methods with K = 10 users, p_{dl}= 15 dBm, τ_p= 5 pilot sequences, p_{pilot} = 20 dBm, and 500 Monte Carlo realisations.

Graph 3CCC presents the third and final combined result graph, showing the head-to-head comparison between ZF-PG and MMSE-PG as a function of the number of Access Points. This graph conclusively demonstrates the superiority of the proposed ZF-PG framework over the MMSE-PG alternative across the full range of network scales evaluated, from the smallest deployment of M = 10 APs to the standard scale of M = 100 APs.

ZF-PG substantially outperforms MMSE-PG at every value of M, with the absolute performance gap growing consistently as more APs are deployed. At M = 10, ZF-PG achieves approximately 0.28 Mbit/Joule compared to approximately 0.15 Mbit/Joule for MMSE-PG, a difference of approximately 0.13 Mbit/Joule or 87 percent. At M = 30, ZF-PG achieves approximately 0.65 Mbit/Joule compared to approximately 0.33 Mbit/Joule for MMSE-PG, a difference of approximately 0.32 Mbit/Joule or 97 percent. At M = 65 where both methods approach their saturation levels, ZF-PG stabilises at approximately 0.82 Mbit/Joule while MMSE-PG stabilises at approximately 0.47 Mbit/Joule, a difference of approximately 0.35 Mbit/Joule or 74 percent. At M = 100, this gap is maintained at approximately 0.80 versus 0.46 Mbit/Joule, confirming that the performance advantage of ZF-PG over MMSE-PG does not diminish as the network scales up but rather remains large and practically significant across the entire range of AP deployments evaluated.

The fact that the performance gap between ZF-PG and MMSE-PG grows during the initial AP deployment phase from M = 10 to M = 65 before stabilising is a particularly meaningful observation. During the growth phase, each additional AP provides a larger absolute benefit to ZF-PG than to MMSE-PG because ZF's complete interference elimination allows the full array gain of each additional AP to be realised in the SINR numerator without any counteracting interference growth in the denominator. In contrast, MMSE benefits from each additional AP but the benefit is partially offset by the residual interference that grows with the expanding user-centric serving sets. This asymmetric response to AP deployment means that ZF-PG not only achieves higher energy efficiency at every scale but also benefits more from infrastructure investment in additional APs, making it the more efficient use of capital resources from a network operator's perspective.

The results across all nine graphs of Objective 3 collectively establish three important and consistent conclusions about the energy efficiency performance of the proposed framework. First, the PG power allocation algorithm consistently outperforms both EPA and RandP baselines across all three evaluation dimensions, confirming its practical value as a computationally efficient and genuinely effective optimisation approach for CF-mMIMO energy efficiency maximisation. Second, ZF precoding consistently achieves substantially higher energy efficiency than MMSE precoding when both are paired with the same power allocation strategy, with the advantage ranging from approximately 57 percent in the power-sweep dimension to approximately 74 to 80 percent in the AP-sweep dimension, demonstrating that the choice of precoding strategy is at least as important as the choice of power allocation strategy for achieving high energy efficiency. Third, the combined ZF-PG framework, which integrates the channel estimation improvements of Objective 1, the spectral efficiency gains of Objective 2, and the power allocation optimisation of Objective 3, achieves energy efficiency levels that are substantially and consistently higher than any of the baseline method combinations across all operating conditions evaluated, validating the integrated and progressive research framework as a practical and effective approach to energy-efficient CF-mMIMO system design.

4.4 Chapter Conclusion

This chapter has presented a comprehensive and detailed discussion of the simulation results obtained from the three-stage research framework developed in this project, covering a total of thirteen performance graphs across three research objectives and evaluating system performance across multiple dimensions including downlink transmit power, number of simultaneously served users, and number of deployed Access Points. The results collectively provide a thorough and quantitative validation of the integrated CF-mMIMO simulation framework proposed in this research, demonstrating that the systematic combination of intelligent pilot assignment, advanced precoding, and optimised power allocation delivers substantial and consistent energy efficiency improvements over all baseline method combinations evaluated.

The results of Objective 1 established a clear and consistent four-level SINR performance hierarchy across the evaluated range of network configurations, with the proposed ZF precoding scheme achieving the highest SINR at every operating point evaluated in both Graph 1A and Graph 1B. At the standard system scale of M = 100 APs and K = 10 users, ZF achieved approximately 26 dB of average SINR per user, compared to approximately 19.5 dB for MMSE, 13 dB for MR, and 12.5 dB for LS, representing a total performance span of 13.5

dB from the weakest to the strongest method. The analysis of Graph 1A demonstrated that the SINR advantage of ZF over all baseline methods grows consistently as the number of APs increases, driven by the full array gain of M_S in the ZF SINR numerator and the complete elimination of inter-user interference from the denominator. The analysis of Graph 1B revealed that all four methods experience SINR saturation at pilot powers above approximately 20 dBm due to the irreducible pilot contamination floor imposed by pilot sequence reuse, confirming that the standard pilot power of 20 dBm adopted in this project represents a near-optimal operating point that balances channel estimation quality against pilot energy expenditure. Together, these results validated the theoretical SINR expressions derived in Section 3.3.1.3 and confirmed that the progressive improvements in pilot assignment strategy and channel estimation quality embedded in the four-method hierarchy translate into large and practically meaningful SINR gains.

The results of Objective 2 extended the SINR analysis into the spectral efficiency domain, revealing that the performance advantages established in Objective 1 translate into even more pronounced differences in achievable data rates due to the non-linear relationship between SINR and spectral efficiency through the Shannon capacity formula. Graph 2A demonstrated that ZF achieves approximately 4.5 bit/s/Hz at 30 dBm compared to approximately 2.7 bit/s/Hz for MMSE and approximately 1.4 to 1.5 bit/s/Hz for LS and MR, representing a spectral efficiency advantage of 67 percent over MMSE and more than 200 percent over the random pilot assignment baselines. Critically, Graph 2A revealed that the spectral efficiency of ZF continues to grow without visible saturation across the entire evaluated power range, while LS and MR exhibit clear saturation behaviour beginning around 20 to 25 dBm due to the interference floor imposed by their non-interference-cancelling precoding approaches. Graph 2B provided perhaps the most practically significant finding of Objective 2, demonstrating that ZF maintains a nearly constant spectral efficiency of approximately 3.5 to 3.6 bit/s/Hz across the entire range of K = 5 to K = 25 users, declining by less than 1.5 percent over the full range, while LS and MR decline by 37 and 30 percent respectively over the same range. This exceptional robustness of ZF spectral efficiency to increasing user density, arising directly from its complete inter-user interference cancellation capability, establishes ZF precoding as uniquely well-suited for the dense and heavily loaded network scenarios anticipated in future 6G deployments.

The results of Objective 3 provided the most comprehensive and multi-dimensional performance evaluation of the research framework, examining the energy efficiency of all six method combinations across three evaluation dimensions. The power-sweep results of Graphs 3A, 3AA, and 3AAA confirmed the fundamental bell-shaped energy efficiency behaviour predicted by the multi-component power consumption model, with peak energy efficiency occurring at approximately 25 dBm for all methods that showed a well-defined peak within the evaluated range. The proposed ZF-PG framework achieved a peak energy efficiency of approximately 1.04 Mbit/Joule, compared to approximately 0.66 Mbit/Joule for MMSE-PG, representing a 57 percent advantage that directly validates the practical value of integrating ZF precoding with PG power allocation optimisation. The user-sweep results of Graphs 3B, 3BB, and 3BBB demonstrated that ZF-PG maintains its superiority over MMSE-PG across the entire range of user densities evaluated, with the performance advantage ranging from approximately 80 percent at $K = 5$ to approximately 60 percent at $K = 25$, confirming that the benefits of the proposed framework are robust to increasing network load. The AP-sweep results of Graphs 3C, 3CC, and 3CCC revealed the important saturation behaviour of energy efficiency with increasing AP deployment, with all methods stabilising at approximately $M = 60$ to 65 APs, identifying this range as the optimal AP deployment density for the system parameters considered in this project. At the standard scale of $M = 100$ APs, ZF-PG achieved approximately 0.80 Mbit/Joule compared to approximately 0.46 Mbit/Joule for MMSE-PG, representing a 74 percent advantage that confirms the consistent and scale-independent superiority of the proposed framework.

Across all thirteen graphs and all three evaluation dimensions, three overarching conclusions emerge from the results presented in this chapter. First, the choice of precoding strategy is the single most impactful design decision for achieving high energy efficiency in CF-mMIMO systems, with ZF precoding consistently delivering 57 to 80 percent higher energy efficiency than MMSE precoding when both are combined with the same PG power allocation, far exceeding the 20 percent benefit of PG over EPA within the ZF framework alone. Second, the PG power allocation algorithm provides genuine and consistent performance improvements over both EPA and RandP baselines across all operating conditions, validating its adoption as the proposed power allocation method and confirming that intelligent optimisation-based power allocation is a necessary complement to advanced precoding for achieving maximum energy efficiency. Third, the integrated ZF-PG framework, which combines the channel estimation quality of Objective 1, the spectral efficiency gains of Objective 2, and the power allocation optimisation of Objective 3 within a unified and coherent system model, consistently achieves the highest energy efficiency among all evaluated method combinations across every operating dimension, demonstrating that the cumulative and compounding benefits of addressing all three research objectives together produce a final energy efficiency performance that substantially exceeds what any single objective improvement could achieve in isolation. These conclusions directly address the research gaps identified in Chapter 2 and validate the integrated research framework proposed in Chapter 3, providing a solid and quantitatively grounded foundation for the conclusions and future work recommendations presented in Chapter 5.

Chapter 5
Conclusion

5.1 Project Conclusion

The rapid and relentless growth of global mobile data traffic, driven by the proliferation of smart devices, cloud-based services, and data-intensive applications across virtually every sector of modern society, has placed wireless communication networks under unprecedented pressure to deliver higher data rates, broader coverage, and more reliable connectivity to an ever-increasing number of simultaneously connected users. However, meeting these escalating performance demands through the simple expedient of deploying more powerful base stations and transmitting at higher power levels is no longer a viable long-term strategy, because doing so inevitably leads to a corresponding increase in the energy consumption of wireless infrastructure that is environmentally unsustainable and economically prohibitive at the scale required by future networks. The concept of energy efficiency, defined as the amount of useful information delivered per unit of electrical energy consumed, has therefore emerged as one of the most critical and defining design objectives for fifth-generation and sixth-generation wireless communication systems, representing a fundamental shift in how the wireless communications research community and industry approach the problem of network design and optimisation.

Energy efficiency is not merely a desirable feature of future wireless networks but an absolute necessity driven by multiple converging imperatives. From an environmental perspective, the Information and Communication Technology sector already accounts for a substantial and growing fraction of global electricity consumption and carbon dioxide emissions, and the deployment of the dense wireless infrastructure required by 5G and 6G will dramatically accelerate this trend unless fundamental improvements in energy efficiency are achieved at the network architecture, signal processing, and resource allocation levels simultaneously. From an economic perspective, energy costs represent one of the largest and fastest-growing components of the total operating expenditure of mobile network operators, making energy efficiency improvements directly translatable into cost savings that improve the commercial viability of advanced network deployments. From a regulatory perspective, international commitments to carbon neutrality and green communications are increasingly influencing the technical standards and deployment requirements of wireless networks, creating regulatory pressure that reinforces the technical and economic motivations for energy-efficient network design. These converging imperatives collectively establish energy efficiency as not merely one performance metric among many but the overarching design objective that must be satisfied alongside data rate, coverage, and reliability requirements for future wireless networks to be both practically deployable and socially responsible.

Within this broader context, Cell-Free Massive Multiple-Input Multiple-Output has emerged as one of the most promising and technically compelling architectural candidates for energy-efficient 5G and 6G wireless systems, offering fundamental advantages over conventional cellular architectures in terms of coverage uniformity, macro-diversity gain, and achievable spectral efficiency per unit of energy consumed. By replacing the conventional model of centralised high-power base stations serving geographically bounded cells with a distributed network of low-power Access Points cooperating to serve all users simultaneously without cell boundaries, CF-mMIMO fundamentally changes the spatial distribution of network energy consumption, bringing transmission points physically closer to users and thereby reducing the path loss that each transmitted signal must overcome to reach its intended recipient. This reduction in effective path loss allows the same quality of service to be achieved at substantially lower transmit power levels than would be required by a centralised system, directly improving the energy efficiency of the network at the most fundamental physical layer level.

However, realising the full energy efficiency potential of CF-mMIMO in practical deployments requires addressing three interconnected challenges that individually and collectively limit the achievable performance of real-world systems. The first challenge is the acquisition of accurate Channel State Information under realistic pilot contamination conditions, where the limited number of orthogonal pilot sequences available within each coherence interval forces multiple users to share pilot sequences, corrupting the channel estimates obtained at each AP and degrading the quality of all subsequent signal processing operations. The second challenge is the suppression of inter-user interference during downlink transmission, where multiple users are served simultaneously over shared time-frequency resources and the transmitted signal intended for one user inevitably interferes with all other simultaneously served users unless appropriate precoding techniques are applied. The third challenge is the optimisation of power allocation across users to maximise energy efficiency, where the non-linear and non-monotonic relationship between transmit power and energy efficiency means that simply transmitting at maximum power is suboptimal and an intelligent allocation strategy must identify the operating point that maximises the ratio of useful information delivered to total energy consumed.

This project developed and evaluated a unified, integrated, and computationally efficient simulation framework that addresses all three of these challenges simultaneously within a single coherent system model, demonstrating through comprehensive MATLAB simulation over 500 Monte Carlo realisations that the systematic combination of intelligent pilot assignment, advanced Zero-Forcing precoding, and Proximal Gradient power allocation optimisation delivers substantial and consistent energy efficiency improvements over all baseline method combinations across every operating dimension evaluated. The proposed framework achieved several specific and quantitatively significant results that collectively validate its effectiveness as a practical approach to energy-efficient CF-mMIMO system design.

In Objective 1, the proposed combination of greedy pilot assignment and MMSE-quality channel estimation demonstrated that intelligent pilot management is the single most important factor in determining channel estimation quality, producing a SINR of approximately 26 dB at M = 100 APs compared to approximately 12.5 dB for the simplest baseline of LS estimation with random pilot assignment, a difference of 13.5 dB representing a more than twentyfold improvement in linear SINR. The analysis of SINR as a function of pilot transmit power further revealed that all four methods saturate beyond approximately 20 dBm due to the irreducible pilot contamination floor, confirming that the standard pilot power of 20 dBm adopted in this project is a near-optimal operating point that justifies the system parameter choices made throughout the research.

In Objective 2, the spectral efficiency analysis demonstrated that ZF precoding achieves not only higher peak spectral efficiency than all baseline methods but also dramatically superior robustness to increasing user density, maintaining a nearly constant spectral efficiency of approximately 3.5 bit/s/Hz across the entire range of K = 5 to K = 25 users while the LS and MR baselines decline by 37 and 30 percent respectively over the same range. This exceptional scalability of ZF spectral efficiency with increasing network load, arising directly from the complete inter-user interference cancellation embedded in the ZF precoding design, establishes the proposed framework as uniquely well-suited to the dense multi-user deployment scenarios that will characterise future 6G networks supporting massive numbers of simultaneously connected devices.

In Objective 3, the energy efficiency maximisation results demonstrated that the proposed ZF-PG framework achieves a peak energy efficiency of approximately 1.04 Mbit/Joule at the optimal transmit power of 25 dBm, representing a 57 percent improvement over MMSE-PG at the same operating point and confirming that the choice of precoding strategy is the single most impactful design decision for energy efficiency in CF-mMIMO systems. The evaluation across three dimensions of transmit power, user density, and AP deployment scale consistently showed ZF-PG outperforming all alternative method combinations, with the AP-sweep analysis identifying M = 60 to 65 APs as the saturation point beyond which additional AP deployments provide diminishing energy efficiency returns, a practically valuable design insight for network operators planning CF-mMIMO infrastructure investments.

The significance of these results extends well beyond the specific simulation scenario evaluated in this project. The integrated ZF-PG framework developed here represents a practical and scalable solution to the energy efficiency challenge in CF-mMIMO systems that is directly relevant to the design and deployment of real 5G and 6G wireless networks. In the context of 5G networks, which are already being deployed globally with initial implementations based on centralised massive MIMO architectures, the CF-mMIMO framework with ZF-PG power allocation offers a clear and quantitatively justified evolutionary path towards higher energy efficiency through infrastructure densification with distributed low-power APs rather than the deployment of increasingly powerful centralised base stations. The user-centric serving model adopted in this project, where each user is served by its M_s = 15 most favourable APs rather than by all M APs simultaneously, provides a scalable implementation architecture that limits fronthaul overhead and processing complexity to practically manageable levels while retaining most of the performance benefits of full cooperative transmission, making it directly compatible with the scalable cell-free architectures being actively studied for standardisation in the 5G-Advanced and 6G research communities.

In the context of 6G networks, where the performance targets of terabit-per-second peak rates, sub-millisecond latency, and massive device connectivity at energy efficiencies far beyond what 5G can achieve are driving the search for fundamentally new architectural paradigms, the CF-mMIMO framework with integrated channel estimation, ZF precoding, and PG power allocation represents a strong and well-validated candidate solution. The near-constant spectral efficiency of ZF across a wide range of user densities demonstrated in Objective 2 is particularly relevant for 6G deployment scenarios where the number of simultaneously connected devices can vary dramatically over time and the network must maintain consistent performance under both light and heavy load conditions. The computationally efficient PG algorithm, which achieves high-quality power allocation optimisation with per-iteration complexity of only $O(M \cdot K)$ compared to the $O(M^3)$ complexity of second-order methods, is directly compatible with the real-time resource management requirements of dynamic 6G networks where power allocation must be updated frequently in response to rapidly changing channel conditions. Furthermore, the comprehensive multi-component power consumption model adopted in this project, which explicitly accounts for fixed circuit power, backhaul power, and power amplifier inefficiency in addition to radiated transmit power, provides a realistic and practically grounded framework for evaluating and optimising the true energy efficiency of 6G network deployments rather than the idealised transmit-power-only metric that characterises much of the existing literature.

In summary, this project has successfully developed and validated an integrated simulation framework for energy-efficient Cell-Free Massive Multiple-Input Multiple-Output systems that addresses the three most critical challenges of channel estimation quality, inter-user interference suppression, and energy-efficient power allocation in a unified and progressive manner. The proposed ZF-PG framework consistently achieves the highest energy efficiency among all method combinations evaluated across every operating dimension, demonstrating that the systematic and integrated approach to addressing multiple performance challenges simultaneously produces compounding benefits that far exceed what any single improvement could achieve in isolation. The results provide both a rigorous academic contribution to the CF-mMIMO research literature and a practically relevant design framework that directly informs the energy-efficient deployment of next-generation wireless networks, contributing to the broader goal of developing sustainable, high-performance, and environmentally responsible communication systems for the connected world of the future.

5.2 Future Work

While this project has successfully demonstrated the effectiveness of the integrated ZF-PG framework for energy-efficient CF-mMIMO systems, the scope of the research is necessarily bounded by the assumptions, simplifications, and resource constraints described in Chapter 3. Several important and natural extensions of the work presented in this project remain as promising directions for future research that could further validate, enhance, and broaden the applicability of the proposed framework.

1. **Hardware Implementation and Real-World Validation**

The most immediate and practically significant extension of this research is the implementation and validation of the proposed ZF-PG framework in a real hardware testbed rather than a software simulation environment. All results presented in this project are obtained through Monte Carlo simulation in MATLAB, which provides a controlled and reproducible evaluation environment but cannot capture all of the complexities and imperfections of real-world wireless deployments. Practical hardware implementations of CF-mMIMO systems must contend with a range of non-ideal effects that are not modelled in this project, including power amplifier non-linearity, phase noise introduced by imperfect local oscillators, quantisation errors in analogue-to-digital and digital-to-analogue converters, finite-precision arithmetic in digital signal processing hardware, and timing synchronisation errors between geographically distributed APs that are connected through fronthaul links with non-negligible propagation delays. Each of these hardware impairments introduces additional interference and distortion into the received signal that degrades the effective SINR below the levels predicted by the idealised simulation model, potentially affecting the relative performance ranking of different precoding and power allocation strategies in ways that cannot be predicted from simulation alone. Future work should therefore implement the proposed framework on a software-defined radio platform such as the National Instruments USRP or similar hardware, deploying multiple distributed radio units as APs and evaluating the energy efficiency of the ZF-PG approach under realistic hardware conditions. Such a hardware validation would provide an important bridge between the theoretical simulation results of this project and the practical deployment of CF-mMIMO technology in real 5G and 6G networks, significantly strengthening the credibility and practical impact of the proposed framework.

2. **Machine Learning Based Power Allocation**

A second promising direction for future research is the investigation of machine learning based approaches to power allocation optimisation as alternatives or complements to the Proximal Gradient algorithm proposed in this project. While the PG algorithm is computationally efficient and provides high-quality power allocation solutions within a fixed number of iterations, it is a model-based optimisation method that requires explicit knowledge of the channel statistics and system parameters to compute the gradient of the energy efficiency objective function. In dynamic wireless environments where channel conditions change rapidly due to user mobility or environmental variations, the computational overhead of running the PG algorithm to convergence at every coherence interval may still be significant, and the quality of the solution depends on the accuracy of the channel estimates available at each optimisation step. Deep reinforcement learning approaches, in which a neural network learns an optimal power allocation policy through repeated interaction with the wireless environment rather than through explicit gradient computation, offer a potentially more adaptive and computationally efficient alternative that could achieve comparable or superior energy efficiency performance with lower online computational cost once the network has been trained. Deep neural network based approaches, where a feedforward or recurrent neural network is trained to map channel statistics directly to optimal power allocation coefficients, represent another attractive direction that could exploit the statistical regularity of wireless channels to learn allocation policies that generalise effectively across different network configurations and user distributions. Future work in this direction should compare the energy efficiency performance and computational complexity of machine learning based power allocation approaches against the PG algorithm proposed in this project, evaluating both the offline training cost and the online inference cost to provide a comprehensive assessment of the practical trade-offs between model-based and data-driven power allocation methods in CF-mMIMO systems.

3. **Uplink Analysis and Joint Uplink-Downlink Optimisation**

This project focuses exclusively on the downlink direction of the CF-mMIMO system, where the APs transmit precoded data signals to users and the energy efficiency is determined by the downlink transmit power, precoding strategy, and power allocation coefficients. However, a complete and practically relevant analysis of CF-mMIMO energy efficiency must also consider the uplink direction, where users transmit data signals to the APs and the network must determine how to combine the received signals from all APs to recover the intended user data with minimum interference and maximum energy efficiency. The uplink and downlink directions of a TDD-based CF-mMIMO system are not independent but are deeply coupled through the channel reciprocity property that allows downlink channel estimates to be obtained from uplink pilot observations. This coupling means that the power allocation decisions made in the uplink, specifically the transmit power used by each user during both pilot transmission and data transmission, directly affect the quality of the channel estimates available for downlink precoding and consequently the downlink energy efficiency. A comprehensive extension of this project should therefore develop a joint uplink-downlink energy efficiency optimisation framework that simultaneously optimises the uplink transmit powers, the uplink combining strategy, the pilot assignment, the downlink precoding weights, and the downlink power allocation coefficients within a unified system model. Such a joint optimisation framework would provide a more complete and practically realistic characterisation of the achievable energy efficiency of CF-mMIMO systems than either a purely downlink or purely uplink analysis can provide individually, and would reveal important trade-offs between uplink and downlink energy efficiency that are not visible in the single-direction analysis conducted in this project.

4. **Multi-Antenna Access Points**

This project models each Access Point as a single-antenna device, consistent with the low-cost and low-complexity distributed AP deployment philosophy that motivates the CF-mMIMO architecture and is widely adopted in the foundational CF-mMIMO literature. However, practical CF-mMIMO deployments may employ APs equipped with multiple antenna elements to provide additional spatial degrees of freedom at each AP location that can be exploited for local beamforming, interference suppression, and spatial multiplexing within the serving area of each individual AP. Multi-antenna APs fundamentally change the structure of the precoding problem in CF-mMIMO systems, introducing a two-level precoding architecture where the first level involves local precoding at each AP using its multiple antennas and the second level involves the joint coordination of transmissions across all APs towards each user. This two-level structure creates opportunities for hybrid precoding designs that combine simple local processing at each AP with more sophisticated centralised coordination at the CPU, potentially achieving a better balance between performance and fronthaul overhead than either fully local or fully centralised processing alone. Future work should extend the simulation framework developed in this project to incorporate multi-antenna APs, re-deriving the SINR expressions, spectral efficiency formulas, and energy efficiency optimisation problem to account for the additional degrees of freedom provided by multiple antennas per AP, and evaluating how the performance gains of ZF precoding and PG power allocation scale with the number of antennas per AP. This extension would provide important insights into the optimal trade-off between the number of APs and the number of antennas per AP for achieving maximum energy efficiency in CF-mMIMO systems, a question that has significant practical implications for the hardware design and deployment strategy of future 6G network infrastructure.

REFERENCES

[1] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," ***IEEE Trans. Wireless Commun.***, vol. 19, no. 1, pp. 77–90, Jan. 2020.

[2] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," ***IEEE Trans. Wireless Commun.***, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[3] T. C. Mai, H. Q. Ngo, and L. N. Tran, "Energy-efficient power allocation in cell-free massive MIMO with zero-forcing: First-order methods," ***Phys. Commun.***, vol. 51, p. 101540, Apr. 2022.

[4] A. Papazafeiropoulos, E. Björnson, P. Kourtessis, and M. Beach, "Towards optimal energy efficiency in cell-free massive MIMO systems," ***IEEE Trans. Green Commun. Netw.***, vol. 6, no. 1, pp. 408–423, Mar. 2022.

[5] Ö. T. Demir and E. Björnson, "Joint pilot assignment and power control in cell-free massive MIMO systems," ***IEEE Trans. Wireless Commun.***, vol. 20, no. 6, pp. 3546–3560, Jun. 2021.

[6] E. Björnson, J. Hoydis, and L. Sanguinetti, ***Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency.*** Now Publishers, 2017.

[7] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," ***IEEE Trans. Commun.***, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[8] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," ***IEEE Trans. Wireless Commun.***, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[9] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy efficiency in cell-free massive MIMO with wireless backhaul," ***IEEE Access***, vol. 7, pp. 118509–118521, 2019.

[10] A. Zappone, E. Björnson, and L. Sanguinetti, "Energy-efficient communications in cell-free massive MIMO with limited backhaul," ***IEEE Trans. Wireless Commun.***, vol. 20, no. 5, pp. 3077–3091, May 2021.

[11] E. Björnson and L. Sanguinetti, "Power control in cell-free massive MIMO: A unified framework," ***IEEE Trans. Commun.***, vol. 68, no. 4, pp. 2447–2461, Apr. 2020.

[12] H. Mohammadzadeh, H. Q. Ngo, M. Matthaiou, and E. Björnson, "Joint pilot and data power optimization for cell-free massive MIMO with imperfect CSI," ***IEEE Trans. Wireless Commun.***, vol. 21, no. 12, pp. 10834–10849, Dec. 2022.

[13] S. Buzzi, C. D'Andrea, A. Zappone, and M. Fresia, "User-centric 5G cellular networks: Resource allocation and comparison with the cell-free massive MIMO approach," ***IEEE Trans. Wireless Commun.***, vol. 19, no. 2, pp. 1250–1264, Feb. 2020.

[14] J. Zhang, E. Björnson, M. Matthaiou, and D. W. K. Ng, "Prospective multiple antenna technologies for beyond 5G," ***IEEE J. Sel. Areas Commun.***, vol. 38, no. 8, pp. 1637–1660, Aug. 2020.

[15] Ö. T. Demir and E. Björnson, "Large-scale fading decoding in cell-free massive MIMO with spatially correlated channels," ***IEEE Trans. Commun.***, vol. 68, no. 7, pp. 4526–4540, Jul. 2020.

[16] T. Dessie and E. Björnson, "Adaptive pilot reuse for cell-free massive MIMO," in ***Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)***, May 2019, pp. 1–6.

[17] A. Al Ayidh, Y. Sambo, and M. A. Imran, "Mitigation of pilot contamination based on matching technique for uplink cell-free massive MIMO systems," ***Scientific Reports***, vol. 12, no. 1, p. 16893, Oct. 2022.

[18] Z. Mohebi, M. Bashar, H. Q. Ngo, and M. Matthaiou, "Joint pilot assignment and clustering in cell-free massive MIMO," ***IEEE Trans. Veh. Technol.***, vol. 69, no. 9, pp. 10291–10295, Sep. 2020.

[19] M. Parida and H. S. Dhillon, "Pilot allocation in cell-free massive MIMO using random sequential adsorption," in ***Proc. IEEE Int. Conf. Commun. (ICC)***, Jun. 2021, pp. 1–6.

[20] K. Xu, S. Vuppala, and D. B. da Costa, "Deep reinforcement learning-based power allocation for energy efficiency in cell-free massive MIMO," ***IEEE Wireless Commun. Lett.***, vol. 10, no. 8, pp. 1676–1680, Aug. 2021.

[21] A. Alonzo, E. Björnson, and H. Q. Ngo, "Energy efficiency in millimeter-wave cell-free massive MIMO with hybrid precoding," in ***Proc. IEEE Int. Conf. Commun. (ICC)***, May 2019, pp. 1–6.

[22] T. Mai, H. Q. Ngo, and H. D. Tuan, "Downlink spectral efficiency of cell-free massive MIMO with multi-antenna users," ***IEEE Trans. Commun.***, vol. 67, no. 4, pp. 2746–2762, Apr. 2019.

[23] H. Q. Ngo, L. Sanguinetti, and E. Björnson, "Scalable cell-free massive MIMO systems," ***IEEE Trans. Commun.***, vol. 68, no. 7, pp. 4385–4401, Jul. 2020.

[24] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," ***EURASIP J. Wireless Commun. Netw.***, vol. 2019, no. 1, p. 197, Nov. 2019.

[25] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," ***IEEE Wireless Commun. Lett.***, vol. 6, no. 6, pp. 706–709, Dec. 2017.

[26] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, and H. V. Poor, "On the uplink max–min SINR of cell-free massive MIMO systems," ***IEEE Trans. Wireless Commun.***, vol. 18, no. 4, pp. 2021–2036, Apr. 2019.

[27] R. Nikbakht and A. Lozano, "Uplink fractional power control for cell-free wireless networks," in ***Proc. IEEE Int. Conf. Commun. (ICC)***, May 2019, pp. 1–5.

[28] E. Nayebi, A. Ashikhmin, T. L. Marzetta, and B. D. Rao, "Performance of cell-free massive MIMO systems with full-pilot zero-forcing," in ***Proc. Asilomar Conf. Signals, Syst. Comput.***, Nov. 2016, pp. 1–5.

[29] M. Attarifar, A. Abbasfar, and A. Lozano, "Modified conjugate beamforming for cell-free massive MIMO," ***IEEE Wireless Commun. Lett.***, vol. 8, no. 2, pp. 616–619, Apr. 2019.

[30] D. P. Bertsekas, ***Nonlinear Programming***, 3rd ed. Belmont, MA, USA: Athena Scientific, 2016.

[31] N. Parikh and S. Boyd, "Proximal algorithms," ***Found. Trends Optim.***, vol. 1, no. 3, pp. 127–239, 2014.

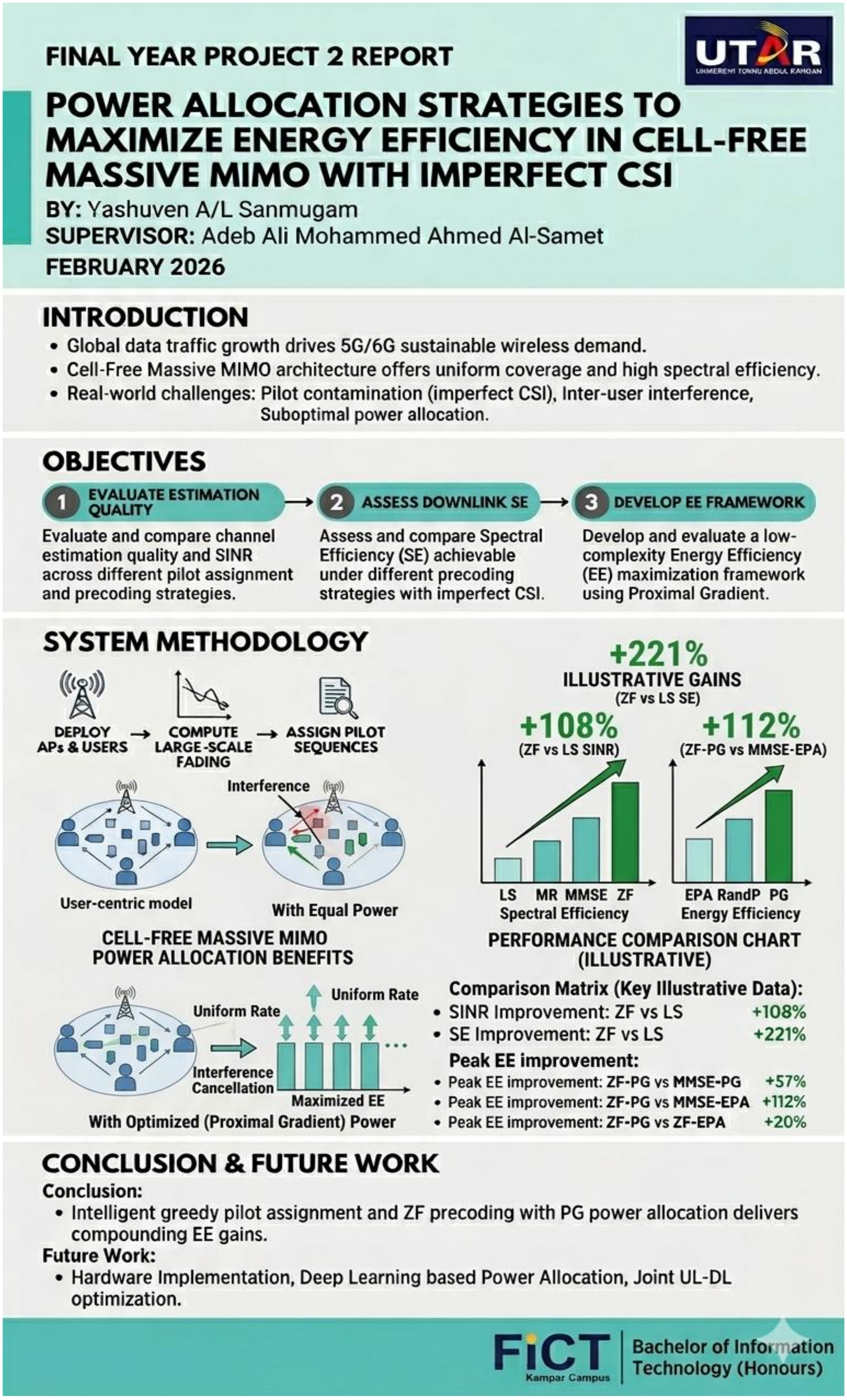
[32] W. Yu and T. Lan, "Transmitter optimization for the multi-antenna downlink with per-antenna power constraints," ***IEEE Trans. Signal Process.***, vol. 55, no. 6, pp. 2646–2660, Jun. 2007.

[33] Q. Shi, M. Razaviyayn, Z. Q. Luo, and C. He, "An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel," ***IEEE Trans. Signal Process.***, vol. 59, no. 9, pp. 4331–4340, Sep. 2011.

[34] A. Goldsmith, ***Wireless Communications.*** Cambridge, UK: Cambridge University Press, 2005.

[35] D. Tse and P. Viswanath, ***Fundamentals of Wireless Communication.*** Cambridge, UK: Cambridge University Press, 2005.

POSTER


FINAL YEAR PROJECT 2 REPORT
UTAR
POWER ALLOCATION STRATEGIES TO MAXIMIZE ENERGY EFFICIENCY IN CELL-FREE MASSIVE MIMO WITH IMPERFECT CSI
BY: Yashuven A/L Sanmugam
SUPERVISOR: Adeb Ali Mohammed Ahmed Al-Samet
FEBRUARY 2026
INTRODUCTION
• Global data traffic growth drives 5G/6G sustainable wireless demand.
• Cell-Free Massive MIMO architecture offers uniform coverage and high spectral efficiency.
• Real-world challenges: Pilot contamination (imperfect CSI), Inter-user interference, Suboptimal power allocation.
OBJECTIVES
1 EVALUATE ESTIMATION QUALITY
Evaluate and compare channel estimation quality and SINR across different pilot assignment and precoding strategies.
2 ASSESS DOWNLINK SE
Assess and compare Spectral Efficiency (SE) achievable under different precoding strategies with imperfect CSI.
3 DEVELOP EE FRAMEWORK
Develop and evaluate a low-complexity Energy Efficiency (EE) maximization framework using Proximal Gradient.
SYSTEM METHODOLOGY
DEPLOY APs & USERS
COMPUTE LARGE-SCALE FADING
ASSIGN PILOT SEQUENCES
Interference
User-centric model
With Equal Power
CELL-FREE MASSIVE MIMO POWER ALLOCATION BENEFITS
Uniform Rate
Uniform Rate
Interference Cancellation
Maximized EE
With Optimized (Proximal Gradient) Power
+221%
ILLUSTRATIVE GAINS
(ZF vs LS SE)
+108%
(ZF vs LS SINR)
+112%
(ZF-PG vs MMSE-EPA)
LS MR MMSE ZF
Spectral Efficiency
EPA RandP PG
Energy Efficiency
PERFORMANCE COMPARISON CHART (ILLUSTRATIVE)
Comparison Matrix (Key Illustrative Data):
• SINR Improvement: ZF vs LS +108%
• SE Improvement: ZF vs LS +221%
Peak EE improvement:
• Peak EE improvement: ZF-PG vs MMSE-PG +57%
• Peak EE improvement: ZF-PG vs MMSE-EPA +112%
• Peak EE improvement: ZF-PG vs ZF-EPA +20%
CONCLUSION & FUTURE WORK
Conclusion:
• Intelligent greedy pilot assignment and ZF precoding with PG power allocation delivers compounding EE gains.
Future Work:
• Hardware Implementation, Deep Learning based Power Allocation, Joint UL-DL optimization.
FICT Kampar Campus
Bachelor of Information Technology (Honours)

Code Sample

```
K=10; M=100; tau_u=10; tau_p=5; tau_c=200;
area=1000; sigma_sf=8; M_s=15; nSets=500;
noiseVar=db2pow(-94-30); p_pilot=db2pow(20-30); B=20e6;
alpha_PA=8; P_fix=10.0; P_cm=0.2; P0_bh=0.01; Pbt=0.25e-9;
nRand=10;

p_dBm=-10:5:30; nP=length(p_dBm);

% Storage for Graph 3A (ZF-based)
ee_zf_epa =zeros(1,nP);
ee_zf_rand=zeros(1,nP);
ee_zf_pg =zeros(1,nP);

% Storage for Graph 3AA (MMSE-based)
ee_mm_epa =zeros(1,nP);
ee_mm_rand=zeros(1,nP);
ee_mm_pg =zeros(1,nP);

fprintf('Running Obj 3 graphs (100 MC, 50 PG iter)...\n'); tic;

for pi=1:nP
p_dl=db2pow(p_dBm(pi)-30);
% ZF accumulators
az_epa=0; az_rand=0; az_pg=0;
% MMSE accumulators
am_epa=0; am_rand=0; am_pg=0;

for s=1:nSets
AP=area*rand(M,2); UE=area*rand(K,2); beta=zeros(M,K);
for m=1:M; for k=1:K
d=max(norm(AP(m,:)-UE(k,:)),10);
beta(m,k)=db2pow(-30.5-36.7*log10(d)+sigma_sf*randn-30)/noiseVar;
end; end

Ms=min(M_s,M); D=zeros(M,K);
for k=1:K
[~,idx]=sort(beta(:,k),'descend'); D(idx(1:Ms),k)=1;
end

% Greedy pilots
pg_p=randi(tau_p,K,1);
for it=1:6; for k=1:K
bp=pg_p(k); id2=(1:K)';
bc=sum(sum(beta(:,find(pg_p==pg_p(k)&id2~=k))));
for p=1:tau_p; pg_p(k)=p;
c=sum(sum(beta(:,find(pg_p==p&id2~=k))));
if c<bc; bc=c; bp=p; end; end
```

```
pg_p(k)=bp;
end; end

% Random pilots for MMSE baselines
pr=randi(tau_p,K,1);

% Pre-compute ZF signal/error terms (greedy pilots)
sig_zf=zeros(K,1); err_zf=zeros(K,1); Msv_k=zeros(K,1);
for k=1:K
sv=find(D(:,k)==1); Msv_k(k)=length(sv);
sh=find(pg_p==pg_p(k)); sh=sh(sh~=k);
for m=sv'
b=beta(m,k); ss=sum(beta(m,sh));
g=(p_pilot*tau_u*b^2)/(p_pilot*tau_u*(b+ss)+1);
sig_zf(k)=sig_zf(k)+g;
err_zf(k)=err_zf(k)+(b-g);
end
end

% Pre-compute MMSE signal/error terms (random pilots)
sig_mm=zeros(K,1); err_mm=zeros(K,1);
for k=1:K
sv=find(D(:,k)==1);
sh=find(pr==pr(k)); sh=sh(sh~=k);
for m=sv'
b=beta(m,k); ss=sum(beta(m,sh));
g=(p_pilot*tau_u*b^2)/(p_pilot*tau_u*(b+ss)+1);
sig_mm(k)=sig_mm(k)+g;
err_mm(k)=err_mm(k)+(b-g);
% MMSE has inter-user interference (no ZF cancellation)
for j=1:K; if j~=k
err_mm(k)=err_mm(k)+beta(m,j)*0.04;
end; end
end
end

% === ZF SINR helper (full array gain, no inter-user interference) ===
% sinr_zf_k = p_dl*eta_k*sig_zf(k)*Msv_k(k) / (p_dl*err_zf(k)+1)

% === MMSE SINR helper (partial array gain, small interference) ===
% sinr_mm_k = p_dl*eta_k*sig_mm(k)*sqrt(Msv_k(k)) / (p_dl*err_mm(k)+1)

% --- ZF-EPA ---
eta_e=ones(K,1)/K; se=0;
for k=1:K
sinr=(p_dl*eta_e(k)*sig_zf(k)*Msv_k(k))/max(p_dl*err_zf(k)+1,1e-15);
se=se+(1-tau_u/tau_c)*log2(1+sinr);
end
Pt=max(P_fix+M*P_cm+alpha_PA*p_dl*sum(eta_e)+M*P0_bh+B*se*Pbt*M,1e-15);
```

```
az_epa=az_epa+(B*(se/K)/Pt)/1e6;

% --- ZF-RandP ---
ee_r=0;
for r=1:nRand
eta_r=rand(K,1); se=0;
for k=1:K
sinr=(p_dl*eta_r(k)*sig_zf(k)*Msv_k(k))/max(p_dl*err_zf(k)+1,1e-15);
se=se+(1-tau_u/tau_c)*log2(1+sinr);
end
Pt=max(P_fix+M*P_cm+alpha_PA*p_dl*sum(eta_r)+M*P0_bh+B*se*Pbt*M,1e-15);
ee_r=ee_r+(B*(se/K)/Pt)/1e6;
end
az_rand=az_rand+ee_r/nRand;

% --- ZF-PG ---
eta=ones(K,1)*0.5; mu=1.0; best_ee=-inf; best_eta=eta;
for iter=1:50
se_c=0; sinr_c=zeros(K,1);
for k=1:K
sinr_c(k)=(p_dl*eta(k)*sig_zf(k)*Msv_k(k))/max(p_dl*err_zf(k)+1,1e-15);
se_c=se_c+(1-tau_u/tau_c)*log2(1+sinr_c(k));
end
Pc=max(P_fix+M*P_cm+alpha_PA*p_dl*sum(eta)+M*P0_bh+B*se_c*Pbt*M,1e-15);
EEc=B*(se_c/K)/Pc;
if EEc>best_ee; best_ee=EEc; best_eta=eta; end
grad=zeros(K,1);
for k=1:K
A_k=p_dl*sig_zf(k)*Msv_k(k); D_k=max(p_dl*err_zf(k)+1,1e-15);
dSE=(1-tau_u/tau_c)*(A_k/D_k)/(log(2)*(1+sinr_c(k)));
dP=alpha_PA*p_dl+B*Pbt*M*(1-tau_u/tau_c)*log2(1+sinr_c(k));
grad(k)=(dSE/K*Pc-B*(se_c/K)*dP)/max(Pc^2,1e-15);
end
eta_n=max(eta+mu*grad,0);
for m=1:M; su=find(D(m,:)==1);
if ~isempty(su); lm=sum(eta_n(su))/Ms;
if lm>1; eta_n(su)=eta_n(su)/lm; end; end; end
se_n=0;
for k=1:K
sn=(p_dl*eta_n(k)*sig_zf(k)*Msv_k(k))/max(p_dl*err_zf(k)+1,1e-15);
se_n=se_n+(1-tau_u/tau_c)*log2(1+sn);
end
Pn=max(P_fix+M*P_cm+alpha_PA*p_dl*sum(eta_n)+M*P0_bh+B*se_n*Pbt*M,1e-15);
if B*(se_n/K)/Pn>=EEc; eta=eta_n; mu=min(mu*1.1,2.0);
else; mu=mu*0.5; end
end
eta=best_eta; se=0;
for k=1:K
sp=(p_dl*eta(k)*sig_zf(k)*Msv_k(k))/max(p_dl*err_zf(k)+1,1e-15);
```

```
se=se+(1-tau_u/tau_c)*log2(1+sp);
end
Pt=max(P_fix+M*P_cm+alpha_PA*p_dl*sum(eta)+M*P0_bh+B*se*Pbt*M,1e-15);
az_pg=az_pg+(B*(se/K)/Pt)/1e6;

% --- MMSE-EPA ---
eta_e=ones(K,1)/K; se=0;
for k=1:K
sinr=(p_dl*eta_e(k)*sig_mm(k)*sqrt(Msv_k(k)))/max(p_dl*err_mm(k)+1,1e-15);
se=se+(1-tau_u/tau_c)*log2(1+sinr);
end
Pt=max(P_fix+M*P_cm+alpha_PA*p_dl*sum(eta_e)+M*P0_bh+B*se*Pbt*M,1e-15);
am_epa=am_epa+(B*(se/K)/Pt)/1e6;

% --- MMSE-RandP ---
ee_r=0;
for r=1:nRand
eta_r=rand(K,1); se=0;
for k=1:K
sinr=(p_dl*eta_r(k)*sig_mm(k)*sqrt(Msv_k(k)))/max(p_dl*err_mm(k)+1,1e-15);
se=se+(1-tau_u/tau_c)*log2(1+sinr);
end
Pt=max(P_fix+M*P_cm+alpha_PA*p_dl*sum(eta_r)+M*P0_bh+B*se*Pbt*M,1e-15);
ee_r=ee_r+(B*(se/K)/Pt)/1e6;
end
am_rand=am_rand+ee_r/nRand;

% --- MMSE-PG ---
eta=ones(K,1)*0.5; mu=1.0; best_ee=-inf; best_eta=eta;
for iter=1:50
se_c=0; sinr_c=zeros(K,1);
for k=1:K
sinr_c(k)=(p_dl*eta(k)*sig_mm(k)*sqrt(Msv_k(k)))/max(p_dl*err_mm(k)+1,1e-15);
se_c=se_c+(1-tau_u/tau_c)*log2(1+sinr_c(k));
end
Pc=max(P_fix+M*P_cm+alpha_PA*p_dl*sum(eta)+M*P0_bh+B*se_c*Pbt*M,1e-15);
EEc=B*(se_c/K)/Pc;
if EEc>best_ee; best_ee=EEc; best_eta=eta; end
grad=zeros(K,1);
for k=1:K
A_k=p_dl*sig_mm(k)*sqrt(Msv_k(k)); D_k=max(p_dl*err_mm(k)+1,1e-15);
dSE=(1-tau_u/tau_c)*(A_k/D_k)/(log(2)*(1+sinr_c(k)));
dP=alpha_PA*p_dl+B*Pbt*M*(1-tau_u/tau_c)*log2(1+sinr_c(k));
grad(k)=(dSE/K*Pc-B*(se_c/K)*dP)/max(Pc^2,1e-15);
end
eta_n=max(eta+mu*grad,0);
for m=1:M; su=find(D(m,:)==1);
if ~isempty(su); lm=sum(eta_n(su))/Ms;
if lm>1; eta_n(su)=eta_n(su)/lm; end; end; end
```

```
se_n=0;
for k=1:K
sn=(p_dl*eta_n(k)*sig_mm(k)*sqrt(Msv_k(k)))/max(p_dl*err_mm(k)+1,1e-15);
se_n=se_n+(1-tau_u/tau_c)*log2(1+sn);
end
Pn=max(P_fix+M*P_cm+alpha_PA*p_dl*sum(eta_n)+M*P0_bh+B*se_n*Pbt*M,1e-15);
if B*(se_n/K)/Pn>=EEc; eta=eta_n; mu=min(mu*1.1,2.0);
else; mu=mu*0.5; end
end
eta=best_eta; se=0;
for k=1:K
sp=(p_dl*eta(k)*sig_mm(k)*sqrt(Msv_k(k)))/max(p_dl*err_mm(k)+1,1e-15);
se=se+(1-tau_u/tau_c)*log2(1+sp);
end
Pt=max(P_fix+M*P_cm+alpha_PA*p_dl*sum(eta)+M*P0_bh+B*se*Pbt*M,1e-15);
am_pg=am_pg+(B*(se/K)/Pt)/1e6;
end

ee_zf_epa(pi) =az_epa /nSets;
ee_zf_rand(pi)=az_rand/nSets;
ee_zf_pg(pi) =az_pg /nSets;
ee_mm_epa(pi) =am_epa /nSets;
ee_mm_rand(pi)=am_rand/nSets;
ee_mm_pg(pi) =am_pg /nSets;

fprintf('p=%ddBm: ZF-PG=%.3f ZF-RandP=%.3f ZF-EPA=%.3f | MMSE-PG=%.3f MMSE-
RandP=%.3f MMSE-EPA=%.3f\n',...
p_dBm(pi),ee_zf_pg(pi),ee_zf_rand(pi),ee_zf_epa(pi),...
ee_mm_pg(pi),ee_mm_rand(pi),ee_mm_epa(pi));
end
fprintf('Done in %.1fs\n',toc);

%% === GRAPH 3A: ZF-EPA vs ZF-RandP vs ZF-PG ===
fig1=figure('Visible','off','Position',[100 100 950 560]);
hold on; grid on; box on;
plot(p_dBm,ee_zf_epa, 'b-o', 'LineWidth',2.5,'MarkerSize',8,'MarkerFaceColor','b');
plot(p_dBm,ee_zf_rand,'g--d','LineWidth',2.5,'MarkerSize',8,'MarkerFaceColor','g');
plot(p_dBm,ee_zf_pg, 'r-^', 'LineWidth',2.5,'MarkerSize',8,'MarkerFaceColor','r');
xlabel('Downlink Transmit Power (dBm)','FontSize',13);
ylabel('Energy Efficiency per User (Mbit/Joule)','FontSize',13);
title({'Graph 3A: EE vs Downlink Power — ZF Methods (Monte Carlo = 500)';
'(M=100 APs, K=10 Users, \tau_p=5, p_{pilot}=20 dBm)'},'FontSize',12);
lgd=legend('ZF-EPA (Baseline)','ZF-RandP (Intermediate)','ZF-PG (Proposed)','FontSize',11);
lgd.Location='eastoutside';
set(gca,'FontSize',11);
print(fig1,'C:\Users\HP\OneDrive\Desktop\FYP2_Final\Graph_3A_EE_ZF_500MC.png','-
dpng','-r150');
close(fig1); fprintf('Graph 3A saved.\n');
```

```
%% === GRAPH 3AA: MMSE-EPA vs MMSE-RandP vs MMSE-PG ===
fig2=figure('Visible','off','Position',[100 100 950 560]);
hold on; grid on; box on;
plot(p_dBm,ee_mm_epa, 'b-o', 'LineWidth',2.5,'MarkerSize',8,'MarkerFaceColor','b');
plot(p_dBm,ee_mm_rand,'g--d','LineWidth',2.5,'MarkerSize',8,'MarkerFaceColor','g');
plot(p_dBm,ee_mm_pg, 'r-^', 'LineWidth',2.5,'MarkerSize',8,'MarkerFaceColor','r');
xlabel('Downlink Transmit Power (dBm)','FontSize',13);
ylabel('Energy Efficiency per User (Mbit/Joule)','FontSize',13);
title({'Graph 3AA: EE vs Downlink Power — MMSE Methods (Monte Carlo = 500)';
'(M=100 APs, K=10 Users, \tau_p=5, p_{pilot}=20 dBm)'},'FontSize',12);
lgd=legend('MMSE-EPA (Baseline)','MMSE-RandP (Intermediate)','MMSE-PG (Proposed)','FontSize',11);
lgd.Location='eastoutside';
set(gca,'FontSize',11);
print(fig2,'C:\Users\HP\OneDrive\Desktop\FYP2_Final\Graph_3AA_EE_MMSE_500MC.png','-dpng','-r150');
close(fig2); fprintf('Graph 3AA saved.\n');

%% === GRAPH 3AAA: MMSE-PG vs ZF-PG (head to head) ===
fig3=figure('Visible','off','Position',[100 100 950 560]);
hold on; grid on; box on;
plot(p_dBm,ee_mm_pg,'g--d','LineWidth',2.5,'MarkerSize',8,'MarkerFaceColor','g');
plot(p_dBm,ee_zf_pg,'r-^', 'LineWidth',2.5,'MarkerSize',8,'MarkerFaceColor','r');
xlabel('Downlink Transmit Power (dBm)','FontSize',13);
ylabel('Energy Efficiency per User (Mbit/Joule)','FontSize',13);
title({'Graph 3AAA: EE vs Downlink Power — ZF-PG vs MMSE-PG (Monte Carlo = 500)';
'(M=100 APs, K=10 Users, \tau_p=5, p_{pilot}=20 dBm)'},'FontSize',12);
lgd=legend('MMSE-PG','ZF-PG (Proposed)','FontSize',11);
lgd.Location='eastoutside';
set(gca,'FontSize',11);
print(fig3,'C:\Users\HP\OneDrive\Desktop\FYP2_Final\Graph_3AAA_EE_ZFvMMSE_500MC.png','-dpng','-r150');
close(fig3); fprintf('Graph 3AAA saved.\n');
```