Optimized Small Cell Selection for Minimizing Power Consumption in 5G Radio Access Network

Enabling Energy-Efficient Operation in 5G Radio Access Networks

Prajakta Udas

Delft University of Technology

Optimized Small Cell Selection for Minimizing Power Consumption in 5G Radio Access Network

Enabling Energy-Efficient Operation in 5G Radio Access Networks

by

Prajakta Udas

to obtain the degree of Master of Science

at the Delft University of Technology,

to be defended publicly on Friday August 29, 2025 at 02:00 PM.

Student number:	5752396	
Project Duration:	September, 2024 – August, 2025	
Faculty:	Electrical Engineering, Mathematics and Computer Science, Delft	
Thesis Committee:	Dr. Ir. Eric Smeitink	Main Supervisor
	Dr. Ir. Edgar van Boven	Daily Supervisor
	Ir. Rogier Noldus	Daily Co-Supervisor
	Dr. Ir. Qing Wang	External Committee Member

Abstract

The energy consumption of mobile networks, particularly the 5G Radio Access Network (RAN), is becoming a growing concern due to its environmental and economic implications. As the demand for higher data rates and low-latency services intensifies, 5G networks, integrating macro cells and small cells, are emerging as critical infrastructures. Although small cells improve coverage and capacity, their increased deployment could lead to a significant rise in the overall power consumption of 5G systems.

Current small cell selection strategies by User Equipments (UEs), although effective in some cases, do not fully account for the dynamic nature of traffic conditions and the specific data requirements of users. Moreover, current techniques such as the maximum Signal-to-Interference-plus-Noise Ratio (max-SINR) and Cell Range Expansion (CRE) purely consider the signal strength of the link between the user and the base station to allocate users to the base station. However, this leads to inefficient utilization of base station resources and uneven distribution of load, causing congestion at some base stations while leaving others underutilized.

In order to address these gaps, this thesis proposes a Traffic Distribution Orchestrator (TDO) to manage the distribution of users between cells dynamically, and optimize energy efficiency without compromising network performance. The proposed cell selection model developed in this thesis also accounts for user mobility and dynamic traffic conditions. The model estimates instantaneous power consumption and informs a real-time algorithm user equipment-base station (UE-BS) association algorithm to dynamically allocate users to the cell which will enhance the energy efficiency of the network while ensuring the required Quality of Service (QoS) requirements. Complementing this, an adaptive sleep mode mechanism puts underutilized small cells in a low power mode and reactivates them when demand rises, using hysteresis to prevent state flapping and reduce idle power.

Through MATLAB simulations, the effectiveness of the model and algorithm is validated, with results indicating a significant reduction in network power consumption in heterogeneous 5G deployments. The proposed UE-BS association algorithm is compared with the max-SINR, CRE and a representative association method from the previous studies, whereas the proposed adaptive sleep mode mechanism is compared with fixed threshold sleep mode mechanism under both bursty and steady traffic. The proposed UE-BS association algorithm combined with the adaptive sleep mode mechanism reduces total network power consumption relative to baseline strategies. This research contributes to the advancement of sustainable 5G network architectures and offers insights into energy efficiency optimization in real-world scenarios.

Preface

This thesis, "Optimized Small Cell Selection for Minimizing Power Consumption in Radio Access Network", marks the culmination of my master's journey at Technical University of Delft - a journey that has tested me not just academically, but emotionally and personally.

First and foremost, I would like to express my sincere gratitude to my supervisors Dr. Ir. Eric Smeitink, Dr. Ir. Edgar van Boven, and Ir. Rogier Noldus, for their guidance, patience, and encouragement throughout this project. Their insights and support have been instrumental in shaping this thesis. Moreover, I would like to thank Dr. Ir. Qing Wang for completing my thesis committee and for taking the time to evaluate this thesis and be part of my thesis defense adding his expertise to my evaluation.

These three years have been extraordinarily difficult. I faced significant financial struggles, personal losses and periods of deep emotional hardship. I lost both my grandparents during this time, with the passing of my grandfather just months before my graduation. This has left a void that words cannot fill. I had always dreamed of showing him my master's degree, a dream that will now remain unfulfilled.

I am deeply grateful to my parents, who have supported me both emotionally and financially, even in the face of our own family's financial challenges. Their strength, love and quiet sacrifices gave me the foundation to keep going when everything felt uncertain.

I am profoundly grateful to Dhanush, who stood by me during my lowest points and offered steady support through my darkest days. Without your presence and understanding, I'm not sure I would have made it through. To Manav, thank you for your unwavering kindness, your quiet strength, and for helping me find confidence when I had nearly lost it. To all my friends who checked in, reached out, and made sure I was okay, your efforts, compassion, and friendship carried me forward more than you know.

And above all, I am grateful for the faith I have in God. In moments when nothing and no one seemed within reach, that faith was the only thing that kept me moving forward.

Completing this thesis is more than just an academic milestone. It is a symbol of perseverance, of healing and of the love and support that surrounded me when everything else felt uncertain.

Prajakta Udas
Delft, August 2025

List of Definitions

Access Point (AP): In the context of cell-free architecture, AP is a radio unit equipped with a few antennas, connected via fronthaul to a central processing unit. In the context of this thesis, the term to be used hereafter is the Radio Unit (RU). [1]

Backhaul: The backhaul connects the Centralized Unit (CU) to the Core Network (CN) [2].

Cell-Free Architecture: Cell-free architecture refers to the type of network architecture which combines the advantages of massive MIMO (Multiple Input Multiple Output) and distributed MIMO, where every UE is served by multiple geographically distributed APs (RUs) in the same time-frequency resources. Therefore, unlike cellular architecture, there are no cell boundaries [3].

Cellular Architecture: Cellular architecture comprises of the coverage area divided into a number of cells where each cell is served by a base station. All base stations are connected to the CN through backhaul links [4].

Centralized Unit (CU): The CU handles higher-layer protocol processing, including Radio Resource Control (RRC), mobility management, and IP packet routing. It connects to multiple DUs over the midhaul[5].

Distributed MIMO: Distributed MIMO refers to the deployment of multiple antennas across a wide area, which are distributed spatially and work together as if they were part of a single MIMO system [6].

Distributed Unit (DU): The DU performs real-time baseband processing tasks such as scheduling, beamforming, and HARQ. It interfaces with the RU via the fronthaul and may be deployed closer to the edge for low-latency processing [5].

Fronthaul: Fronthaul is the network that links the radio unit (RU) to the distributed unit (DU) [2].

Handovers: Handovers refer to the process of transferring an active user connection from one base station to another without service interruption. This is essential for maintaining seamless connectivity as users move through the network [7].

Macro Cells: Macro cells are base stations in a mobile network that provide a wide coverage area which can range from a few hundred meters up to a few kilometers [7].

Massive MIMO: Massive MIMO refers to the use of a large number of antennas at a base station to simultaneously serve many users in the same frequency band [8].

Midhaul: Midhaul connects the DU and the CU [2].

Open RAN: Open Radio Access Network (Open RAN) is an industry initiative to define open and standardized interfaces between the RU, DU, and CU. It allows multi-vendor interoperability, network flexibility, and cost-effective deployment through disaggregation and virtualization [2].

Radio Unit (RU): The RU is the component of a base station responsible for transmitting and receiving radio signals via the antennas. It handles functions like analog-to-digital conversion, RF amplification, and filtering, and is typically located near or integrated with the antenna [5].

Small Cells: Small cells are low-power, short-range base stations (e.g., femto cells, pico cells) used to increase coverage and capacity in dense areas. They offload traffic from macro cells and are crucial for 5G HetNets and energy-efficient deployment [2].

X-haul Network: The fronthaul, midhaul and backhaul together form the x-haul network [9].

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

Abbreviation	Definition
5G NR	5G New Radio
AAS	Active Antenna System
AAU	Active Antenna Unit
ABS	Almost Blank Subframes/Slots
AIR	Antenna Integrated Radio
ANN	Artificial Neural Network
ASM	Advanced Sleep Mode
BBU	BaseBand Unit
BN	Backhaul Network
BS	Base Station
BW	Bandwidth
CAPEX	CAPital EXpenditure
CIO	Cell Individual Offset
CN	Core Network
CRAN	Cloud-Radio Access Network
C-RAN	Centralized Radio Access Network
CRE	Cell Range Expansion
CU	Centralized Unit
DAC	Digital-to-Analog Converter
DL	Downlink
D-RAN	Distributed Radio Access Network
DRL	Deep Reinforcement Learning
E2E	End to End
EC	Energy Consumption
EE	Energy Efficiency
EPB	Energy Per Bit
EPON	Ethernet Passive Optical Network
FFT	Fast Fourier Transform
GPON	Gigabit-capable Passive Optical Network
H-CRAN	Heterogeneous Centralized Radio Access Network
HetNet	Heterogeneous Network
HLS	Higher Layer Split
HRAN	Heterogeneous RAN
HUDN	Heterogeneous Ultra-Dense Network
ICR	Interference Contribution Ratio
IFFT	Inverse Fast Fourier Transform
IoT	Internet of Things
LD	Load Distribution
LHLA	Low-High Load Association

Abbreviation	Definition
LLS	Lower Layer Split
LTE	Long Term Evolution
MC	Macro Cell
MCBS	Macro Cell Base Station
MDP	Markov Decision Process
ML	Machine Learning
mMIMO	massive Multiple Input Multiple Output
mmWave	millimeter Wave
NFV	Network Function Virtualization
NG-PON	Next-Generation Passive Optical Network
NSA	Non Standalone
OLT	Optical Line Terminal
ONU	Optical Network Unit
O-RAN	Open Radio Access Network
PA	Power Amplifier
PC	Power Consumption
PL	Path Loss
PON	Passive Optical Network
PRB	Physical Resource Block
PtmP	Point-to-multi-Point
PtP	Point-to-Point
QoS	Quality of Service
RAN	Radio Access Network
RF	Radio Frequency
RL	Reinforcement Learning
RLC	Radio Link Control
RLF	Radio Link Failure
RSRP	Reference Signal Received Power
RSSI	Received Signal Strength Indicator
RU	Radio Unit
SA	Standalone
SARSA	State Action Reward State Action
SC	Small Cell
SCBS	Small Cell Base Station
SE	Spectral Efficiency
SFR	Soft Frequency Reuse
SINR	Signal-to-Interference-plus-Noise Ratio
SNR	Signal-to-Noise Ratio
STAA	SINR and Throughput-Aware Association algorithm (Proposed algorithm)
TVWS	Television White Space
UE	User Equipment
UL	Uplink
VDSL	Very-high-bit-rate Digital Subscriber Line

Literature-Specific Abbreviations

Abbreviation	Full Form
ADA-CS	ADAptive Cell Selection
AHP	Analytic Hierarchy Process
BRETS	Bayesian Response Estimation and Threshold Search
DDLB	Dynamic Distance-Based Load Balancing
DMA	Decoupled Multiple Association
GC	Global Controllers
JUBSM	Joint UE-Backhaul Switch-off Model
LC	Local Controllers
LSCBS	Load-Sharing-based SCBS Sleep
MCUAA	Multi Connectivity-based User Association Algorithm
RBP	Resource Block Pair
RUBSM	Robust UE-Backhaul Switch-off Model
VTSP	Variable Threshold Sleep Process

List of Symbols in Literature Review (Explained in chapter 2)

Symbol	Definition
P_{BS}	Total power consumption of a base station
P_{PA}	Power consumed by the Power Amplifier per antenna
P_{RF}	Power consumed by the transceiver (RF) chain per antenna
P_{BB}	Power consumed by the baseband processing unit
N_{TRX}	Number of transceiver chains (antennas)
σ_{DC}	Power loss due to DC-DC conversion
σ_{MS}	Power loss due to main supply
σ_{cool}	Power loss due to cooling
P_t	Transmit power
η	Power amplifier efficiency
P_{cir}	Circuit power consumption
$P_{\text{load-independent}}$	Load-independent base station power
$P_{\text{load-dependent}}$	Load-dependent base station power
P_{DAC}	Power consumed by Digital-to-Analog Converter
P_{mix}	Power consumed by mixer
P_{filt}	Power consumed by filter
P_{syn}	Power consumed by synthesizer
Δp	Slope factor for load-dependent power
P_{out}	Output power from the PA
P_{sleep}	Power consumed in sleep mode
N_{RB}	Total number of available PRBs at the BS
f_u	Data rate required by UE u
x_u	Binary indicator of whether UE u is associated with the BS
N_{au}	Number of antennas used by UE u
$SINR_u$	SINR between UE u and the serving BS
ρ	Load slope factor
C_u	Capacity (in Mbps) required to serve UE u
C_{max}	Maximum capacity of the BS
N_u	Number of PRBs used by UE u
$P_{switching}$	Power consumed during switching states
$s_{\text{off}}, s_{\text{sleep}}$	Binary indicators for OFF and SLEEP transitions
$P_{on-off}, P_{on-sleep}$	Switching power from ON to OFF/SLEEP
P_0	Baseline BS power (to keep BS operational)
P_{Tran}	Power consumed by transceivers
M_{ac}	Number of active RF chains
D_{PA}	Static PA power per RF chain
C	Number of carriers
P_{AAU}	Total power consumption of an Active Antenna Unit
T	Number of time slots or intervals
$M_{av,t}$	Number of active carriers in time t
$D_{Tran,t}$	Transceiver power per active carrier in time t
P_{PON}^{BH}	Power consumed by PON-based backhaul
N_{BS}	Number of base stations connected to backhaul
P_{ONU}	Power consumed by ONU (optical network unit)
N_{port}^{GPON}	Number of ports in GPON OLT
P_{port}^{GPON}	Power consumed by a single GPON port

Symbol	Definition
$N_{UL/DL}$	Number of uplink/downlink interfaces
P_{SFP+}	Power consumed by the SFP+ transceiver
C_{port}^{GPON}	Capacity per GPON port
D_{ONU}^{max}	Maximum data rate per ONU
$N_{\text{splitter-ratio}}$	Passive splitter ratio in PON
C_t	Total traffic load to be carried by the backhaul
D_{int}^{max}	Maximum data rate per electrical interface
P_{BH}^{l}	Power consumption of backhaul link l
T_{BH}^{l}	Number of transceiver chains in link l
S_{BH}^{l}	State of the backhaul link (e.g., sleep or active)
P_{0l}^{BH}	Static power of backhaul link l
ΔP_{l}^{BH}	Slope of dynamic power consumption for link l
P_{dl}^{BH}	Dynamic power consumption of link l
G_{rx}^{l}	Receiver antenna gain of link l
L_{rx}^{l}	Receiver link losses
PL^{l}	Path loss for link l
LM^{l}	Link margin
T_{n}^{l}	Thermal noise for link l
NF^{l}	Receiver noise figure
R_{BS}^{l}	Data rate for BS over link l
BW^{l}	Bandwidth allocated for link l
P_{fixed}^{l}	Circuit power of backhaul transceiver
λ	Power consumed per bps of traffic
t_s	Traffic to be sent per second

List of Symbols in Proposed System Model (Explained in chapter 3)

Symbol	Definition
$\mathcal{U}$	Set of all UEs
$\mathcal{B}$	Set of all BSs (macro cells and small cells)
$\mathcal{M}$	Set of macro cells
$\mathcal{S}$	Set of small cells
i	Index for UE
j	Index for base stations (both macro cells and small cells)
s	Index for small cells
m	Index for macro cells
$a_{i,j} \in \{0,1\}$	Represents association between BS j and UE i (1=associated, 0=not associated)
$\alpha_s \in \{0,1\}, s \in \mathcal{S}$	Sleep state of small cell (1=sleep, 0=active)
PL_{SCBS}, PL_{MCBS}	Path loss for macro cell and small cell respectively
d	Distance between UE and BS in kilometers
$\text{SINR}_{i,j}$	SINR of user i from BS j
$\text{SINR}_{threshold}$	Minimum SINR required for UE-BS association
$P_{tx_j}^{max}$	Maximum transmission power for BS j
$G_{i,j}$	Channel gain between UE i and BS j
N_0	Thermal Noise
R_i	Throughput of UE i
D_i	Throughput demand of UE i
B_i	Bandwidth allotted to UE i
L_j	Load at base station j
$PRB_{i,j}$	Number of PRBs required by UE i from BS j
B_{PRB}	Bandwidth of one PRB
P_{total}	Total network power consumption
P_{MCBS} and P_{SCBS}	Power consumption of macro cell and small cell BS respectively
$P_{switching}$	Power consumed by the network while switching between on and sleep states
P_{active}	Load independent power consumption of a BS in active state
P_{sleep}	Load independent power consumption of a small cell during sleep state
$P_{on-sleep}$	Power consumed by individual small cell while switching between on and sleep states
η_{PA}	Efficiency of the power amplifier
$T_s \in \{0,1\}$	Transition between on and sleep states for small cell s(1=transition, 0=no transition)
EE_{total}	Energy efficiency of the network
$score_s$	Computation of cost function governed by load and power consumption metrics
P_{max} and P_{min}	Maximum and minimum power that the network can consume, respectively
epb_{total}	Average energy per bit values for all users
$T_{threshold_s}$	Threshold value for deciding the transition of small cell s between sleep and active state

1
Introduction

5G networks mark a transformative shift in wireless communication, designed to meet the increasing demands of a connected, data-driven world. Beyond just faster data rates, 5G supports diverse applications, including ultra-reliable low-latency communication and massive connectivity for billions of devices. It enables next-generation technologies such as autonomous vehicles, smart cities, and immersive experiences such as augmented reality (AR) and virtual reality (VR), revolutionizing industries and reshaping how people and devices interact.

1.1. Problem statement and research gap

In order to meet the growing traffic demands, 5G introduces the concept of small cells, which are low-power, short-range base stations that help extend coverage and increase the capacity of mobile networks. Unlike traditional macro cells, which have much larger coverage areas, small cells have a smaller coverage area and are designed to be deployed in areas with high user density such as urban centers, stadiums, shopping malls, and other public places. They are crucial for high speed, low latency, and efficient usage of network resources. A paper published by Ericsson [10] compares the power consumption of LTE (Long Term Evolution, also called 4G) versus that of 5G. Contrary to the notion, the study suggests that the complete shift from LTE to 5G New Radio (NR) could lead to even higher energy savings while significantly enhancing the performance in terms of speed and latency, which can be seen in Figure 1.1 [10].

The deployment of 5G has not completely replaced 4G networks. Instead, both technologies coexist, with 4G still serving as the backbone for coverage and reliability, while 5G introduces enhanced capacity, low latency, and higher data rates. In fact, most operators have rolled out 5G in Non-StandAlone (NSA) mode, where 5G radio access is anchored to the 4G core. This ensures seamless mobility and efficient use of existing infrastructure. As the transition toward StandAlone (SA) 5G progresses, coexistence with 4G remains essential, both to maintain service continuity and to maximize spectrum utilization. Consequently, network design, resource allocation, and energy efficiency strategies must account for the dual operation of 4G and 5G systems.

Figure 1.1 demonstrates the energy saving potential of 5G NR. As seen in Figure 1.1, the addition of LTE small cells leads to a substantial increase in power consumption. The small cells are referred to as 'micro' base stations in Figure 1.1. Unlike LTE, adding NR small cells results in a slight increase in power consumption. This trend is observed due to long sleep

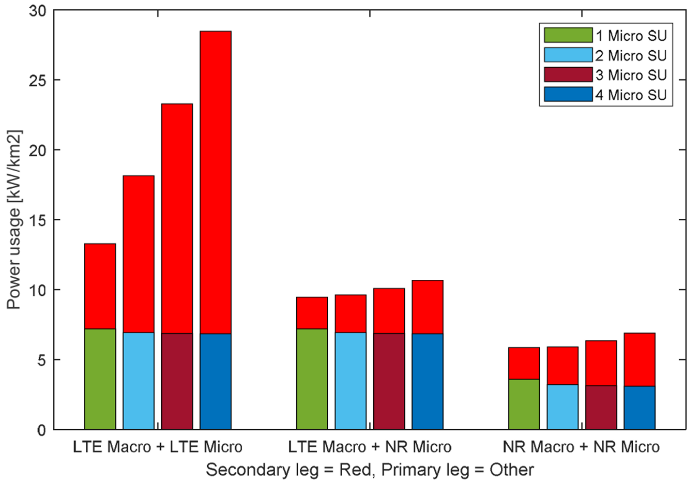


Figure 1.1: Super dense urban Heterogeneous Network daily power consumption

modes that are accessible to NR compared to LTE. The study concludes that the addition of NR small cells to LTE macro cells alone can lead to almost 50% reduction in power consumption. Moreover, upgrading LTE macro cells to NR macro cells could lead to energy savings of up to 70%. While the technologies employed in 5G significantly outperform those of previous generations, direct comparisons may not be entirely fair due to the differences in the number of devices, the continuously increasing traffic demands, and the higher densities of small cells in 5G networks.

According to a report published in 2021 [11], the global energy consumption of the telecommunications industry accounts for 2-3% of the total global energy consumption, making it one of the most energy-intensive industries. Energy costs, comprising fuel and electricity consumption, contributed to more than 90% of the total energy consumption. Majority of this energy was consumed by RANs while a smaller share was consumed by data centers and fiber transport. However, this increase in energy consumption cannot only be attributed to the advent of newer technologies such usage of millimeter wave frequency bands or massive MIMO used in 5G. The operation of LTE and 5G networks, coupled with exponential growth in the data traffic are projected to triple the energy consumption. As per Ericsson Mobility Report published in 2019 [12], the average data usage per smartphone in 2015 was 2GB per month. This figure rose to 10GB per month by 2019 and it was expected to rise to 42GB by 2029. However, the recent Mobility Report published by Ericsson [13] has mentioned that the market has experienced a decline in the mobile data growth rate since the second quarter of 2024. The network data growth rate is expected to decline from 21% in 2024 to 16% in 2030. The 5G subscriptions are expected to surpass that of 4G by 2027 and 5G is expected to carry approximately 80% of the total traffic by 2030. Despite the decline in growth rate, unprecedented data traffic is one of the major contributors to high energy consumption.

To manage such vast amounts of data, a dense network of small cells with varying coverage areas and both indoor and outdoor scenarios must be implemented. The definitions of dense and ultra-dense networks vary from literature to literature. However, according to [14], a density of more than 10^3 cells/km^2 or more than 600 active users/km^2 can constitute an ultra-dense network. With such a dense network of small cells, energy consumption is expected to surpass that of macro cell-only deployment scenarios in LTE.

Additionally, although 5G utilizes advanced technologies, it still relies on the same UE-BS association strategy, which connects the UE to the strongest cell [15]. Macro cells transmit at much higher power than small cells, so their signals typically provide higher received signal strength at the UE. Under strongest-cell association, this biases attachment toward macro cells, causing more UEs to connect there even when small cells are nearby. This leads to UEs prioritizing macro cells over small cells for association, thereby resulting in inefficient use of new technologies. Despite extensive research on UE-BS association in previous networks, significant gaps remain in 5G, particularly in integrating sleep mode mechanisms with realistic user throughput demands and mobility patterns into a unified and energy-efficient association framework.

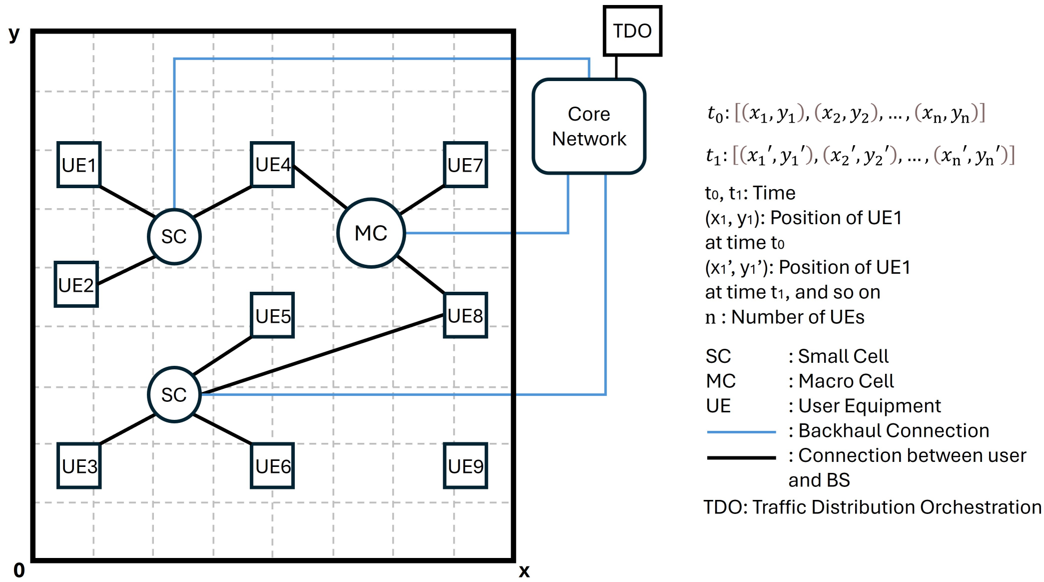


Figure 1.2: 5G Heterogeneous Macro Cell and Small Cell Access Network

Figure 1.2 illustrates the envisioned heterogeneous network architecture, highlighting the interactions between small cells, macro cells, user equipments, and the Traffic Distribution Orchestrator (TDO). There is a need for an intelligent coordination mechanism that can dynamically evaluate the energy efficiency (EE) of activating or deactivating small cells based on real-time network conditions. This thesis addresses this gap by designing and evaluating a Traffic Distribution Orchestrator (TDO) that adaptively assigns UEs to base stations and determines small cell sleep states by jointly considering power consumption, user mobility, and per-user traffic requirements. The proposed solution aims to minimize energy per bit and enhance energy efficiency while ensuring consistent Quality of Service (QoS) across heterogeneous 5G deployments.

1.2. Research domain and relevance

This research contributes to the research domain of energy-efficient optimization in 5G heterogeneous RAN, with a specific emphasis on UE-BS association strategies. With the deployment of dense small cell networks and the rapid growth in data traffic, minimizing power consumption while maintaining the network performance has become a critical challenge in 5G and beyond networks.

The relevance of this study lies in its contribution to green communication strategies aimed at reducing RAN power consumption without compromising Quality of Service (QoS). By leveraging context-aware, UE-driven association mechanisms, this study proposes a dynamic and scalable solution aligned with the energy efficiency objectives of modern 5G and future 6G networks. This research supports ongoing efforts in sustainable network design and contributes to advancements in adaptive RAN management techniques.

1.3. Research objective

The objective of this research is to develop a novel approach to minimize power consumption in 5G RANs by implementing a small cell selection strategy. This strategy aims to optimize small cell usage based on traffic demands, user mobility, and network load, ensuring energy savings while maintaining the required service quality.

The thesis also investigates how the deployment of small cells in conjunction with macro cells enhances the overall energy efficiency of 5G networks. Specifically, it examines the trade-offs and benefits of heterogeneous network architectures compared to macro cell-only deployments, highlighting how small cells can reduce transmission distances, offload traffic from macro cells, and enable more targeted resource utilization. By analyzing these configurations under varying load and traffic conditions, the study aims to quantify the energy savings achievable through the integration of small cells in 5G RAN.

1.4. Research questions

In order to achieve the research objective as described in section 1.3, a set of research questions and sub-questions is formulated as follows:

1. What are the optimal deployment strategies for small cells to maximize energy efficiency without compromising network performance in 5G networks?
    1.1 What are the limitations of the current power consumption models and how can they be improved to better reflect real-world network conditions?
    1.2 What are the key parameters that impact the power consumption of small cells compared to macro cells in a 5G heterogeneous network?
    1.3 How do different (RAN-related) topologies affect the power consumption in the network?
    1.4 What trade-offs arise between minimizing power consumption and ensuring seamless connectivity in RAN with time-varying traffic and user mobility?
    1.5 What are the current energy efficiency maximization techniques developed and what are their limitations?
2. What simulation scenarios are most effective in evaluating the proposed algorithms performance?
    2.1 Which existing algorithms should the proposed algorithm be compared with to evaluate its performance in terms of energy efficiency, user association, and Quality of Service (QoS) in heterogeneous 5G networks?
    2.2 What key performance metrics should be used to assess the algorithms effectiveness across different scenarios?
    2.3 What urban, sub-urban, and rural deployment scenarios should be simulated to capture a diverse range of user densities, mobility patterns (stationary, pedestrian,

vehicular), and traffic loads?

3. What approach enables the development of a dynamic user association algorithm to optimize energy efficiency and maintain Quality of Service (QoS) in heterogeneous 5G networks with macro cells and small cells?

   3.1 What criteria should be used to determine the optimal distribution of high-volume and low-volume data traffic between macro cells and small cells to maximize energy efficiency?

   3.2 How can the algorithm optimize the selection of small cells to minimize energy consumption per bit for various traffic loads and service requirements?

   3.3 How can the trade-off between minimizing energy consumption and maintaining user experience be effectively managed in the algorithm's decision-making process?

   3.4 How can the algorithm dynamically adapt to changes in UE movement and service requirements, ensuring continuous QoS and satisfactory energy efficiency?

1.5. Methodology

Figure 1.3 illustrates the overall research methodology adopted in this thesis. It outlines a structured approach comprising three main phases: theoretical analysis, algorithm development, and performance evaluation. The first phase involves an extensive literature review and the formulation of energy consumption models tailored for heterogeneous networks. The second phase focuses on the design and implementation of a small cell selection algorithm aimed at optimizing power consumption. Finally, the third phase evaluates the proposed algorithm under various simulation scenarios to analyze its effectiveness in improving energy efficiency while maintaining network performance. This methodology provides a comprehensive framework for systematically addressing the research objective.

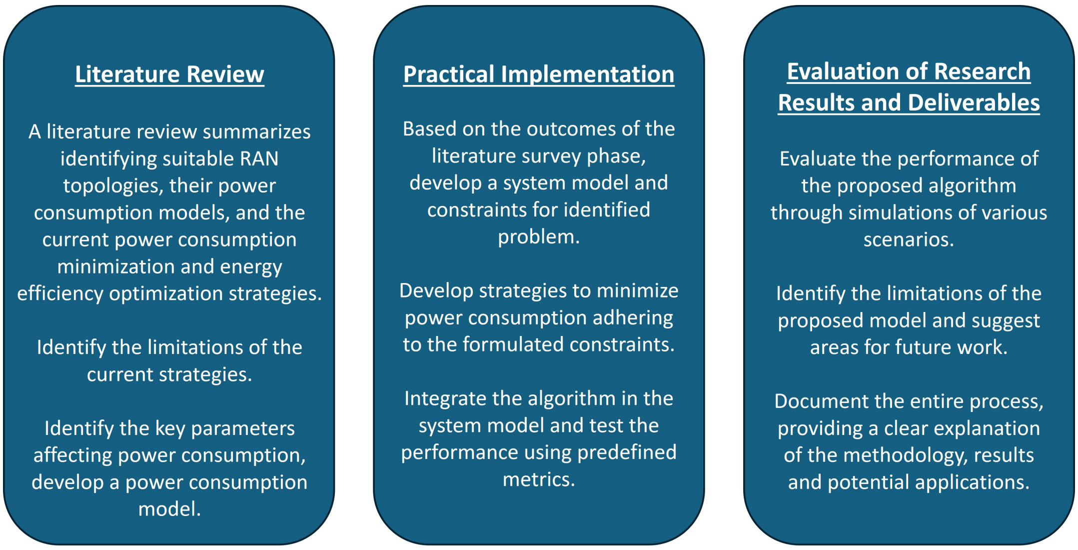


Figure 1.3: Thesis methodology

1.6. Structure

This thesis is structured as follows:

Chapter 2 reviews power consumption models for base stations and the backhaul, presents state-of-the-art techniques for reducing power use or improving energy efficiency in 5G RAN,

and identifies the key gaps this thesis addresses. The chapter answers the research question 1.
Chapter 3 then specifies the chosen RAN topology and full system model, introduces the proposed sleep-mode strategy and UE-BS association algorithm, and describes the simulation framework used to integrate them. This chapter answers the research question 3.
Chapter 4 defines the evaluation metrics and baselines, details the simulation scenarios, and reports results on power consumption, energy efficiency, and energy per bit. The chapter also answers the research question 2.
Chapter 5 concludes by summarizing the main contributions, findings, and implications, discussing limitations, and outlining directions for future work.

2

Literature Review

This chapter provides an overview of the research performed in the field of 5G networks regarding energy efficiency enhancement and energy savings. The chapter begins with the current issues in the 5G networks related to power consumption, the power consumption models and then delves into the current solutions that are developed to tackle the issues. These solutions are categorized into sleep mode strategies, UE-BS association algorithms and load management. Finally, the limitations of current methods are summarized, highlighting gaps that motivate the next chapter.

2.1. Overview of power consumption models

The RAN contributes to approximately 75% of the total energy consumption. The NGMN Network Energy Efficiency Phase 2 report [16] breaks down this RAN power consumption into more detailed network components. Out of this 75%, the Base Station (BS) equipment comprising Radio Units (RUs), BaseBand Units (BBUs), and main control, consumes about 50% of the power. Moreover, almost 80 to 90% of the power in the BS is consumed by the RU. Therefore, it is only reasonable to quantify the energy consumption of the AN accurately in order to develop strategies to improve the energy efficiency of the network. However, these models vary based on the network architecture, underlying technologies and the hardware. Hence, it is essential to analyze the current models and develop an energy consumption model tailored for a specific network architecture type and technologies. This includes the power consumption of Macro Cell Base Stations (MCBSs), Small Cell Base Stations (SCBSs) as well as that of the supporting infrastructure such as the x-haul networks. The system structure of a cellular network as given by Busch et al. [17] is shown in Figure 2.1.

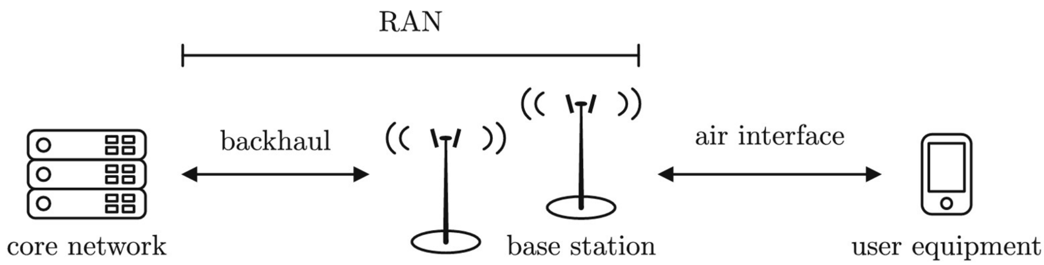


Figure 2.1: Cellular network system structure

2.1.1. Power consumption modeling of base station

Energy consumption in 5G networks is a critical consideration, given the rising demand for data, the proliferation of devices, and the environmental impact of increasing energy use. Various models are developed to study, predict, and optimize energy consumption in 5G networks. These models consider the energy consumed by network components, the influence of operational strategies, and the trade-offs between energy efficiency and performance.

The paper "Comparison of Power Consumption Models for 5G Cellular Network Base Stations" by Busch et al. [17] provides a detailed comparative analysis of existing power consumption models used for 5G cellular networks. The models are broadly categorized into three categories: analytical, empirical, and machine learning-based models. Analytical models employ mathematical equations to estimate power consumption, empirical models rely on real-world measurements, while machine learning-based models use artificial intelligence techniques to predict BS power consumption dynamically. Among these, analytical models are particularly relevant for this research as their theoretical foundations make them well suited for simulations.

Although the study by Busch et al. [17] provides valuable insight into the comparative performance and accuracy of different modeling approaches, it solely focuses on the quantitative evaluation of these models under varying assumptions such as BS types and parameter settings. It does not explicitly present the mathematical formulations of these models. Nevertheless, across the surveyed models, the main components contributing to power consumption are consistently categorized in the same way. Figure 2.2, adopted from [18], shows a traditional non-disaggregated BSs, where radio and baseband processing are co-located.

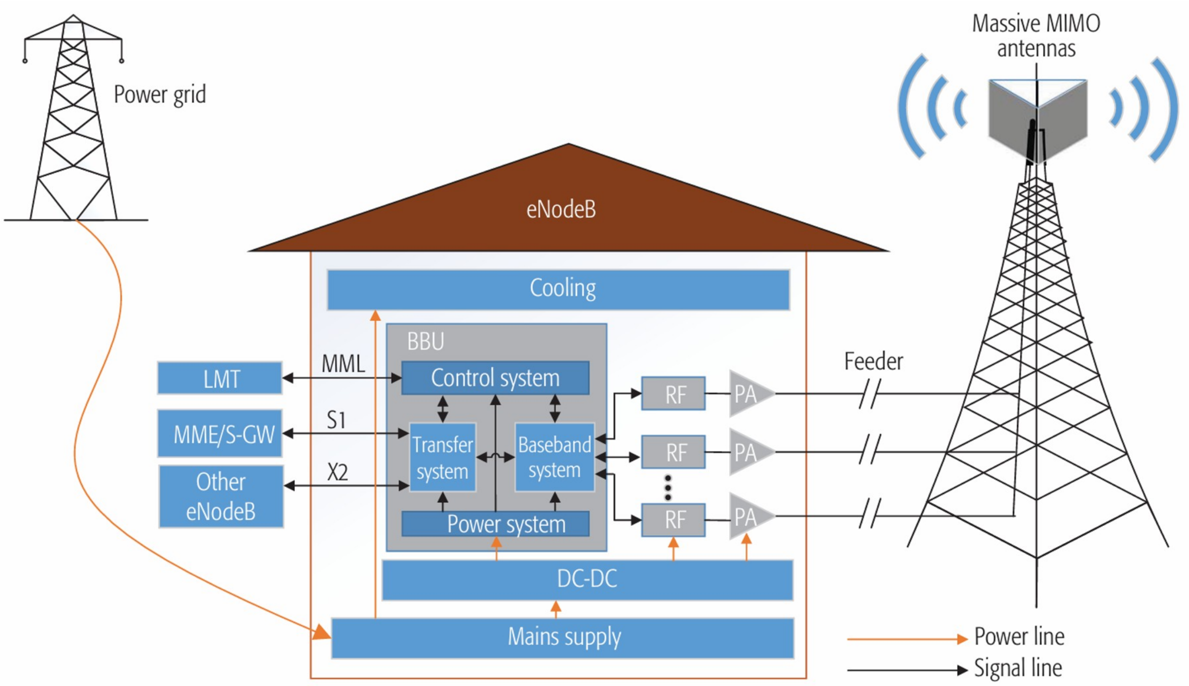


Figure 2.2: Architecture of a traditional base station

Thus, the primary contributors to power consumption include:

1. **Power Supply and Cooling**: Includes power consumed by cooling units and DC-DC converters.

2. **Baseband Processing**: Comprises the power consumed by digital signal processing functions such as channel coding and decoding, equalization, Fast Fourier Transform (FFT) and Inverse FFT (IFFT), and linear processing.
3. **Analog Frontend**: Comprises the RF (Radio Frequency) transceivers and power consumed by circuit components such as converters, mixers, filters attached to each antenna, and
4. **Transmit Power**: Account for most of the power consumption on the transmit side.

Based on these components, the power consumption of i^{th} base station is often modeled using formulations such as the one presented by Ge et al. [18] is:

$$P_{BS_i} = \frac{P_{PA} \cdot N_{TRX} + P_{RF} \cdot N_{TRX} + P_{BB}}{(1-\sigma_{DC})(1-\sigma_{MS})(1-\sigma_{cool})} \tag{2.1}$$

where P_{PA} is the power consumed by the Power Amplifier (PA) per antenna,
P_{RF} is the transceiver power per antenna,
P_{BB} is the power consumed by the BBU,
N_{TRX} is the number of antennas on the transmitter side, and
σ_{DC}, σ_{MS} and σ_{cool} are the power loss rates of the DC-DC converter, the alternating current supply and the cooling, respectively.

Building upon Equation 2.1 presented in [18], Daehan et al. [19] extend the framework to incorporate the impact of massive MIMO technology, with a particular emphasis on circuit power consumption associated with the mMIMO systems which is shown in Figure 2.3.

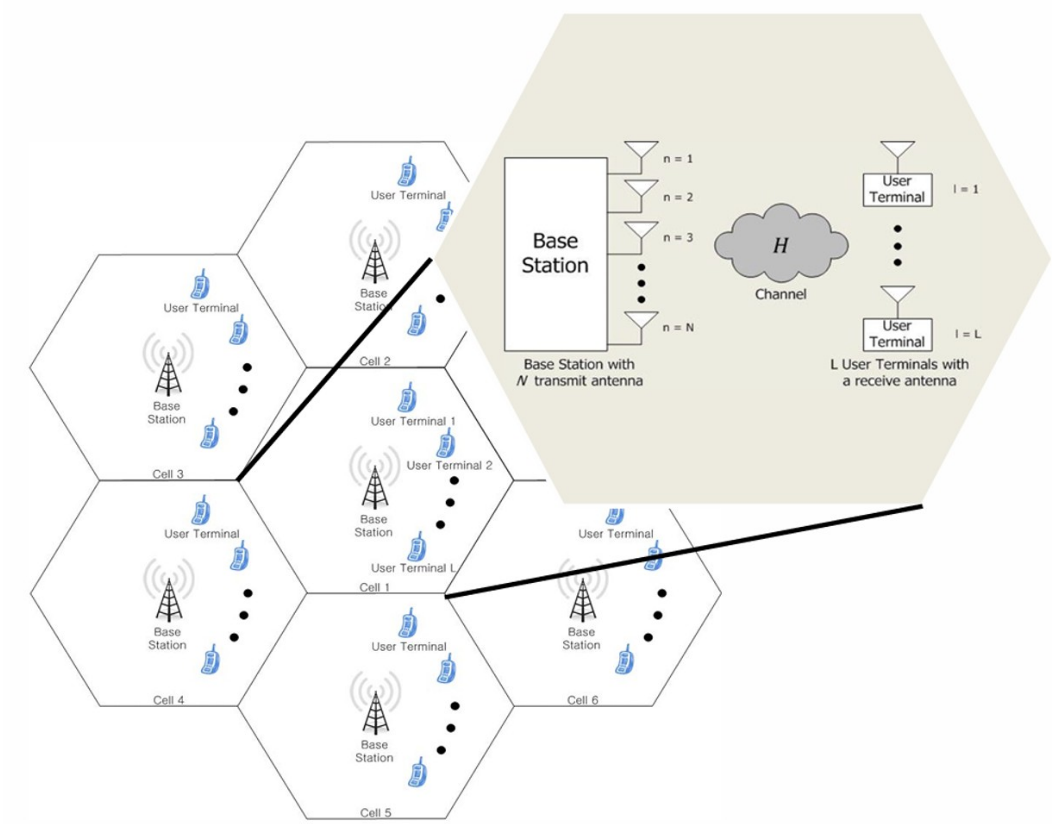


Figure 2.3: A multi-cell architecture with massive Multiple Input Multiple Output (mMIMO) systems at base stations

Accordingly, the power consumption for the i^{th} BS is expressed as:

$$P_{BS_i} = \frac{\frac{P_t}{\eta(1-\sigma_{feed})} + P_{cir} + P_{load-independent_i}}{(1-\sigma_{DC})(1-\sigma_{MS})(1-\sigma_{cool})} \tag{2.2}$$

where P_t is the transmit power,
η is the PA efficiency,
P_{cir} is the circuit power consumption,
σ_{feed} is the lossy factor of the antenna feeder,
$P_{load-independent_i}$ is the baseline power consumption of the i^{th} BS while the rest of the terms are the same as explained for Equation 2.1. In this model, the circuit power, which refers to the power consumed to operate the circuit at the transmitter end is expressed as, $P_{cir} = N(P_{DAC} + P_{mix} + P_{filt}) + P_{syn}$. The circuit power comprises P_{DAC}, P_{mix}, P_{filt} and P_{syn} that represent the power consumed by the DAC, mixer, filter, and frequency synthesizer, respectively. Thus, the circuit power scales linearly with the number of antennas N, with the exception of components such as the frequency synthesizer, whose power consumption remains constant. The model proposed by Daehan et al. [19] includes only the Digital-to-Analog Converter (DAC), mixer, filter, and frequency synthesizer as part of the circuit power, thereby treating it as the power consumed by the RF or transceiver chains. However, this simplification omits the significant power consumed by complex baseband processing circuitry, which involves both analog and digital components.

To address this limitation, Bjornson et al. [20] propose a more comprehensive circuit power model for multi-user MIMO systems. Their model accounts for the power consumed not only by the transceiver chains but also by key processing tasks such as channel estimation, channel coding/ decoding, linear precoding, and backhaul communication. Importantly, these contributions scale with parameters such as the number of antennas, number of active UEs, and data rates, often linearly or non-linearly, depending on complexity of the operations involved. The paper provides a detailed mathematical framework that decomposes each of these components and quantifies their individual power consumption. Numerous such complex power consumption models exist, varying based on underlying technologies such as massive MIMO and carrier aggregation, or on the level of detail in sub-component modeling. Many of these models are derived from frameworks like the Energy Aware Radio and neTwork tecHnologies (EARTH) project or the 3GPP reference model [17].

Although these models offer deep insights into base station power consumption, incorporating this level of granularity is beyond the scope of this thesis, since the primary objective is to develop strategies to reduce power consumption rather than to perform an exhaustive analysis of each power-consuming component. Consequently, many studies on the enhancement of energy efficiency adopt simplified power consumption models to facilitate analysis and reduce computational complexity. While these models may differ slightly in how they represent base station power consumption, they generally follow a common structure: modeling the total power as the sum of load-independent and load-dependent components. Auer et al. [21] investigated the interdependence of power consumption among components in a BS transceiver, using empirical data from various BS types, including macro, micro, pico and femto BSs. Their study concludes that transmit power is almost linearly related to overall BS power consumption. The power model proposed by Auer et al. [21] simplifies the expression given in Equation 2.2 as follows:

$$P_{BS_i} = \begin{cases} N_{TRX} \cdot P_{load-independent_i} + P_{load-dependent_i}, & 0 < P_{load-dependent} \leq P_{max} \\ N_{TRX} \cdot P_{sleep_i}, & P_{load-dependent} = 0 \end{cases} \tag{2.3}$$

where $P_{load-dependent} = \Delta_p P_{out}$, $0 < P_{out} < P_{max}$
Δ_p is the multiplier for load-dependent power consumption,
P_{out} is load-dependent transmit power,
P_{max} is the maximum transmit power available at the transceiver,
P_{sleep} is the BS power consumption in sleep mode while the rest of the terms are the same as explained for Equation 2.1 and Equation 2.2. In the literature, load-dependent power consumption per base station i is modeled in various ways, depending on the system assumptions and level of abstraction, as follows:

1. Ryu et al. [22] formulate it as:

$$P_{load-dependent_i} = \Delta_p \cdot P_{max} \cdot \frac{1}{N_{RB}} \sum_{u \in U} \left\lceil \frac{f_u x_{iu}}{N_{a_u} \cdot B_{RB} \cdot log_2(1 + SINR_{iu})} \right\rceil \tag{2.4}$$

   where N_{RB} is the maximum number of Physical Resource Blocks (PRB) at the i^{th} BS,
   f_u is the data rate for each UE 'u' belonging to the set of UEs which is denoted by U,
   x_{iu} is an integer value {0,1} representing the association between the u^{th} user and the i^{th} BS,
   N_{a_u} is the number of antennas participating in MIMO transmission between the u^{th} UE and the i^{th} BS link if the user u is associated to the i^{th} BS, and
   $SINR_{iu}$ is the Signal-to-Interference-plus-Noise Ratio (SINR) for the link between the u^{th} UE and i^{th} BS.
   In LTE (4G) and 5G systems, a PRB is the smallest unit of resources that can be allocated to a UE in the frequency-time domain. Because a UEs rate scales with the PRBs it receives, PRB allocation serves as a practical measure of cell load for power consumption models.

2. Fourati et al. [23] formulate it as:

$$P_{load-dependent_i} = \rho \cdot P_{max} \cdot \sum_{u \in U} x_{iu} \times \frac{C_u}{C_{max}} \tag{2.5}$$

   where ρ is the slope of load-dependent power consumption,
   C_u represents the capacity (in Mbps) needed to serve UE 'u',
   C_{max} is the maximum capacity of BS in Mbps while the remaining notations are the same as those explained for Equation 2.3 and Equation 2.4.

3. Whereas, Kooshki et al. [1] formulate it as:

$$P_{load-dependent_i} = P_{max} \cdot \frac{1}{\eta} \cdot \sum_{u \in U} \frac{N_{iu}}{N_{RB}} \tag{2.6}$$

   where η is the efficiency of the power amplifier,
   N_{iu} is the number of PRBs of BS i used by the UE u, and the remaining terms are the same as those explained for Equation 2.3, Equation 2.4 and Equation 2.5.

Although load-dependent power consumption models may differ in their details, they are fundamentally based on the same principle: evaluating how many resources are utilized by each base station. Equation 2.4 and Equation 2.6 represent these resources in terms of the number of PRBs allocated to the UEs whereas Equation 2.5 uses capacity expressed as the total data rate provided to the UEs. However, the data rate offered to each UE depends on the SINR of the link between the UE and the BS, which varies dynamically and cannot be predetermined as a fixed capacity value. Therefore, a BSs capacity is more accurately represented in terms of

PRBs rather than total data rate.

Moreover, although some models express load-dependent power using PA efficiency η as given in Equation 2.6, others consolidate multiple overheads into a multiplier Δ_p on output power as given in Equation 2.4. Both approaches are valid and commonly published; they represent the same underlying physical relationship but differ in parameter definitions and modeling granularity.

The model described in Equation 2.3 is based on the power consumption values for LTE BSs. While 3GPP NR Release 16 defines a power consumption model for 5G UEs, it does not provide a corresponding model for 5G networks. To address this gap, Piovesan et al. [24] proposed a realistic power consumption model based on data obtained from a 5G multi-carrier mMIMO Active Antenna Unit (AAU). In typical 5G network deployments, the mMIMO BSs are divided into three components: the CU, the DU and the RU, which is also called as the AAU. The specific functions performed by each unit depend on the chosen functional split, which defines the deployment architecture. Functional splits are standardized ways of partitioning the RAN protocol stack across RU, DU, and CU, defining which functions run where and the fronthaul/midhaul interface requirements, trading off centralization versus latency/-transport capacity.

The AAUs considered by Piovesan et al. [24] include the baseband processing, the RF chains, the PAs and the antennas. The proposed model is based on power measurements collected over a 12-day period from thousands of AAUs spanning 25 different types, in a real-world deployment. It captures the impact of advanced technologies such as carrier aggregation and mMIMO, as well as energy-saving mechanisms including carrier shutdown, channel shutdown, symbol shutdown and sleep modes. The power consumed by an active AAU is defined as:

$$P_{\mathrm{AAU}} = P_0 + P_{\mathrm{BB}} + \underbrace{\sum_{t=1}^{T} M_{\mathrm{av},t} D_{\mathrm{Tran},t}}_{P_{\mathrm{Tran}}} + \underbrace{M_{\mathrm{ac}} D_{\mathrm{PA}}}_{P_{\mathrm{PA}}} + \underbrace{\frac{1}{\eta} \sum_{c=1}^{C} P_{\mathrm{TX},c}}_{P_{\mathrm{out}}} \tag{2.7}$$

where,
P_0 is the baseline power consumption which is the power required to keep essential BS functions active,
P_{BB} is the power consumed by the baseband processing,
P_{Tran} is the power consumed by the transceivers,
P_{PA} is the power consumed by the power amplifiers,
P_{out} is the power used to generate the transmit power, which scales linearly with the number of PRBs used,
M_{ac} is the number of active RF chains,
D_{PA} static power consumed by PA per active RF chain,
$P_{TX,c}$ represents the transmit power for carrier c,
η is the efficiency of PA, and
C is the number of carriers

Importantly, while this model introduces enhancements tailored to 5G architectures, it remains consistent with the fundamental structure of earlier models by distinguishing between load-independent and load-dependent power components. This alignment is supported by empirical evidence from real-world data, validating that the power consumption still scales predictably with load, as observed in earlier LTE-based models.

These models represent the power consumption of a BS when it is in the active state. However, in practice, BSs often transition into low-power or sleep modes during periods of low traffic to reduce energy consumption. Therefore, it is essential to account not only for the power consumed in the active state, but also for the power consumed in sleep mode, as well as the energy required during transitions between active and sleep states. Furthermore, while some studies consider only two operational states-active (ON) and sleep. Other studies introduce an additional OFF state, which represents a deeper low-power mode with even lower energy consumption but potentially longer reactivation times. The choice of states included in the power model depends on the desired balance between energy savings and system responsiveness. Thus, the power consumption of switching between the on, off and sleep states at any time t given by Fourati et al. [23] is:

$$P_{switching} = s_{off} \times (1 - s_{sleep}) \times P_{on-off} + s_{sleep} \times (1 - s_{off}) \times P_{on-sleep} \tag{2.8}$$

where,
s_{off} and $s_{sleep} \in 0, 1$ are binary variables indicating whether the BS transitions between on and off, or between on and sleep modes respectively, and
P_{on-off} and $P_{on-sleep}$ are the power consumption values for transition between on and off, or on and sleep modes respectively.

Depending on the specific modeling requirements, the relevant power consumption terms can be included, while those deemed insignificant or irrelevant to the analysis may be excluded. Thus, the total power consumption can be formulated as $P_{total} = P_{BS} + P_{switching}$. Adding more detail, if the network includes M macro cells and S small cells, the power consumption model can be given as $P_{total} = \sum_{m \in M} P(m)_{macrocell} + \sum_{s \in S} P(s)_{smallcell} + P_{switching}$, considering that the macro cell is always active.

Power use in macro cells and small cells is driven by the same factors (such as the transmit power, number of active transceiver chains, operating bandwidth, the processing performed by the base station, load on the cell and site cooling), but with different magnitudes. This answers the sub-question 1.2 of the research question 1.

2.1.2. Power consumption modeling of x-haul network

Most research aiming to improve energy efficiency or reduce power consumption in 5G networks predominantly focus on the base station, often overlooking the energy footprint of the x-haul network. The x-haul network comprises the backhaul, midhaul, and fronthaul segments, depending on the chosen architectural framework. Notably, the selected functional split directly impacts not only the power consumption but also the data rates handled by each segment of the x-haul network. Figure 2.4 shows an x-haul network adopted from [2]. RU is the Radio Unit, DU is the Distributed Unit and CU is the Centralized Unit.

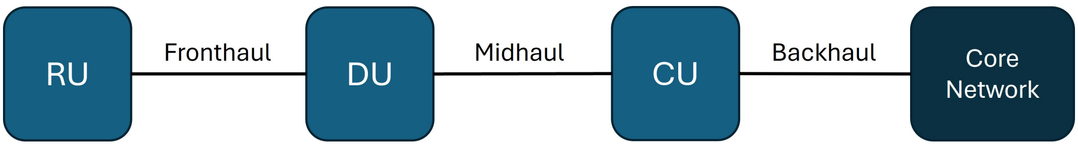


Figure 2.4: X-haul network

Furthermore, x-haul energy consumption is influenced by both the underlying transmission technology and the deployment topology. Thus, there are three main factors influencing the power consumption of x-haul networks:

1. **Underlying Technology:** The x-haul network can be implemented using wired, wireless, or hybrid transmission technologies. Wired solutions include fiber optics, Passive Optical Networks (PONs), and VDSL2, whereas wireless alternatives utilize microwave, satellite, Television White Space (TVWS), or millimeter Wave (mmWave) links [25]. In practice, it is often infeasible to connect all small cells via optical fiber due to high CAPital EXpenditure (CAPEX) and deployment constraints, particularly for small cells located in remote or inaccessible areas. Moreover, existing microwave and copper/fiber technologies may not be sufficient to handle the growing traffic demands of 5G. Notably, PON-based backhauling is shown to be more energy-efficient at high traffic loads, whereas mmWave solutions are preferable under low-load conditions. Therefore, hybrid backhaul architectures are considered a practical solution. For example, Munjure et al. [26] propose a hybrid model where small cells are connected to an aggregate small cell using either PON or mmWave based on traffic demand, while the macro cell connects to the core network via PON.
2. **Deployment Topology:** The aforementioned technologies can be deployed using either Point-to-Point (PtP) or Point-to-multiPoint (PtmP) topologies. PtP topologies can take the form of chain, ring, mesh, or tree structures. However, depending on the topology, additional delays and energy overheads may arise due to increased hop counts and congestion at aggregation points. In contrast, PtmP topologies may offer reduced complexity and improved energy efficiency in certain scenarios but require careful consideration of traffic patterns and latency requirements [25].
3. **Functional Splits:** Functional splits significantly impact the energy consumption of x-haul networks because they determine the volume and location of processing tasks, thereby affecting the required data rates across the fronthaul, midhaul, and backhaul links. While several functional split options exist, industry practice commonly adopts split options 2, 6, and 7 due to their practical trade-offs between performance, latency, and deployment complexity [5].

From a broader perspective, the power consumption modeling of x-haul networks can be categorized into wired and wireless. While both types support multiple underlying technologies, as previously discussed, fiber optics and mmWave backhaul have emerged as the most promising solutions for 5G Heterogeneous Network (HetNet) deployment. Fiber optics is generally preferred due to its high capacity, low latency, and reliable performance. However, advancements in mmWave technology have demonstrated its potential as a high-capacity wireless backhaul solution.

Power consumption of Passive Optical Network (PON)-based backhaul

Among fiber-based technologies, Passive Optical Network (PON) is widely adopted due to its support for PtmP communication and the use of passive splitters, which significantly reduce operational costs compared to active optical systems. Common PON standards include GPON, NG-PON, and EPON. A typical PON architecture comprises Optical Network Units (ONUs) located at BSs and an Optical Line Terminal (OLT) situated at the service providers central office, which interfaces with the core network. The power consumption of a PON-based backhaul network can thus be modeled as [26]:

$$P_{BH}^{PON} = N_{BS}P_{ONU} + N_{GPON}^{port}P_{GPON}^{port} + N_{UL/DL}P_{SFP+} \tag{2.9}$$

where,
N_{BS} is the number of BSs that can include small cells as well as macro cells,

P_{ONU} is the power consumed by the ONUs
N_{GPON}^{port} is the number of GPON ports in the OLT and P_{GPON}^{port} is its corresponding power consumption,
$N_{UL/DL}$ is the number of uplink or downlink interfaces, and
P_{SFP+} is the power consumed by the Small Form-factor Plug-in (SFP) interfaces, which are responsible for converting electrical signals into optical signals and vice versa.

The number of GPON ports in an OLT is computed as:

$$N_{GPON}^{port} = \left\lceil \frac{N_{BS}}{N_{ONU}^{max}} \right\rceil \tag{2.10}$$

where N_{ONU}^{max} is given as $N_{ONU}^{max} = max\left(\frac{C_{GPON}^{port}}{D_{ONU}^{max}}, N_{splitter-ratio}\right)$.
C_{GPON}^{port} denotes the GPON capacity per port in bps, and
D_{ONU}^{max} is the maximum data rate for each ONU.
$N_{splitter-ratio}$ represents te splitter ratio which can be 1:8, 1:16, or 1:32.
The number of uplink or downlink interfaces to transmit the aggregated traffic to the core network is given by:

$$N_{UL/DL} = max\left(N_{GPON}^{port}, \frac{C_t}{D_{int}^{max}}\right) \tag{2.11}$$

where, C_t is the total aggregated traffic requirements to be backhauled, and
D_{int}^{max} is the maximum transmission rate of an interface.

Power consumption of mmWave-Based backhaul
Nevertheless, as discussed earlier, it is not feasible to connect all RUs using fiber optics alone, particularly in dense or hard-to-reach urban areas. In such scenarios, mmWave technology presents a feasible alternative, especially for fronthaul links. The mmWave backhaul power model follows a structure comparable to that of the access link, comprising both static (load-independent) and dynamic (load-dependent) power components. The backhaul power consumption is modeled as follows:

1. The power consumption for the l^{th} backhaul link using the linear approximation model given by Kuna et al. [27] is:

   $$P_l^{BH} = T_l^{BH}(S_l^{BH} P_{0l}^{BH} + \Delta_{Pl}^{BH} P_{dl}^{BH}) \tag{2.12}$$

   where, T_l^{BH} is the number of active transceiver chains,
   P_{0l}^{BH} is the non-zero power consumption depending on the state S_l^{BH} which can be sleep or active,
   Δ_{Pl}^{BH} is the slope-dependent power utilization of the l^{th} link, and
   P_{dl}^{BH} is the dynamic power consumption which is given by $P+dl^{BH} = SINR_l^{min} - G_{rx_l} + L_{rx_l} + PL_l + LM_l + T_{n_l} + N_{f_l}$.
   Here, G_{rx_l} and L_{rx_l} represent the receiver antenna gain and losses of the link l.
   $SINR_l^{min}$ represents the minimum SINR for successful transmission of backhaul traffic,
   PL_l is the Path Loss (PL),
   LM_l is the link margin,
   T_{n_l} is the thermal noise, and
   N_{f_l} is the receiver noise figure for link l.
   The SINR is given by $SINR_l^{min} = 10log_{10}\left(2^{\frac{R_l^{BS}}{BW_l}} - 1\right)$, where R_l^{BS} is the data traffic to be backhauled from base station BS and BW_l is the bandwidth allocated for link l.

2. In a simpler way, Hyebin et al. [28] model it as:

$$P_l^{BH} = P_l^{fixed} + l \cdot t_s \tag{2.13}$$

where P_l^{fixed} represents the circuit power of the transceiver and the switch,
l is power consumption per bps,
t_s is the traffic to be transmitted per second.

However, despite supporting high data rates, the mmWave backhaul faces certain limitations. It requires Line-of-Sight (LOS) communication, which can be challenging to maintain in dense urban environments due to obstruction such as buildings and foliage. Furthermore, the use of very high frequency bands limits its effective transmission range, making the radio range significantly shorter than that of fiber optic backhaul.

2.2. Strategies for energy consumption minimization in Radio Access Network

The energy consumption minimization strategies in 5G RAN can be broadly classified into sleep mode strategies and UE-BS association algorithms, which will be explained in the following subsections. Limitations of each study are highlighted in gray for clarity. The following subsections answer the sub-question 1.4 and 1.5 of the research question 1.

2.2.1. Sleep mode strategies

Sleep mode strategies play a pivotal role in reducing energy consumption in 5G RAN by dynamically deactivating underutilized BS hardware components during periods of low traffic. These strategies vary from light to advanced deep sleep modes, each offering various trade-offs between energy savings and service latency. The Advanced Sleep Modes (ASMs) were initially introduced by IMEC [29] within the EARTH project [30], in which four ASMs were identified by categorizing the BS components into groups with similar activation/deactivation times [31]. Thus, the ASM management strategizes shutting down of BS components gradually from the ones with lower activation times to the ones with higher activation times. These ASMs were later standardized by 3GPP in TR38.864 Release 18 [32] document which focuses on standardizing energy-saving mechanisms, especially sleep modes to enhance energy efficiency in 5G networks. Sleep modes are classified based on the depth of power savings and the wake-up latency. Thus, deeper the sleep mode, higher the components deactivated, energy savings as well as the transition times. Therefore, it is important to strategize the use of these sleep modes accurately so as to provide users with the required QoS.

This subsection explores the evolution of sleep mode strategies, their implementation challenges, and the contribution of key studies in advancing this critical area of research. The sleep mode strategies are further classified into 'non-adaptive', 'adaptive' and 'ML-based' strategies.

Non-adaptive sleep mode strategies

Non-adaptive sleep mode strategies use fixed, context-independent rules to control base station states. In sleep control, they rely on static on/off thresholds and fixed timers. Although these methods are straightforward and introduce low overheads, they perform poorly in dynamic environments where traffic demands or user mobility fluctuate, often causing unnecessary toggling, energy waste and QoS degradation. They serve mainly as baselines for measuring the benefit of adaptive mechanisms.

The study done by Kooshki et al. [1] proposes an EE enhancement algorithm for cell-free architectures that aims to maximize throughput per unit energy while maintaining service quality. UEs are first associated to the base station with the highest Reference Signal Received Power (RSRP). The sleep control mechanism then proceeds in two phases, namely the "Certain Phase" and the "Conditional Phase". In the Certain Phase, each base station's Interference Contribution Ratio (ICR) is evaluated. The concept of the ICR first introduced by Shen et al. in [33] is computed as the ratio of the aggregate RSRP observed by non-served UEs (interference) to the aggregate RSRP of the base station's served UEs (desired signal). Thus, higher the ICR of a RU, lesser is the significance of the signal provided by that base station towards the network. The base stations with high ICRs are put to sleep. In the Conditional Phase, additional checks based on network-level EE and base station load opportunistically switch more base stations to sleep while meeting the global performance targets. **Although the base station states are updated over time, the key thresholds (ICR, load, and minimum rate limits) are preset, making the approach suboptimal under rapidly varying environment.**

Building on Kooshki et al. [1], Osama et al. [34] combine a static Soft Frequency Reuse (SFR) plan with a threshold-based on/off policy driven by the ICR to jointly mitigate interference and reduce power in dense HetNets. Small cells are partitioned into center or edge zones with fixed sub-band allocations, while cells whose ICR exceeds a preset limit are switched off. The resulting partitioning and SFR pattern as illustrated in [34] is shown in Figure 2.5.

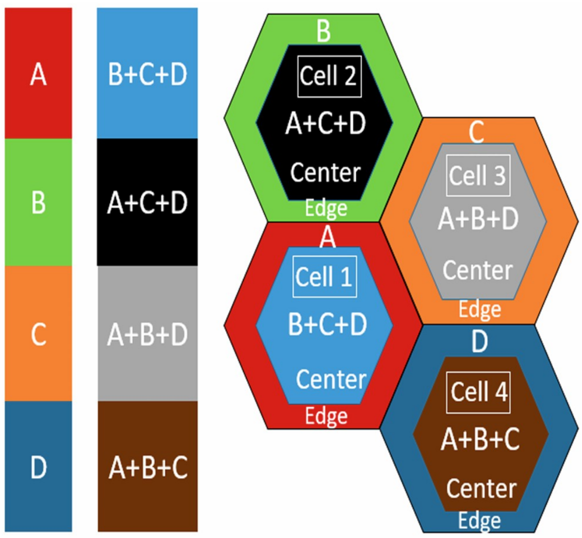


Figure 2.5: Cell zones and Soft Frequency Reuse

As shown in Figure 2.5, the frequency bands A-D are assigned across neighboring cells so that each cells edge zone uses a different band, reducing inter-cell interference. **Both the SFR partitioning, and the ICR and minimum throughput thresholds are preset, making the scheme non-adaptive.**

Similarly, Sudhakar et al. [35] propose a threshold-based scheme that monitors the number of UEs per base station in a 3x3 cluster and switches off any base station whose load falls below a preset limit. The associated UEs are then handed over to neighboring base stations subject to two fixed capacity constraints: a maximum allocating limit per neighbor and a per-base station maximum serving limit. **This is a non-adaptive fixed threshold on/off policy where decisions depend on static limits rather than real-time traffic dynamics or QoS feedback. Furthermore,**

the analysis is limited to a toy 3x3 homogeneous grid without a macro cell overlay, making the scenario unrealistic and excluding macro-tier interference.

Unlike the previous non-adaptive schemes, the research done by Salem et al. [36] goes deeper by explicitly modeling timing parameters such as user request arrival instant, activation/ deactivation delays, sleep window length and signaling burst periodicity to quantify the energy latency trade-off. The study investigates ASMs by extending idle periods and coordinating periodic wake-ups for signaling. Base stations progressively enter deeper sleep levels as inactivity grows and periodically wake up to transmit signaling bursts. User requests arriving during inactivity are buffered to curb delay, with buffering time determined by the phase (activation, sleep or deactivation) and the chosen sleep level. **However, as the sleep depth and the wake-up periodicity are preset, this serves as a non-adaptive strategy.**

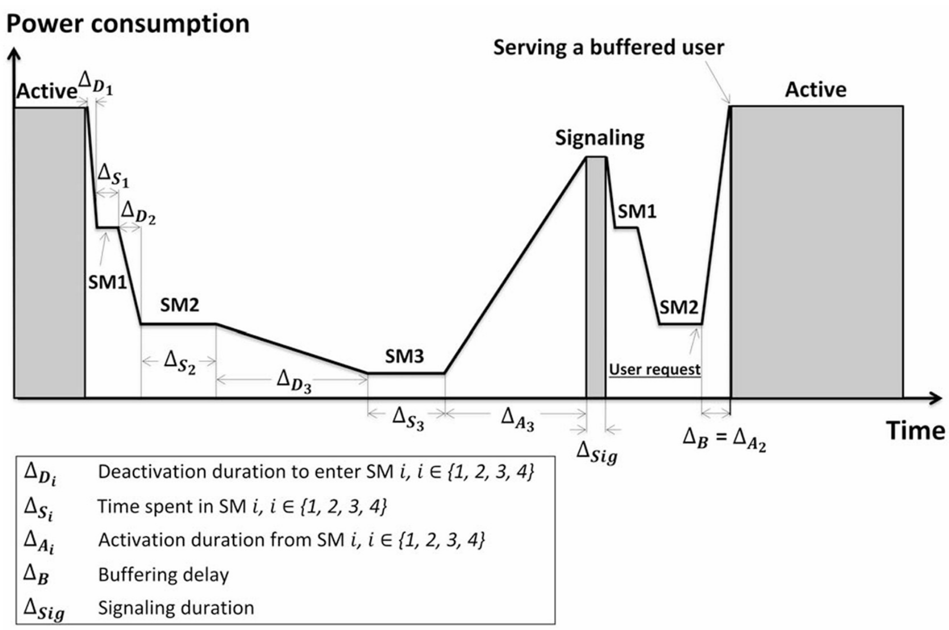


Figure 2.6: Advanced Sleep Mode implementation strategy

Fixed threshold approaches are simple with less computational complexity. However, setting low sleep thresholds could lead to the cells being active even for lower loads and thus consume unnecessary power. On the other hand, if the threshold is too high, the cells will be in sleep modes for long times and save more energy but at the cost of QoS of the users.

Adaptive sleep mode strategies

Prior studies on BS sleep strategies largely relied on predefined thresholds, which lack real-time adaptability. The research done by X. Ma et al. [37] introduces a variable threshold-based sleep mode scheme to incorporate dynamic changes in the environment. The scheme builds on a simple association step, which is to connect to the highest rate small cell and fall back to macro cell to satisfy QoS. Then a Variable Threshold Sleep Process (VTSP) is applied which monitors spatio-temporal UE distribution and adjusts each cell's sleep threshold as a fraction of its UE-capacity and consequently switches off the lightly loaded cells. Compared with fixed threshold algorithms, this approach cuts the number of active base stations and total power while maintaining the required QoS. **However, the paper considers load only in terms of the number of UEs attached to a cell, while excluding the PRBs or the traffic demands of the UEs. Moreover, it also does not consider the power consumed by the base stations while switching**

from one state to another.

While X. Ma. et al. [37] adapt per cell sleep thresholds to traffic dynamics, Habibi et al. [38] go further by combining adaptive sleep with transmit power zooming and a joint controller that targets minimum energy efficiency under QoS constraints. The paper [38] proposes three adaptive schemes for ultra-dense HetNets that maximize the minimum EE of SCBSs under QoS constraints on RSRP and Spectral Efficiency (SE). UEs first associate by highest RSRP. Then, an iterative sleeping policy turns of the base stations with lower efficiency if the network's minimum EE improves, while a zooming policy adjusts the transmit power to meet the SE targets and curb interference. Among the three policies studied, sleeping-only, zooming-only, and joint sleeping plus zooming - the joint scheme yields the lowest total power and highest energy efficiency, while zooming-only attains the highest spectral efficiency. This approach as illustrated in [38] is shown in Figure 2.7. Thus, cells that reduce coverage or enter sleep save power, while neighboring cells that extend coverage serve the offloaded users. **The goal maximizes EE of individual SCBSs which may not reflect the network wide EE.**

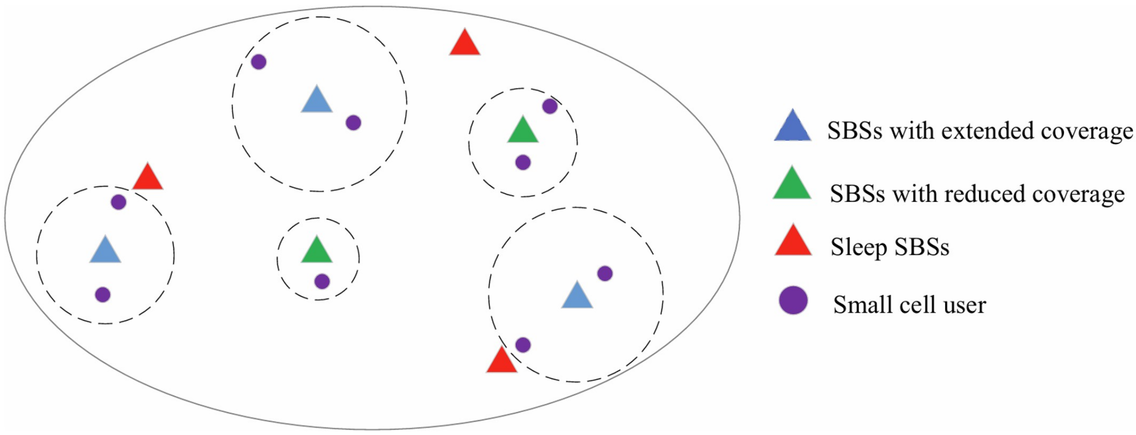


Figure 2.7: Cell zooming and sleep mode mechanism

The preceding methods adapt parameters but still depend on fixed design choices. Adaptive approaches can also increase the number of handovers and control overheads if not carefully tuned. In contrast, the following group of studies replaces these rules with data-driven policies. These are essentially Machine Learning (ML)-based models that learn when to sleep, wake, or perform actions to enhance EE from observed spatial and temporal user patterns and QoS feedback.

Machine Learning-based sleep mode strategies

To overcome the challenges of the non-ML-based approaches as mentioned above, recent research has increasingly turned to ML techniques which offer the ability to learn from past network behavior, predict future traffic loads and optimize energy-saving strategies dynamically. Unlike traditional approaches, ML models can process large datasets, recognize patterns, and autonomously adjust system parameters to maximize EE while maintaining network performance.

Among various ML-based approaches, Reinforcement Learning (RL) is gaining significant traction in optimizing EE in 5G networks. RL is a branch of ML in which an agent learns to make decisions by interacting with an environment to achieve a specific goal. The agent takes actions in various states and obtains feedback in the form of rewards, aiming to maximize cumulative rewards over time. This learning paradigm is distinct from supervised or unsupervised learning where the learning is achieved through a set of training data which can be labeled or unlabeled, respectively. Instead, RL focuses on optimal behaviors through trial and error. A foundational concept in RL is the Markov Decision Process (MDP) is a mathematical approach to represent an environment in decision-making scenarios, where outcomes are influenced both by randomness and choices made by a decision-maker. MDP consists of the following elements as explained by Wiering et al. [39] are:

1. **States:** A set of states is a complete description of the environment at a decision point, containing enough information to predict what will happen next under any action. Thus, in the context of this thesis, the states can be described as the set of values of load, status of the cells (on/off/sleep) traffic demands of the users, SINR values, mobility parameters, and power consumption values at any given point of time.
2. **Actions:** Actions are the choices available to the decision maker in a given state, and selecting one influences how the environment evolves. Thus, the set of actions could comprise keeping a cell on, turning it to sleep mode, waking it up from sleep mode or keeping it in sleep mode.
3. **Transitions between states:** By applying an action in a state, the system transitions from that state to the next state. Based on the one of these actions being applied to a state s at time t, the system transitions to a new state. Thus, the new state will comprise the updated values of the 'states'.
4. **Reward function definition:** The reward function specifies the rewards for being in a state or for doing some action in a state. Thus, the reward function implicitly decides the goal for 'learning'. For instance, if the new state consumes more power than the previous state or if it fails to provide the minimum required throughput to the users, the action is penalized. Consequently, reduction in power consumption while ensuring required throughput to users will be rewarded.

With these MDP elements, RL can be viewed as the closed-loop interaction between a learning agent, which is the decision maker, and the environment modeled by the MDP.The agent observes the current state, selects an action, and receives the next state and the reward. This decision making is influenced by the 'policy'. The policy is the strategy used to determine the next best action that should be taken based on the current state in order to maximize the future reward. The interaction among these elements is depicted in Figure 2.8 which is adopted from [40]. The observation is the state of the environment that is observed after applying an action. Two fundamental RL algorithms to solve the MDP are State Action Reward State Action (SARSA) and Q-Learning.

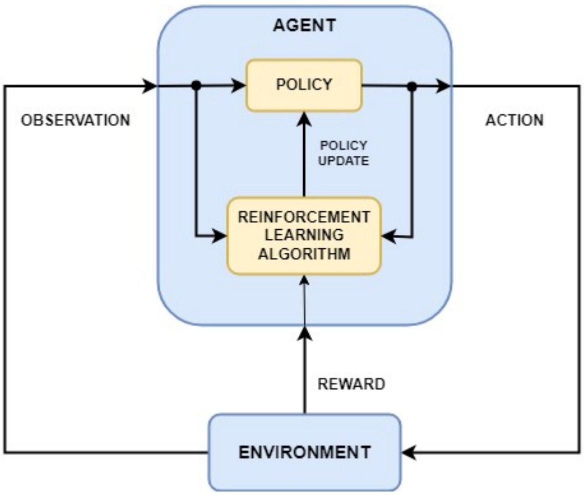


Figure 2.8: Reinforcement Learning

The Table 2.1 summarizes the RL papers at a glance: what each tries to achieve, which learning rule it uses, what the agent observes, what actions it can take, how it is rewarded or penalized, and the key limitations. In brief, almost all papers follow the same pattern. The agent sees a very small state. The action is to choose the next sleep level or how long to remain in a level. The reward encourages power savings and penalizes delay or wake-up cost. Exploration means occasionally trying actions that are not currently preferred so the agent can discover seek higher-return policies. SARSA learns from the actions it actually took, which makes it more cautious during exploration. Q-learning learns as if it always picked the best next action, which makes it more aggressive. This choice affects learning behavior, not the basic problem setup (state, actions, and reward).

A common limitation across every paper in Table 2.1 is that none of the studies use a realistic RAN topology. Most are single-cell or toy co-coverage, without multi-cell interference, handovers, or neighbor coordination. Because of this gap, their headline energy gains are not directly comparable to the earlier adaptive and non-adaptive strategies reviewed under more realistic multi-cell assumptions. Therefore, in this thesis, the results of these RL-based algorithms are treated as conceptual references rather than performance baselines.

Unlike the previously discussed ML-based approaches that do not use a realistic multi-cell topology and treat energy saving and interference separately, Romero et al. [41] formulate sleep control for a multi-cell cluster. They cast the problem as a constrained contextual bandit in which, a controller chooses cluster settings from the context information such as traffic level, QoS indicators, and current energy usage. The cluster settings comprise turning small cells on/off and tuning system parameters such as Almost Blank Subframe (ABS) and Cell Range Expansion (CRE) bias in order to reduce power consumption. Almost Blank Subframes/Slots (ABS) is a time-domain muting technique where the macro cell stays silent for a configured fraction of time so nearby small cells can serve edge users with less interference; this fraction is the ABS ratio. Cell Range Expansion (CRE) bias is a positive offset added to a small cells received power during association to steer more UEs onto small cells for load balancing, which can raise interference unless paired with coordination such as ABS. The proposed Bayesian Response Estimation and Threshold Search (BRETS) algorithm combines:

- **Threshold Search:** Finding traffic levels where a setting would violate the QoS requirements, and

- **Bayesian Response Estimation:** Learning how energy responds to traffic for each setting.

The BRETS methodology involves a two-level framework:

- **Global Controller (GC):** Manages a group of macro cells and determines the optimal IC-ES configurations based on network traffic data
- **Local Controllers (LC):** For every macro cell, the local controllers translate the global control strategy into local decisions

Although the approach improves energy efficiency under QoS constraints, it is computationally heavy, which may limit real-time deployment at larger scales.

Another study by Foivos et al. [42] proposes an ML-based sleep mode scheme with three BS states-sleep (no users), active (medium load), and full load (approximately 100%) which derived from per-cell user count, traffic, throughput, and energy. It compares a greedy policy (puts a cell to sleep based only on its own state) with a neighbor-aware policy (also considers neighboring states and offloads users to active neighbors). The greedy policy suits uniformly distributed traffic (e.g., households/offices), whereas the neighbor-aware policy is more suitable for clustered urban hotspots. **However, the study ignores mobility, context, and inter-/intra-tier interference. Its sole QoS metric is users still connected (omitting throughput/latency/reliability), and the user-transfer procedure is unspecified. Greedy can save energy but degrades QoS, and neither policy guarantees full QoS satisfaction, motivating richer algorithms that jointly optimize sleep states and service quality.**

Aim	Algorithm	States	Actions	Rewards	Limitations	Ref.
Find durations/usage of ASM levels under an energydelay trade-off	Q-Learning	BS mode: idle, active, SM1−SM3	Pick how many times to repeat SM3, then SM2, then SM1 before waking	Power savings are rewarded; delays are penalized	Single cell No mobility No neighboring cells No interference modeling	[43]
Activate sleep modes to meet energy delay objectives	Q-Learning	Load: inter-arrival time of user requests BS mode: active, SM1−SM3	Choose repetitions or durations for SM3→SM2→SM1; wake on user arrival	Power savings are rewarded; delays are penalized	Single cell No mobility No neighboring cells No interference modeling	[44]
Decide the best sleep mode over time to cut BS energy with acceptable delay	SARSA	Load level: High/Low Last chosen mode: active or SM1−SM3	Go to the next mode; wait for transition time	Power savings are rewarded; delays are penalized	Single-site co-coverage (two cells, different bands) No multi-BS coordination No handovers	[45]
Reduce BS energy while meeting 5G use-case end-to-end latencies (eMBB/mMTC/URLLC)	SARSA	BS mode: awake, SM1−SM3 Buffer status: Low/High	Choose next mode (awake/SM1/SM2/SM3)	Power savings are rewarded; wake-up delay and high buffer are penalized; a weight tunes energy versus latency	Single cell No mobility No neighboring cells No interference modeling	[46]
Save energy with multi-level sleep while keeping delay acceptable using location/speed awareness	Q-Learning	User zone/position Mobility context: path and velocity (setup)	Set BS state: Active, SM1−SM3	Energy gain is rewarded; added delay is penalized	Single cell No inter-cell coordination No interference modeling	[47]

Table 2.1: Comparison of Reinforcement Learning-based sleep mode strategies

2.2.2. User Equipment - Base Station association mechanisms

Efficient association of UEs to BS is a key lever for improving EE while lowering power consumption. Good association balances traffic, reduces transmit power, limits interference, and improves resource use in heterogeneous networks. In contrast, legacy rules such as strongest RSSI or maximum SINR often overload a few cells while others remain underused, increasing interference and wasting energy. As a result, some base stations run near peak capacity while neighbors sit idle, driving up PC and reducing EE.

Signal-based association algorithms

The Max-SINR association technique is a widely used method for connecting UE to the optimal BS in a cellular network [48]. Max-SINR connects each UE to the base station with the highest SINR, which boosts throughput and QoS. **However, it often creates congestion and macro cell bias in heterogeneous networks. As a result, small cells remain underused while macro cells run hot, increasing interference and power consumption.**

Cell Range Expansion (CRE) addresses this by artificially adding a bias to small-cell signals so more UEs attach to small cells, improving load balance and spectral efficiency [49]. In practice, CRE works best with dynamic bias tuning and interference coordination techniques such as ABS setting.**While CRE yields a more even distribution than max-SINR, it does not explicitly account for real-time cell load, traffic demand, or QoS.**

Beyond signal-only rules, an Analytic Hierarchy Process (AHP)-based association presented by Achki et al. [50] assigns weights to criteria such as SINR, distance, and energy, and selects the serving cell by minimizing a weighted cost. AHP provides a transparent multi-criteria framework with a built-in consistency check for the chosen weights.**The paper considers a simple topology consisting of one macro cell and just two small cells. Moreover, the assigned weights to different parameters are fixed and subjective, limiting adaptability and making results sensitive to arbitrary weight choices.**

Load balancing algorithms

CRE improves offloading relative to Max-SINR but does not react to real-time load, so it can misallocate users and raise energy use under changing conditions. This motivates load-aware association. Hassan et al. [51] propose a simple load-balancing method that minimizes the standard deviation of BS load. Candidate UE-BS pairs are ranked by SINR and current load; each UE is attached to the least-loaded viable BS, the network-wide load deviation is recomputed, and UEs are reassigned until the deviation no longer improves or falls below a threshold. Compared with Max-SINR and CRE, this yields more even load and higher energy efficiency and data rates. **However, the method uses fixed SINR and deviation thresholds that require retuning as conditions change, ignores user mobility, and treats each UE as an equal unit of load without accounting for heterogeneous traffic demands.**

Uneven user distribution not only leads to poor allocation of resources but also leads to inefficiencies in handovers which consequently lead to failures or ping-pong effects. Alkalsh et al [52] propose Dynamic Distance-Based Load Balancing (DDLB), which adapts load-balancing and handover decisions from real-time conditions. It monitors PRB usage to detect congestion, computes per-cell dynamic thresholds from neighbor loads, selects offload candidates by signal strength and proximity, and then adjusts Cell Individual Offset (CIO) values in overloaded and neighboring cells to steer handovers. This reduces handover failures, radio link failures, and overload duration. CIO is an offset added to a cells RSRP used in selection and handover; raising a neighbors CIO makes it more likely to be chosen. **The proposed DDLB is evaluated in**

a small cell-only topology. The authors note future research on small-macro coordination for heterogeneous networks. Performance is reported via mobility and congestion metrics, with no explicit energy model. Thus energy effects are indirect rather than quantified.

Building on that mobility-driven load balancing, Gures et al. [53] go a step further by combining load and signal strength with an explicit preference for millimeter-wave cells and careful tuning of handover margin and time-to-trigger. First, the algorithm prefers millimeter-wave cells as targets when they meet a load limit and a signal-strength threshold; otherwise it chooses the macro cell with the strongest signal. If the target is millimeter-wave, the handover is triggered immediately; if it is a macro cell, a handover margin is applied to avoid flapping. They also study the effect of the handover margin and the time-to-trigger: shorter times improve load balance, throughput, and spectral efficiency by offloading users faster, while larger margins push more traffic to neighbors but must be tuned to avoid new overloads. **The study evaluates load-balancing and mobility metrics such as throughput, spectral efficiency, call-drop rate, and overload duration without an explicit power or energy-efficiency model. Thus, any energy gains are therefore indirect.**

As networks densify, many small cells will not connect directly to the core; traffic may hop via gateways through neighboring cells. Mesodiakaki et al. [54] therefore use a backhaul-aware association that maximizes energy efficiency rather than focusing only on the RAN. Their method first finds candidate base stations for each UE that can meet its throughput with the fewest resource-block pairs (RBPs) (estimated from the UEs SNR; an RBP is one for UpLink (UL) plus one for DownLink (DL) resource block). Candidates are then ranked by fewer backhaul hops to the core and lower backhaul load. Each UE is attached to the highest-ranked BS that has enough RBPs, and resources are updated iteratively. This balances load and favors short, uncongested backhaul paths, outperforming max-SINR in both energy efficiency and load balance.**The study ignores interference between cells by simply considering SNR instead of SINR. Additionally, the power consumption model does not include load independent power consumption values, making the EE calculations unrealistic.**

Recent research adds mobility considerations to cell association policies. ADAptive Cell Selection (ADA-CS), a scheme proposed by Alablani et al. [55] replaces RSSI-only selection by using RSSI, vehicle speed, travel direction, and cell load, with the goal of maximizing dwell time to cut handovers. Slow vehicles are attached to small cells whereas fast vehicles are steered to macro cells to avoid ping pong. The method filters candidate cells by speed, load, and a direction cone, then selects the best match. The approach significantly reduces unnecessary handovers and failures compared to the highest RSRP approach. **Although it targets mobility and load balance rather than energy, longer and more stable connections can indirectly lower energy use by reducing signaling and improving resource use.**

Building on their own research Alablani et al. [56] propose an Artificial Neural Network (ANN)-based cell selection for ultra-dense 5G networks. Trained on actual data, it uses vehicle location, speed and direction to predict the serving cell that maximizes dwell time, reducing handovers. **A trade-off is that prioritizing dwell time can lead to UE picking cells with lower RSSI, which may reduce peak throughput. The method also relies on fixed thresholds for RSSI, speed, and dwell time, which can limit adaptability when conditions change quickly.**

Traffic offloading can suffer from heavy signaling, adding latency, energy use, and inefficiency. To address these issues, Alqerm et al. [57] proposes a conditional offloading scheme with online

RL with an 'intuitive' component where macro cells anticipate neighbors offloading behavior without direct communication, cutting coordination overhead. The controller dynamically switches small cells on/off and offloads users to keep small-cell load within thresholds, preserving QoS and avoiding power drain. Results show higher energy efficiency and a more even load balance, with faster convergence than standard RL. **The method assumes static user demands and mobility. Thus, rapid swings in traffic or mobility can degrade offloading decisions. Also, the threshold for system load to balance EE and QoS trade-off is predefined.**

Load-balancing schemes like dynamic load balancing or standard-deviation methods spread traffic well but are largely heuristic and rule-based. They struggle with the multi-objective trade-offs of heterogeneous 5G (energy efficiency, spectral efficiency, and QoS), especially under fast mobility, shifting demand, and backhaul constraints.

Energy and performance-driven algorithms-Power/Throughput/EE
Load-driven association focuses on relieving congestion and equalizing utilization across cells, which can indirectly help energy use and throughput. However, these methods do not explicitly optimize a target metric. This subsection turns to objective-driven approaches that make the goal function explicitgrouping papers that aim specifically at power reduction, throughput maximization, or energy-efficiency (EE) enhancement.

Li et al. [58] propose Decoupled Multiple Association (DMA) for ultra-dense HetNets which solely focuses on enhancing the network throughput. UEs can attach to different sets of base stations for uplink and downlink, chosen using context such as proximity, signal level, device type, and application needs. To avoid the combinatorial explosion of many-to-many choices, the method forms predefined BS groups for uplink and for downlink and solves a many to one matching problem. A utility that blends data rate, delay, and packet-error guides assignments, and the algorithm iterates group choices until no further improvement, yielding a higher number of UEs meeting QoS requirements than coupled association. Figure 2.9 shows the illustration of the problem conversion from many-to-many to many-to-one association. **Although the algorithm significantly enhances the throughput, it does not consider the power consumption.**

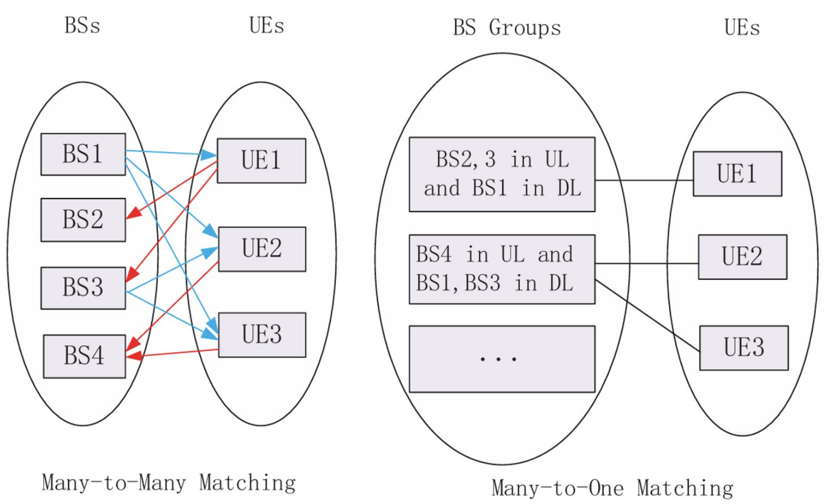


Figure 2.9: An illustration of problem conversion from many-to-many to many-to-one matching

The paper presented by Merve Saimler and Sinem Coleri. [59] formulates user association in a heterogeneous Cloud-RAN as a power-minimization problem with UL/DL decoupling and multi connectivity, so each user may be served by a small set of cooperating base stations and may use different serving stations in UL and DL. The Multi Connectivity-based User Association Algorithm (MCUAA), proceeds in two stages. First, it solves a relaxed version of the problem to obtain a fast draft of which base stations should remain active and which userstation links are promising; this draft is rounded to a feasible on/off and association plan. Second, it refines the plan with a lightweight assignment step that reattaches users to the most power-efficient candidates and toggles base stations on or off only when total network power decreases while all rate and QoS constraints remain satisfied. The procedure iterates small changes until no further reduction is possible. MCUAA is compared with a method that repeatedly makes small random changes to the current UE-BS assignment and keeps the change only if it reduces the total power. In comparison to this baseline, MCUAA achieves lower total power, faster convergence and higher UL/DL data rates. **However, the main assumptions of the study are static user positions and user demands during each run which perform well in dynamic environments. While multi connectivity increases reliability, the added signaling and pilot overhead can degrade achievable rates, which is not considered in the study.**

2.2.3. Strategies for energy consumption minimization in x-haul networks

There are studies considering joint UE association as well as backhaul routing that are developed to enhance EE and SE. In joint UE association, backhaul routing, and switch off model (JUBSM), the UEs are connected to SCBSs satisfying their throughput requirements. Backhaul routes are selected based on the least power-consuming paths [60]. The least loaded SCBSs are switched off and the UEs are re-associated with neighboring BSs. The backhaul links carrying no traffic are turned off. Similarly, the robust UE association, backhaul routing and switch off model (RUBSM) associates the UEs with SCBSs using a robust uncertainty value which accounts for variations in resource demands [61]. This value balances between the minimum PRBs required and the maximum deviation allowed for demand satisfaction. Backhaul routes are selected based on the availability of multiple paths and their respective energy costs. Unused SCBSs and backhaul links are switched off to save power. **However, these studies have not considered the minimum power consumption path which rather leads to increased congestion and power consumption. One major drawback of both JUBSM and RUBSM is the large computation time. The JUBSM runs up to ten times slower than baselines at peak hours while the RUBSM could take up to two days to reach optimality, making real-time use impractical.**

Venkateswararao et al. [27] target the often-ignored backhaul side of energy use. They propose an intelligent backhauling scheme for UE-BS association, paired with a Load-Sharing-based SCBS Sleep algorithm (LSCBS), to minimize total power. The association/routing problem is cast as minimum-cost flow in which the cell and backhaul links are chosen such that the power consumption is minimized. For each UE, candidate BSs must satisfy PRB needs derived from its rate and SINR, and paths are chosen by a cost that combines actual power so far with an estimate to the destination. This is repeated for all UEs to build energy-efficient routes from small cells to macro cells. For sleeping, BSs are sorted by load and under-utilized small cells are put to sleep only if neighbors can absorb their traffic without harming service. The process iterates until no further deactivations are possible while meeting QoS. This algorithm outperforms all the other algorithms in terms of active SCBSs, backhaul links and load balancing. **However, LSCBS is only approximately 10% slower than JUBSM and 17% slower than RUBSM, making it still impractical for realistic deployments.**

In the paper presented by Javad-Kalbasi et al. [62] power-cost function includes coupling between users and links: assigning one user changes which base stations and backhaul links are active and how heavily they are loaded, so costs are interdependent rather than separable per user. This makes the objective non-linear and computationally hard. The authors therefore use a quadratic upper bound that slightly overestimates power but is easy to optimize; minimizing this bound keeps the true cost controlled. **In simulation, the method improves energy efficiency by up to 8% over range expansion, minimum path loss, and SNR baselines, and it also outperforms advanced schemes such as JUBSM and RUBSM on power, backhaul selection, and load balance, at the cost of a modest increase in runtime.**

The novel approach proposed by Mowla et al. [26] targets backhaul energy in 5G small-cell networks with a hybrid PON-mmWave design. PON is more energy efficient at high load, while mmWave suits low load. This model accounts for both spatial and temporal traffic variations, allowing backhaul links to enter sleep modes during off-peak periods and reactivating them during peak demand. **However, the paper does not include the power consumption cost of switching between PON-based and mmWave-based backhaul and that of the unused backhaul infrastructure when one technology is employed.**

2.3. Limitations of the existing power consumption models and minimization strategies

Despite significant advancements in power consumption modeling and energy-efficient strategies for 5G RANs, several gaps persist, necessitating further research. The literature review has offered several key insights that influenced the approach of this thesis.

Several base-station power-consumption models exist, reflecting different technologies and levels of detail. Although fine-grained models are available, adopting that granularity here is impractical given data requirements and scope. Most studies therefore use simplified formulations with an idle (baseline) term and a load-dependent term, with baseline values taken from prior literature. Because vendor-specific, function-level measurements are not publicly disclosed, published baselines are used, and results are best read as comparative trends rather than precise absolute values. This answers the sub-question 1.1 of the research question 1.

Firstly, the non-adaptive strategies, although simple, lack the ability to adapt to highly dynamic environments. On the other hand, adaptive approaches adjust sleep levels or associations using current load and QoS, yielding lower power at equal QoS, fewer overloads, and reduced delay compared to fixed, non-adaptive threshold mechanisms. However, frequent optimizations can cause ping-pong handovers without careful hysteresis or timers. Moreover, the RL-based approaches mentioned in Table 2.1 adapt over time but most are evaluated on a single-cell, omitting multi-tier interference and neighbor coordination. Hence, comparisons to other multi-cell adaptive baselines are not fair. An exception is [41], which uses a realistic multi-cell topology but its computation is high and unsuitable for real-time deployments.

Another drawback is that the backhaul power consumption is often ignored. The few works that co-optimize access and backhaul (explained in subsection 2.2.3) report energy efficiency gains but at the cost of heavy computation.

In terms of RAN topology, many studies rely on oversimplified and often unrealistic assumptions. Sleep-mode mechanisms are frequently evaluated on isolated single cells rather than multi-cell networks Table 2.1, removing inter-cell coupling effects. Homogeneous small cell deployments are commonly modeled on idealized grids, which overlook irregular site layouts and traffic hotspots [35]. Several works also assume a single macro cell, thereby eliminating macro-macro interference and biasing results.

A pervasive limitation is the unrealistic modeling of user throughput demand. Many studies ([51][59] to name a few) assume identical per-UE requirements or draw from a small set of fixed rates randomly assigned to users [23][36], ignoring application-level traffic profiles and QoS constraints (e.g., web browsing, adaptive video streaming, interactive gaming, file transfer) as well as temporal variability and user-to-user correlation. Consequently, the induced load distribution and scheduling behavior are distorted. Moreover, with few exceptions [55][56], users are treated as static (no mobility model).

2.4. Summary

This chapter reviewed how 5G energy use is modeled and optimized across RAN comprising the base station and the x-haul network, and synthesized the main techniques and their limitations that motivate this thesis. The RAN dominates network energy with base station being the highest contributor of energy consumption. Base station power is commonly decomposed into load-independent and load-dependent parts, with AAU data confirming this structure and the need to account for sleep and switching overheads.

Concerning power consumption models:

Analytical models remain the most simulation-friendly and typically express base station power as baseline plus a load term driven by resource usage, while more detailed formulations include power consumption by the processing performed at the base station. This thesis adopts a simplified yet standards-aligned form to enable algorithmic exploration rather than component-level accounting.

Energy on x-haul network is shaped by the functional split, transport technology and deployment topology. Fiber PON is typically more efficient at high load, while mmWave is attractive at low load, suggesting hybrid designs.

Concerning energy consumption minimization techniques:

Most studies related to energy consumption minimization fall under the categories of putting base stations to sleep and steering users to different cells. They are then further categorized based on the 'technologies/methods' applied to achieve the objective. Fixed threshold-based strategies are simple but fail when traffic changes. Adaptive strategies can save more energy but can hurt QoS or trigger extra handovers and control overheads leading to ping–pong unless carefully tuned. Complex schemes often save the most, yet they are heavy to run and coordinate.

Concerning gaps and research direction:

The literature often:

1. omits realistic multi-cell topologies and thus, interference computation
2. simplifies per-UE throughput demand distributions, and
3. neglects x-haul energy consumption,
4. ignores switching overheads, control signaling overheads, and handover costs
5. omits user mobility

These gaps motivate the thesis focus on a tractable power model that spans macro/small cells and x-haul, combined with adaptive UE-BS association and sleep control evaluated in realistic urban scenarios.
Of these gaps, this thesis addresses realistic per-UE demands, a realistic multi-cell urban topology, sleep/wake switching overheads, and user mobility. The x-haul network topology optimization and energy consumption form a substantial research area on their own. Hence, the x-haul network is treated as provisioned and outside the scope of energy consumption accounting.

3

Joint base station sleep mode and selection strategy for energy-efficient small cell networks

This chapter presents the overall system architecture and the methodological framework used in the simulation. It details the modeling of network components, including macro cells, small cells, and UEs. The chapter also introduces the power consumption modeling based on the system model, the simulation workflow, and the proposed adaptive sleep mode mechanism and the proposed UE-BS association algorithm.

3.1. Selection of optimal topologies and technologies

The architecture of a Radio Access Network (RAN) fundamentally defines how radio and baseband processing functions are deployed, coordinated, and scaled. Consequently, RAN architecture plays a pivotal role in determining how efficiently a mobile network can deliver performance, scalability, and energy savings. Traditionally, mobile networks employed Distributed RAN (D-RAN) architectures, where each RU was paired with a dedicated BBU located at the cell site. While D-RAN offered simplicity and independence for each base station, it lacks the scalability and bandwidth efficiency required for emerging 5G services [63].

To overcome these limitations, the industry introduced Centralized RAN (C-RAN) or Cloud RAN, in which BBUs from multiple sites are pooled at a centralized location. This centralization allows for better coordination, improved resource utilization, and enhanced energy efficiency. However, as 5G networks became denser and more heterogeneous, particularly with the proliferation of small cells, Heterogeneous C-RAN (H-CRAN) architectures emerged. H-CRAN combines the advantages of heterogeneous network topologies with centralized processing, and introduces the decoupling of the control plane and user plane to improve flexibility, scalability, and load balancing across the RAN [63].

This evolution continued with the introduction of Network Function Virtualization (NFV) initially in the core network and subsequently extended to the RAN, enabling software-defined, cloud-native deployment of network components. Building upon these developments, Open RAN (O-RAN) has gained significant momentum in both academia and industry. O-RAN further disaggregates the BBU into two logical components: the Distributed Unit (DU) and the

Centralized Unit (CU). This modularization enables flexible functional splits across network layers and promotes interoperability through standardized, open interfaces [63]. The evolution of RAN architectures is shown in Figure 3.1.

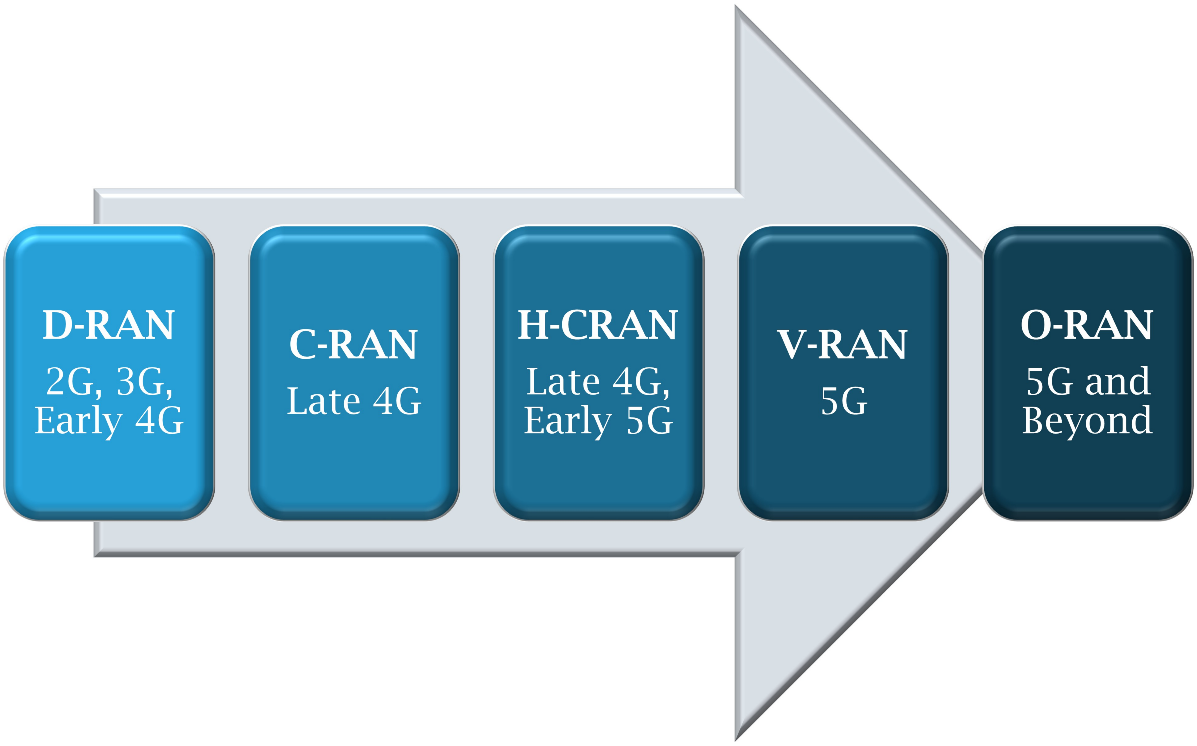


Figure 3.1: Evolution of RAN architectures

Functional splits define the functions executed by each unit, which in turn determine the data carried over the x-haul links (fronthaul, midhaul, backhaul) and in turn, the power consumption of each unit and each x-haul link. According to 3GPP, there are eight standardized functional split options. Among these, option 7.2x has emerged as a practical compromise, offering a good trade-off between RU simplicity, the fronthaul data rate and latency requirements. Under this configuration, the RU handles functions such as cyclic prefix addition/removal, FFT, beamforming, port expansion, and partial precoding. The DU, on the other hand, completes the remaining lower-layer processing tasks, including parts of precoding, modulation, scrambling, layer mapping, and mapping to PRBs. This layered, virtualized, and open architecture enables RAN deployments to be highly adaptive to diverse performance, cost, and energy efficiency requirements in 5G and beyond which is why it is chosen for this thesis [64] [5].

Earlier RAN architectures required operators to source all components RUs, DUs and CUs - from the same vendor, leading to vendor lock-in. In contrast, the O-RAN Alliance defines open interfaces between the RU, DU and CU, enabling multi-vendor interoperability. Additionally, O-RAN also supports advanced technologies such as massive MIMO, enhancing network capacity and spectral efficiency, making it an optimal choice for urban outdoor environments, which is the chosen scenario for this thesis [2].

Figure 3.2 compares the protocol stack division of the traditional BS, the C-RAN architecture and the O-RAN architecture. As illustrated in Figure 3.2 the O-RAN framework introduces a split within the physical layer, separating it into Low-PHY and High-PHY segments. Consequently, a portion of the physical layer processing is handled at the RU, while the remaining tasks are offloaded to the DU [9].

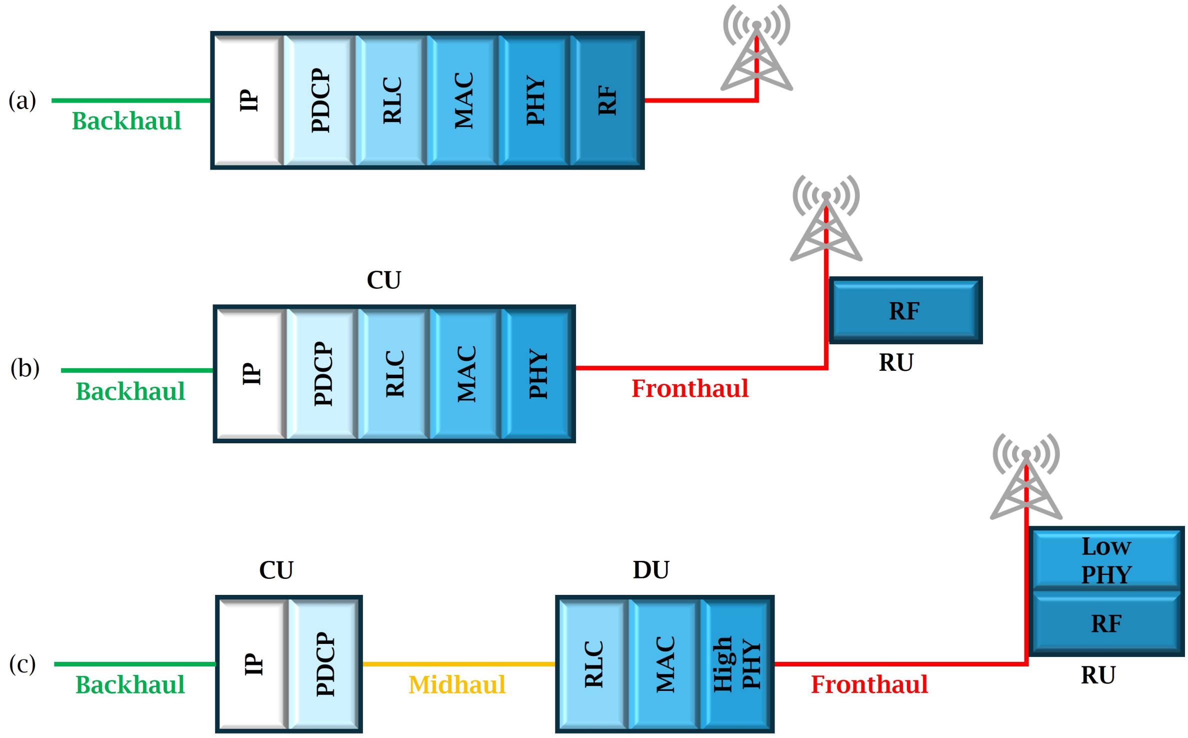


Figure 3.2: (a) Traditional base station with a Radio Access Network (RAN) protocol stack, (b) Cloud-RAN architecture, (c) Open-RAN architecture

3.2. Development of the energy consumption model

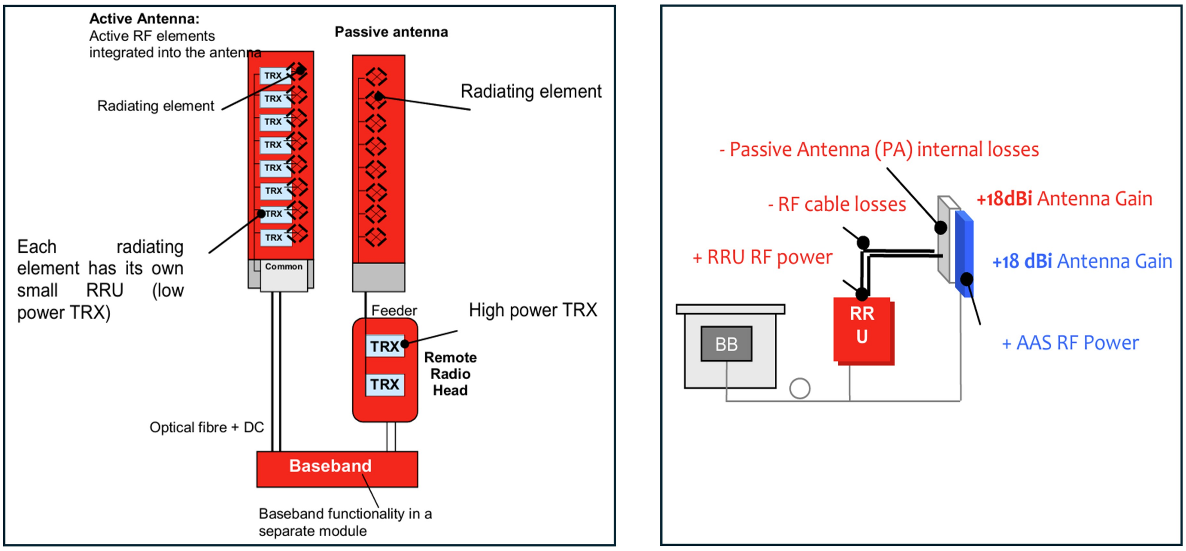
As explained in section 3.1, O-RAN architecture with split option 7.2 has been selected for this thesis. As per split option 7.2, apart from RF transmission, the RU performs a few processing functions such as adding cyclic prefixes, performing Fourier Transforms and beamforming, while the rest of the functions are performed by the DU. The RU performs comparatively less computationally intensive functions than the DU. However, the exact location of this split is still under discussion by 3GPP [5]. This ambiguity makes it difficult to incorporate the models that disintegrate the power consumption based on the hardware and the key processing tasks.

In modern 5G BS deployments, as shown in Figure 3.3, especially those supporting massive MIMO, AAUs [65] and Antenna Integrated Radio (AIR) [66] units are increasingly being adopted. These components combine the RU and the antennas in a single light-weight compact enclosure that can easily be deployed on rooftops or poles. These units support massive MIMO, high bandwidths, multiple frequency bands and advanced beamforming capabilities while minimizing power consumption. The integration of RUs and antennas leads to a significant reduction in RF cable losses making them more energy efficient. Moreover, these units also provide support for O-RAN architecture thereby facilitating 5G HetNet deployments.

(a) Ericsson Antenna Integrated Radio (AIR) (b) Huawei Active Antenna Unit (AAU)

Figure 3.3: (a) Ericsson (Antenna Integrated Radio) and (b) Huawei (Active Antenna Unit)

These units employ Active Antenna Systems (AAS), which replaces a single power amplifier by a matrix of smaller power amplifiers. This matrix is then integrated into the antenna which in turn reduces the cable losses. Figure 3.4 compares the traditional remote radio head with the AAS [67]. As can be seen from Figure 3.4, the AAS technology fits perfectly into the O-RAN architecture where the RU includes both processing elements as well as transmission hardware. However, the detailed technical specifications of these AAUs and AIRs are limited or proprietary, making it challenging to model power consumption based on real-world hardware.



(a) Remote Radio Head Versus Active Antenna Design (b) Inefficiencies Avoided by the Active Antenna

Figure 3.4: Active Antenna System

With functional split architectures such as option 7.2, physical layer processing is distributed between the RU and DU. However, the RU still performs the most power-intensive RF and analog tasks, whereas baseband processing in the DU and CU can be offloaded to centralized cloud infrastructure. Additionally, several studies in the literature treat the power consumption of the BBU (comprising the DU and the CU) as largely load-independent. In contrast, the RU power consumption is typically modeled as traffic dependent. This makes the RU the primary on-site power consumer, especially relevant in edge deployments like small cells. This approach is consistent with previous studies such as [37], which isolate RU power consumption for energy efficiency analysis in RAN deployments.

To estimate the power consumption of the network components realistically, a bottom-up modeling approach has been adopted, beginning with radio propagation characteristics and culminating in the total power consumption of the access and backhaul elements.

3.2.1. System model

Figure 3.5 shows the system model used in this thesis. The system consists of one serving macro cell base station (MCBS) with multiple small cell base stations (SCBSs) deployed inside its coverage. Es are placed within the highlighted macro area. or interference, the first-tier neighboring macro cells (left panel, surrounding hexes) are included in the SINR calculations. However, only the central macro and its SCBSs participate in the algorithms (association and sleep control). All BSs operate co-channel, so downlink interference comes from active SCBSs in the central cell plus the neighboring macro cells. The right panel zooms into the central site: each SCBS/MCBS radio unit connects over fronthaul to a centralized DU/CU, which in turn connects over backhaul to the core network. The Traffic Distribution Orchestrator (TDO) runs near the DU/CU and issues the UE-BS association and SCBS sleep/wake decisions. The macro cell remains always on to provide coverage; only SCBSs are subject to sleeping. DU/CU power is assumed constant (no load scaling or sleep), since short-timescale variations are minor and detailed vendor data are unavailable [37].

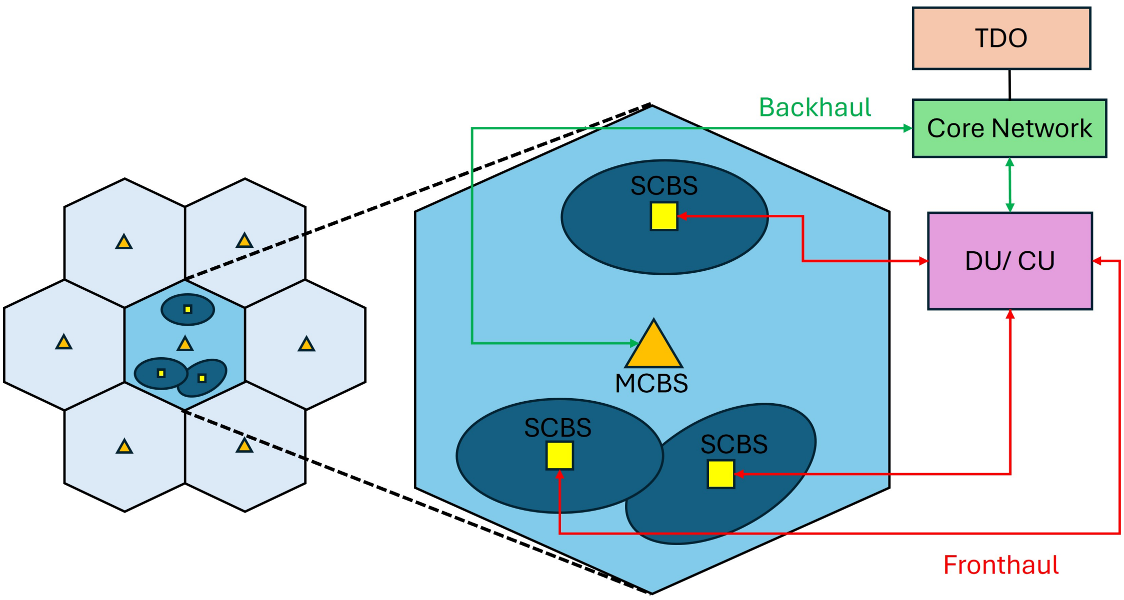


Figure 3.5: System model

The list of symbols and notations used in the energy consumption model are given in Table 3.1. To ensure a realistic and comprehensive simulation environment, a set of standardized parameters has been defined for both macro cell and small cell configurations. These parameters form the foundation of the system model and are consistently used throughout the simulation to evaluate performance metrics such as SINR, throughput, and power consumption. Table 3.2 presents the key simulation parameters, including transmission power levels, operating frequencies, idle and sleep mode power consumption, PRB capacities, and the adopted path loss

Symbol	**Description**
$\mathcal{U}$	Set of all UEs
$\mathcal{B}$	Set of all BSs (macro cells and small cells)
$\mathcal{M}$	Set of macro cells
$\mathcal{S}$	Set of small cells
i	Index for UE
j	Index for base stations (both macro cells and small cells)
s	Index for small cells
m	Index for macro cells
$a_{i,j} \in \{0,1\}$	Represents association between BS j and UE i (1=associated, 0=not associated)
$\alpha_s \in \{0,1\}, s \in \mathcal{S}$	Sleep state of small cell (1=sleep, 0=active)
PL_{SCBS}, PL_{MCBS}	Path loss for macro cell and small cell respectively
d	Distance between UE and BS in kilometers
$\text{SINR}_{i,j}$	SINR of user i from BS j
$\text{SINR}_{threshold}$	Minimum SINR required for UE-BS association
$P_{tx_j}^{max}$	Maximum transmission power for BS j
$G_{i,j}$	Channel gain between UE i and BS j
N_0	Thermal Noise
R_i	Throughput of UE i
D_i	Throughput demand of UE i
B_i	Bandwidth allotted to UE i
L_j	Load at base station j
$PRB_{i,j}$	Number of PRBs required by UE i from BS j
B_{PRB}	Bandwidth of one PRB
P_{total}	Total network power consumption
P_{MCBS} and P_{SCBS}	Power consumption of macro cell and small cell BS respectively
$P_{switching}$	Power consumed by the network while switching between on and sleep states
P_{active}	Load independent power consumption of a BS in active state
P_{sleep}	Load independent power consumption of a small cell during sleep state
$P_{on-sleep}$	Power consumed by individual small cell while switching between on and sleep states
η_{PA}	Efficiency of the power amplifier
$T_s \in \{0,1\}$	Transition between on and sleep states for small cell s(1=transition, 0=no transition)
EE_{total}	Energy efficiency of the network
$score_s$	Computation of cost function governed by load and power consumption metrics
P_{max} and P_{min}	Maximum and minimum power that the network can consume, respectively
epb_{total}	Average energy per bit values for all users
$T_{threshold_s}$	Threshold value for deciding the transition of small cell s between sleep and active state

Table 3.1: List of symbols belonging to the proposed system model

models. These values are selected based on widely accepted benchmarks in the literature and reflect typical urban deployment scenarios for 5G heterogeneous networks [23].

Parameter	Value
Simulation Configuration	
Simulation area	500 m × 500 m
Bandwidth of a PRB	180 kHz
Noise power density	−174 dBm/Hz
Power amplifier efficiency	0.39
Frequency and Transmission Power	
Macro cell operating frequency	2.6 GHz
Small cell operating frequency	3.5 GHz
Macro cell transmit power	46 dBm
Small cell transmit power	30 dBm
Power Consumption Parameters	
Macro cell idle power	780 W
Small cell idle power	21.6 W
Small cell sleep mode power	2 W
Small cell on-sleep switching power	13.5 W
Resource Capacity	
Macro cell capacity (Number of PRBs)	500
Small cell capacity (Number of PRBs)	100
Path Loss Models	
Macro cell path loss model	$128.1 + 37.6 \log d$ [km]
Small cell path loss model	$140.7 + 36.7 \log d$ [km]

Table 3.2: System model parameters

Path Loss estimation

The first step involves computing the path loss between the transmitter and the receiver, which directly influences the received signal power. The path loss is modeled as a function of distance, frequency and propagation environment. This determines the signal degradation over the wireless link. The path loss models for macro cell and small cell, denoted by PL_{MCBS} and PL_{SCBS} respectively are:

$$PL_{MCBS} = 128.1 + 37.6 \times log_{10}(d) \tag{3.1}$$

$$PL_{SCBS} = 140.7 + 36.7 \times log_{10}(d) \tag{3.2}$$

where D represents the distance between the transmitter and the receiver in kilometers [33] [22].

Signal-to-Interference-plus-Noise Ratio calculation

The received signal power obtained from the path loss model is then used to compute the SINR at the receiver. This accounts for both co-channel interference from neighboring transmitters and thermal noise. The SINR quantifies the quality of the received signal and directly influences the achievable data rate over the link. The SINR for the downlink for i^{th} UE, transmitted by j^{th} BS is given by:

$$SINR_{i,j} = \frac{P_{tx_j}^{max} G_{i,j}}{\sum_{k \neq j, k \in B} P_{tx_k}^{max} G_{i,k} + N_0} \tag{3.3}$$

Throughput estimation
Based on SINR, achievable throughput is estimated using Shannons capacity theorem. The effective data rate achieved is obtained as a function of the SINR and allocated bandwidth. This throughput estimation forms the basis for calculating the network load, which directly impacts the dynamic portion of the power consumption.

$$R_i = B_i log_2(1 + SINR_{i,j}) \tag{3.4}$$

Load estimation
The network load is determined by the percentage of PRBs that are used. Every RU/ BS has a fixed capacity, that is, the total number of PRBs. The number of PRBs required by every user is given as: The higher the SINR, the higher the achieved data rate and the lower is the number of PRBs used. The number of PRBs required by the UE i from BS j is given by $PRB_{i,j}$ as follows:

$$PRB_{i,j} = \left\lceil \frac{D_i}{B_{PRB} log_2(1 + SINR_{i,j})} \right\rceil \tag{3.5}$$

Thus, based on Equation 3.2.1, the load at BS j can be calculated as $L_j = \frac{PRB_i}{PRB_{total_j}}$.

Power Consumption
The total power consumption of a BS is modeled as the sum of load-dependent and load-independent components, as explained in subsection 2.1.1. The total power consumption of the network is calculated as:

$$P_{total} = P_{MCBS} + \sum_{s \in S} \alpha_s P_{SCBS_s} + P_{switching} \tag{3.6}$$

The power consumption of a macro cell BS, which is always active, is:

$$P_{MCBS} = P_{active} + P_{tx_{MCBS}}^{max} \left(\frac{1}{\eta_{PA}} \right) L_i \tag{3.7}$$

Similarly, the power consumption of a small cell BS considering active and sleep states is given by:

$$P_{SCBS} = \begin{cases} P_{active} + P_{tx_{SCBS}}^{max} \left(\frac{1}{\eta_{PA}} \right) L_s, \alpha_s = 0 \\ P_{sleep}, \alpha_s = 1 \end{cases} \tag{3.8}$$

The switching power is given by:

$$P_{switching} = \sum_{s \in S} T_s P_{on-sleep} \tag{3.9}$$

where T_s is 1 if the state of the small cell s changes from on to sleep or vice versa, and 0 otherwise. Thus, $T_s(t) = 1$ if $\alpha_s(t) \neq \alpha_s(t-1)$ and 0, otherwise, where t is the current timestep and $(t-1)$ is the previous timestep.

Energy Efficiency
The energy efficiency of the network is given as the ratio of total throughput to the total power consumption, given as:

$$EE_{total} = \frac{\sum_{i \in U} Ri}{P_{total}} \tag{3.10}$$

3.3. Problem formulation

The objective of this thesis is to minimize the power consumption without compromising the QoS. Thus, it can also be expressed as maximizing the energy efficiency.

$$max\ EE = \frac{\sum_{i=1}^{U} R_i(t)}{P_{total}} \tag{3.11}$$

Subject to:

$$R_i \geq D_i, \qquad \forall i \in \mathcal{U} \tag{C1}$$

$$a_{i,j} \in \{0,1\}, \qquad \forall i \in \mathcal{U},\ j \in \mathcal{B} \tag{C2}$$

$$\sum_{j=1}^{\mathcal{B}} a_{i,j} \leq 1, \qquad \forall i \in \mathcal{U} \tag{C3}$$

$$\alpha_s \in \{0,1\}, \qquad \forall j \in \mathcal{B} \tag{C4}$$

$$\mathrm{SINR}_{i,j} \geq \mathrm{SINR}_{\mathrm{threshold}} \qquad \forall i \in \mathcal{U},\ j \in \mathcal{B} \tag{C5}$$

The constraint C1 ensures that the data rate R_i achievable by UE i is at least equal to its data demand D_i. This is essential for maintaining minimum quality of service (QoS) for all users in the network. It guarantees that the network does not under-provision any user. The variable $a_{i,j}$ in C2 indicates whether UE i is associated with BS j, where $\mathcal{B}$ is the total number of base stations. A value of 1 implies that the user is connected to that base station, and 0 otherwise. The third constraint C3 ensures that every UE is connected to at the most one BS. Every small cell s_j can either be in active or sleep state as indicated by C4. Lastly, constraint C5 ensures that the wireless link between UE i and BS j maintains acceptable reliability and signal quality.

3.4. Proposed sleep mode mechanism

To minimize power consumption while preserving Quality of Service (QoS), an adaptive sleep mode mechanism was implemented for small cells in the HetNet. This mechanism dynamically determines whether each small cell should remain active or enter sleep mode based on a weighted cost function that considers both network load and energy efficiency.

Small cells consume non-negligible power even under light traffic conditions. However, simply turning off underutilized cells can lead to service degradation, particularly when user demand is highly dynamic or unevenly distributed. Therefore, a mechanism is required that intelligently balances load-driven performance with energy-aware operation. The goal is to deactivate (put in sleep state) small cells during underutilization while ensuring sufficient capacity during load surges. Thus, deactivating small cells solely based on load values or power consumption is undesirable. Hence, a cost function-based adaptive sleep mode mechanism has been developed.

3.4.1. Weighted cost function

The decision to transition a small cell to sleep mode is governed by a cost function formulated as a weighted sum of load and power consumption metrics. For every small cell s, a score $score_j$ is computed as:

$$score_s = w_1(s) \cdot L_s + w_2(s) \cdot \frac{P_{total}}{P_{max}} \tag{3.12}$$

where,
$L_s \in$ [0,1] is the PRB load of small cell s, $s \in \mathcal{S}$,

P_{total} is the total instantaneous network power consumption,
P_{max} is the maximum power that can be consumed by the network when all base stations are on and all PRBs of all base stations are allocated,
Thus, $\frac{P_{total}}{P_{max}} \in [0,1]$, and

$w_1(s)$ and $w_2(s)$ are the adaptive weights assigned to the load and power terms, respectively such that $w_1(s) + w_2(s) = 1$.

The adaptive updating of these weights considers three factors: load on base stations, network power consumption and energy per bit value. Energy per bit is how much energy the network spends to deliver one data bit. A target load $L_{threshold}$, power threshold $P_{threshold}$ and energy per bit threshold $epb_{threshold}$ are set as below:

- $L_{threshold}$ is a predefined constant load value between 0 and 1. The actual value is explained in chapter 4.
- $P_{threshold} = 0.6 * P_{max} + 0.4 * P_{min}$,
  where P_{min} is the power consumed by the network when all small cells are in sleep mode and the macro cell is on, but not serving any traffic. $P_{threshold}$ favors curbing power when the network tends toward its upper value.
- $epb_{threshold}$ is a predefined constant energy per bit value. The actual value is explained in chapter 4.

3.4.2. Adaptive weight updating

The weights $w_1(s)$ and $w_2(s)$ are updated dynamically at every time step to reflect current network conditions. The weight $w_1(s)$ is incremented by 0.1 (capped at 1) when:

- the small cell's load exceeds a target load value ($L_s > L_{threshold}$),
- the macro cell has high load ($L_m > 0.7$) , or
- any UE i in coverage of small cell base station s is unsupported (does not meet the QoS requirements), that is, $(\sum_{j \in \mathcal{B}} a_{i,j} = 0)$.

Conversely, when the network power consumption exceeds a threshold $P_{threshold}$, the mechanism shifts emphasis towards minimizing power consumption by decrementing the weight $w_1(s)$ by 0.1 (floored at 0) when:

- $P_{total} > P_{threshold}$, or
- $epb_{total} > epb_{threshold}$

. Thus, the weights are updated using the following logic:

$$w_1(s) \leftarrow \begin{cases} \min\left(w_1(s) + 0.1,\ 1.0\right), & \text{if } L_s > L_{threshold} \text{ or} \\ & L_m > 0.7 \text{ or} \\ & \sum_{j \in \mathcal{B}} a\{i,j\} = 0 \text{ for UE } i \text{ in coverage area of small cell} \\ & \text{base station } s \\ \max\left(w_1(s) - 0.1,\ 0.0\right), & \text{if } \mathrm{P}_{total} > \mathrm{P}_{\text{threshold}} \text{ or} \\ & epb_{total} > epb_{threshold} \end{cases} \tag{3.13}$$

$$w_2(s) = 1 - w_1(s) \tag{3.14}$$

where the $P_{threshold}$ is computed as $P_{threshold} = 0.6 * P_{max} + 0.4 * P_{min}$.

3.4.3. State transition logic

The transitions between sleep and active states are performed based on a threshold value which is computed as:

$$T_{threshold_s} = w_1(s) \cdot L_{threshold} + w_2(s) \cdot \frac{P_{min}}{P_{max}} \tag{3.15}$$

where,
P_{min} is the minimum network power conditions under ideal conditions.

A small cell transitions from active to sleep state if:

$$score_s < T_{threshold_s}$$

Consequently, a small cell transitions from sleep mode to active mode if:

$$score_s > 1.2 * T_{threshold_s}$$

A hysteresis margin is applied to avoid frequent switching of states. Upon entering sleep mode, the small cell deactivates its transmission chain and is removed from the UE associations and SINR calculations.

By adaptively adjusting the decision criteria, the mechanism avoids over-prioritizing either load or energy, leading to improved overall energy efficiency while maintaining user service levels. This approach offers a practical and computationally efficient method for enabling energy-aware control of small cells in 5G heterogeneous networks.

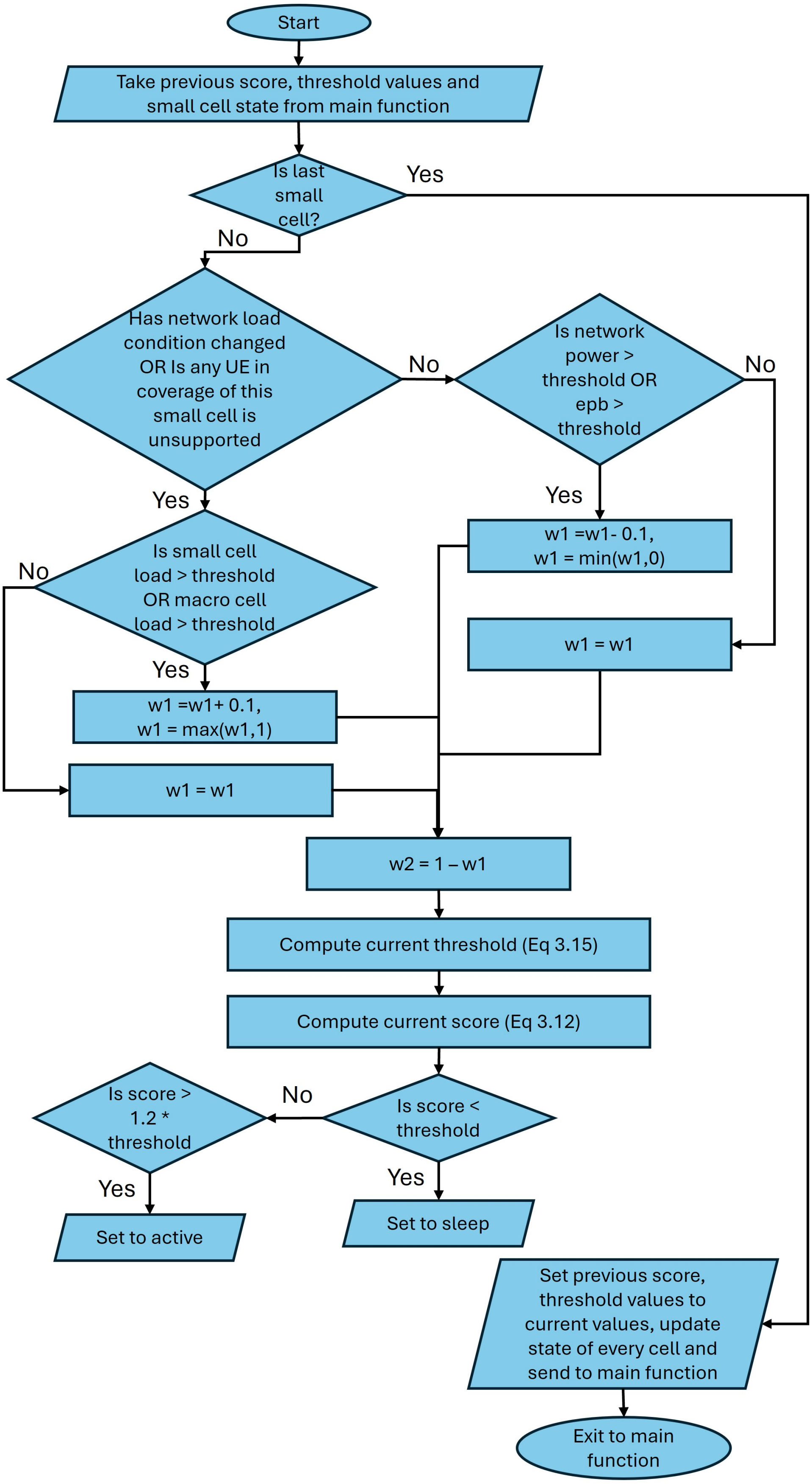


Figure 3.6: Proposed sleep mode algorithm flowchart

3.5. Proposed User Equipment-Base Station association algorithm

According to the 3GPP specifications TS 36.304 [68] and related procedural definitions in TS 38.331 [69], the UE connects to the cell that satisfies the selection criteria and that typically offers the highest received signal strength or quality. While signal strength remains the primary criterion in standardized cell selection and reselection, it alone is insufficient for optimal network performance.

In this thesis, additional parameters are incorporated into the cell selection process. Specifically, the cell load is considered for determining the activation and deactivation of small cells within the proposed sleep mode mechanism as explained in section 3.4. SINR and throughput demands are used for defining the UE-BS association strategy. Although cell load and throughput demand are closely related, since higher throughput demand from UEs increases the cell load, they represent distinct factors. Load represents the current utilization of cell resources while throughput demand characterizes the traffic requirements of individual UEs.

Thus, relying solely on signal strength-based association is insufficient for achieving both energy efficiency and balanced resource utilization in HetNets. While SINR remains the primary determinant of link quality, its integration with throughput demand enables a more informed association process that accounts not only for the connectivity quality but also for the energy cost per bit and distribution of traffic load across cells. This combined consideration forms the basis for the proposed UE-BS association algorithm, which is designed to reduce energy consumption and ensure fair utilization of macro cell and small cell resources.

The macro cell, due to its significantly higher transmit power, inherently consumes more energy per bit compared to a small cell. Since energy per bit is directly proportional to the transmit power of the serving base station, it is more energy-efficient to serve low throughput demand users via the macro cell and high-throughput demand users via small cells.

The approach in [23] adopts a binary classification of users into high-load and low-load categories where low-load UEs are directly associated with the macro cell. For the purpose of this study, this algorithm is referred to as **'Low-High Load Association (LHLA)'** algorithm throughout the thesis to facilitate comparison with the proposed algorithm. However, this methodology can result in inefficient resource utilization. In particular, users within small cell coverage areas may still be connected to the macro cell, leaving small cell resources underutilized. This necessitates a more refined mechanism to ensure efficient utilization of available cells.

The algorithm proposed in this thesis still uses signal strength (SINR) as the main factor for association, because low SINR can limit both high and low demand applications. However, it also takes the users throughput into account. The macro cell is treated as a fallback option, only serving users when small cells cannot meet their resource requirements. However, this leads to the users with low macro cell SINR values unserved. Thus, considering these aspects, the proposed algorithm is:

1. **Initial Allocation:** UEs whose SINR with the macro cell falls below a defined threshold are first redirected to small cells where they meet the minimum SINR requirement. This ensures that users with poor macro cell link quality are not associated with it, even if macro cell PRBs are available.
2. **Throughput-Based Ordering:** The remaining UEs are sorted in descending order accord-

ing to their throughput demands. This prioritization ensures that high-demand users are considered earlier in the allocation process, increasing the likelihood that they receive service from cells with sufficient resources and good link quality.

3. **SINR-Based Scheduling:** For each UE in the prioritized list, the algorithm identifies the candidate cell offering the highest SINR, but applies a Cell Range Expansion (CRE) bias of 6 dB to small cells. This bias artificially boosts the measured SINR of small cells in the decision process, increasing their selection likelihood over the macro cell when both have comparable link quality The value '6dB' is adopted from [70].

4. **Iterative Scheduling** UEs are allocated one by one to the highest-ranked cell from the adjusted SINR list (with CRE applied), subject to available PRBs. Small cells are preferred whenever they can provide adequate SINR after biasing. Only when small cell resources are insufficient or SINR is inadequate will the macro cell be selected.

The flowchart of the proposed algorithm is shown in Figure 3.7.

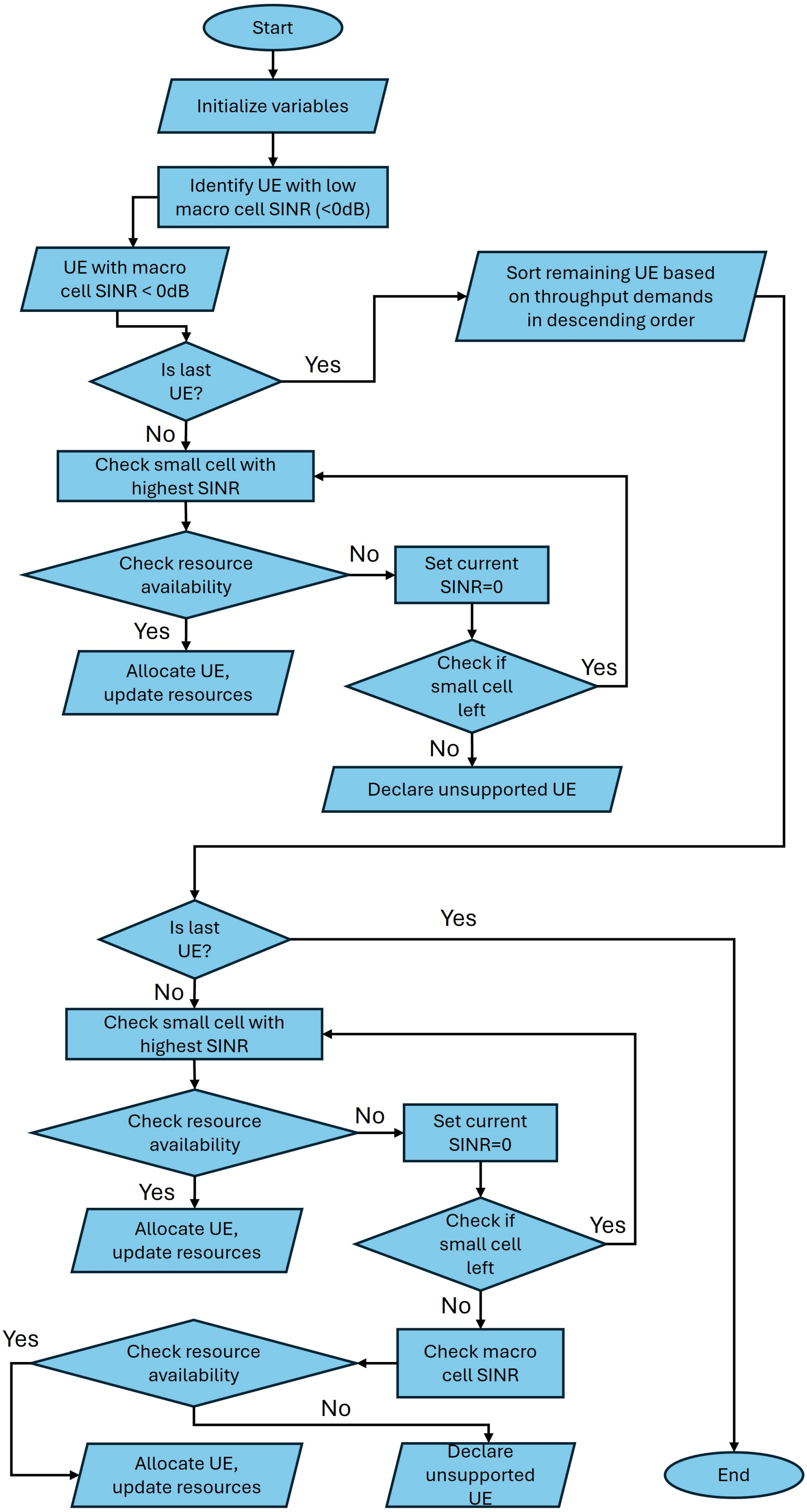


Figure 3.7: Proposed association algorithm flowchart

3.6. Simulation framework

There exist several software platforms for simulating 5G networks, including NS-3 [71], OMNeT++ [71], MATLAB [72], and Python-based frameworks [73][71], each offering distinct advantages depending on the simulation objectives. These tools support different levels of abstraction, ranging from protocol-level modeling to system-level behavior. Detailed protocol-level simulators such as NS-3 and OMNeT++ are well-suited for evaluating control signaling, protocol stack implementations, and lower-layer interactions in communication networks [71]. However, the primary objective of this thesis is to analyze energy efficiency through adaptive mechanisms such as sleep mode control and UE-BS association, which rely on metrics like SINR, load distribution, and power consumption. These aspects can be effectively modeled at a system level without requiring full protocol stack simulation. The choice of simulation tool thus depends on the trade-off between modeling granularity, computational efficiency, and the ease of use. In contrast to other 5G simulation tools, MATLAB and Python-based tools are more suited for system-level performance analysis. In this thesis, MATLAB was selected due to its strong support for numerical computation, real-time variable inspection [74], and widespread adoption in the literature for energy efficiency studies in wireless networks [54] [51] [20] [34].

One of MATLAB's major strengths lies in its interactive debugging environment, which allows users to set breakpoints, step through code execution line-by-line, and inspect or modify variables in real time [74]. The workspace panel provides a centralized view of all variables at any simulation stage, offering a level of interactivity and transparency that is particularly valuable when debugging complex adaptive algorithms or validating intermediate results. Although Python allows for variable inspection via debuggers or inline cells, it lacks MATLAB's centralized variable workspace and requires more manual configuration.

The simulation workflow is shown in Figure 3.8. The process begins with initialization of the system parameters which are given in Table 3.2. The simulation framework developed in this thesis aims to evaluate the energy efficiency of 5G heterogeneous networks under varying traffic loads and control mechanisms. The simulation proceeds in discrete time steps, and each time step represents a snapshot of the network operation where key decisions such as small cell sleep mechanism and UE-BS association mechanism are dynamically made based on the current network conditions.

After initialization of parameters, the simulation then proceeds to channel estimation through the computation of path loss and SINR values for the link between every UE-BS pair. Based on channel estimation, an estimate of achievable throughput and PRB demands is made. These metrics are then mapped to determine the load at every BS.

Based on the calculated load and current power consumption, an adaptive mechanism determines whether the small cell should remain active or enter a sleep mode to save energy without compromising the user experience. The association algorithm then determines the appropriate BS to serve every UE based on SINR and coverage. The evaluation metrics such as power consumption, energy per bit, and network energy efficiency are then updated based on a realistic power consumption model that accounts for baseline power, load dependent transmit power, and switching power.

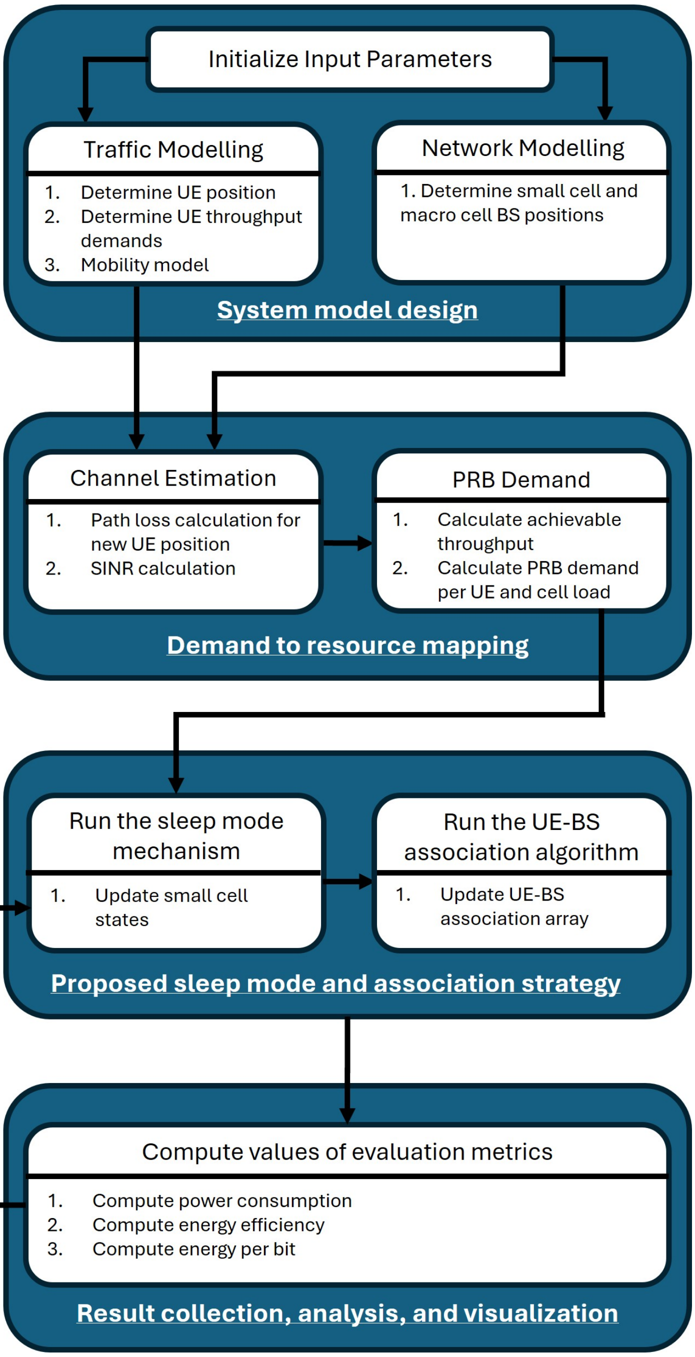


Figure 3.8: Simulation workflow

The process is repeated iteratively over multiple time steps, enabling the simulation to capture the dynamic behavior of the network under time-varying traffic conditions. To emulate realistic usage patterns, the simulation incorporates throughput requirements derived from commonly used applications. Table 3.3 presents representative data rate ranges for services such as WhatsApp, YouTube, and Netflix, allowing the evaluation to reflect practical load scenarios encountered in modern mobile networks.

Based on the throughput requirements of various applications presented in Table 3.3, users are classified into four categories according to their overall data consumption patterns, as shown in Table 3.4. For a given target cell load, user throughput demands are assigned such that

Application	Data Rate Requirement (Mbps)	Citation
WhatsApp (Text Only)	0.1–0.2	[75]
WhatsApp (Images)	0.5–1.5	[75]
WhatsApp (Voice Messages)	0.3–0.6	[75]
WhatsApp (Voice Call)	0.3–0.6	[75]
WhatsApp (Video Call)	0.5–1.5	[75]
Instagram	0.5–6	[75]
Email	~10 KB per message	[75]
Emails with Attachments	0.005–2	[75]
Zoom Meetings	60–100 Kbps, up to 3–4 Mbps	[76]
Microsoft Teams	10Kbps up to 1.5Mbps	[77]
Web Browsing (HTML)	~10 KB per page	[75]
YouTube	SD (480p): 1–1.5, 720p HD: 2.5, 1080p: 5, 4K: 20	[78]
Netflix	SD: $\geq$3, HD: $\geq$5, Ultra HD: $\geq$15	[79]
Real-time Video Gaming	720p: 2–6, 1080p: 6–9, 4K: 25–35	[75]

Table 3.3: Data rate requirements for common applications

the distribution of user types aligns with the load condition. For instance, under low-load scenarios, a larger proportion of users are categorized as light users, and their data rates are randomly selected within the corresponding range defined in Table 3.4. Furthermore, user

User Type	Typical Usage Pattern	Throughput Range (Mbps)
Light User	WhatsApp messaging, web browsing, email	0.1 – 0.5
Moderate User	Social media browsing, occasional video calls	0.5 – 2
Heavy User	HD video streaming, frequent video conferencing	2 – 10
Very Heavy User	4K streaming, gaming, multiple devices	10 – 25

Table 3.4: User categories based on throughput demand

mobility is modeled by assigning varying movement speeds to users located in hotspot regions. This approach reflects realistic urban environments, such as city centers and public gathering spaces, where user mobility patterns are highly dynamic. Incorporating such elements enhances the fidelity of the simulation and ensures that the proposed energy efficiency mechanisms are evaluated under realistic and heterogeneous operating conditions. To model user mobility, a random direction model was implemented.

At each time step, users are assigned a random direction and move with a predefined speed.

If a user reaches the boundary of the simulation area, their direction is reflected, simulating a bounce-back effect to keep them within bounds. This approach emulates realistic, unconstrained urban movement within a macro cell.

3.7. Summary

This chapter specifies the urban 5G radio access network under study, with a macro layer overlaid by small cells, and defines traffic and mobility models using realistic per-user demand profiles, hotspot placement, and user movement. It introduces a base-station power model with a baseline component and a load-dependent component, and it includes explicit switching overheads for transitions between sleep and active states. Load is represented through physical resource block utilization, which links signal-to-interference-plus-noise ratio, achievable rate, and scheduler decisions.

The chapter then details two control mechanisms: a load-aware controller for sleep and wake decisions with clear state transitions, and a user-to-base-station association procedure that extends the strongest-cell rule by ordering users by demand, applying a cell range expansion bias toward small cells when beneficial, enforcing feasibility with respect to available physical resource blocks, and using the macro cell as a fallback. The simulation workflow, parameter choices, and evaluation pipeline are described to ensure reproducibility.

4

Simulation Results

This chapter outlines the simulation environment, parameters, and traffic configurations used to evaluate the performance of the proposed mechanisms. Different deployment scenarios, including varying UE densities, mobility patterns, and small cell placements, are defined to reflect realistic network conditions and test the scalability of the approach.

4.1. Simulation scenarios

To evaluate the performance of the proposed energy-efficiency mechanisms and understand the impact of different deployment strategies, a series of simulation scenarios are constructed. Each scenario is designed to isolate and analyze specific aspects of network behavior in a 5G heterogeneous environment comprising a macro cell and multiple small cells.

The thesis focuses on an urban scenario since small cells are designed for capacity densification and spatial reuse under high UE density, whereas in rural low-density settings, capacity is not the bottleneck. Key system parameters are configured according to typical urban 5G deployment values, as summarized in Table 3.2.

The choice of simulation cases is carefully structured to follow a logical progression. Initially, the impact of small cell deployment is studied to understand the theoretical benefits in terms of energy and capacity. However, to reflect practical deployment concerns, further simulations are conducted to analyze how these benefits are influenced by the actual distribution of users. To gain deeper insight into how user distribution influences the effectiveness of small cell deployment, a SINR contour analysis is conducted to visualize coverage dominance and signal quality across the simulation area. Building on this, additional scenarios vary the UE distribution to reflect different traffic load levels and evaluate how such distributions impact power consumption and energy efficiency.

Further simulations examine the break-even points of energy efficiency, determining at what load levels the HetNet outperforms the macro cell-only configuration. Finally, the effectiveness of the proposed adaptive sleep mode mechanism is evaluated under dynamic conditions, where small cells transition between active and sleep states based on traffic load and power consumption, thereby demonstrating real-time energy savings without compromising user experience. This section answers the sub-question 2.3 of the research question 2.

4.2. Evaluation metrics and testing parameters

To quantitatively evaluate the performance of the proposed energy-efficient mechanisms in heterogeneous cellular networks, several metrics are employed. These metrics assess the trade-off between throughput, energy usage, and user experience. The key evaluation metrics are described below.

1. **Total Power Consumption:** Total power consumption represents the aggregate electrical power consumed by all active components in the network, including macro cells, small cells, and any associated backhaul infrastructure. This metric reflects the energy cost of operating the network infrastructure and is a key determinant in designing energy-aware sleep mechanisms.

2. **Energy Efficiency (EE):** Energy efficiency quantifies how efficiently the network converts energy into useful data transmission. It is defined as the ratio of total throughput to total power consumed. This is the primary metric for evaluating the trade-off between energy and capacity. A higher EE implies that the network is more power-efficient while still meeting traffic demands.

3. **Energy per Bit (EPB):** Energy per bit measures the amount of energy consumed to transmit a single bit of data across the network. Thus, energy per bit computation considers only the transmit power while excluding the processing power. Energy per bit for every user as given in [80] can be computed as $E_{b_u} = \frac{P_{tx_u}}{R_u}$. P_{tx_u} is the total transmit power required for a user 'u' and R_u is the achieved data rate for user 'u'. Then the average of all the EPB values is considered as the network EPB. Although [80] considers only transmit power for energy per bit computation, it is more realistic metric also include circuit power which includes baseband processing as explained in subsection 2.1.1. However, such detailed component-level data is not available due to which only transmit power has been considered in energy per bit calculation.

This section answers the sub-question 2.2 of the research question 2.

4.3. Evaluated algorithms

Both the proposed sleep mode strategy as well as the association strategy are compared against baseline algorithms, which are explained in the following subsections. This section answers the sub-question 2.1 of research question 2.

4.3.1. Evaluated User Equipment-Base Station association algorithms

In order to evaluate the effectiveness of the proposed UE-BS association strategy, four algorithms with varying levels of complexity and decision criteria are compared under identical simulation conditions. These include:

1. **Max-SINR:** A baseline approach where each UE connects to the BS providing the highest SINR, without considering network load or throughput demands.

2. **Cell Range Expansion (CRE):** A biased Max-SINR approach that applies a fixed positive offset to small cell SINR values, encouraging more UEs to connect to small cells.

3. **Low-High Load Association (LHLA):** A load-tiered strategy where UEs are classified into low-load and high-load groups based on throughput demands, with low-load UEs assigned to the macro cell and high-load UEs prioritized for small cells.

4. **Proposed SINR-Throughput Aware Association (STAA):** The developed algorithm which integrates SINR-based prioritization, throughput demand awareness, CRE bias, and resource availability to balance energy efficiency and performance.

Algorithm	SINR	CRE Bias	Throughput Demand	Capacity Awareness
Max-SINR	✓	×	×	×
CRE	✓	✓ (6dB)	×	×
LHLA	✓	✓ (6dB)	✓ (Low/ High Demand split)	✓
Proposed STAA	✓	✓ (6 dB)	✓	✓

Table 4.1: Decision criteria used in compared user–base station association algorithms

Table 4.1 summarizes the decision criteria used in the four UE-BS association algorithms evaluated in this thesis. The Max-SINR algorithm represents the baseline approach, where UEs connect solely to the base station offering the highest instantaneous SINR, without considering resource availability or traffic demand [51] [70] [81]. The Cell Range Expansion (CRE) algorithm builds upon Max-SINR by introducing a fixed bias in favor of small cells, thereby increasing their coverage area; however, it still ignores cell load and throughput demands. The Low-High Load Association (LHLA) method classifies UEs into low- and high-demand categories and associates low-demand UEs to the macro cell and high-demand UEs to small cells. While this improves load balancing, it can underutilize small cells if UEs in their coverage areas are classified as low demand. Finally, the Proposed SINR-Throughput Aware Association (STAA) algorithm integrates SINR, CRE bias, throughput demand, and resource availability to optimize energy efficiency and power consumption.

4.3.2. Evaluated sleep mode algorithms

In addition to evaluating UE-BS association strategies, this thesis compares two sleep mode mechanisms for small cells: the Fixed Threshold Sleep Mode and the Proposed Adaptive Sleep Mode. The fixed threshold method relies solely on a pre-defined load value (which is set to 40% for every small cell) to determine activation and deactivation, whereas the proposed method dynamically adjusts decision thresholds based on both load and energy efficiency indicators. The adaptive mechanism employs a weighted cost function, where the relative importance of load and Energy Per Bit (EPB) changes according to network conditions, enabling more responsive and context-aware state transitions. Table 4.2 summarizes the main differences between the two approaches.

4.4. Analysis of key findings

This section sums up the key results and explains their significance. It highlights how energy efficiency, network power consumption and energy per bit change. It compares a macro cell-only setup with a mix of macro and small cells, pinpoints the break-even point where adding small cells with sleep mode starts to save energy, and notes how bursty versus steady traffic affects the outcome.

4.4.1. Effect of adding small cells in macro cell-only configuration

The deployment of small cells is a widely proposed technique to enhance the capacity and energy efficiency of next-generation mobile networks. Small cells are typically characterized by lower transmit power and smaller coverage areas in comparison to macro cells. Theoretically, introducing small cells into a macro cell-dominated environment can offload traffic from heavily loaded macro cells, reduce transmission distances, and improve energy efficiency.

Feature	**Fixed Threshold Sleep Mode**	**Proposed Adaptive Sleep Mode**
Decision Criteria	Single fixed load threshold for activating/deactivating small cells.	Weighted cost function combining cell load and Energy Per Bit (EPB). Weights (w_1, w_2) are adaptively updated.
Load Awareness	Yes (binary: above or below threshold).	Yes (continuous influence via w_1; load changes influence score).
Energy Awareness	No explicit consideration of EPB or power consumption.	Explicit EPB awareness; prioritizes energy saving when EPB or total power is high.
Adaptivity	No; same threshold in all traffic scenarios.	Yes; thresholds dynamically adapt based on previous load, EPB, and power consumption trends.
Inter-cell Awareness	No; decisions made per cell independently.	Partially; macro cell load is also considered in weight updates.
Objective	Reduce energy consumption by turning off low-load cells.	Jointly optimize energy savings and load handling without sacrificing QoS.

Table 4.2: Comparison of sleep mode mechanisms

To evaluate the effectiveness of small cell deployment, the simulation initially investigates the impact of gradually increasing the number of small cells within the macro cell coverage area. A uniform user distribution (with 250 users) is assumed throughout the simulation domain, and small cells are incrementally added while keeping user positions and throughput demands fixed. This ensures a fair comparison between the HetNet scenario and the macro cell-only configuration. At every iteration a new small cell is added in the coverage area of the macro cell. The location of small cells is random. User association is governed by the Max-SINR algorithm, whereby each user connects to the cell offering the highest SINR. Figure 4.1 shows the comparison of network energy efficiency for the macro cell-only scenario and the HetNet scenario. Since the user distribution and user demands remain constant throughout this simulation, the energy efficiency of the macro cell-only scenario remains constant. With users uniformly distributed, each new small cell picks up only a few users. That gives little throughput gain, but each small cell adds idle power and some interference. Thus, as more small cells are added, more of them are lightly loaded and their baseline power dominates. Hence, the total power rises faster than throughput and energy efficiency falls.

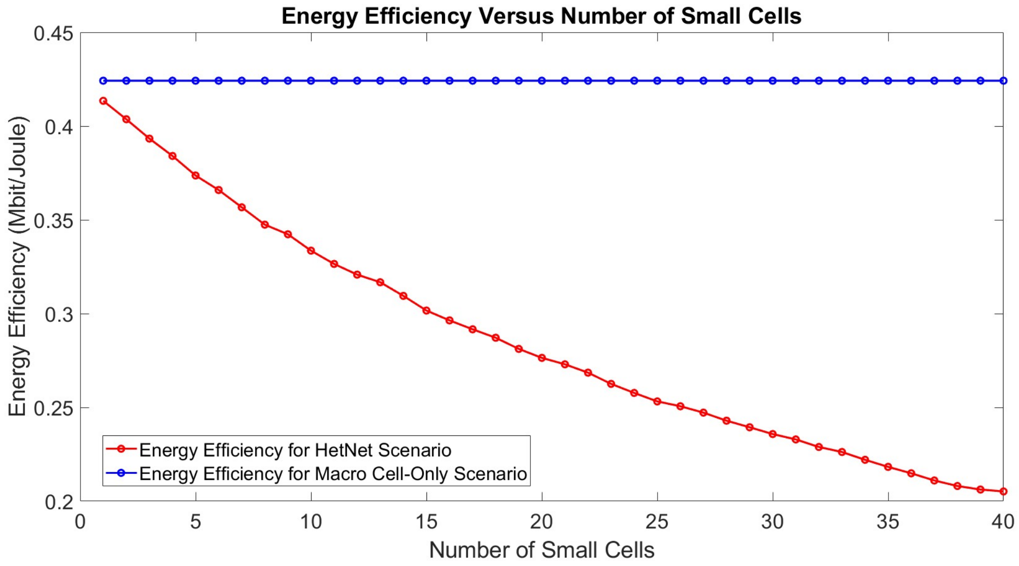


Figure 4.1: Effect of adding small cells on Energy Efficiency of the network

As anticipated, the addition of small cells leads to an increase in UE associations with small cells. However, the rise in small cell associations is less pronounced than expected due to the uniform nature of user distribution. Despite this, even a modest increase in small cell usage significantly impacts the networks EPB. Since the transmit power of macro cells is substantially higher than that of small cells, offloading even a few users to small cells results in a noticeable reduction in energy consumed per bit transmitted.

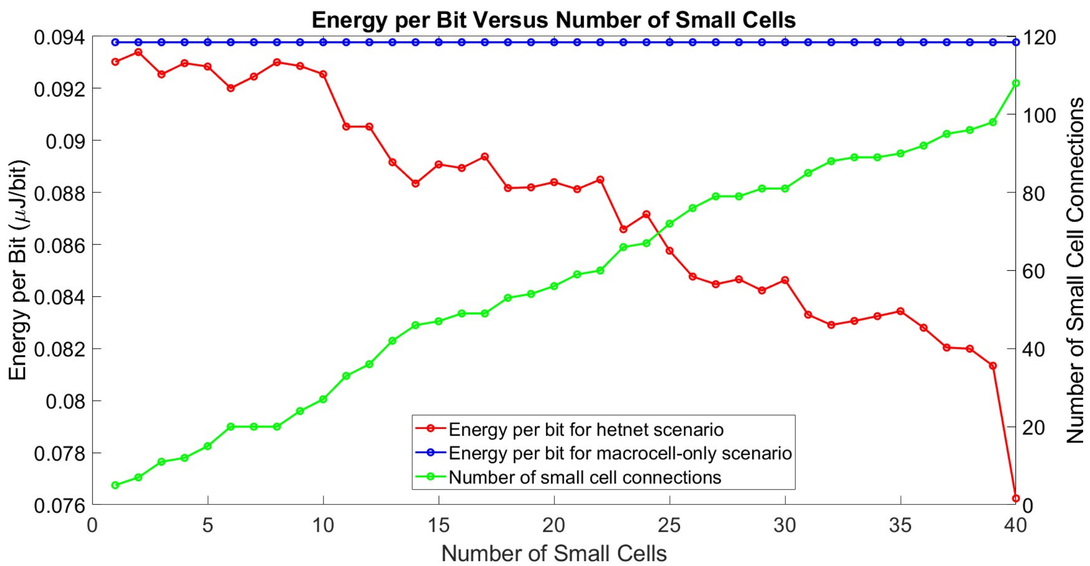


Figure 4.2: Effect of adding Small Cells on the energy per bit values

Thus, densifying with uniform demand and no sleep control is energy inefficient. To understand the distribution for which small cells improve the capacity and network energy efficiency, an SINR contour plot is studied as explained in subsection 4.4.2.

4.4.2. Visualization of cell dominance regions

To fully exploit the advantages of a HetNet deployment, it is crucial to understand the user distribution patterns that can most efficiently utilize the available network resources. Among the different factors that affect performance, signal strength indicators, especially the SINR, play an important role because they directly influence key aspects like network load. For example, higher SINR values correspond to lower resource block requirements to meet a given throughput demand. Moreover, SINR can be calculated independently of specific application requirements, making it a fundamental metric for determining the effective coverage regions of macro and small cells.

Figure 4.3 shows a contour plot representing the difference in SINR values between the macro cell and each of the small cells across the simulation area. The macro cell base station is located at the center (shown as a red triangle), and small cell base stations are marked as blue squares. The circular boundary corresponds to the macro cell's coverage radius of 500 meters.

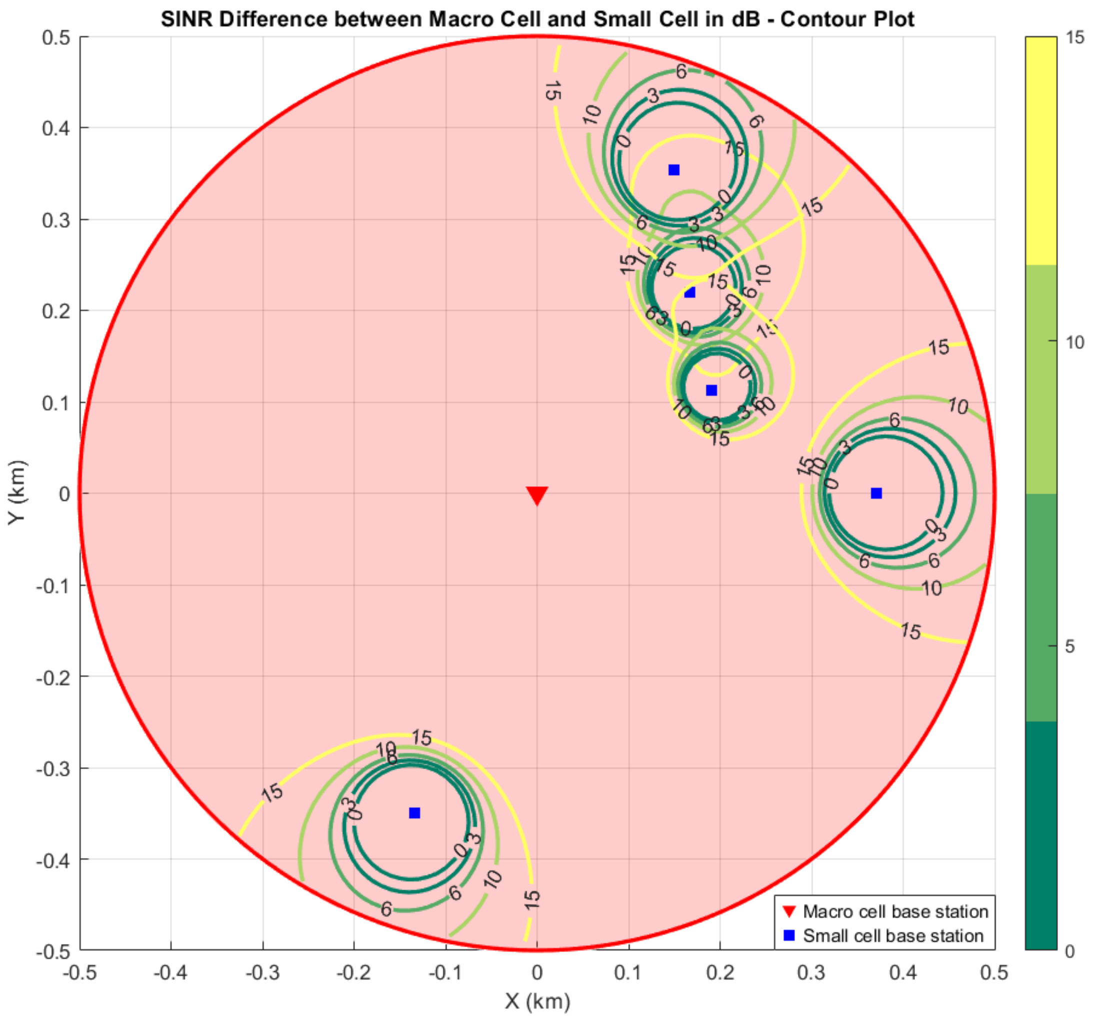


Figure 4.3: Signal-to-Interference-plus-Noise Ratio contour plot

The figure shows the SINR advantage of the macro cell over the best small cell at each location (x,y) in dB, defined as

$$\Delta SINR(x,y) = SINR_{MCBS}(x,y) - SINR_{SCBS}(x,y)$$

The color scale on the right indicates the difference in SINR in dB scale between the macro cell and the strongest small cell at each location. Thus, the color bar indicates the $\Delta SINR(x,y)$ values in dB. The first contour is plotted at regions where the macro cell SINR is equal to the strongest small cell SINR at that location. Hence the color bar begins from 0 and ranges till 15. Although the very high $\Delta SINR(x,y)$ regions are not used for small cell prioritization, they are still plotted only to understand how SINR values are influenced by distance from macro cell or neighboring small cells.

The first boundary in dark green color, labeled as '0' is the 0 dB locus where the macro cell and the (plotted) small cell provide equal SINR. Inside this contour, SINR of small cell is greater than that of the macro cell. Thus, inside the first contour, the $\Delta SINR(x,y)$ is negative since at those points, SINR of the small cell is greater than that of the macro cell (only the boundaries are plotted, not the inner regions). Similarly, the contour labeled as '3' indicates that the macro cell SINR is greater than the small cell SINR by 3dB. The region between the '0' and '3' boundary indicates that $0 < SINR_{MCBS} - SINR_{SCBS} < 3$ dB. This interpretation applies to all contours.

It can be observed from Figure 4.3 that the boundaries between macro cell and small cell coverage regions form elliptical contours, rather than perfect circles. These contours indicate locations where the SINR values from the macro cell and a nearby small cell are comparable. The elliptical shape arises due to the asymmetric transmit power between macro and small cells, with the macro cell having a significantly higher transmit power, resulting in a wider SINR dominance region for the macro cell.

Furthermore, it is evident that small cells located farther away from the macro cell exhibit larger green regions-areas where the SINR difference between small cell and macro cell is low, suggesting a better potential for offloading traffic. In contrast, small cells placed closer to the macro cell or to each other show smaller effective SINR regions, primarily due to interference from neighboring base stations and the overwhelming signal strength of the macro cell in close proximity. These observations emphasize that interference and relative placement significantly influence the effectiveness of small cell deployments and highlight the importance of strategic planning in heterogeneous network design.

Thus, it becomes evident that to fully harness the benefits of a HetNet deployment, user distribution should be concentrated within regions offering higher SINR from small cells. These regions lie closer to the small cell base stations and experience lower interference from the macro cell. Such user concentration patterns are commonly observed in urban hotspot environments, such as city centers, public transport hubs, and event venues, where large groups of users tend to cluster within confined areas. This reinforces the need for strategic placement of small cells in high-demand zones to maximize network performance and energy efficiency. Also, it is more logical to place small cells closer to the edge of the macro cell coverage area.

This plot is particularly useful for visualizing relative coverage strength. Rather than absolute SINR, it emphasizes where small cells provide competitive or superior signal quality compared to the macro cell, guiding intelligent UE association strategies. Green regions suggest effective coverage dominance by small cells, which helps offload users from the macro cell and can lead to improvements in both capacity and energy efficiency. Such analysis is crucial for optimizing small cell placement and understanding the spatial influence of signal strength on network performance.

4.4.3. Power and Energy Efficiency trade-offs in macro cell-only and HetNet deployments

Since the primary focus of this thesis is to minimize power consumption in 5G Radio Access Networks (RAN), it is essential to compare the power consumption of a macro cell-only network with that of a HetNet configuration. However, evaluating power consumption in isolation does not provide a complete picture of network performance. A more comprehensive and meaningful metric is energy efficiency, which accounts for both the power consumed and the throughput delivered to users. This allows for an assessment that balances energy usage with the quality of service, making EE a more suitable indicator for evaluating the effectiveness of energy-saving mechanisms.

Power Consumption break-even point

Although small cells are deployed to enhance network capacity and energy efficiency, their dense deployment can also lead to increased power consumption due to the additional baseline power required to keep them active. Therefore, mechanisms to minimize small cell power usage, such as sleep mode strategies and efficient user association algorithms, are critical. It is important to note that small cells are not always power-efficient; when underutilized, they may consume significant energy without contributing meaningfully to network capacity.

Figure 4.4 compares the total power consumption of three scenarios: macro cell-only, HetNet with sleep mode enabled, and HetNet without sleep mode. The overall power consumption is decomposed into four components: the baseline and transmit power of both the macro cell and small cells. Excluding the macro cell baseline power (which remains constant), it is evident that the macro cell-only scenario is often as power-efficient as, or even more efficient than, the HetNet scenario under low load conditions. For example, at low user densities, the HetNet without sleep mode consumes the highest power because the small cells remain active while serving very few users, leading to unnecessary baseline power consumption.
In contrast, the HetNet scenario with sleep mode significantly reduces baseline power at low loads by switching off underutilized small cells. As the load increases, more small cells are activated, effectively offloading users from the macro cell. This not only improves the utilization of small cell resources but also reduces the macro cell transmit power. For higher loads, the power consumption of the macro cell-only scenario rises sharply because all users are served by the macro cell. Consequently, a break-even point is observed: for the HetNet without sleep mode, power efficiency surpasses the macro cell-only scenario around 50% load (approximately 80 UEs), while the HetNet with sleep mode achieves this break-even point earlier, at around 35% load (approximately 60 UEs).

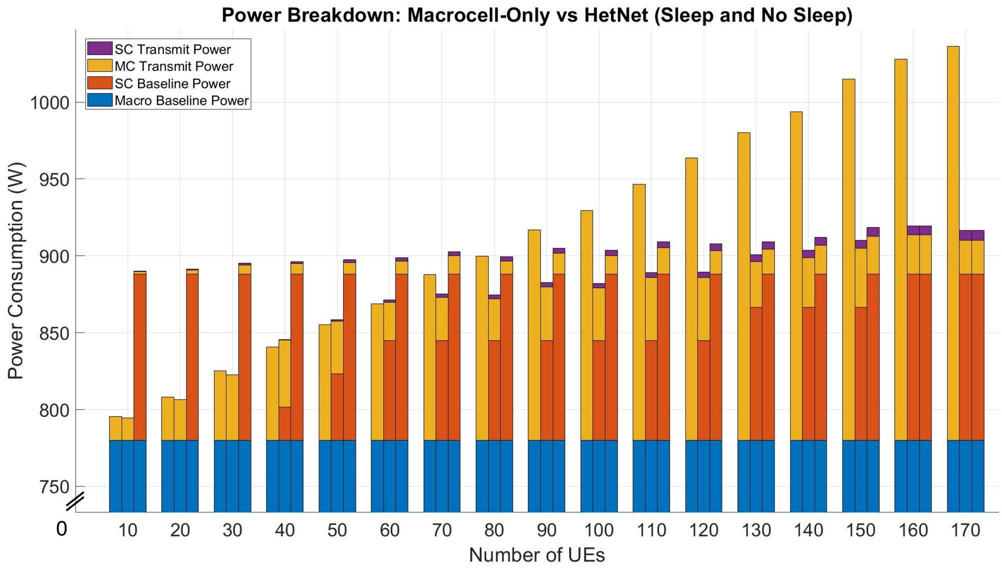


Figure 4.4: Power Consumption comparison for macro cell-Only, HetNet with and without Sleep Mode Scenarios

This trend is clearly depicted in Figure 4.5. The upper subplot shows the variation in total network power consumption with respect to the number of UEs, while the lower subplot illustrates the number of small cell connections in HetNet scenarios with and without sleep mode. It can be observed that the number of small cell connections in the HetNet scenario without sleep mode is consistently higher, since all small cells remain active regardless of the load. In contrast, the sleep mode-enabled scenario activates only the necessary small cells based on network load. As the load increases, more small cells transition to the active state, leading to a gradual increase in small cell connections.

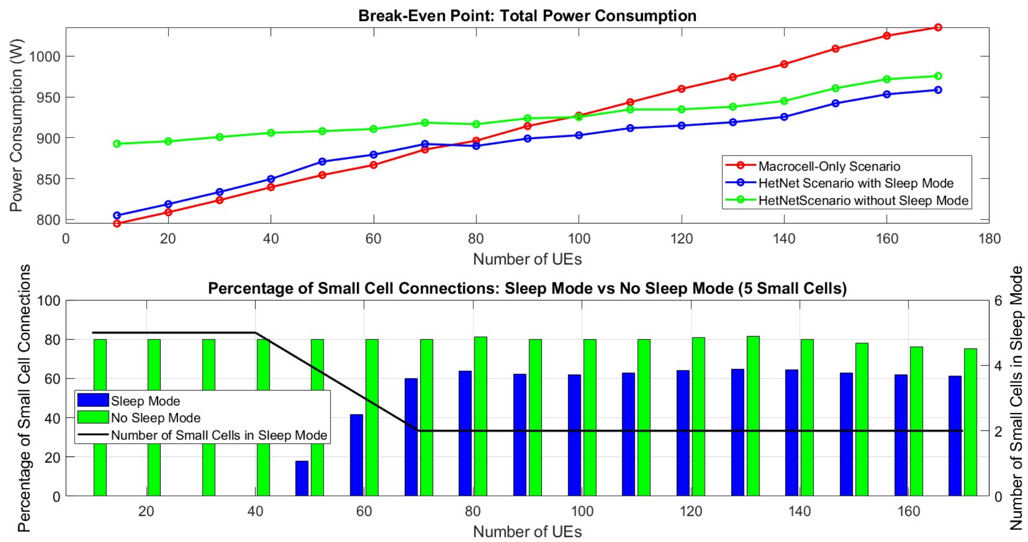


Figure 4.5: Power Consumption break-even point

Energy Efficiency break-even point of macro cell-only and heterogeneous network scenario
Based on the contour plot explained in subsection 4.4.2, the user locations follow a clustered hotspot pattern with most users concentrated near designated hotspot centers. The number of UEs is gradually increased to understand the energy efficiency at increasing network loads. The network contains 5 small cells. The energy efficiency versus offered load is plotted to identify the break-even load at which activating small cells becomes worthwhile. Each UEs target rate is 2 Mbps so that the load scales directly with the number of UEs and the break-even point is easier to see. The sleep mode algorithm used in this is the fixed threshold sleep mode algorithm. The findings from this graph shaped the development of the proposed sleep mode algorithm.

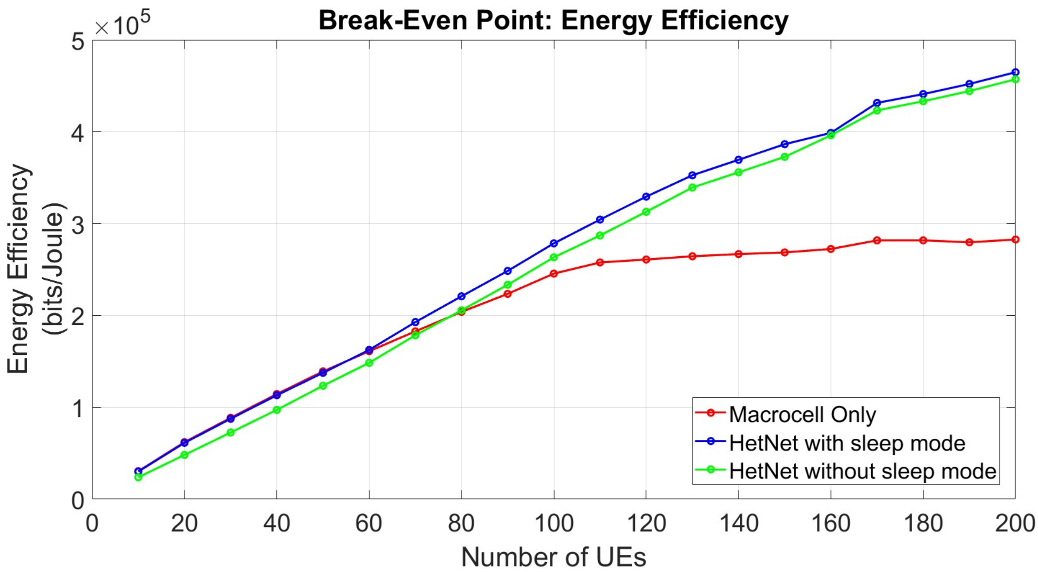


Figure 4.6: Energy Efficiency break-even point

The EE trend closely mirrors the network power consumption behavior. At lower loads, where small cells remain largely underutilized, the energy efficiency of the HetNet scenario is comparable to that of the macro cell-only setup. As the load increases and small cell resources are more effectively utilized, EE improves significantly. Notably, the energy efficiency curve for the macro cell-only scenario plateaus beyond a certain load level, indicating that the macro cell reaches its capacity limit, leaving many UEs unserved. This highlights a critical advantage of the HetNet configurationnot only does it enhance energy efficiency at higher loads, but it also substantially increases overall network capacity. The break-even point for HetNet with sleep mode is achieved at number of UEs equal to 60 (which corresponds to approximately 35% network load) whereas that for HetNet without sleep mode is achieved at 80 UEs ($\tilde{5}0$% network load) with respect to macro cell-only scenario. In order to find the break-even point for one small cell, the same simulation with one small cell is performed. This is done to check the load on small cell at which the small cell becomes more energy efficient than the macro cell-only scenario. Figure 4.7 shows the plot of energy efficiency of HetNet with one small cell (without sleep mode) versus the macro cell-only scenario.

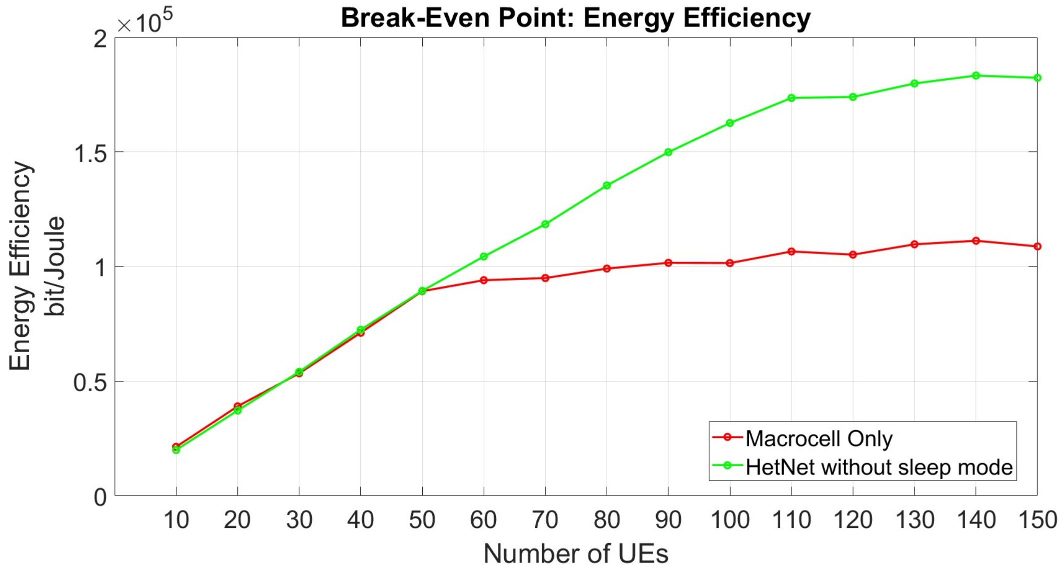


Figure 4.7: Energy Efficiency break-even point for only one small cell in HetNet scenario

In this case, the break-even point is achieved at approximately 50% load on the small cell (not the total network load). Thus, from the results achieved from the two simulations, the load target value, $L_{threshold}$ (as explained in subsection 3.4.2), is set to 40%.

Comparison of Energy per Bit for macro cell-only and HetNet scenario

As discussed earlier in subsection 4.4.1, the Energy Per Bit (EPB) decreases in the HetNet scenario as the number of active small cell connections increases. In Figure 4.8, at low loads, when all small cells remain in sleep mode, the EPB values are identical to those of the macro cell-only scenario, as all UEs are served exclusively by the macro cell. However, as the network load increases and more small cells transition to active mode, the EPB values begin to drop due to shorter transmission distances and lower transmit power requirements.

Notably, the lowest EPB values are observed in the HetNet scenario without sleep mode, since it allows the maximum number of UEs to associate with small cells, reducing reliance on the macro cell. Nonetheless, as load increases and all small cells in the sleep mode-enabled scenario eventually become active, the EPB values of both HetNet configurations converge.

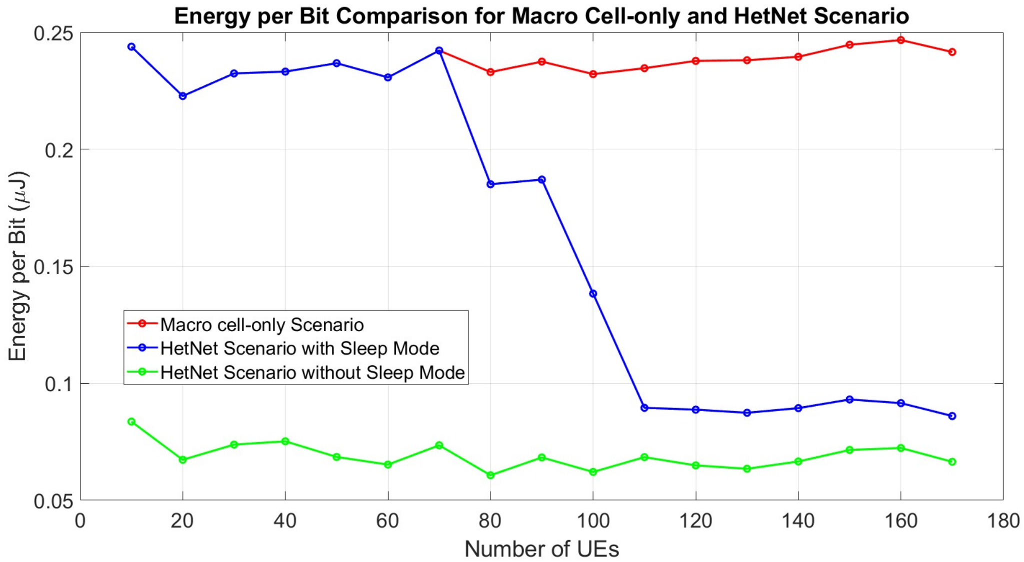


Figure 4.8: Energy per Bit Comparison for Macro Cell-Only and HetNet Deployment with Sleep Mode

While EPB effectively reflects how efficiently the transmit power is used to deliver data, it can be a somewhat misleading metric as it excludes other significant power components, such as baseband processing and circuit power. Despite this limitation, EPB remains a valuable indicator for highlighting the role of small cells in enhancing network performance and energy efficiency. Based on this plot, the $epb_{threshold}$ value is set to 0.1uJ since the average energy per bit values for HetNet scenarios vary from 0.05 to 0.1 uJ.

4.4.4. User Equipment – Base Station association algorithm performance

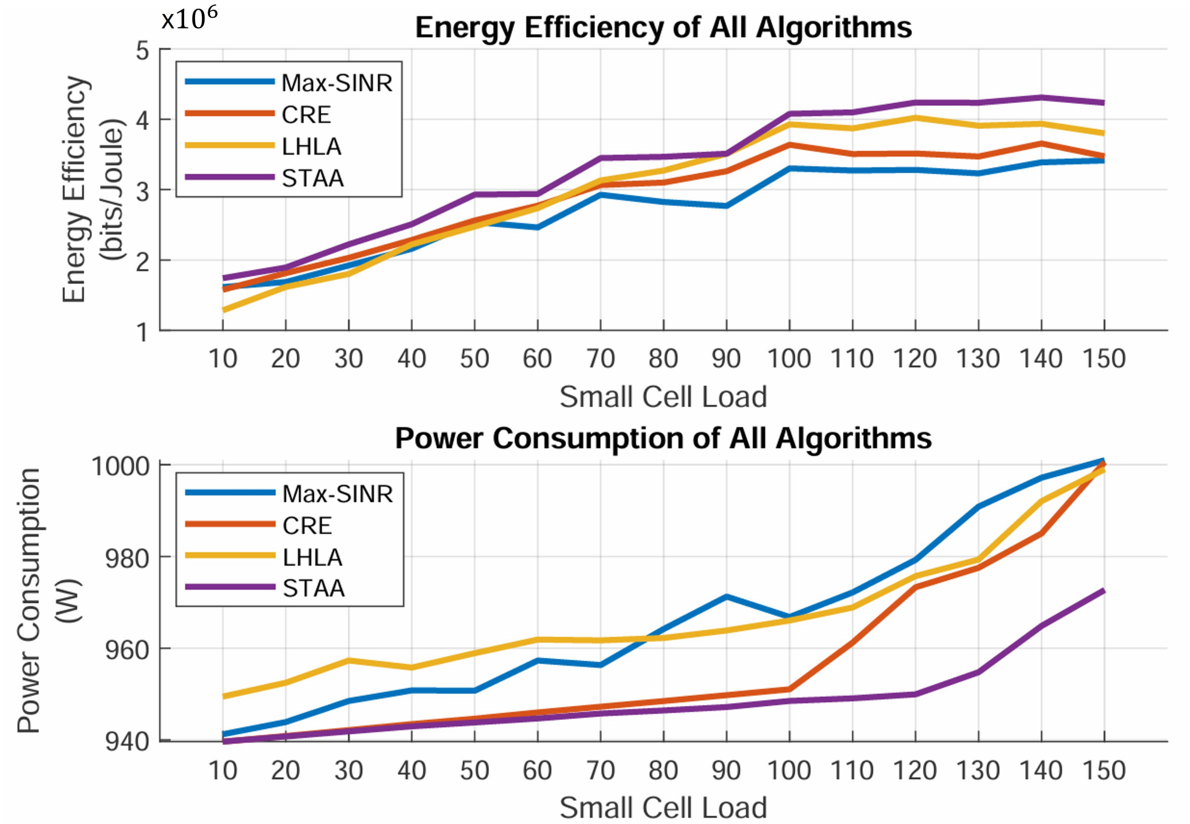


Figure 4.9: Performance comparison of User Equipment - Base Station association algorithms in terms of Energy Efficiency and power consumption

Figure 4.9 compares the performance of four UE-BS association algorithms Max-SINR, CRE, Low-High Load Association (LHLA), and the proposed SINR-Throughput Aware Association (STAA) in terms of energy efficiency (top plot) and power consumption (bottom plot) for varying small cell load conditions. The load on every small cell is gradually increased from 10% to 150%. Loads greater than 100% indicate that the small cells are overloaded. The users and their demands are set such that all of these users will not be able to be served by that small cell. This is included to replicate realistic conditions in urban hotspots where a large number of users are concentrated.

In the energy efficiency plot, all algorithms show an upward trend as the small cell load increases, reflecting the higher utilization of energy-efficient small cells. However, the proposed STAA consistently achieves the highest energy efficiency across the entire load range, with noticeable gains at high loads. This improvement stems from STAAs joint consideration of SINR and throughput demands, which ensures high-load UEs are preferentially served by small cells, reducing the average energy per bit.

The power consumption plot shows that while Max-SINR and CRE consume similar or higher total power, particularly at high loads due to suboptimal small cell utilization. STAA maintains significantly lower power consumption throughout. The gap is most evident at high loads, where STAAs targeted small cell activation strategy avoids unnecessary macro cell transmissions, reducing overall energy draw. LHLA achieves slightly better efficiency than CRE at lower loads, but its performance plateaus as the load increases due to underutilization of small cells in some scenarios.

Overall, these results confirm that while Max-SINR and CRE rely primarily on signal strength or fixed bias, the proposed STAAs load- and throughput-aware association strategy achieves both higher energy efficiency and lower power consumption, especially under heavy network load conditions.

Moreover, the LHLA algorithm performs better than Max-SINR and CRE in terms of both EE and power consumption for loads exceeding around 50%. However, its performance is worse than both Max-SINR and CRE for loads lower than 50%. This suggests that inclusion of throughput demands is necessary for improving the algorithm. However, aggressively allotting low throughput users to macro cells consumes unneccessary power while providing throughput similar to max-SINR or CRE. This is where the proposed STAA algorithm outperforms LHLA approach as well as the traditional Max-SINR and CRE approaches.

4.4.5. Sleep mode mechanism algorithm performance

As explained in subsection 4.3.2, the proposed adaptive sleep mode algorithm is compared with the fixed threshold sleep mode algorithm. The algorithms are compared in terms of power consumption as well as EE. In this study, the Adaptive Threshold Sleep Mode Mechanism and the Fixed Threshold Sleep Mode Mechanism are evaluated under two distinct traffic patterns: bursty traffic, characterized by sudden and irregular spikes in demand, and steady traffic growth, representing a gradual and sustained increase in network load. These traffic scenarios are chosen to assess the robustness and adaptability of each mechanism in handling both highly variable and progressively increasing loads. The algorithms are compared over a sequence of timesteps where each timestep represents a point at which the throughput demands and in turn, the small cell loads fluctuate.

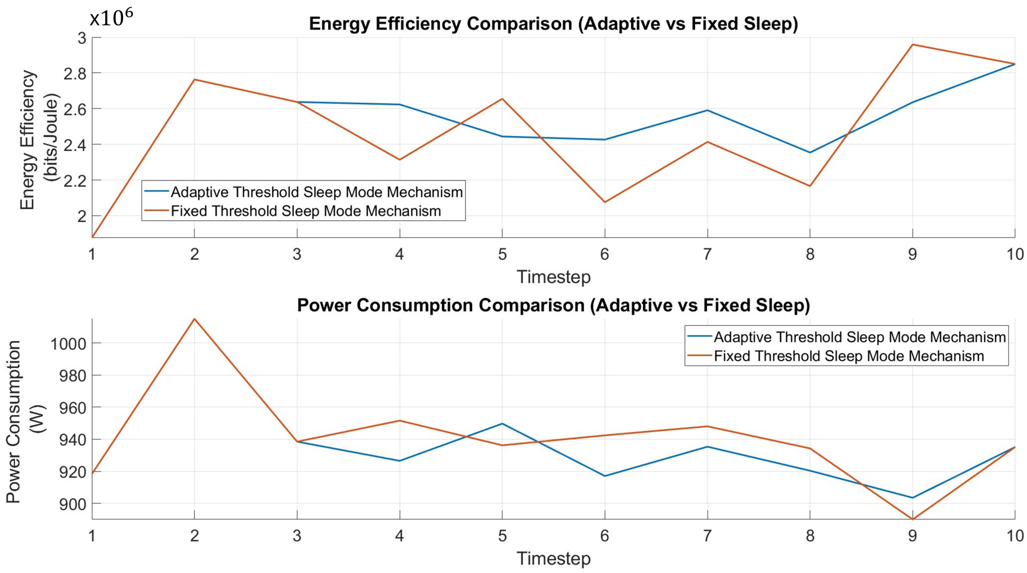


Figure 4.10: Performance comparison of sleep mode algorithms in terms of Energy Efficiency and power consumption for bursty traffic

Figure 4.10 compares the performance of the adaptive threshold sleep mode mechanism with fixed threshold sleep mode mechanism over different small cell loads. As it can be seen from both the subplots, the adaptive mechanism maintains a relatively stable performance throughout the simulation, avoiding sharp fluctuations. In contrast, the fixed threshold mechanism exhibits significant variability, with peaks at certain timesteps (timesteps 2 and 9) but also noticeable drops (timesteps 4 and 6). While the fixed threshold approach occasionally achieves higher efficiency than the adaptive mechanism, the gains are inconsistent.

The lower subplot represents the corresponding power consumption for both the algorithms. The adaptive strategy generally consumes less power during most timesteps. Overall, the results highlight that the adaptive threshold mechanism provides more consistent performance across varying network conditions whereas the fixed threshold algorithm may yield occasional short-term gains but lacks stability. This consistency is particularly advantageous in maintaining the required QoS while ensuring energy savings in dynamic network environments.

Table 4.3 presents the simulation results for ten consecutive timesteps, showing the load percentage experience by each small cell (SC1 to SC5) along with the corresponding number of small cells placed in sleep mode and the number of transitions for both adaptive and fixed threshold sleep mode mechanism. The table also includes cumulative totals for the number of sleep instances and the overall power consumption for each approach.

The data points highlighted in yellow in Table 4.3 show the differences in the number of small cell connections or the number of transitions for both the algorithms. The data points that are not highlighted in the 'Proposed Adaptive Algorithm' and 'Fixed Threshold Algorithm' show that both the algorithms respond in the exact same way to the loads at the corresponding timesteps. The cells highlighted in green and red color show the results obtained in terms of number of small cells in sleep mode, number of transitions of small cell modes, and the network power consumption for proposed adaptive threshold sleep mode mechanism and fixed threshold sleep mode mechanism respectively. Focusing on timestep 4, small cells SC3 and SC4 are in sleep mode since the load is below 40%. However, for the adaptive mechanism, SC1 is also in sleep mode since the decision depends not only on the current load but also on the previous load value. For timestep 3, SC1 is in sleep mode. However, for SC1 to turn on in timestep 4, its score computed as per Equation 3.4.1 must be greater than the threshold which also includes the hysteresis margin. This condition is not satisfied due to which the SC1 remains in sleep mode even when the load is 50%. Consequently, SC5 is in on mode since it was already on in timestep 3. Thus, the adaptive mechanism prolongs the state of a small cell in order to avoid multiple switching of the small cell between on and sleep states. Although the fixed threshold algorithm also incorporates a hysteresis margin of 20%, it performs more transitions than the adaptive algorithm since the previous state is not taken into consideration.

Moreover, to make a fair comparison between the two algorithms, the total number of small cells in sleep mode, the total number of transitions between steps and total network power consumption for all timesteps is considered. Thus, the adaptive algorithm has more small cells in sleep modes as well as performs less number of transitions between on and sleep states as compared to the fixed threshold mechanism. This leads to a reduction in power consumption of approximately 100 W.

Similar to bursty traffic, the same comparison between two algorithms is done for gradually increasing small cell load which can be seen in Figure 4.11. The small cell load is gradually increased from 10% to 90% at every timestep. At timestep 5, when the load on all small cells is 50%, all small cells turn on from sleep mode for the fixed threshold algorithm, leading to a

Timestep	Load Percentage					Proposed Adaptive Algorithm		Fixed Threshold Algorithm	
	SC 1	SC 2	SC 3	SC 4	SC 5	N_sleep	N_switch	N_sleep	N_switch
1	10	10	10	10	10	5	5	5	5
2	90	90	90	90	90	0	5	0	5
3	10	50	40	70	40	1	1	1	1
4	50	90	30	10	50	3	2	2	3
5	90	20	80	10	30	3	4	3	3
6	10	60	10	50	50	4	3	2	5
7	80	10	70	50	30	3	3	2	4
8	10	60	10	40	40	4	3	3	3
9	10	30	40	50	60	4	2	3	2
10	90	70	10	90	30	2	4	2	3
Total						29	32	23	34
Total PC (W)						9407.7		9507.6	
N_sleep represents number of small cells in sleep mode at each timestep									
N_switch represents number of transitions of small cells to and from sleep mode at each timestep									
Total PC represents the total power consumed by the network for all timesteps									

Table 4.3: Results comparing adaptive and fixed threshold sleep mode algorithms in terms of load distribution, small cell activity, transitions, and total power usage

sudden spike in power consumption due to inclusion of switching power. However, the adaptive algorithm gradually performs this transition leading to lower spikes in power consumption. Thus at lower loads (till 40%) and higher loads (above 80%), the results produced by both algorithms is the same. The total power consumption of the network for all timesteps for adaptive algorithm is 9203.6 W and that for the fixed threshold algorithm is 9268.4 W. This difference in power consumption is caused since the adaptive algorithm lets cells be in sleep mode for longer times than the fixed threshold algorithm. This comparison reveals that the adaptive algorithm outperforms fixed threshold algorithm in terms of both EE and power consumption for bursty as well as steadily changing traffic.

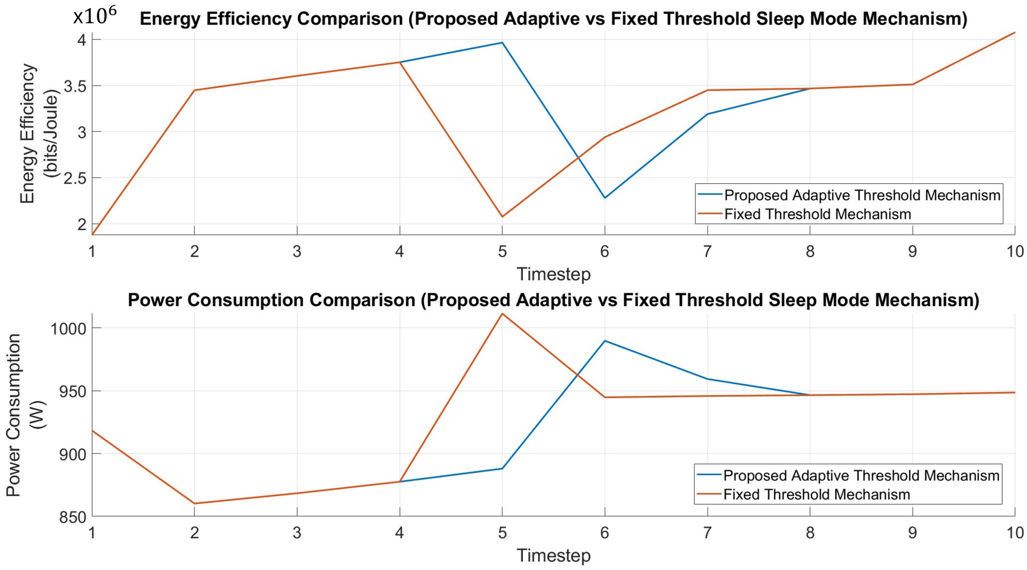


Figure 4.11: Performance comparison of sleep mode algorithms in terms of Energy Efficiency and power consumption for gradually increasing traffic

4.5. Summary

This chapter reports the quantitative evaluation of the proposed sleep control and UE - BS association against fixed-threshold and strongest-cell baselines across urban scenarios with realistic traffic demands and mobility. It presents results for total power consumption, energy efficiency, and energy per bit, alongside service quality indicators such as the number of UEs meeting QoS targets, small cell utilization, and sleep/wake transitions.

The initial results show how the user distribution and small cell placements strongly affect the performance of the network. User distribution and small cell placement are interdependent. The spatial demand determines where capacity is needed and placement decides whether small cells actually serve that demand. True small cell potential is unlocked when deployment and activation aligns with persistent user hotspots.

The experiments integrating the proposed strategies show that coupling association with adaptive sleep control reduces network power and energy per bit in moderate-to-high load and hotspot conditions while maintaining a higher count of QoS-satisfied UEs. Under very light load, the gains are narrow as small-cell idle overheads dominate.

5
Conclusion

This chapter presents the thesis' conclusions, limitations and recommendations for future research. The answers to the research questions and sub-questions that are formulated in section 1.4 are summarized in section 5.1 and the limitations are presented in section 5.2. This chapter concludes with section 5.3 providing directions for future research.

5.1. Conclusions

This thesis addresses energy consumption in the 5G radio access network by focusing on how small cells are deployed and adaptively managed in a heterogeneous network. A MATLAB simulation framework evaluates the impact of small-cell placement, user distribution, and adaptive control on power use and energy efficiency. We study the theoretical motivations for small cells as traffic offloaders from the macro layer, review power-consumption models and minimization strategies, and then build a tractable system model that we can use to test algorithms under realistic conditions.

Power models in the literature range from simple baseline-plus-load formulations to detailed component models that separate baseband, RF, and power-amplifier contributions. Although detailed component values exist in vendor data, they are not publicly available in a consistent form, and embedding an exhaustive model into a system-level simulator adds complexity with limited benefit for our questions. The thesis therefore uses a common core model for both macro and small cells with a baseline and a load-dependent term, and we account explicitly for the energy and delay overheads of transitions between sleep and active states. The topology of the RAN still matters: where functions are placed determines what each base station processes, which in turn influences its power.

The thesis adopts an urban multi-cell scenario because small cells are designed to increase capacity, especially where user density and obstruction are high. Hotspots such as city centers and train stations dominate demand, so we include heterogeneous per-user throughput demands and model pedestrian mobility. Unlike many prior studies that assume idealized grids, no interference, homogeneous traffic, or static users, our evaluation uses a realistic topology with interference, realistic demand distributions, mobility, and switching overheads. The x-haul energy consumption is excluded on purpose: transport-network design and topology optimization are a large topic of their own, so the transport network is treated as given and outside our energy accounting.

The results show that user distribution and small-cell placement are interdependent. When small cells are placed where hotspots persist and are activated only when needed, they carry more of the load, total power and energy per bit fall, and more users meet QoS. If placement is misaligned with demand, small cells sit idle while the macro layer is overloaded, which raises power and harms QoS. We also observe a clear break-even: at very light load, macro-only or deep sleep is preferable because idle and switching costs dominate; as user density or per-user demand rises, the advantage shifts to the heterogeneous network once the savings from shorter links outweigh the fixed costs.

In city-like multi-cell networks, the proposed method cuts total power consumption and energy per bit especially at medium and high loads and in hotspot regions. It does this by sending heavy users to nearby small cells so that the transmit power drastically reduces and letting idle small cells sleep until the loads increase.

At low loads, small cells are not worth waking since idle power consumption dominates the network power consumption without increasing the network throughput. As user density or throughput demand increases, serving bits on short links saves energy. Moreover, the proposed sleep mode mechanism avoids waking up cells even after a certain threshold, until the QoS of any user is compromised. Thus, it lets the small cell stay in sleep mode for longer time than the fixed threshold mechanism, thereby saving more energy.
The proposed association algorithm, STAA, performs better than the max–SINR, CRE, and LHLA algorithms in terms of energy efficiency as well as power consumption, without compromising QoS of any users. It still relies on the signal strength as the primary factor for association. However, the manner in which the users are assigned to base stations ensures efficient allocation of base station resources. Moreover, the proposed association algorithm has execution time comparable to that of the max-SINR algorithm, which makes it deployable in realistic scenarios.

In conclusion, the thesis has made contributions to...

5.2. Limitations of this thesis

The following are limitations of this thesis:

- One of the main limitations is that although the system model includes mobility of the users, the mobility parameter is not involved in the decision making of the proposed sleep mode mechanism or the association algorithm.
- The thesis does not incorporate association of a user to multiple base stations, which is also known as multi–connectivity. Without dual or multi-connectivity, the system cannot split traffic accross base stations for robustness, smoother handovers or energy-aware load sharing.
- The data rates are computed from SINR–rate mapping. Beamforming, precoding and other functionalities are not simulated. This simplifies evaluation but can over- or underestimate achievable rates.
- The x-haul network capacity, latency, topology, and energy are treated as given. The x-haul network topology and technology could provide a more realistic understanding of the power consumption and latency.
- Moreover, this thesis proposes that the decision-making take place in at the network side,

precisely, at the TDO. However, the control signaling overheads caused due to this are not considered.

5.3. Recommendations for future research

While the proposed strategies demonstrate promising energy savings in 5G HetNets, several directions remain open for future exploration. The foremost extensions could be to create a more realistic system model by incorporating the following:

1. Model 3D geometry by using BS/UE heights and true 3D distances (and 3D path loss/blockage).
2. Use map-based deployment with obstacles (e.g., buildings) to place small cells.
3. Co-model the x-haul for capacity, latency, topology, and energy, and make sleep/association transport-aware.
4. Add dual-/multi-connectivity (and carrier aggregation) to split traffic for robustness and energy/QoS trade-offs.
5. Introduce spectrum assignment/reuse (e.g., graph/game-theoretic channel selection among neighbors) to mitigate interference. One way to do this is using game theory considering only a few neighboring cells. The cells compete with each other such that they select different frequency bands to avoid interference.
6. Consider different mobility models to incorporate vehicles and pedestrians, with dwell-time or handover-risk prediction in the control loop.
7. Include modulation schemes and transmit power control mechanisms.

Another promising direction is to develop machine-learningbased sleep/association controllers on these more realistic, large-scale models and assess not only energy efficiency but also computational cost, stability, and deployability in live networks.

References

[1] F. Kooshki, A. G. Armada, M. M. Mowla, A. Flizikowski, and S. Pietrzyk, "Energy-efficient sleep mode schemes for cell-less ran in 5g and beyond 5g networks," *IEEE Access*, vol. 11, pp. 1432–1444, 2023. DOI: 10.1109/ACCESS.2022.3233430.

[2] S. Pal, "The potential of small cell deployment in theory and practice," Available at TU Delft Repository: https://resolver.tudelft.nl/87a280f3-72d0-4516-aa3a-17fb4b3f0484, MSc thesis, Delft University of Technology, Delft, The Netherlands, Oct. 2024.

[3] J. Zong, Y. Liu, H. Liu, Q. Wang, and P. Chen, "6g cell-free network architecture," in *2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)*, 2022, pp. 421–425. DOI: 10.1109/ICETCI55101.2022.9832308.

[4] H. A. Ammar, R. Adve, S. Shahbazpanahi, G. Boudreau, and K. V. Srinivas, "User-centric cell-free massive mimo networks: A survey of opportunities, challenges and solutions," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 611–652, 2022. DOI: 10.1109/COMST.2021.3135119.

[5] L. M. P. Larsen, A. Checko, and H. L. Christiansen, "A survey of the functional splits proposed for 5g mobile crosshaul networks," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 146–172, 2019. DOI: 10.1109/COMST.2018.2868805.

[6] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, *Cell-free massive mimo versus small cells*, 2017. arXiv: 1602.08232 [cs.IT]. [Online]. Available: https://arxiv.org/abs/1602.08232.

[7] 3rd Generation Partnership Project (3GPP), "Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; 5G; Vocabulary for 3GPP Specifications," ETSI, Technical Report (TR) ETSI TR 121 905 (3GPP TR 21.905), version V17.1.0, May 2022, Release 17. [Online]. Available: https://www.etsi.org/deliver/etsi_tr/121900_121999/121905/17.01.00_60/tr_121905v170100p.pdf.

[8] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014, ISSN: 1558-1896. DOI: 10.1109/mcom.2014.6736761. [Online]. Available: http://dx.doi.org/10.1109/MCOM.2014.6736761.

[9] L. M. Larsen, S. Ruepp, M. S. Berger, and H. L. Christiansen, "Ran design guidelines for energy efficient 5g mobile xhaul networks," in *2022 14th International Conference on Communications (COMM)*, 2022, pp. 1–6. DOI: 10.1109/COMM54429.2022.9817244.

[10] P. Frenger and R. Tano, "More capacity and less power: How 5g nr can reduce network energy consumption," in *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*, 2019, pp. 1–5. DOI: 10.1109/VTCSpring.2019.8746600.

[11] G. Insights, *Market data: Distributed generation and energy storage in telecom networks*, Retrieved October 28, 2024, from https://www.businesswire.com/news/home/20210209005153/en/Guidehouse-Insights-Report-Finds-Telecom-Networks-Are-Expected-to-Install-122-GW-of-New-Distributed-Generation-and-Distributed-Energy-Storage-Capacity-from-2021-2030, n.d.

[12] Ericsson, "Ericsson mobility report, november 2019," Ericsson, Tech. Rep., 2019, Accessed: 2025-01-23. [Online]. Available: https://www.ericsson.com/4ac657/assets/local/reports-papers/mobility-report/documents/2019/emr-november-2019.pdf.

[13] Ericsson, "Ericsson mobility report, november 2024," Ericsson, Tech. Rep., 2024, Accessed: 2025-01-23. [Online]. Available: https://www.ericsson.com/4adb7e/assets/local/reports-papers/mobility-report/documents/2024/ericsson-mobility-report-november-2024.pdf.

[14] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2522–2545, 2016. DOI: 10.1109/COMST.2016.2571730.

[15] 3GPP, "Nr; overall description; stage-2," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 138 300, version V15.8.0, 2020. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138300_138399/138300/15.08.00_60/ts_138300v150800p.pdf.

[16] NGMN Alliance, *Network energy efficiency phase 2*, Accessed: 2025-01-23, n.d. [Online]. Available: https://www.ngmn.org/publications/network-energy-efficiency-phase-2.html.

[17] A. M. Busch, K. Eger, and B. Richerzhagen, "Comparison of power consumption models for 5g cellular network base stations," in *International Symposium on Intelligent and Distributed Computing*, Springer, 2023, pp. 199–212.

[18] X. Ge, J. Yang, H. Gharavi, and Y. Sun, "Energy efficiency challenges of 5g small cell networks," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 184–191, 2017. DOI: 10.1109/MCOM.2017.1600788.

[19] D. Ha, K. Lee, and J. Kang, "Energy efficiency analysis with circuit power consumption in massive mimo systems," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2013, pp. 938–942. DOI: 10.1109/PIMRC.2013.6666272.

[20] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user mimo systems: Is massive mimo the answer?" *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3059–3075, 2015. DOI: 10.1109/TWC.2015.2400437.

[21] G. Auer, V. Giannini, C. Desset, *et al.*, "How much energy is needed to run a wireless network?" *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40–49, 2011. DOI: 10.1109/MWC.2011.6056691.

[22] K. Ryu and W. Kim, "Multi-objective optimization of energy saving and throughput in heterogeneous networks using deep reinforcement learning," *Sensors*, vol. 21, no. 23, 2021, ISSN: 1424-8220. DOI: 10.3390/s21237925. [Online]. Available: https://www.mdpi.com/1424-8220/21/23/7925.

[23] H. Fourati, R. Maaloul, N. Trabelsi, L. Chaari, and M. Jmaiel, "An efficient energy saving scheme using reinforcement learning for 5g and beyond in h-cran," *Ad Hoc Networks*, vol. 155, p. 103 406, 2024, ISSN: 1570-8705. DOI: https://doi.org/10.1016/j.adhoc.2024.103406. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570870524000179.

[24] N. Piovesan, D. López-Pérez, A. De Domenico, X. Geng, H. Bao, and M. Debbah, "Machine learning and analytical power consumption models for 5g base stations," *IEEE Communications Magazine*, vol. 60, no. 10, pp. 56–62, 2022. DOI: 10.1109/MCOM.001.2200023.

[25] M. Jaber, M. A. Imran, R. Tafazolli, and A. Tukmanov, "5g backhaul challenges and emerging research directions: A survey," *IEEE Access*, vol. 4, pp. 1743–1766, 2016. DOI: 10.1109/ACCESS.2016.2556011.

[26] M. M. Mowla, I. Ahmad, D. Habibi, and Q. V. Phung, "Energy efficient backhauling for 5g small cell networks," *IEEE Transactions on Sustainable Computing*, vol. 4, no. 3, pp. 279–292, 2019. DOI: 10.1109/TSUSC.2018.2838116.

[27] K. Venkateswararao, P. Swain, S. S. Jha, I. Ioannou, and A. Pitsillides, "A novel power consumption optimization framework in 5g heterogeneous networks," *Computer Networks*, vol. 220, p. 109 487, 2023, ISSN: 1389-1286. DOI: https://doi.org/10.1016/j.comnet.2022.109487. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128622005217.

[28] H. Park and Y. Lim, "Reinforcement learning for energy optimization with 5g communications in vehicular social networks," *Sensors*, vol. 20, no. 8, 2020, ISSN: 1424-8220. DOI: 10.3390/s20082361. [Online]. Available: https://www.mdpi.com/1424-8220/20/8/2361.

[29] imec, *Interuniversity Microelectronics Centre*, 2024. [Online]. Available: https://www.imec.be/nl.

[30] M. Gruber, O. Blume, D. Ferling, D. Zeller, M. A. Imran, and E. C. Strinati, "Earth energy aware radio and network technologies," in *2009 IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications*, 2009, pp. 1–5. DOI: 10.1109/PIMRC.2009.5449938.

[31] B. Debaillie, C. Desset, and F. Louagie, "A flexible and future-proof power model for cellular base stations," in *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, 2015, pp. 1–7. DOI: 10.1109/VTCSpring.2015.7145603.

[32] 3rd Generation Partnership Project (3GPP), "3GPP TR 38.864: Study on Network Energy Savings for NR," 3GPP, Tech. Rep. V1.0.0, 2022, Accessed: 2024-02-13. [Online]. Available: https://www.3gpp.org/ftp/Specs/archive/38_series/38.864/.

[33] B. Shen, Z. Lei, X. Huang, and Q. Chen, "An interference contribution rate based small cells on/off switching algorithm for 5g dense heterogeneous networks," *IEEE Access*, vol. 6, pp. 29 757–29 769, 2018. DOI: 10.1109/ACCESS.2018.2841044.

[34] M. Osama, S. El Ramly, and B. Abdelhamid, "Interference mitigation and power minimization in 5g heterogeneous networks," *Electronics*, vol. 10, no. 14, 2021, ISSN: 2079-9292. DOI: 10.3390/electronics10141723. [Online]. Available: https://www.mdpi.com/2079-9292/10/14/1723.

[35] D. Sudhakar and D. Sunehra, "A traffic-aware and power-saving base station switch off method for cellular networks," in *2022 7th International Conference on Communication and Electronics Systems (ICCES)*, 2022, pp. 12–15. DOI: 10.1109/ICCES54183.2022.9835851.

[36] F. E. Salem, A. Gati, Z. Altman, and T. Chahed, "Advanced sleep modes and their impact on flow-level performance of 5g networks," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, 2017, pp. 1–7. DOI: 10.1109/VTCFall.2017.8288125.

[37] X. Ma, Y. Mu, Z. Liu, X. Jiang, J. Zhang, and Y. Gao, "Energy consumption optimization of 5g base stations considering variable threshold sleep mechanism," *Energy Reports*, vol. 9, pp. 34–42, 2023, 2023 the 7th International Conference on Energy and Environmental Science, ISSN: 2352-4847. DOI: https://doi.org/10.1016/j.egyr.2023.04.026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352484723003876.

[38] S. Habibi, V. Solouk, and H. Kalbkhani, "Adaptive energy-efficient small cell sleeping and zooming in heterogeneous cellular networks," *Telecommunication Systems*, vol. 77, pp. 23–45, 2021.

[39] M. A. Wiering and M. Van Otterlo, "Reinforcement learning," *Adaptation, learning, and optimization*, vol. 12, no. 3, p. 729, 2012.

[40] MathWorks. "What is reinforcement learning?" Reinforcement Learning Toolbox Documentation. (2025), [Online]. Available: https://nl.mathworks.com/help/reinforcement-learning/ug/what-is-reinforcement-learning.html (visited on 08/14/2025).

[41] J. A. Ayala-Romero, J. J. Alcaraz, A. Zanella, and M. Zorzi, "Online learning for energy saving and interference coordination in hetnets," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1374–1388, 2019. DOI: 10.1109/JSAC.2019.2904362.

[42] A. Foivos, B. Christos, and K. Vasileios, "A machine learning mechanism for improved energy consumption in 5g small cells," in *2022 IEEE Symposium on Computers and Communications (ISCC)*, 2022, pp. 1–6. DOI: 10.1109/ISCC55528.2022.9913060.

[43] F. E. Salem, Z. Altman, A. Gati, T. Chahed, and E. Altman, "Reinforcement learning approach for advanced sleep modes management in 5g networks," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, 2018, pp. 1–5. DOI: 10.1109/VTCFall.2018.8690555.

[44] F. E. Salem, T. Chahed, Z. Altman, and A. Gati, "Traffic-aware advanced sleep modes management in 5g networks," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, 2019, pp. 1–6. DOI: 10.1109/WCNC.2019.8886051.

[45] M. Masoudi, M. G. Khafagy, E. Soroush, D. Giacomelli, S. Morosi, and C. Cavdar, "Reinforcement learning for traffic-adaptive sleep mode management in 5g networks," in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, 2020, pp. 1–6. DOI: 10.1109/PIMRC48278.2020.9217286.

[46] S. Malta, P. Pinto, and M. Fernández-Veiga, "Using reinforcement learning to reduce energy consumption of ultra-dense networks with 5g use cases requirements," *IEEE Access*, vol. 11, pp. 5417–5428, 2023. DOI: 10.1109/ACCESS.2023.3236980.

[47] A. El-Amine, H. A. Haj Hassan, M. Iturralde, and L. Nuaymi, "Location-aware sleep strategy for energy-delay tradeoffs in 5g with reinforcement learning," in *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2019, pp. 1–6. DOI: 10.1109/PIMRC.2019.8904155.

[48] E. Altman, A. Kumar, C. Singh, and R. Sundaresan, "Spatial sinr games of base station placement and mobile association," *IEEE/ACM Transactions on Networking*, vol. 20, no. 6, pp. 1856–1869, 2012. DOI: 10.1109/TNET.2012.2186980.

[49] P. Bhuvaneswari, S. Indu, N. L. Shifana, D. Arjun, and A. S. Priyadharshini, "An analysis on cell range expansion in 4g lte networks," in *2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN)*, 2015, pp. 1–6. DOI: 10.1109/ICSCN.2015.7219824.

[50] S. Achki, L. Aziz, F. Gharnati, and A. Ait Ouahman, "User association strategy for energy efficiency and interference mitigation of heterogeneous networks," *Advances in Materials Science and Engineering*, vol. 2020, no. 1, p. 7 018 727, 2020.

[51] N. Hassan and X. Fernando, "An optimum user association algorithm in heterogeneous 5g networks using standard deviation of the load," *Electronics*, vol. 9, no. 9, 2020, ISSN: 2079-9292. DOI: 10.3390/electronics9091495. [Online]. Available: https://www.mdpi.com/2079-9292/9/9/1495.

[52] M. M. H. Alkalsh and A. Kliks, "Enhancing wireless network efficiency with the techniques of dynamic distributed load balancing: A distance-based approach," *Sensors*, vol. 24, no. 16, 2024, ISSN: 1424-8220. DOI: 10.3390/s24165406. [Online]. Available: https://www.mdpi.com/1424-8220/24/16/5406.

[53] E. Gures, I. Shayea, M. Sheikh, M. Ergen, and A. A. El-Saleh, "Adaptive cell selection algorithm for balancing cell loads in 5g heterogeneous networks," *Alexandria Engineering Journal*, vol. 72, pp. 621–634, 2023, ISSN: 1110-0168. DOI: https://doi.org/10.1016/j.aej.2023.04.012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1110016823002855.

[54] A. Mesodiakaki, F. Adelantado, L. Alonso, and C. Verikoukis, "Energy-efficient context-aware user association for outdoor small cell heterogeneous networks," in *2014 IEEE International Conference on Communications (ICC)*, 2014, pp. 1614–1619. DOI: 10.1109/ICC.2014.6883553.

[55] I. A. Alablani and M. A. Arafah, "An adaptive cell selection scheme for 5g heterogeneous ultra-dense networks," *IEEE Access*, vol. 9, pp. 64 224–64 240, 2021. DOI: 10.1109/ACCESS.2021.3075324.

[56] I. A. Alablani and M. A. Arafah, "Enhancing 5g small cell selection: A neural network and iov-based approach," *Sensors*, vol. 21, no. 19, 2021, ISSN: 1424-8220. DOI: 10.3390/s21196361. [Online]. Available: https://www.mdpi.com/1424-8220/21/19/6361.

[57] I. AlQerm and B. Shihada, "Energy efficient traffic offloading in multi-tier heterogeneous 5g networks using intuitive online reinforcement learning," *IEEE Transactions on Green Communications and Networking*, vol. 3, no. 3, pp. 691–702, 2019. DOI: 10.1109/TGCN.2019.2916900.

[58] Z. Li, K. Zhu, R. Wang, and Y. Xu, "Context-aware decoupled multiple association in ultra-dense networks," in *2018 IEEE Global Communications Conference (GLOBECOM)*, 2018, pp. 1–6. DOI: 10.1109/GLOCOM.2018.8647512.

[59] M. Saimler and S. Coleri, "Multi-connectivity based uplink/downlink decoupled energy efficient user association in 5g heterogenous cran," *IEEE Communications Letters*, vol. 24, no. 4, pp. 858–862, 2020. DOI: 10.1109/LCOMM.2020.2967050.

[60] A. Mesodiakaki, E. Zola, R. Santos, and A. Kassler, “Optimal user association, backhaul routing and switching off in 5g heterogeneous networks with mesh millimeter wave backhaul links,” *Ad Hoc Networks*, vol. 78, pp. 99–114, 2018, ISSN: 1570-8705. DOI: https://doi.org/10.1016/j.adhoc.2018.05.008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1570870518302397.

[61] E. Zola and I. Martin-Escalona, “A robust user association, backhaul routing, and switching off model for a 5g network with variable traffic demands,” *IEEE Access*, vol. 8, pp. 96 714–96 726, 2020. DOI: 10.1109/ACCESS.2020.2992330.

[62] M. Javad-Kalbasi, Z. Naghsh, M. Mehrjoo, and S. Valaee, “A new heuristic algorithm for energy and spectrum efficient user association in 5g heterogeneous networks,” in *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, 2020, pp. 1–7. DOI: 10.1109/PIMRC48278.2020.9217199.

[63] M. A. Habibi, M. Nasimi, B. Han, and H. D. Schotten, “A comprehensive survey of ran architectures toward 5g mobile communication system,” *IEEE Access*, vol. 7, pp. 70 371–70 421, 2019. DOI: 10.1109/ACCESS.2019.2919657.

[64] M. Polese, L. Bonati, S. DOro, S. Basagni, and T. Melodia, “Understanding o-ran: Architecture, interfaces, algorithms, security, and research challenges,” *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376–1411, 2023. DOI: 10.1109/COMST.2023.3239220.

[65] Huawei, *Huawei’s ultra-lean site series accelerates scaled 5g deployment*, Accessed: 14-07-2025.

[66] Ericsson, *Massive mimo for high-performing networks*, https://www.ericsson.com/en/ran/massive-mimo/solutions, Accessed: 14-07-2025.

[67] C. Serrano, M. Diaz, A. Gomez, M. Arranz, and S. Tenorio, “Active antenna - efficient mobile broadband networks,” in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, 2011, pp. 1–4. DOI: 10.1109/VETECS.2011.5956751.

[68] 3rd Generation Partnership Project (3GPP), *TS 36.304 V16.3.0: Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) procedures in idle mode*, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2425, Release 16, Mar. 2020.

[69] 3rd Generation Partnership Project (3GPP), *TS 38.331 V16.3.1: NR; Radio Resource Control (RRC); Protocol specification*, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3197, Release 16, Jan. 2021.

[70] M. Fall, Y. Balboul, M. Fattah, S. Mazer, M. E. Bekkali, and A. D. Kora, “Towards sustainable 5g networks: A proposed coordination solution for macro and pico cells to optimize energy efficiency,” *IEEE Access*, vol. 11, pp. 50 794–50 804, 2023. DOI: 10.1109/ACCESS.2023.3278209.

[71] C. Bouras, A. Gkamas, G. Diles, and Z. Andreas, “A comparative study of 4g and 5g network simulators,” *International Journal on Advances in Networks and Services*, vol. 13, no. 1, 2020.

[72] P. K. Gkonis, P. T. Trakadas, and D. I. Kaklamani, “A comprehensive study on simulation techniques for 5g networks: State of the art results, analysis, and future challenges,” *Electronics*, vol. 9, no. 3, 2020, ISSN: 2079-9292. DOI: 10.3390/electronics9030468. [Online]. Available: https://www.mdpi.com/2079-9292/9/3/468.

[73] Network Simulation Tools, *Python 5g simulation*, https://networksimulationtools.com/python-5g-simulation/, Accessed: 2025-07-24, 2023.

[74] LinkedIn Contributors, *What are some of the most useful libraries and tools for simulation in matlab and python?* https://www.linkedin.com/advice/1/what-some-most-useful-libraries-tools-simulation, Accessed: 2025-07-24, 2024.

[75] SmoothTalker, *Internet speed requirements for different system apps*, https://www.smoothtalker.com/internet-speed-requirements-for-different-system-apps/, Accessed: July 29, 2025, 2024.

[76] Zoom Support, *Zoom system requirements: Windows, macos, linux*, https://support.zoom.com/hc/en/article?id=zm_kb&sysparm_article=KB0060748, Accessed: July 29, 2025, 2024.

[77] Microsoft Docs, *Prepare your organization's network for microsoft teams*, https://learn.microsoft.com/en-us/microsoftteams/prepare-network, Accessed: July 29, 2025, 2024.

[78] YouTube Support, *System requirements & supported devices for youtube*, https://support.google.com/youtube/answer/78358?hl=en, Accessed: July 29, 2025, 2024.

[79] Netflix, *Internet connection speed recommendations*, https://help.netflix.com/en/node/306, Accessed: July 29, 2025, 2023.

[80] M. Younes and Y. Louet, "Energy consumption management under quality-of-service constraints in the context of small cell networks," in *2022 29th International Conference on Systems, Signals and Image Processing (IWSSIP)*, vol. CFP2255E-ART, 2022, pp. 1–4. DOI: 10.1109/IWSSIP55020.2022.9854453.

[81] A. Alizadeh and M. Vu, "Time-fractional user association in millimeter wave mimo networks," in *2018 IEEE International Conference on Communications (ICC)*, 2018, pp. 1–6. DOI: 10.1109/ICC.2018.8422540.

A

Pseudo Code of the proposed sleep mode mechanism

Algorithm: Adaptive Sleep Mode Decision Function

function SLEEPMODE(state, num_sc, load, epb, sinr_sc, unsupported_ues, num_ue, w1, w2, pc_total, eb_real, ue_dr, mcload, enable_sleep, previous_load, previous_power, previous_epb)
    $ideal\_l \leftarrow 0.4$
    $ideal\_epb \leftarrow 0.1$
    Convert epb and $previous\_epb$ to μJ/bit using num_ue
    $epb\_max \leftarrow 0.2$
    $p_{target} \leftarrow 0.6 \cdot p_{max} + 0.4 \cdot p_{min}$
    for $i = 1$ to num_sc **do**
        if $|load[i] - previous\_load[i]| \geq 0.1$ **then**
            if $unsupported\_ues > 0$ **or** $load[i] > ideal\_l$ **or** $mcload > 0.7$ **then**
                $w1[i] \leftarrow \min(w1[i] + 0.1,\ 1.0)$
            end if
        else if $pc\_total > previous\_power$ **or** $epb > previous\_epb$ **then**
            if $epb > ideal\_epb$ **or** $pc\_total > p_{target}$ **then**
                $w1[i] \leftarrow \max(w1[i] - 0.1,\ 0.0)$
            end if
        end if
        $w2[i] \leftarrow 1 - w1[i]$
        $score\_threshold[i] \leftarrow w1[i] \cdot ideal\_l + w2[i] \cdot \frac{p_{min}}{p_{max}}$
    end for
    Initialize $score$ as a zero vector of length num_sc
    if $enable\_sleep == 1$ **then**
        for $i = 1$ to num_sc **do**
            $score[i] \leftarrow w1[i] \cdot load[i] + w2[i] \cdot \frac{pc\_total}{p_{max}}$
            if $score[i] < score\_threshold[i]$ **then**
                $state[i] \leftarrow 1$ ▷ Sleep mode
                $sinr\_sc[:, i] \leftarrow 0$
            else if $score[i] > 1.2 \cdot score\_threshold[i]$ **then**
                $state[i] \leftarrow 0$ ▷ Active mode
            end if

```
        end for
    end if
    return state, score, sinr_sc, score_threshold, w1, w2
end function
```

B

Pseudo Code of the proposed sleep mode mechanism

Algorithm 1 UE–BS Association with CRE and Throughput-Aware Scheduling

function NewAssociationAlgorithm(num_ue, num_sc, prbs_required, ue_throughput_demands, sinr_sc, sinr_mc, ue_enb_index, ue_dr, bw_prb, downlink_subcarriers, mc_xpos, mc_ypos, sc_xpos, sc_ypos, ue_xpos, ue_ypos, mc_R, sc_R, prb_mc, prb_sc)
    $sc\_connections \leftarrow 0$, $mc\_connections \leftarrow 0$, $unsupported\_ues \leftarrow 0$
    $association \leftarrow 0$, $sinr\_connected\_ue \leftarrow 0$, $bw\_allotted \leftarrow 0$
    $available\_prbs \leftarrow [prb\_sc, prb\_mc]$, $bias\_factor \leftarrow 4$ ▷ 6 dB CRE bias
    $low\_sinr\_ues \leftarrow \{u \mid sinr\_mc[u] < 1\}$
    for u **in** low_sinr_ues **do**
        $(sinr, sc\_idx) \leftarrow \max(sinr\_sc[u] \cdot bias\_factor)$
        $prb\_demand \leftarrow prbs\_required[u]$
        if $sinr > sinr\_mc[u]$ **and** $available\_prbs[sc\_idx] \geq prb\_demand$ **then**
            Assign u to sc_idx, update PRBs, SINR, throughput
        else
            $unsupported\_ues \leftarrow unsupported\_ues + 1$
        end if
    end for
    $remaining\_ues \leftarrow$ all UEs $\setminus$ low_sinr_ues
    Sort $remaining\_ues$ by $ue\_throughput\_demands$ (desc.)
    for u **in** $remaining\_ues$ **do**
        $(sinr, sc\_idx) \leftarrow \max(sinr\_sc[u] \cdot bias\_factor)$
        $prb\_demand \leftarrow \lceil \frac{ue\_throughput\_demands[u]}{bw\_prb \log_2(1+sinr)} \rceil$
        if $sinr > sinr\_mc[u]$ **and** $available\_prbs[sc\_idx] \geq prb\_demand$ **then**
            Assign u to sc_idx
        else if $available\_prbs[mc] \geq \lceil \frac{ue\_throughput\_demands[u]}{bw\_prb \log_2(1+sinr\_mc[u])} \rceil$ **then**
            Assign u to macrocell
        else
            $unsupported\_ues \leftarrow unsupported\_ues + 1$
        end if
    end for
    return $association$, $sc\_connections$, $mc\_connections$, ue_dr
end function