Addis Ababa University
Addis Ababa Institute of Technology
School of Electrical and Computer Engineering
Telecommunication Network Engineering Graduate Program

Traffic-Aware Band-Level Cells On Off for Energy Saving in LTE-Advanced Networks with Inter-Band Carrier Aggregation

By

Yilma Melaku

Adviser

Dr. -Ing. Dereje Hailemariam

A Thesis Submitted to the School of Graduate Studies of Addis Ababa University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Telecommunication Network Engineering

September 2023

Addis Ababa, Ethiopia

Addis Ababa University

Addis Ababa Institute of Technology

School of Electrical and Computer Engineering

Telecommunication Network Engineering Graduate Program

Traffic-Aware Band-Level Cells On Off for Energy Saving in LTE-Advanced Networks with Inter-Band Carrier Aggregation

By

Yilma Melaku

Dr. –Ing. Dereje Hailemariam	_____
Adviser	Signature
_____	_____
Internal Examiner	Signature
_____	_____
External Examiner	Signature
_____	_____
Chairperson	Signature

Declaration

I, the undersigned, declare that this thesis is my original work and does not include any content from any previous submissions for a degree or diploma at any other university or institute of higher learning, without proper acknowledgment. To the best of my knowledge, it does not contain any previously published or authored material by another individual, unless explicitly referenced within the text.

Yilma Melaku _____
Name Signature

Addis Ababa _____/_____/ 2023
Place Date of Submission

The above candidate has carried out research for the master's thesis under my supervision.

Dr. -Ing. Dereje Hailemariam _____
Name Signature

Abstract

Mobile network operators employ Carrier Aggregation (CA) in Long Term Evolution (LTE)-Advanced (LTE-A) networks to meet demand for high-rate mobile data from smartphones. CA allows users to use multiple LTE carriers, including fragmented component carriers (CCs) in different bands, called inter-band CA, which can increase bandwidth. However, inter-band CA requires more radio frequency (RF) chains that rely on inefficient power amplifiers. Our survey of real LTE-A network operated by ethio telecom, Ethiopia, found that although mains power from renewable sources, outages lead to a significant reliance on non-green diesel use. Additionally, the high daily traffic load variance observed in the survey indicates the need for adaptive solutions to achieve potential power savings.

Previous works have focused on switching on/off cells separately, transferring users to active cells (which is a non-CA scenario), or deactivating/activating CCs at the user level showing limitations to consider both network and end-user devices. This thesis proposes a novel traffic load adaptive band-level cells on/off (BLCOO) approach for the CA scenario. BLCOO optimizes the number of serving CCs to save power for RF units and user devices during off-peak hours.

The energy-saving problem is formulated as a Markov decision process (MDP) with uncertain network conditions. Deep reinforcement learning algorithms, specifically Deep Q-Networks and proximal policy optimization, are trained to solve the MDP problem for discrete and continuous evolved node B (eNB) load states. Operator data, including resource blocks usage, energy consumption, and CA configuration are used to build RF power consumption models and a custom simulation environment. The proposed algorithms are evaluated for a one-day hourly performance. The results show that, on average, 72.0% of CCs are sufficient to meet the actual traffic demand, resulting in a maximum of 18.71% and average of 14.62% reduction in RF power consumption.

***Keywords**: Inter-Band CA, DQN, Energy Saving, BLCOO, MDP, Reinforcement Learning, PPO*

Acknowledgment

First and foremost, I thank the almighty God for giving me all the help, courage, and strength to accomplish the thesis. This journey has been long and challenging.

Second, I would like to deeply thank **Dr. -Ing. Dereje Hailemariam** for his guidance, encouragement, and constructive comments throughout my thesis. It would not have been possible to complete and achieve the ultimate target of this research without his invaluable support, enthusiasm, and commitment.

I am also grateful to my examiners, **Dr. –Eng. Yihenew Wondie** and **Dr. Beneyam Berehanu,** for their helpful feedback and suggestions. Their comments have helped me to improve my thesis immeasurably.

I would like to thank my colleagues, **Anteneh Atnafu**, **Woretaw Chanie** and **Yonas Alemayehu**, for their support and encouragement. They have been a great source of ideas and inspiration, and I am lucky to have them as colleagues.

I am thankful to **Fitsum Mergia** and **Solomon Kedir** for making my working environment more enjoyable.

I am also grateful to my brother, **Endalew Melaku** for his love and support. He has always been there for me, no matter what.

I love you my daughter, **Halyot Yilma**, my front gear. You are my inspiration.

My Mom, **Ayehu Alem**, I am lucky to have you always been there to support me and catch me when I fall.

This thesis would not have been possible without the support of all of these amazing people. I am truly grateful for their contributions.

Table of Contents

List of Figures

List of Tables

List of Acronyms

3GPP	Third Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
BCCH	Broadcast Control Channel
BLCOO	Band-Level Cells On Off
CA	Carrier Aggregation
CC	Component Carrier
CQI	Channel Quality Indicator
DDPG	Deep Deterministic Policy Gradient
DL-SCH	Downlink Shared Channel
DQN	Deep Q-Network
DRL	Deep Reinforcement Learning
DRX	Discontinuous Reception
DTX	Discontinuous Transmission
eNB	Evolved Node B

E-UTRAN	Evolved Universal Terrestrial Radio Access Network
FDD	Frequency Division Duplex
GSMA	Global System for Mobile Communications Association
HARQ	Hybrid Automatic Repeat Request
HetNet	Heterogeneous Network
ICI	Inter-Cell Interference
IMT	International Mobile Telecommunication
LTE	Long Term Evolution
LTE-A	LTE-Advanced
MAC	Medium Access Control
MCS	Modulation and Coding Scheme
MDP	Markov Decision Process
MIMO	Multiple Input Multiple Output
MNO	Mobile Network Operator
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PA	Power Amplifier
PAPR	Peak-to-Average Power Ratio
PCC	Primary Component Carrier
PCell	Primary Cell
PDCCH	Physical Downlink Control Channel
PDCP	Packet Data Convergence Protocol
PDSCH	Physical Downlink Shared Channel
PPO	Proximal Policy Optimization
PRB	Physical Resource Block
PUCCH	Physical Uplink Control Channel
PUSCH	Physical Uplink Shared Channel
QoS	Quality of Service
RAT	Radio Access Terminal
RF	Radio Frequency
RL	Reinforcement Learning
RLC	Radio Link Control
RRC	Radio Resource Connection
RRU	Remote Radio Unit
SCC	Secondary Component Carrier
SCell	Secondary Cell
SINR	Signal to Interference Plus Noise Ratio
UE	User Equipment

1. Introduction

1.1. Boost Capacity or Save Power

Currently, the number of mobile connections, particularly in Fourth Generation (4G) Long Term Evolution (LTE) and beyond mobile networks, is growing rapidly showing that the demand for high data rate will remain an issue in the years to come. According to the Global System for Mobile Communications Association (GSMA) mobile Economy 2022 report [1], by the end of 2025, 4G will account 55% of the total 8.8 billion mobile connections whereas Fifth Generation (5G) will account for around 25%. In Sub-Saharan Africa countries, the growth of 4G LTE mobile connections is forecasted to reach 33% by 2025. In the same report and region, 5G will reach 41 million (4% of total mobile connections in the region) by 2025. This increases the demand for higher bandwidth as well as more efficient spectrum management and utilization techniques in mobile networks.

Carrier Aggregation (CA), as capacity enhancement feature, is employed in LTE-A networks to alleviate the anticipated capacity demands. Moreover, LTE-A is expected to remain the dominant mobile broadband network in the global market for the years to come due to its backward compatibility with the standard LTE and also its co-existence with 5G in non-stand-alone (5G NSA) deployment scenario [2], [3].

Likewise, ethio telecom, the dominant mobile operator in Addis Ababa, Ethiopia, has deployed LTE-A network and employed inter-band CA, primarily to serve packet switching (PS) data traffic. The operator is using Global System for Mobile Communication (GSM) and Universal Mobile Telecommunication Service (UMTS) networks to support primarily the circuit switched voice traffic. On the other hand, the 4G LTE-A network takes the highest share (*Figure 1.1*) in the increasing PS traffic and accounting 83% of the total PS traffic as of January 2023.

The operator has implemented inter-band CA technique by using the below 3.5 GHz frequency bands; namely, 2600 MHz (Band 7), 1800 MHz (Band 3) and a recently re-

farmed 800 MHz (Band 20) to boost its user throughput and able to support the high PS traffic demand [5].

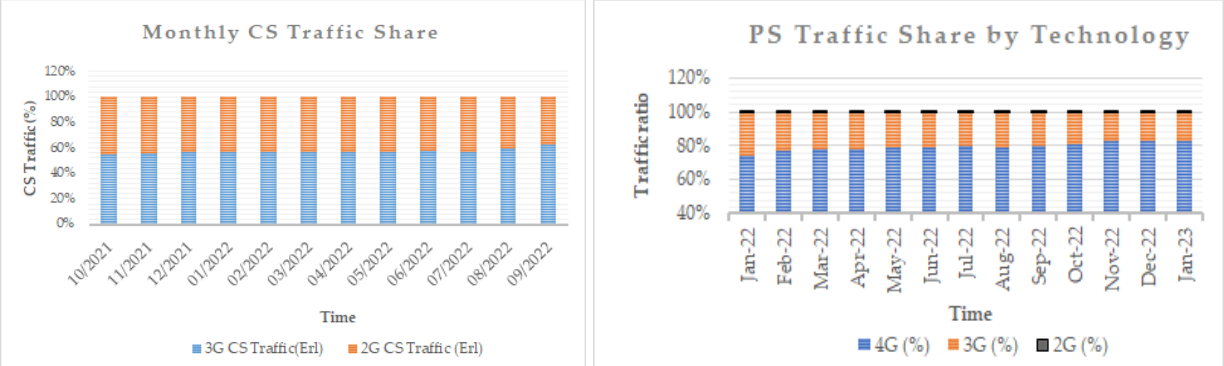

Figure 1.1: 4G PS traffic share [4]

However, employing inter-band CA requires additional RF chain provisioning for each band and CC added. An RF chain consists power amplifier (PA) and filter to process band specific radio signals. RF chains or units are considered to be the base station components that consume the highest power [6], [7]. To be more specific, the power consumptions of RF units varies mainly based on manufacturer, radio access technology and operating frequency. E.g., for LTE network RF units operating at the frequency bands 900/1800 MHz, 2100 MHz, and 2600 MHz, the maximum power consumptions are 990 W, 990 W, 1380 W, and 1980 W, respectively. This includes both static and traffic-dependent power consumption [8].

We have analyzed the hourly variations in the power consumption of RF units operating in LTE-A mode with frequency bands 1800 MHz and 2600 MHz over a period of over 30 days. RF units for legacy networks (GSM and UMTS) operating modes with frequencies (in MHz) 900, 1800 and 2100 have been included for our comparative analysis

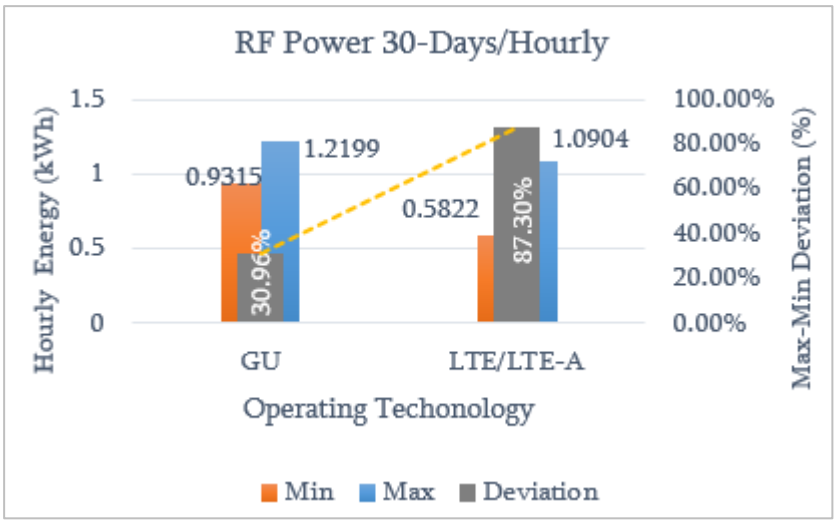

Figure 1.2: RF dynamic power variations for different technologies [9]

In the above figure, *Figure 1.2,* the RF units hourly energy consumption for 30-days operating in three different technologies are grouped in two categories. The base power consumption is significant in GSM and UMTS (GU) case. On the other hand, for LTE/LTE-A case, the dynamic power variation is 87.30%, which is higher compared to 30.96% in GU. In the latter case, hardware design improvements can led to significant energy saving, whereas in the former case, operational energy saving and optimization solutions can improve the traffic dependent power. The higher variance in the LTE-A daily dynamic traffic load coincides to our premise to explore traffic-aware energy saving problems.

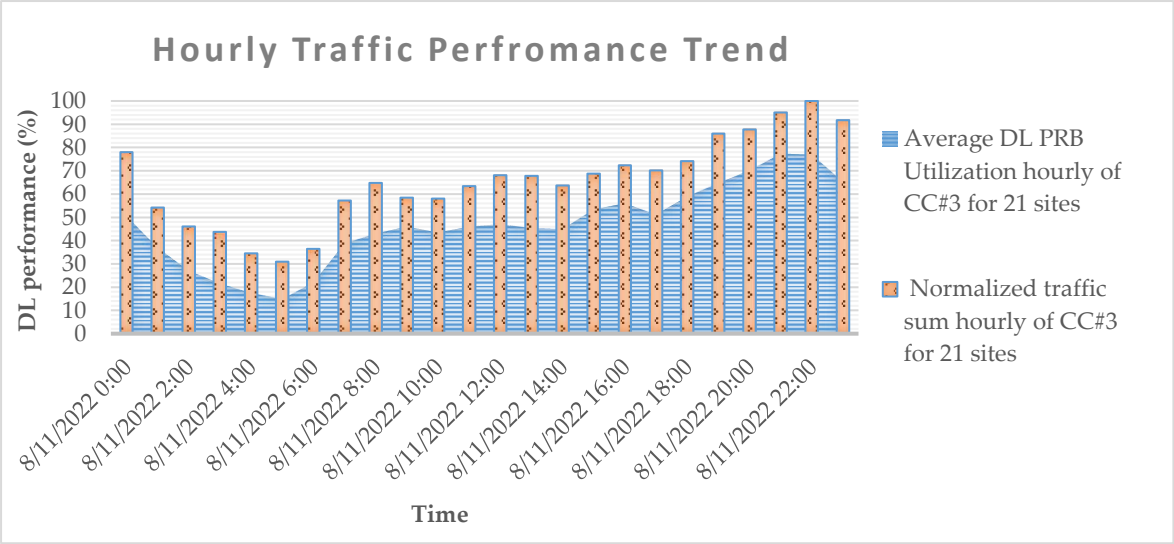

Figure 1.3: Dynamic traffic and capacity demand trends for operational sites [4]

Moreover, the traditional global common practice radio planning and resources provisioning based on peak-hour traffic load statically for the dynamically varying traffic in actual operation is non-optimal. This led to low resource utilization during off-peak hours, illustrated for a randomly selected 21 operational sites (*Figure 1.3)* while the relying RF units, also the higher power consuming components of eNBs, are operating at always-on states.

The inefficiency of PAs in those RF units leads to high power wastage, especially at low traffic load conditions. This is because the PAs' performance drops to 5% [31] at these conditions. The impact of this inefficiency is further compounded by the increasing number CCs from different bands. This is because each CC requires its own PA, which further increases power consumption even at light load states. The battery life of RRC connected user terminals with CA capability is also impacted by this inefficiency, as they have to monitor multiple control channels due to the increasing number of serving CCs. As a result, operators' relief from inter-band CA for high transmission capacity gain at peak hours is challenged by high power wastages at off-peak hours.

Therefore, further research is required to improve the operational efficiency of LTE-A networks targeting on both maximizing power savings at lower traffic hours and benefiting CA capacity at peak loads.

In green and more profitable mobile industry, energy inefficiencies are minimized mainly by saving energy with cautious turning off underutilized devices or components during off-peak hours while guaranteeing the quality of service (QoS). This includes different kinds of sleep modes like micro and deep sleep modes. In micro sleep mode, base station transmission is suspended in milliseconds long and components wake up immediately before transmission resumes. In deep sleep mode, so many circuits may be completely switched off and it takes longer to wake up but the number of daily switched off times is reduced [10].

The limitations on reactive solutions as well as the retransmission time 0.2 ms for LTE/LTE-A to switch off and save energy hungry components, inevitably need a

proactive and longer sleep time solutions. Thus, it saves more energy and minimizes frequent hardware on/off for improved equipment service life. This is achievable with prior knowledge of the actual dynamic network traffic demand for a more accurate decision-making.

In recent researches, proactive base station energy saving solutions have been proposed focusing on deactivating capacity cells with increased sleeping duration and less QoS impact [12]–[14]. Authors therein solved energy saving problems under probabilistic network state transitions for heterogeneous networks (HetNets) and multi-cell eNB network scenarios by using Deep Reinforcement Learning (DRL) algorithm preferred for its trial-and-error self-learning capability and improved inferences. In [15], DRL based secondary CC (SCC) activation/deactivation per scheduling time interval for CA users battery saving is also studied. However, to adopt per cell separate on/off solutions in CA networks induces network instability due to A6 event introduced in LTE-A for intra-eNB and intra-band SCC handover requests to active cells from users located at contiguous cell boundaries of an eNB. Such cell independent on/off decision within a multi-cell sector of an inter-band CA enabled eNB is also prone to multi-bands' load imbalances.

In this research, we have proposed a band-level capacity cells on/off (BLCOO) energy-saving solution trained using DRL. In BLCOO, all same-band cells configured in different sectors of an eNB are to be activated or deactivated simultaneously in determining optimal set of serving CCs. Our optimal set of CCs controlling DRL agent perceives the highest of sectors' cumulative loads expected and the current activation status of CCs responded by our custom network environment for its action selection.

The band-level decision improves network instability due to frequent SCC handover requests and perceiving the sector cumulative traffic load for decision brings a more robust solution to multi-band load balancing. We use operator data and DRL algorithms, Deep Q-Networks (DQN) and Proximal Policy Optimization (PPO), to solve our energy saving problem. This also improves the battery life of end-user terminals due to the lower number of CCs during light traffic load hours.

1.2. Statement of the Problem

CA is used to increase capacity in LTE-A and beyond networks. However, its power consumption outweighs capacity gain during light traffic/loads. This is more significant especially for their OFDMA underlying radio access networks due to high PAPR and inter-band CA scenario which requires independent FFT modules per each carrier added. At low traffic loads, the static power consumption is higher due to lower power efficiency of PA at low load.

We observed up to 87% hourly load variations in RF modules of LTE-A networks in Addis Ababa, signifying the need for load adaptive solutions. This is in contrast to legacy networks, where the hourly load variation is only 30.96%.

Moreover, mains power from a utility company occasional outage caused annual 13.44% of power supply on average to be filled by carbon emitting diesel generators. Thus, reducing power consumption in base station is critical for economic and environmental reasons.

Previous works have shown that CA can waste power, especially when data rate boost is not mandatory. For example, [16] solved for single-carrier achievable boost rates and found that single carrier outperforms CA in some cases. [15] proposed a RL-based approach to deactivate SCCs to allocate the required PRBs per user from the minimum possible number of CCs and minimized CA's impact on UEs' battery life, but this approach ignores RF unit power saving and leaves deactivated resources unutilized.

Network side power saving solutions have also been proposed for multi-carrier and HetNet scenarios, including distributed and centralized cell on/off solutions.

In this aspect, we have observed limitations on energy saving and capacity maximization optimization operational solutions for LTE-A CA scenario targeting both for network and UE power saving.

In this work, we have proposed a DRL-trained BLCOO power saving operational solution to use traffic load adaptive and optimal set of serving CCs for urban macro eNB in inter-band CA enabled LTE-A network. This leverages network side and user terminal separate power saving solutions. The proposed BLCOO further improves the network instability due to intra-eNB SCC handover requests emanating from treating cells independently. Data collected from operator performance and configuration management systems as well as on site measurements are our inputs to: model RF units power consumption trends, build custom environment and train proposed DRL agent.

Lastly, we evaluate per eNB performance improvement for our metrics energy saving (kWh) and saving gain, average number of serving CCs and average achievable rate.

1.3. Objectives

1.3.1. General Objective

The general objective of the research is to save power in inter-band CA employing LTE-A networks by activating the optimal set of CCs based on traffic load, while meeting both network and user-side requirements.

1.3.2. Specific Objectives

The specific objectives to achieve our main objective are:

- To collect a realistic performance data from LTE-A network management and via measurements
- To build RF power consumption models
- To build custom network simulator
- To design multi-objective optimization function
- To train BLCOO energy saving RL agent
- To proactively determine optimal set of CCs
- To ensure CA capacity is sustained when traffic load demands

1.4. Methodology

The methods and procedures to achieve the above specific objectives are:

- Review related works on the topic, do a survey on the operational network, identify problems and address research gaps
- Collect operator data and preprocess to build custom simulation environment to replicate the realistic network environment. This overcomes limitations of using simulation data, and infeasible for risks on actual equipment.
- RF power measurements and performance data inputs to regression and modeling
- Next, we design our objective function problem targeted at using CA to boost capacity at high traffic load and maximize power saving at light traffic loads
- In order to achieve our goal for minimal complexity and to avoid signaling overload from centralized solution, eNB level solution is proposed
- Evaluate and do comparative analysis for RF power consumption of target eNB with and without the proposed solution and for different algorithms
- Evaluate the optimal number of CCs over the total available CCs as a metric for user device monitoring power reduction performance

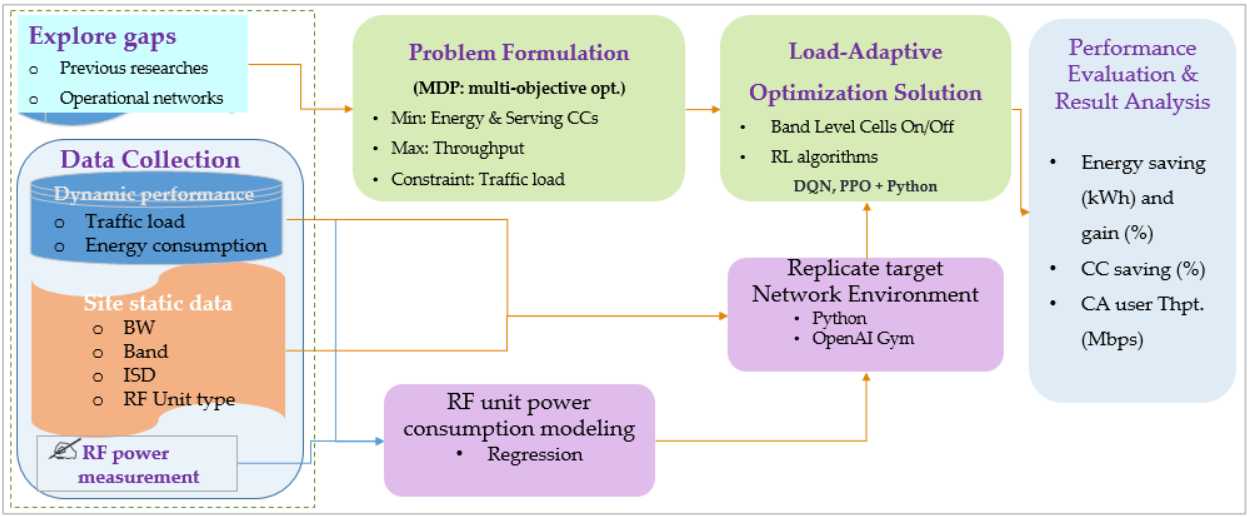

Figure 1.4: General diagram of the research methodology

1.5. Related Works

This section reviews the most relevant works on energy saving approaches for CA networks, both for network (multi-carrier eNBs) and end-user terminals.

In [19], the effect of CA investigated on the total transmitted power of an LTE-A eNB with minimum user data rate as a QoS constraint. They used Shannon capacity and a path loss model (to relate cell radius and user power), assuming equally shared PRBs. The transmitted power of eNB with multiple users significantly reduced by increasing the CA bandwidth. CA and bandwidth expansion were also proposed for energy saving in [11]. However, these are at the expense of scarce spectral resources.

Thus, energy efficiency (EE) and spectral efficiency (SE) in CA system are to be managed in tradeoff. Authors in [20] proposed a variable for performance factor a design phase parameter that reflects the priority level of EE and SE at different configuration preferences to adjust the tradeoff performance.

In [17], the authors proposed an energy minimization scheme for a multiple CC base station by classifying incoming traffic sessions as exponential (real-time) and truncated lognormal (non-real-time) distributions. The authors used a mathematical formulation to calculate the number of RBs required for the minimum data rate of each user type and turns off SCCs when not required. However, the scheme requires prior knowledge of the traffic load in each sub frame time.

The use of machine learning, such as RL in energy saving and resource optimization tasks for wireless networks added capabilities to compromise demands of low QoS impact, highly dynamic network traffic load and increasing number of metrics to consider.

In [12], a centralized DRL-based cell on-off energy saving solution has been proposed for a network region consisting of multiple eNBs and cells. The solution uses throughput and traffic volume maps as observation states. Apart from the computational cost of two-dimensional (2D) maps and updates as an input, the authors recommended distributed alternatives, which have lower signaling overhead. Moreover, the assumption that all

cells in the same band are equally loaded throughout the network region is an oversimplified and coarse generalization that ignores the spatial dynamics of different eNBs loads.

In [14] and [13], DRL-based cell-level on-off, distributed energy saving solutions were proposed using synthetic data in HetNet and realistic data with multi-carrier eNB scenarios, respectively. The authors in [13] extended their solution to consider anticipated load imbalances in multi-carrier sectors and also included an additional metric to control instability induced by excessive handover signaling during user offloading.

CA-user devices battery life is also a concern by researchers on the topic with the fast evolving mobile technologies and provisioned to support data intensive applications.

Authors in [21] compared the power consumption of two CA capable UE of 2012 and 2019 products. The UEs were connected to 10 MHz and 20 MHz single band carriers and aggregated 20 MHz from 2CCs (10 MHz + 10 MHz). Though improvement for the latest products, the power consumption in CA scenario was still 15% higher.

Moreover, [17] with the increased number of activated CCs and their bandwidths, radio frequency and baseband components of UEs consumed more energy to receive and decode the signal. Joint CC selection and PRB allocation (activate SCC during PRB allocation) schemes demonstrated energy saving improvement with no impact on user data rate and delay.

In [15], the authors proposed an RL-based UE power saving method by deactivating SCCs per user so long as its PRBs can be sufficiently allocated from a single CC. While all CCs or cells of the eNB were kept always active, unutilized spectral resources cannot be deactivated during off-peak hours and save network-side power.

Thus, in this work, energy saving for both network and UE in CA scenario is considered.

1.6. Scope and Limitations

1.6.1. Scope of the Research

We have collected 3 months real LTE-A network eNB DL traffic load performance and different RF units power consumption data, eNB configuration parameters to build our custom simulation environment and RF units power consumption models. RL training is triggered for our energy saving algorithm to learn for the optimal policies of activation/deactivation inferring. We compared the power consumption of optimized CA networks with always-active CCs cases. We also evaluated the users' achievable rates, RF power consumption reduction, and CCs usage based on 24 hour eNB traffic trend. The evaluation was done for a single urban macro base station and 30 stationary users in an interference limited network environment.

1.6.1. Limitations of the Research

Though the proposed energy saving solution is scalable to higher number of CCs, results presented in this work are based on evaluation for inter-band CA with a maximum of three serving CCs due to real traffic data input limitations for higher number of CCs.

1.7. Contributions of the Research

In this thesis work, new power consumption models for RF units operating at different frequency bands have been built using power consumption performance realistic data and on-site measurements. In addition, a new custom network simulation environment is built based on OpenAI Gym API to replicate the realistic inter-band CA enabled LTE-A network physical layer and for RL agent interactive learning. Moreover, a new multi-objective function capturing metrics for target performances of user and network is designed. The converging parameters for adopted DRL algorithms (DQN and PPO) are fine tuned. Finally, a DRL trained BLCOO energy saving solution is provided.

1.8. Thesis Outline

The remaining part of the document is organized as follows. An overview of CA, power consumption and saving techniques from LTE-A network aspect, including surveys on operator ethio telecom power consumption and traffic performances, Addis Ababa have been presented in Chapter 2. MDP and RL optimization tool are discussed in Chapter 3. Then description of system model, assumptions and parameter values for the BLCOO energy saving and optimization evaluations are presented in Chapter 4 which is followed by presentation of results of the energy saving performances in Chapter 5. Finally, concluding remarks and future directions are provided in Chapter 6.

2. Overview of Carrier Aggregation in LTE-Advanced

2.1. LTE Overview

Before heading to LTE-A, it is better to discuss the basis technology, LTE. LTE technology was first specified by the 3rd Generation Partnership Project (3GPP) group as 3GPP R8 and commercially deployed by late 2009. Since then, the number of commercial networks increased dramatically worldwide and underwent continuous technological improvements, making it the fastest mobile system in terms of development. 3GPP groups used the term "Release" (we use "R" for this document) to denote enhancements achieved by adding technology components, with R9 marking the initial enhancement. However, it was R10, also known as *LTE-Advanced* that brought significant improvements, surpassing even the requirements of International Mobile Telecommunication (IMT). Since then, 3GPP has continued to use the name LTE-Advanced for the technology family, with subsequent evolutions, including 3GPP R12, despite different marketing terms.

LTE uses six standard carriers with six bandwidth options. Each carrier is mapped as time-frequency OFDMA resource grid for downlink resource allocation. It also supports multiple antenna streams with a maximum of 4 for diversity and multiplexing purposes. A PRB is the smallest unit radio resource that can be allocated by packet scheduler for data transmission to an LTE user. In OFDMA, its defined by a time-frequency 2-dimensional resource grid. It encompasses 12 sub-carriers with 15 kHz spacing each for a total of 180 KHz (one sub-channel) in frequency domain and a time slot of 0.5 ms in time domain. With the inclusion of multiple antenna streams for spatial multiplexing, it can further extended to time – frequency – space, and considered as 3-dimensional resource grid. Thus, a RB pair in time domain can be allocated for an LTE subframe duration of 1 ms.

The transmitted signal is organized into subframes of 1 ms duration with ten subframes forming a radio frame as illustrated in *Figure 2.1*. Each downlink subframe consists of a

control region of one to three orthogonal frequency division multiplexing (OFDM) symbols, used for control signaling from the base station to the terminals, and a data region comprising the remaining part and used for data transmission to the terminals. The data transmissions in each subframe are dynamically scheduled by the base station. As seen in *Figure 2.1*, cell-specific reference signals are also transmitted in each downlink subframe. These reference signals are used for data demodulation at UE, and for measurement purposes, e.g. for channel-status reports sent from the terminals to the base station.

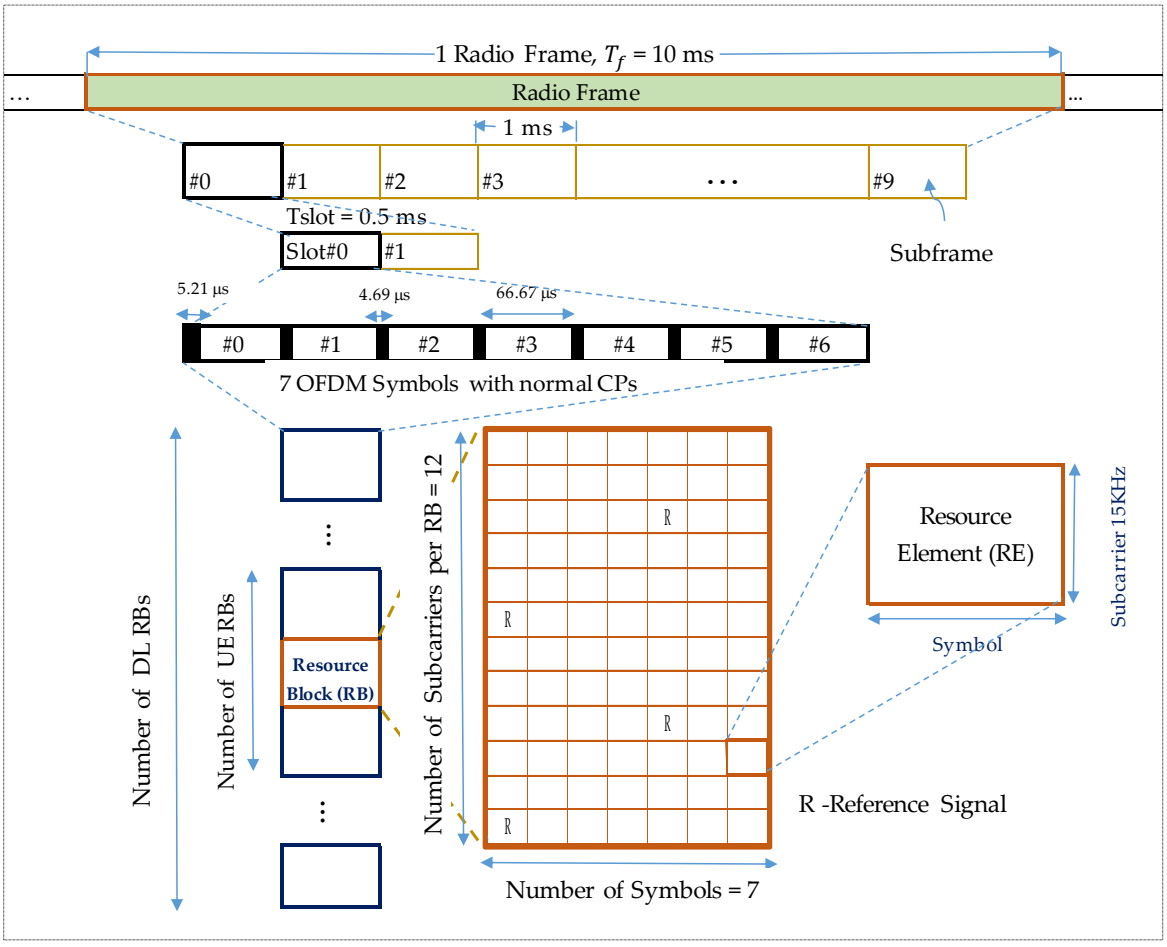

Figure 2.1: LTE Resource grid for downlink FDD [22], [23]

Spectrum flexibility is one of the key properties of the LTE radio-access technology.

A wide range of different bandwidths is defined and both frequency division duplex (FDD) and time division duplex (TDD) modes-of-operation are supported, allowing for operation in both paired and unpaired spectrum. According to 3GPP, LTE systems can

operate on variable bandwidth for a given band ranging from 1.4 MHz to 20 MHz, denoted by 1.4, 3, 5, 10, 15 or 20 (in *MHz)*, which determines the maximum number of PRBs. The total number of PRBs (*Figure 2.1*) depends on the carrier bandwidth in use. An important requirement in the LTE design has been to avoid un-necessary fragmentation and strive for commonality between the FDD and TDD modes-of-operation while still maintaining the possibility to fully exploit duplex-specific properties such as channel reciprocity in TDD. Aligning the two duplex schemes to the extent possible does not only increase the momentum in the definition and standardization of the technology but also further improves the economy-of-scale of the LTE radio-access technology.

To support for multi-antenna transmission is an integral part of LTE from the first release. Downlink multi-antenna schemes supported by LTE include transmit diversity, spatial multiplexing (including both so-called single-user Multiple-Input Multiple-Output (MIMO) as well as multi-user MIMO and beam forming.

The International Telecommunications Union (ITU) defined a set of requirements for radio-access technologies beyond IMT-2000, called IMT-Advanced (IMT-A). In response to this invitation, 3GPP initiated a study item on LTE-Advanced in March 2008 [24]. The goal of this study item was to define requirements and investigate potential technology components for the LTE evolution. LTE R10 was submitted to the ITU in 2010 and approved as one of two IMT-A technologies. It meets or exceeds all of the IMT-A requirements [25], and it includes a number of additional features, such as CA and MIMO. To meet the IMT-A requirements, LTE R8/9 is upgraded to LTE-A (R10+ onwards).

2.2. LTE-Advanced

LTE-A, is not a new radio-access technology but the evolution of LTE to further improve the performance. Being an evolution of LTE, LTE-A includes all the features of R8/9 and adds several new features of which the most important ones – use of CA, enhanced

multi-antenna support, improved support for heterogeneous deployments and relaying. CA will be discussed in next sections. Evolving LTE rather than designing a new radio-access technology is important from an operator perspective as it allows for a smooth introduction of new technologies without jeopardizing existing investments.

A LTE-A terminal can directly connect to a network of an earlier release, as well as a R8/9 terminal can connect to a network supporting the new enhancements. Hence, an operator can deploy a R8 network and later, when the need arises, upgrade to LTE-A functionality where needed. In fact, most of the LTE-A features can be introduced into the network as simple software upgrades.

Table 2.1: IMT-A requirements and LTE-A capability [24], [25]

Capability Requirements		LTE Release 8	IMT-Advanced	LTE-Advanced Release 10	
Transmission bandwidth		1.4 to 20 MHz	Minimum of 40 MHz	Up to 100 MHz	
Spectral efficiency (peak)(bps/Hz)	DL	16.3 (4 layers)	15	16.3 (4 layers)	30.6 (8 layers)
	UL	4.3	6.75	8.1 (2 layers)	16.1 (4 layers)
Latency (ms)	CP	50	Less than 100	50	
	UP	4.9	Less than 10	4.9	

DL: Downlink *CP: Control Plane*

UL: Uplink *UP: User Plane*

For LTE-A network, advanced MIMO, HetNets, coordinated multipoint (CoMP), Relay, self-organizing network (SON) and CA are the main enabling technologies. Hence, in LTE-A mainly by extending the previous 4x4 MIMO capacity to 8x8 (also called Extended-MIMO) for downlink, using small cells and aggregating multiple LTE standard carriers helped to achieve IMT-A requirements. In LTE-A R12, Flexible-CA from multiple-carriers (FDD scenario), higher modulation schemes 256 QAM and 64 QAM supported for DL and UL respectively are among LTE-A feature enhancements.

Our focus for the time is on CA and to be discussed in subsequent sections.

2.2.1. Carrier Aggregation

CA is the use of multiple standard LTE CCs per user so as to increase the transmission bandwidth and peak user data rate. CA permits LTE-A to achieve the goals mandated by IMT-Advanced while maintaining backward compatibility with LTE R8/9.

LTE –A CA permits the LTE radio interface to be configured with any number (up to five) carriers, of any bandwidth, including differing bandwidths, in any frequency band.

In CA, the DL and UL can be configured completely independently, with only the limitation that the number of uplink carriers cannot exceed the number of downlink carriers. Each aggregated carrier refers a CC and can have a bandwidth of 1.4, 3, 5, 10, 15 or 20 (in *MHz*). With a maximum of five CCs, the maximum aggregated bandwidth is 100 MHz.

CA primarily increases peak user throughput and also improve spectral resource management. CA does not increase the spectral efficiency [26], but enhances data rates by assigning more bandwidth to the individual users. CA can be used both in FDD and TDD schemes. However, in R10, CA TDD is limited to intra-band CA whereas CA of TDD of different bands (TDD inter-band) is supported. The enhanced LTE FDD-TDD joint feature is incorporated in R12. In this regard, DC [3] is also considered as a possible solution of this FDD-TDD joint feature. In DC, the UE is simultaneously connected to two different eNB base stations (on different carriers) by aggregating flows of the two eNB despite the latency caused by the backhaul (X2 link). The base stations do not need to be synchronized (and therefore may not be co-located), main difference with the aggregation of carriers where the aggregation of flows is performed on the same eNB in two different radio bands [2].

In LTE-A Pro, R13, CA of up to 32 CCs is supported to achieve a wideband transmission channel of up to 640 MHz by using unlicensed 5 GHz band carriers License-Assisted Access (LAA) CA. In 5G New Radio (NR), CA with UE DL and UL of up to 640 MHz from 16 carriers each 400 MHz is supported.

Thus, CCs can be of different bandwidths and from duplex schemes and inter-band CA with different numerologies is supported. For NR users to operate on narrow or wider bandwidth i.e. depending on data demand, Bandwidth Part (BWP) is defined in NR. That makes NR an energy efficient solution despite the support of wideband operation. Though LTE and NR aggregation is not possible, LTE-NR DC in R15 supports aggregation of throughputs from simultaneous transmission of both LTE and NR carriers. Several combinations of CA components are supported in R16 and in latter releases.

Intra-band contiguous and Intra-band non-contiguous and inter-band are the three modes of CA and discussed in subsequent sections.

A. Intra-Band Contiguous

Intra-band contiguous refers multiple contiguous CCs are aggregated from the same operating frequency band and simplest one for an operator to use. A contiguous bandwidth wider than 20 MHz is not a likely scenario given frequency allocations for the below sub-band 3.5 GHz. The spacing between center frequencies of contiguously aggregated CCs is a multiple of 300 kHz to be compatible with the 100 kHz frequency raster of R 8/9 and preserving orthogonality of the subcarriers with 15 kHz spacing. From an RF perspective, intra-band contiguous aggregated carriers have similar properties to a corresponding wider bandwidth carrier being transmitted and received.

B. Intra-Band and Inter-Band Non-contiguous

Mobile network operators (MNOs) are currently facing the problem of a fragmented spectrum. The non-contiguous allocation has been specified to fit those scenarios, the allocation could either be intra-band, i.e. the CCs belong to the same operating frequency band, but have a gap or gaps in between, or it could be inter-band, in which case the CCs belong to different operating frequency bands.

With the introduction of CA in R10, the aggregation of bands has been specified by 3GPP for specific sets of CA bands which correspond to a combination of Evolved Universal Terrestrial Radio Access (E-UTRA) operating bands.

In general, advancements introduced on CA features in latter releases and generations improved the performance of CA resource utilization though adds complexity to resource schedulers.

C. UE Capability

The UE capability procedure will take place to establish an Evolved Packet Service (EPS) bearer after the Random Access Procedure. An LTE-A capable UE will report extra information to the network regarding its CA bands support capabilities. The capabilities are notified per frequency band, independently for downlink and uplink. It will define the proper CA configuration set to be used. The CA configuration is a combination of operating bands, in association with the CA bandwidth class of the UE, for each band. It determines which band to be used and the channel bandwidth allocated on each operating band. The inter-band architecture represents a major challenge for UE design since multiple simultaneous chains have to coexist.

In general, the number of CCs that can be aggregated is limited by the UE's capabilities. while the CA bandwidth is limited by the available spectrum, its configuration must be coordinated between the UE and the eNodeB.

D. E-UTRA Aspects of Carrier Aggregation

In R10, CA introduced a distinction between a primary cell (PCell) and a secondary cell (SCell) [27]. PCell and SCell are two important concepts in CA. The PCell is the main cell with which the UE communicates, while the SCell is a secondary cell that can be used to provide additional radio resources.

The PCell is defined as the cell with which RRC signaling messages are exchanged. There is always one PCell active in RRC_CONNECTED mode. The PCell is equipped with one physical downlink control channel (PDCCH) and one physical uplink control channel

(PUCCH). The PDCCH is used to transmit control information to the UE, such as scheduling information and handover commands. The PUCCH is used by the UE to transmit uplink control information, such as acknowledgements and non-acknowledgements (NACKs). The measurement and mobility procedures are based on the PCell. This means that the UE measures the signal strength of the PCell and uses this information to determine if it should handover to another cell. The random access procedure is performed over the PCell. This is the overall procedure that the UE uses to establish a connection with the network.

A SCell is not always active, and can only be configured after connection establishment, in CONNECTED mode. A SCell could be equipped with a one PDCCH or not, depending on UE capabilities. A SCell never has a PUCCH.

The MAC layer based dynamic activation/deactivation procedure is supported for SCell for UE battery saving. This means that the SCell can be activated or deactivated based on the UE's needs.

All PCells and SCells are known collectively as serving cells. The CCs on which the PCell and Scell are based are the primary CC (PCC) and secondary CC (SCC) respectively. Different CA capable UEs within a cell can have different PCC on different band.

E. Protocol Layer Impact

CA impacts a limited number of protocol layers. PCell carries Non-Access Stratum (NAS), key exchange and mobility procedures. The Packet Data Convergence Protocol (PDCP) and Radio Link Control (RLC) layers are completely transparent to CA signaling, but the RLC layer has a larger buffer size to support higher data rates.

The CA impact on physical and MAC protocol layers from data aggregation schemes and backward compatibility aspect also discussed.

Firstly, from data aggregation aspect, the transmission blocks (TBs) from different component carriers can be aggregated at either the medium access control (MAC) or physical layer.

In MAC layer data aggregation scheme, each component carrier has its own transmission configuration parameters in the physical layer, as well as an independent hybrid automatic repeat request (HARQ) entity in the MAC layer. This scheme is more flexible and efficient than the physical layer data aggregation scheme, but it requires more control channels.

In physical layer data aggregation scheme, one HARQ entity is used for all the aggregated CCs, and new transmission configuration parameters should be specified for the entire aggregated bandwidth. This scheme is simpler and requires fewer control channels than the MAC layer data aggregation scheme, but it is less flexible and efficient.

Compared to the physical-layer scheme, the transmission parameters are configured independently for each CC under the MAC layer data aggregation scheme. Hence, the latter can support more flexible and efficient data transmissions in both UL and DL, at the expense of multiple control channels.

Secondly, from backward compatibility aspect, the MAC layer data aggregation scheme guarantees backward compatibility with LTE systems, since the same physical layer and MAC layer configuration parameters and schemes can be used.

F. SCC Addition and Removal

RRC signaling based on the information UE capability sends a message for addition or removal of SCCs. The CA additional SCells cannot be activated immediately at the time of RRC establishment. Thus, there is no provision in the RRC Connection Setup procedure for SCells. SCells are added and removed from the set of serving cells through the RRC Connection Reconfiguration procedure. Note that, since intra-LTE handover is treated as an RRC connection reconfiguration, SCell "handover" is supported. The CA-related information sent by the eNodeB pursuant to this the RRC Connection Reconfiguration procedure is summarized below.

- Cross-carrier scheduling configuration – Indicates, among other things, if scheduling for the referenced SCell is handled by that SCell or by another cell.

- SCell PUSCH configuration – Indicates, among other things, whether resource block group hopping is utilized on the SCell.
- SCell uplink power control configuration – Carries a number of primitives related to SCell uplink TPC, including the path loss reference linking parameter.
- SCell CQI reporting configuration – Carries a number of primitives related to CQI measurements reporting for SCells.

G. SCC Activation and Deactivation

The PCell change is performed in the same manner as a R8/9 handover procedure. During the PCell change, the SCell will be deactivated by the serving eNodeB and then selected and activated by the target eNodeB. The DL of a configured SCell can be activated and deactivated dynamically to save power. The UE does not monitor the deactivated SCell for PDCCH.

Cell activation/deactivation is a mechanism aiming to reduce UE power consumption in LTE-A CA on top of discontinuous reception (DRX), which is already supported in LTE R8/9. DRX puts UE into power saving mode when the UE is not expected to receive data from the network. According to network configuration and ongoing HARQ processes, UE determines the DRX on/off duration common to all serving cells.

For a deactivated SCell, UE does not receive any downlink signal; nor does the UE transmit any uplink signal. Conversely, for an activated SCell, UE performs normal activities for DL reception and UL transmission. The Scell activation/deactivation is enabled by a combination of explicit and implicit means where the network can issue an activation/deactivation command in the form of a MAC control element (CE), or the UE autonomously deactivates a serving cell upon timer expiry. Serving cell activation/deactivation is performed independently for each SCell, allowing UE to be activated only on a necessary set of SCells. Activation/deactivation is not applicable for the PCell since the functions provided by the PCell require it to always remain activated when the UE has an RRC connection to the network.

MAC signaling based on the information from UE and upper layers sends a message for activation or deactivation of SCCs.

Table 2.2: Downlink MAC CEs for UE power saving [28]

Release	Application	MAC Control Element
LTE-A Release 10	(De)activates SCells	Activation/Deactivation
LTE Release 8	DRX command	Sends UE to sleep during DRX

H. Scell Handover in LTE-A Inter-Band CA

Handover processing for LTE in R10 is similar to R8/9, with clarifications made to refer to the PCell in the measurement-related RRC signaling messages. In LTE-A R10, Event A6 (neighbor cell becomes offset better than Scell) is introduced for CA to compare the channel quality of a neighboring cell relative to a reference SCell, thereby allowing fast reconfiguration of SCells.

Event A6 occurs when a neighboring cell's strength becomes better than an SCell's strength by an offset. Event A6 reporting indicates that the UE is able to report Event A6, which occurs when a neighbor PCell becomes stronger than a serving SCell by an offset.

In the case of intra-band SCells, this event is less useful, as the strength of the PCell and the SCells usually is very similar. However, with inter-band serving cells, the strength of a neighboring PCell could be significantly different from a serving SCell. Depending on network conditions such as traffic load distribution, it could be advantageous to execute a handover to the cell identified by Event A6.

I. Resource Scheduling in CA

Resource scheduling methods have been modified for CA. In non-CA communications, a dedicated control channel is used for each CC. However, when CA is enabled, SCells can be either scheduled using their own control channel or they can be scheduled using the PCell control channel called same carrier scheduling and cross carrier scheduling respectively. The latter requires the inclusion of a Carrier Field Indicator (CFI) that

specifies which CC the scheduling message, included in PDCCH in the CA case, is referring.

Scheduling strategy, it holds true for traffic split ratio as well, is implementation-specific and not specified by 3GPP.

The UE sends two types of reports to the eNB in resource scheduling: Channel status information (CSI) and sounding reference signals (SRS) reports. CSI report provides information about the channel quality of the different CCs. The eNB uses this information to determine which CCs to schedule the UE on. The other one is SRS signal is used to measure the channel quality. The UE sends SRS on the PCell and the SCells. The eNB uses this information to determine the channel quality of the different CCs. The UE also sends MAC messages to the eNB to report the status of its buffers.

These reports provide the eNB with information about the channel quality of the different CCs, which the eNB uses to determine which CCs to schedule the UE on and how much data to send to the UE.

J. Capacity Enhancement in CA

As we come to the last section of discussions on the changes due to the introduction of CA, we can infer that the overall modifications are targeting capacity enhancement and UE power saving. For its capacity enhancement aspect, we preferred to derive mathematical expressions to quantify the capacity enhancement of CA systems as an extension of non-CA counterparts.

To begin with, adapting non-CA scenario in [14], for P_{TX} cell transmission power, $G^t_{k,i,j,u}$ channel fading between $C_{k,i,j}$ and u^{th} user in $C_{k,i,j}$ cell, $G^t_{k,p,q,u}$ channel fading between interfering cell $C_{k,p,q}$ and u^{th} user in $C_{k,i,j}$ cell, and assuming that all active PRBs collide, or worst-case scenario, the SINR can be computed as follows.

$$SINR^t_{k,i,j,u} = \frac{P_{TX} G^t_{k,i,j,u} \frac{n^t_{k,i,j,u}}{M_{k,i,j}}}{n^t_{k,i,j,u} N_o + P_{TX} \sum_{p\in[3],p\neq i} \sum_{q\in[B],q\neq j} G^t_{k,p,q,u} \frac{n^t_{k,i,j,u}}{M_{k,i,j}} l^t_{k,p,q}} \quad (2.1)$$

Where:

- N_o is white noise power spectral density
- k and B for number of 3-sector eNBs and total cells respectively
- $l^t_{k,p,q}$ total number of PRBs used per cell
- $n^t_{k,i,j,u}$ is the number of resource blocks allocated to u^{th} user from i^{th} carrier or cell or band
- $M_{k,i,j}$ is the total number of resource blocks available from i^{th} band
- $G^t_{k,i,j,u}$ and $G^t_{k,p,q,u}$ refer gains between serving and interfering eNBs and user
- $P_{TX}(\frac{n^t_{k,i,j,u}}{M_{k,i,j}})\, G^t_{k,i,j,u}$ and $P_{TX}(\frac{n^t_{k,i,j,u}}{M_{k,i,j}})\, G^t_{k,p,q,u}$ refer to received power by a user from serving and interfering eNB cells respectively

As cell configured in eNB sector of an LTE-A network, it uses any of LTE standard carriers and bandwidths that defines a CC from CA aspect.

The gain between user and eNB (G^t_{dB}) computed from sum of distance-based pathloss, noise figure, antenna gains, wall penetration loss (indoor user) is obviously in dB. Thus, its linear equivalent is for our SINR calculation:

$$G^t_{linear} = 10^{-G^t_{dB}/10} \tag{2.2}$$

Distance based path loss is determined from propagation models where operators fine-tuned with some empirical measurements for a better customized uses.

The cell throughput $D_{k,i,j}$ of cell $C_{k,i,j}$ can be expressed as:

$$D_{k,i,j} = \sum_{u}^{U^t_{k,i,j}} \min(W * n^t_{k,i,j,u}\ \log_2(1 + SINR^t_{k,i,j,u}), V^t_{k,i,j,u}) \tag{2.3}$$

Where $V^t_{k,i,j,u}$ is the buffered data size of u^{th} user in cell $C_{k,i,j}$ to consider non-full buffer users for the cell throughput calculation. where W represents a factor accounting the overhead and number of layers during the transmission process.

Similarly, the user throughput served by a single carrier or cell $C_{k,i,j}$ will be:

$$r_{k,i,j,u} = \min(W * n^t_{k,i,j,u}\ \log_2(1 + SINR^t_{k,i,j,u}), V^t_{k,i,j,u}) \tag{2.4}$$

For more practical throughput of u^{th} macro-user with target bit error rate (BER),

$$r_{k,i,j,u} = \min(W * n^t_{k,i,j,u}\ \log_2(1 + \xi SINR^t_{k,i,j,u}), V^t_{k,i,j,u}) \tag{2.5}$$

Where, $\xi = -1.5/\ln(\text{BER})$ [29], computed according to the CQI feedback and the number of RBs assigned to the user. For inter-band CA each CC is affected differently hence has different channel quality.

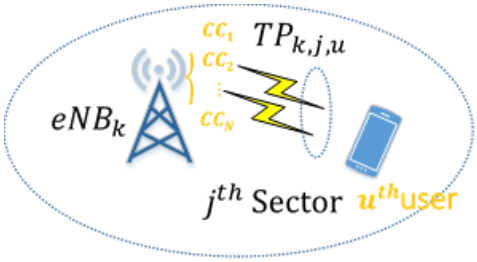

Figure 2.2: CA user and throughput

Hence, for CA user, user throughput $TP_{k,j,u}$ is defined as sum of all throughputs from assigned CCs with scheduled PRBs for that user. No further cell indexing in for same sector users required.

$$TP_{k,j,u} = \sum_{i=1}^{N} (r_{k,i,j,u})\gamma_{i,u} \tag{2.6}$$

Where N is the maximum number of CCs assigned for u^{th} user in K^{th} eNB and j^{th} sector. For CC state indicator $\gamma_{i,u} \in \{0,1\}$, $\gamma_{i,u} = 0$ if i^{th} CC is deactivated and $\gamma_{i,u} = 1$ if activated. The CC with lowest band is usually for coverage while the remaining bands are for capacity purposes. Thus, $\gamma_{1,u}$ is always 1 for as the first carrier is always active.

In LTE-A network, Radio Resource Management (RRM) algorithm assigns CCs, RBs and Modulation and Coding Scheme (MCS) to the user in accordance with their CSI or CQI and CA capabilities to satisfy their QoS requirements.

Regarding full buffer traffic, it refers to a network traffic scenario where the network nodes (e.g., eNBs) always have packets waiting in their buffer queues, and there are always packets to be transmitted. In other words, the buffer queues are always full. This assumption is common in performance analysis studies, as it provides a worst-case scenario for the network.

Non-full buffer traffic, on the other hand, refers a more realistic assumption as it takes into account the fact that not all users are always generating traffic. In non-full buffer traffic, the buffer queues at the network nodes may not be always full. This can lead to underutilization of the network resources, as the buffer queues may not be fully utilized.

The choice of whether to use full buffer or non-full buffer traffic depends on the specific application and the desired results. For example, if the goal is to maximize the network throughput, then full buffer traffic may be a better choice. However, if the goal is to minimize the latency, then non-full buffer traffic may be a better choice.

Due to the demand to support the asymmetric downlink heavy user traffic pattern, much of the research works have prioritized to focus on DL CA.

2.2.2. Towards Greener Telecom Networks

The Information and Communication Technology (ICT) industry is taking an unprecedented step forward in tackling climate change with the release of the first-ever science-based pathway to reduce greenhouse gas emissions across the telecoms sector. This is a major milestone for the industry, as it demonstrates the commitment of mobile network operators to reducing their environmental impact.

Science-based target (SBT) sets emissions trajectory reductions over the decade (2020-2030) for each ICT sub-sector. For example, mobile network operators adopting the SBT are required to reduce emissions by at least 45% over this period. This will require a significant effort from operators, as they will need to find ways to reduce their power consumption while also meeting the growing demand for mobile data.

There are a number of ways that mobile network operators can reduce their power consumption. One is to switch to renewable energy sources. Another is to improve the energy efficiency of their networks. Operators can also encourage their customers to use their devices more efficiently. UE power consumption is also a major factor. Mobile devices are becoming increasingly powerful, which is driving up their power

consumption. Operators can help to reduce this by promoting the use of energy-efficient devices and by developing new technologies that can reduce the power consumption of mobile networks.

2.2.3. Power Saving in LTE/LTE-Advanced Networks

2.2.3.1. User Terminal Power Saving

System designers often fret about power consumption. In battery-powered mobile devices, the talk time from one charge depends on battery capacity and the power consumed by the device. Lower power means longer operating time. At the base station, battery power may not be the issue but the cost of energy in the network can add up. And, of course, wasted power in any device generates heat which must be dissipated.

In this sub section, technological improvements to improve UE power saving focusing on LTE R8 and onwards will be addressed. Moreover, CA impacts on battery life of user terminals and recommendations by researchers have been discussed.

One of UE power saving mechanisms was the implementation of DRX where user saves power when no active transmission to receive. DRX is already supported in LTE R8/9 to save UE power consumption. Hence, secondary cell activation/deactivation for CA capable users and on top of DRX in LTE-A network further improves the battery life of connected users' devices. Activating or deactivating SCCs, but not applicable for PCell, is executed in two options, using command from MAC CE or pre-set timer expiry in user side, depending on resource demand to convey its downlink data traffic or buffer size.

The other is on UE architecture. UE transmitter and receiver aspects of CA output power dynamics are correlated to its architecture, which can be based on a single or multiple PAs. 3GPP considered different options as possible implementations for UE PA architectures to support the different types of CA.

One option is to use a wideband transceiver to cover all bands. This would require the use of expensive wideband RF components and ultra-high performance Analog-to-

Digital and vice versa convertors (ADCs/DACs) with baseband processing bandwidth ≥ 20 MHz. Designing wideband transceivers brings numerous challenges, such as frequency-dependent path loss, Doppler frequency shift, noise power, receiver input signal, nonlinearity of components in the analog receiver, and maximum input signal.

Another option is to use multiple transceiver chains, each of which is dedicated to a specific band. This would reduce the complexity of the UE receiver, but it would also increase the cost and power consumption.

The choice of UE PA architecture depends on a number of factors, such as the desired data rates, the cost, and the power consumption.

Unlike the DL power consumption, UE uplink power consumption is independent of the data traffic volume. The UE's transmit power is controlled by the eNB via a Transmit Power Control (TPC) command. The absence of OFDMA or the use of SC-FDMA also reduces the uplink power consumption.

The introduction of Clustered Discrete Fourier Transform Spread Orthogonal Frequency Division Multiplexing (DFT-S-OFDM) in R10 of 3GPP to support multiple uplink transmission over CA requires a larger power back-off in order to avoid the generation of nonlinearities. This means that UEs will have shorter battery life when transmitting using this waveform. However, Clustered DFT-S-OFDM is still more efficient in terms of PAPR than classical OFDMA, and it is backward compatible with previous versions of 3GPP.

Joint resource allocation and cross-layer scheduling are also among techniques to improve the energy efficiency of UEs in wireless networks.

With all these possible UE power saving, it is still one those areas that needs improvement and including some of the related researches presented here. In [21] real current consumption measurements on UEs have been conducted to evaluate the battery life impact of CA. Smartphone models of the year 2019, a product with the latest

hardware design features by then was selected and its power consumption evaluated to have 15% higher than non-CA case. The authors in [21], [30] and [16] recommend to overload a single user if resource demand is satisfied or minimize the RF chains and amplifiers in increasing number of for set of serving cells and carriers in CA for an improved energy efficiency.

Authors in [30] have strived to demonstrate that CA can be used to save power and scenarios that CA could to led power wastages. While UE fast return to idle state at higher CA data rate boosting, power saving comes from shorter transmission duration. Whereas CA wastes UE power for lower CA rate boost as evaluated for rate boost less than 29%. As per the findings, better rate boosting than in CA can be achieved by overloading single carrier until PRB demand is satisfied with almost no impact on delay. Therein ,and also in *Figure 2.3*, a power consumption of 323 mW is incurred by merely of enabling CA but before allocation of extra PRBs from SCC. Note that the first 100 PRBs belong to the PCC or single carrier case while the second 100 PRBs ranging from 101 to 200 belong to the SCC of CA enabled UE device. Except the CA enabling transition, the power consumption remains same with linear approximate increment of 0.8 mW per PRB.

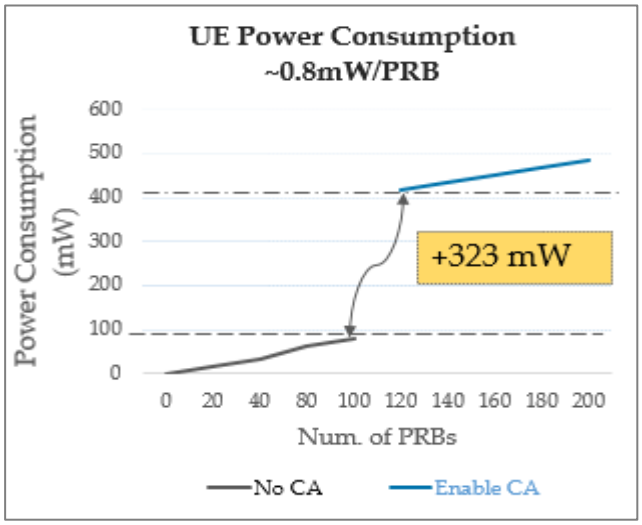

Figure 2.3: UE Power Consumption in CA [30]

In [21], prevailing CA power wastage improvements with smartphone generations, for a high data volume, at 1GB DL data size, compared to 20MHz (no CA scenario), power consumption for 10MHz+10MHz (CA scenario) is evaluated to be higher by 13%.

2.2.3.2.Base Station Power Saving

Current base stations have a distributed architecture in which the radio is divided into two main pieces of equipment: a baseband unit and a radio remote unit (RRU). The baseband unit is the unit that performs radio functions in the digital baseband domain. It can be installed in indoor or outdoor environments. This equipment uses common public radio interfaces (CPRI) and electrical or optical cables to communicate with the radio remote units. Additionally, the radio remote unit modulates and demodulates baseband and RF signals, combines and divides baseband and RF signals, and processes data.

Below are some of the energy saving technologies and mechanisms supported by LTE/LTE-A.

DTX/DRX energy saving -In earlier technologies WCDMA/HSPA, DTX at base stations was not supported. The continuous transmission of pilot signals hindered DRX power saving that is by applying switching off periods on power consuming BS components. In LTE BSs its possible since cell reference signals continuous transmission is not mandatory.

Micro-sleep energy saving -Several short gaps in the data transmissions in the networks traffic patterns, observed even during highly loaded times. These short gaps are utilized to reduce power consumption in the network by quickly putting components into sleep mode and then activating them again before the next transmission occurs. The longer these time gaps are, the more components can be put to sleep, and the lower the energy consumption becomes.

A feature that utilizes these time gaps is known as micro-sleep TX, which puts radio units in sleep mode whenever there is a time gap [10].

In the 4G standard, these time gaps where micro-sleep TX and other energy saving features can operate are always very short, at the most 0.2ms. This is because the 4G LTE standard mandates that a large amount of synchronization signals, reference signals, and system information should always be transmitted over the entire bandwidth. These signals are needed to secure cell coverage and a good connection with users. However, they also limit possible energy savings we can achieve with LTE.

Capacity cell switch off -The power efficiency in the infrastructure and terminal is an essential part of the cost related requirements in LTE-Advanced. There was a strong need to investigate possible network energy saving mechanisms to reduce carbon emission and OpEx of mobile network operators. The basic method is to partly switch off eNodeBs, which cover the same area, when capacity is not needed, e.g. during night times. 3GPP conducted a study on possible solutions and concluded that OAM-based approach and signaling-based approach, as well as hybrid approaches, are feasible, applicable and backward compatible for improving energy efficiency. This includes the enhancement to inter RAT in 3GPP Release 11 which needs information transfer over S1 interface to switch off LTE/LTE-A cells where basic coverage is provided by legacy networks.

The RF unit is the highest transmission power consuming BS component [12], [31]–[33] [34] and discussed separately.

2.2.3.3.RF Unit Energy Consumption

With the increasing number of cells in macro BS in the network, the increased volume of traffic and the adoption of digital power amplifiers with large PAPR, the traffic-dependent portion of energy consumption will become even more significant.

Power Amplifier (PA) -Since PAs are significant consumer of power in a wireless system, so amplifier efficiency is important. The PA consumes most of the power of the RRU.

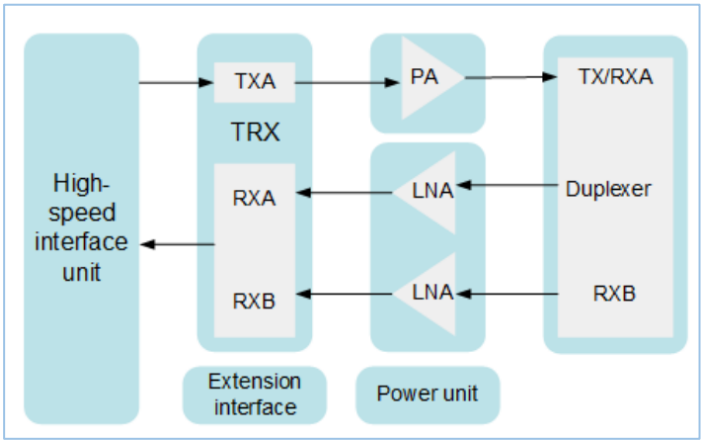

Figure 2.4: PA in RF Unit [8]

RF PA Efficiency - Modern digital modulation methods such as OFDMA in LTE and beyond often result in signals that vary considerably in instantaneous power. For a typical OFDMA signal, the instantaneous power level of a signal with high peak power compared to its average power, the PAPR of the waveform is expressed in decibels. Amplifier efficiency tends to drop at low power levels. The PA must handle the peak power, but the signal's average power mostly determines its practical power efficiency. Thus, high PAPR tends to low performance of PAs and thereby RF chains.

Having the above idea on high PAPR in OFDM symbols and the increasing power inefficiency of PAs in RF chains at low loads discussed in [31], we can have general understanding for LTE/LTE-A or any of same modulation shown below in better illustrative way.

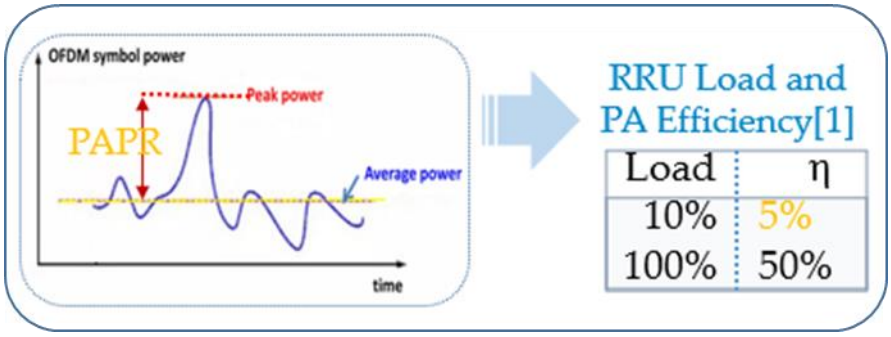

Figure 2.5: PA performances for different RRU loads [31]

Therein the authors addressed for PA performances under different RRU loads. For an RRU load of not more than 10%, the PA efficiency drops to 5%. The best performance at full-load reaches 50%. This inefficiency impact can be considered how it summed up to worsen for multi-carrier and multi-band deployment scenarios such as inter-band CA employed networks operating at their traffic loads.

Targeting the green energy-saving goal for their base stations vendors striving for continuously improved PA efficiency. To achieve this goal, vendors have developed the RRU PA Efficiency Improvement feature to save the power consumed by PA in RRU with a new feature which uses the new-generation.

PA predistortion -PA predistortion algorithm is implemented with higher complexity for RRUs to further improve multi-carrier PA efficiency. After this feature is enabled, the TRX software runs the new PA predistortion algorithm to improve the PA efficiency and save power consumption. Based on a vendor evaluation report, the PA predistortion algorithm reduces RRU power consumption by 2% to 15%. Thus, using the RRU PA efficiency improvement feature on RRUs that support this feature conserves energy and reduces emissions.

Intelligent carrier shutdown -Intelligent carrier shutdown reduces the energy consumption of LTE base stations when the network is under light or no load. However, the energy saving gains are less-than desirable in two scenarios. First, if multiple carriers share the same PA of an RF module, the PA cannot be shut down because the other carriers are still working. Secondly, hardware faults may occur if the temperature differential is excessively large before and after the RF module enters the carrier shutdown state.

2.2.4. Energy Consumption Trend: Operator Data

For operators running hundreds or thousands of capacity-enhanced eNBs, energy billing reduction is inevitable. We also assessed further to address the impacts the non-renewable power supply contribution that incur more cost and carbon emissions.

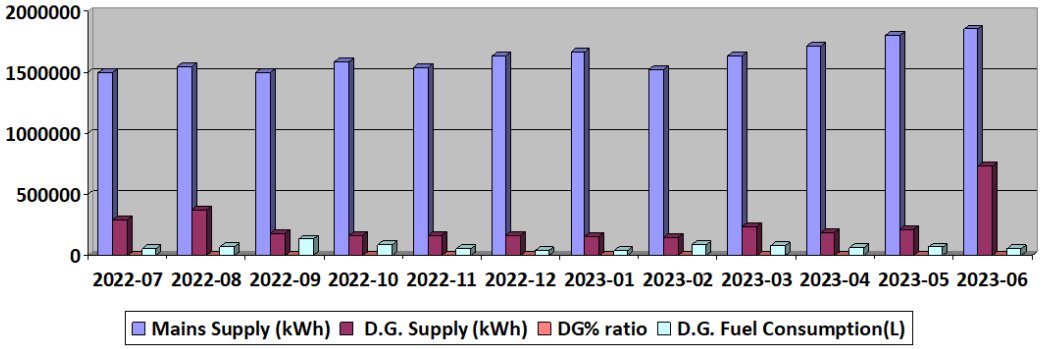

Figure 2.6: Base station power consumption trend [9]

As *Figure 2.6* depicts the annual power consumption trend report in the time period for mobile sites running in the capital city, Addis Ababa, Ethiopia, where the mains power supply is generated from a hydro-electric, renewable source. In such environment, carbon emission is not a concern; instead, operators focus on energy billing reduction. To withstand service outage and increase their quality of service and experience, diesel generators, non-renewable, commonly provisioned to take over mains power outages.

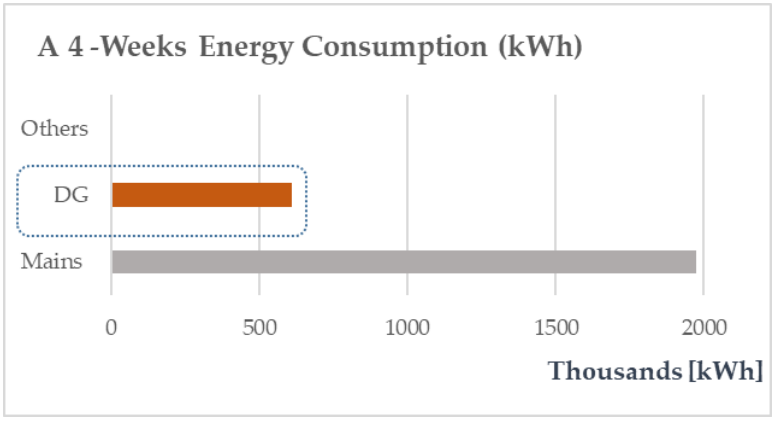

Figure 2.7: Power supply shares based on operator report [9]

Based on our survey findings, the non-green monthly power supply share reaches to 28%. Equivalently, an annual average power supply reported to be of 13.44% from diesel generator. This incurs higher fuel cost and contributes significant carbon emissions.

2.2.5. Network Traffic Trend

Network traffic refers to the amount of data or communication activity on the network at a given time. It is typically measured in terms of volume (the total amount of data transmitted) or density (the amount of traffic relative to the capacity of the network to

carry it). In a cellular network, traffic can come from a variety of sources, such as voice calls, text messages, and data services. It is video content dominated data service that consumes the highest resources share.

Resources refer to the various components and capabilities of the network that are used to support traffic, such as bandwidth, power, and spectrum. These resources are limited and must be carefully managed and allocated to support the desired level of traffic and Grade of Service.

Mobile network traffic has almost doubled in just 2 years, from 66 EB per month in first quarter 2021 to 126 EB in first quarter of 2023 [35]. Addressing the high data traffic growth by itself may not a big concern for researchers. However, deeper understanding on the characteristics of its trends can help to devise for optimal resource provisioning. In this work, we have closely analyzed on realistic data traffic DL for LTE-A network. First, the degree of traffic dependency to the PRB usage is analyzed, and secondly, the busy hour diversities in among eNBs for the feasibility of resource distributions is assessed.

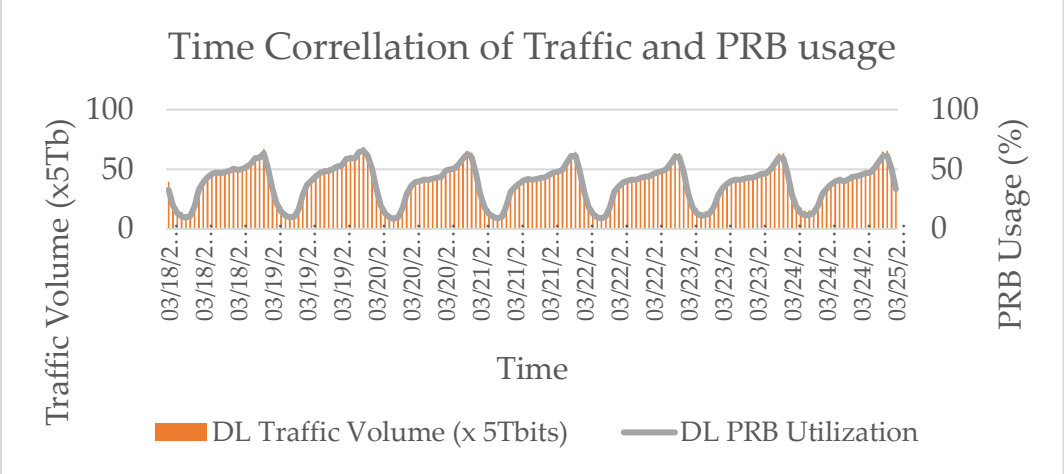

Figure 2.8: One-week plots for LTE/LTE-A traffic volume and PRB usage [4]

In our analysis, the downlink traffic volume and resource blocks usage are 99.48% correlated, one-week data in *Figure 2.8*. Thus, prior information for PRBs usage performance captures the traffic volume without loss of generalization for our traffic-aware problem solving to propose energy saving solutions.

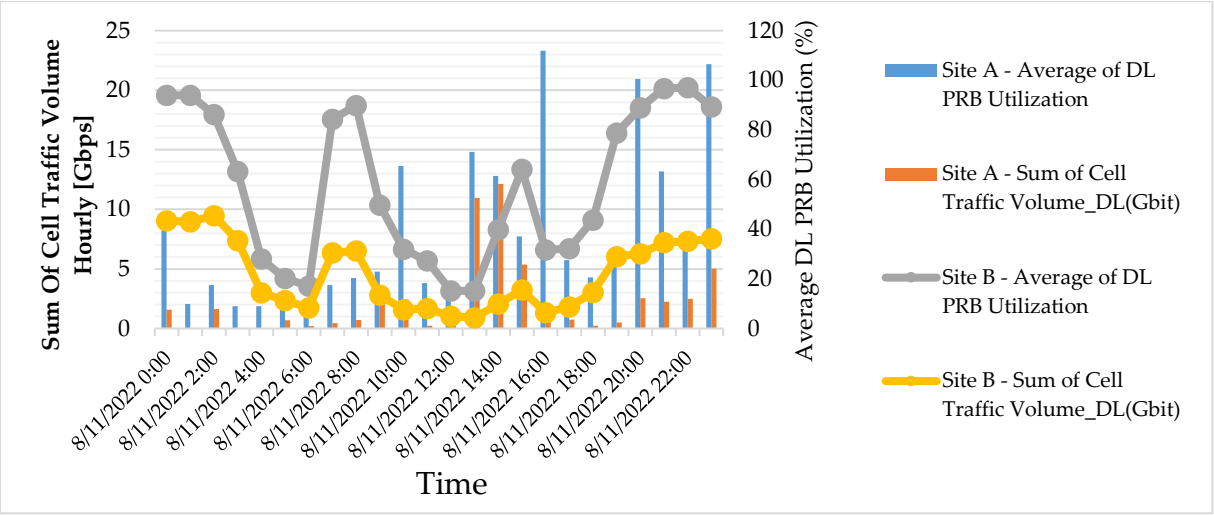

Figure 2.9: eNBs with uncorrelated temporal traffic loads [4]

Further, sites have different peak-hours. It is natural for residential and office or market area sites to have uncorrelated daily performance. The performances plot in *Figure 2.9* is from an operator with two acknowledged peak hours. Redistributing soft resources depending on proactive decisions for deactivated and unutilized duration can improve the network capacity and reduce operator's soft license expenses.

3. Markov Decision Process and Reinforcement Learning

3.1. Markov Decision Process

MDP is based on Markov property that next states rely on the current state or independent on the history.

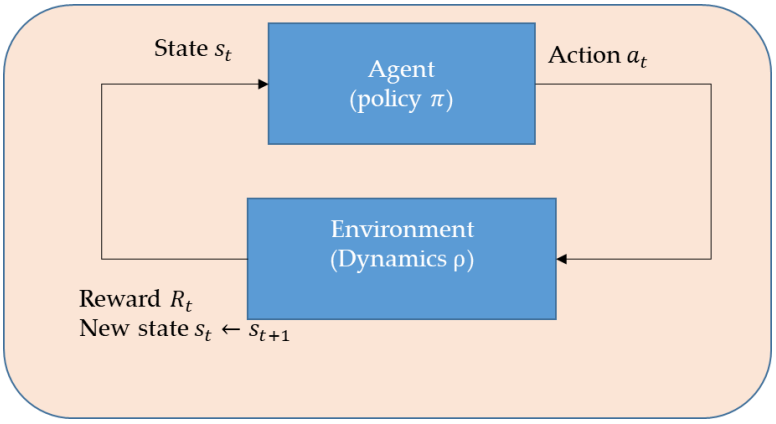

Figure 3.1: Agent-Environment interaction [36]

If the state signal has the Markov property, then the environment's response at t + 1 depends only on the state and action representations at t, in which case the environment's dynamics can be defined by specifying only [37].

$$p(s', r|s, a) = \Pr\{R_{t+1} = r, S_{t+1} = s'|S_t, A_t\} \quad (3.1)$$

for all r, s', S_t and A_t.

In this case, the environment and task as a whole are also said to have the Markov property.

If an environment has the Markov property, then given the current state and action, its one-step dynamics enables us to predict the next state and expected next reward. One can show that, by iterating this equation, one can predict all future states and expected rewards from knowledge only of the current state as well as would be possible given the complete history up to the current time. It also follows that Markov states provide the best possible basis for choosing actions. That is, the best policy for choosing actions as a function of a Markov state is just as good as the best policy for choosing actions as a function of complete histories.

The Markov property is important in reinforcement learning because decisions and values are assumed to be a function only of the current state. In order for these to be effective and informative, the state representation must be informative.

The assumption of Markov state representations is not unique to reinforcement learning but is also present in most if not all other approaches to artificial intelligence.

However, dynamic programing and Reinforcement learning are the two common optimization algorithms in order to solve MDP problem with higher state-action spaces, the transition probability model for all states is unknown.

3.2. Dynamic Programing for Optimization

Dynamic programming is an optimization technique that breaks down a problem into smaller sub-problems and solves them recursively. It requires a mathematical model of the problem, including the state space, the action space, and the transition probabilities.

It can be used to find optimal solutions for problems with multiple objective functions. However, it often finds sub-optimal solutions for problems with large action and state spaces, or problems with dynamically changing transition probabilities. It becomes non-tractable for higher dimensions and continuous states or actions.

3.3. Reinforcement Learning and Optimization

3.3.1. Overview

To begin with, machine learning refers mainly three paradigms, namely; supervised learning, unsupervised learning and reinforcement learning.

a. ***Supervised Learning*** –use of labeled dataset for known actions to which conditions to generalize or extrapolate and accurately act on new dataset.
b. ***Unsupervised Learning*** –finding hidden structure from unlabeled dataset.
c. ***Reinforcement Learning*** –often for interactive problems which is impractical to obtain examples of desired behavior that are both correct and representative of all

the situations in which the agent has to act. The trade-off between exploration and exploitation which refer the agents ability to try diverse actions and the right action respectively, not in other paradigms, is a challenge in RL [36].

RL, as described in [36] involves learning how to make decisions in order to maximize a numerical reward. The learner is not provided with a predetermined set of actions, but instead must discover which actions yield the highest reward through trial and error. In complex scenarios, actions can have an impact not only on immediate rewards but also on future situations and subsequent rewards. These two key characteristics, trial-and-error search and delayed reward, distinguish reinforcement learning.

The use of artificial intelligence or machine learning, such as RL algorithms, is the preferred approach for optimizing more complex and higher dimensional MDP problems. Reinforcement learning is both a problem and a set of solution methods that are effective for addressing this problem.

Markov decision processes aim to encompass these three aspects - perception, action, and goal - in their simplest forms without oversimplifying any of them. Any method that is well-suited for solving such problems is considered a reinforcement learning method.

In RL, an environment is usually the way we refer to a simulator or a control system. Various libraries provide simulation environments for RL, including Gymnasium (previously OpenAI Gym), DeepMind control suite, and many others.

In the last decades, RL has obtained remarkable success in several fields, including autonomous driving, robotic locomotion, and video games, to mention a few [38].

RL is how the agent observes the simulation or real environment and chooses an action that maximizes the cumulative future reward. The optimal control using reinforcement learning has been studied mostly in the area of game playing and real-world control systems [39].

The goal and purpose of RL is maximization of the expected value of the cumulative sum of a received scalar signal (called reward). The use of a reward signal to formalize the idea of a goal is one of the most distinctive features of reinforcement learning.

The rewards, which define of the goal of learning, are computed in the environment rather than in the agent.

The agent tries to select actions so that the sum of the (rewards discounted to avoid infinite cumulative rewards for infinite episode length) rewards it receives over the future is maximized. In particular, it chooses A_t to maximize the expected discounted return (G_t):

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \qquad (3.2)$$

Where: γ is a parameter, $0 \leq \gamma \leq 1$, called the discount rate.

The discount rate determines the present value of future rewards: a reward received $\boldsymbol{k}$ time steps in the future is worth only $\boldsymbol{\gamma^{k-1}}$ times what it would be worth if it were received immediately. If $\boldsymbol{\gamma} < 1$, the infinite sum has a finite value as long as the reward sequence $\{R_k\}$ is bounded. If $\boldsymbol{\gamma} = 0$, the agent is "myopic" in being concerned only with maximizing immediate rewards: its objective in this case is to learn how to choose A_t so as to maximize only R_{t+1}. If each of the agent's actions happened to influence only the immediate reward, not future rewards as well, then a myopic agent could maximize, Equation (3.2), by separately maximizing each immediate reward. But, in general, acting to maximize immediate reward can reduce access to future rewards so that the return may actually be reduced. As $\boldsymbol{\gamma}$ approaches 1, the objective considers future rewards more strongly: the agent becomes more farsighted. It is essential to define some of the functions associated to RL.

A) Value Functions

Almost all RL algorithms involve estimating value functions, functions of states (or of state-action pairs) that estimate how good it is for the agent to be in a given state (or how good it is to perform a given action in a given state). The notion of "how good" here is defined in terms of future rewards that can be expected, or, to be precise, in terms of expected return. Of course, the rewards the agent can expect to receive in the future depend on what actions it will take. Accordingly, value functions are defined with respect to particular policies.

Recall that a policy $\boldsymbol{\pi}$, is a mapping from each state, $\boldsymbol{s} \in \boldsymbol{S}$, and action, $\boldsymbol{a} \in \boldsymbol{A}(\boldsymbol{s})$, to the probability $\boldsymbol{\pi}(\boldsymbol{a}|\boldsymbol{s})$ of taking action a when in state $\boldsymbol{s}$. Informally, the value of a state $\boldsymbol{s}$ under a policy $\boldsymbol{\pi}$, denoted $\boldsymbol{v_\pi}(\boldsymbol{s})$ is the expected return when starting in s and following $\boldsymbol{\pi}$ thereafter. For MDPs, we can define $\boldsymbol{v_\pi}(\boldsymbol{s})$ formally as [37]:

$$v_\pi(s) = E_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = s\right] \tag{3.3}$$

Where $\boldsymbol{E_\pi}[\cdot]$ denotes the expected value of a random variable given that the agent follows policy $\boldsymbol{\pi}$, and $\boldsymbol{t}$ is any time step. Note that the value of the terminal state, if any, is always zero. We call the function $\boldsymbol{v_\pi}$ the state-value function for policy $\boldsymbol{\pi}$.

Similarly, we define the value of taking action $\boldsymbol{a}$ in state s under a policy $\boldsymbol{\pi}$, denoted $\boldsymbol{q_\pi}(\boldsymbol{s}, \boldsymbol{a})$, as the expected return starting from $\boldsymbol{s}$, taking the action $\boldsymbol{a}$, and thereafter following policy $\boldsymbol{\pi}$:

$$q_\pi(s, a) = E_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \mid S_t = s,\ A_t = a\right] \tag{3.4}$$

We call q_π the action-value function for policy π.

The agent would have to maintain $\boldsymbol{v_\pi}$ and $\boldsymbol{q_\pi}$ as parameterized functions and adjust the parameters to better match the observed returns. This can also produce accurate estimates, although much depends on the nature of the parameterized function approximator.

A fundamental property of value functions used throughout reinforcement learning and dynamic programming is that they satisfy particular recursive relationships. For any policy $\boldsymbol{\pi}$ and any state $\boldsymbol{s}$, the consistency condition, Bellman equation for $\boldsymbol{v_\pi}$,[36] holds between the value of s and the value of its possible successor states. The value function $\boldsymbol{v_\pi}$ is the unique solution to its Bellman equation.

B) Optimal Value Functions

Solving a reinforcement learning task means, roughly, finding a policy that achieves a lot of reward over the long run. For finite MDPs, we can precisely define an optimal policy in the following way. There is always at least one policy that is better than or equal to all other policies. This is an optimal policy. Although there may be more than one, we denote all the optimal policies by $\pi*$. They share the same state-value function, called the optimal state-value function, denoted v$*$, and defined as:

$$v_*(s) = \max_\pi v_\pi(s) \qquad (3.5)$$

for all s ∈ S.

Optimal policies also share the same optimal action-value function, denoted q_* and defined as:

$$q_*(\mathrm{s},\mathrm{a}) = \max_\pi q_\pi(s,a), \qquad (3.6)$$

for all s ∈ S and a ∈ A(s). For the state-action pair (s, a), this function gives the expected return for taking action a in state s and thereafter following an optimal policy. Thus, we can write q_* in terms of v_* as follows:

$$q_*(s,a) = E\,[\,R_{t+1} + \gamma v_*(S_{t+1}) \mid S_t = s,\ A_t = a] \qquad (3.7)$$

C) Replay buffer

Replay buffers are a common building piece of off-policy RL algorithms. In on-policy contexts, a replay buffer is refilled every time a batch of data is collected, and its data is repeatedly consumed for a certain number of epochs.

In solving MDP, finding policy (π) which is the prescription and the outcome of the learning process is to be determined as an output of the learning process and used at inferencing phase. The commonly used RL algorithms have been discussed subsequently.

3.3.2. Q-Learning

It is a model-free RL algorithm i.e. it does not require a complete model of the environment. This makes it more applicable to real-world problems than dynamic programming.

Q-learning works by iteratively updating a table of Q-values, which represent the expected rewards for taking an action in a given state. It may slow for large number states and actions. Q-learning also lacks to generalize for unseen state-action value. The use of deep neural networks in RL helps to overcome such limitations, covered in next discussion.

3.3.3. Deep Reinforcement Learning

While the Q-learning algorithm is extremely helpful, it does have some drawbacks. Its main disadvantage is that the requirement of the agent visiting all state-action pairs and storing the observed information becomes impossible to satisfy in practice for problems with sufficiently large state and or action spaces.

3.3.3.1. Deep Q Networks

One major change that the Deep Q-Networks (DQN) made over that of the basic Q-Learning algorithm is that of the introduction of a new 'Target-Q-Network'.

$$Q(s,a) = (1-\alpha)Q(s,a) + \alpha(r + \gamma \max_{a^t} Q(s^t, a^t)) \quad (3.8)$$

So essentially, in this equation, the Q-Function $Q(s,a)$ is being referred twice and each of this reference is for different purposes. The first reference i.e. $(1-\alpha)Q(s,a)$ is mainly to retrieve the present state-action value so as to update its value (use of Q as in: Q (s, a) = (1- α) Q (s, a) + …), and the second is to get the 'target' value for the subsequent Q value for the next state-action (i.e. Q as in: $r + \gamma \max_{a^t} Q(s^t, a^t)$). Though in the basic Q-Learning algorithms both these Q Functions/ Networks (or Q-Tables in case of a tabular Q-Learning approach) were same, it may not necessarily be so always.

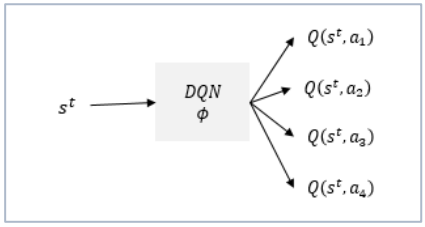

Figure 3.2: DQN with discrete action-state spaces [36]

In DQN the algorithm, there is a distinction between the target Q-network and the one that is continuously updated at each step. This differentiation is implemented to address the limitations associated with using the same Q-network for both continuous updates and referencing target values.

These limitations arise due to two main reasons. Firstly, as mentioned earlier in the basic DQN section, using highly correlated and too-frequent data for training can lead to issues with delayed or sub-optimal convergence. When the targets for training are derived from the same network, they tend to be correlated. Secondly, it is not advisable to utilize target values from the same function to correct its own updates. This can result in an unstable target function.

To enhance the stability of the Q-network, it has been observed that employing two separate Q-networks for these distinct purposes is beneficial. However, if a target action-value is required for training and this target value remains unchanged after initialization (which can even be all zeros in the case of Q-Learning, an off-policy algorithm), the

actively updated network cannot be effectively updated. Therefore, the target Q-network is synchronized and updated with the actively updated Q-network once every T_n number of steps.

Continuous action and state spaces are easy for policy-based methods but awkward or impossible for ϵ-greedy methods and for action-value methods in general [36]. In recent years, attention has returned to actor–critic methods and to policy-gradient methods in general. The two known actor–critic methods have been discussed in subsequent sections.

3.3.3.2. Deep Deterministic Policy Gradient

Deep Deterministic Policy Gradient (DDPG) is a powerful technique in the field of RL that combines the strengths of Q-learning and Policy gradients. As an actor-critic method, DDPG utilizes two models: the Actor and the Critic. The Actor, unlike other policy networks, directly outputs the precise action in a continuous form, rather than providing a probability distribution over actions. On the other hand, the Critic is responsible for estimating the Q-value by taking both the state and action as input.

DDPG is an off-policy method, meaning it can learn from data collected by a different policy. It is specifically designed for environments with continuous action spaces. The term *deterministic* in DDPG refers to the fact that the Actor directly computes the action without the need for an $argmax$ operation or a probability distribution. Instead of relying on $argmax$ over Q-values, the Actor network in DDPG directly outputs the action, even in continuous action spaces.

In summary, DDPG is an RL technique that combines Q-learning and Policy gradients. It excels in continuous action settings by utilizing an Actor network that directly outputs the action, bypassing the need for $argmax$ operations.

3.3.3.3. Proximal Policy Optimization

PPO is a method used in the field of RL that falls under the category of Actor-Critic algorithms. The Actor-Critic system consists of two models: the Actor and the Critic and same role as in DDPG. In PPO, however, the Actor network takes the observation (state) as input and produces a list of probabilities, with each probability corresponding to a specific action. These probabilities form a distribution, and the action is chosen by sampling from this distribution.

The Critic network, on the other hand, takes the state as input and outputs a single number that represents the estimated value of that state.

PPO is a policy-gradient algorithm that involves collecting a batch of data and using it directly to train the policy network. The goal is to maximize the expected return while adhering to certain proximality constraints.

In comparison to other RL algorithms designed for continuous domain environments, PPO has been tested with a clipping function in various environments and has demonstrated superior performance [40]:

$$L^{CLIP}(\theta) = \hat{E}_t[\min(r_t(\theta)\hat{A}_t, clip(r_t(\theta), 1-\varepsilon, 1+\varepsilon)\hat{A}_t)] \quad (3.9)$$

Where: epsilon (ε) is a hyperparameter. The probability ratio, $r_t(\theta) = \frac{\pi_\theta(a_t|\, s_t)}{\pi_{\theta_{old}}(a_t|\, s_t)}$ whose value is unity for $r(\theta_{old})$. Hence, with modification on the second term in Equation (3.9) avoids the incentive for r_t values not in range of $1 \pm \varepsilon$. Depending on the advantage $\hat{A}_t$ is negative or positive, r_t is clipped to $1-\varepsilon$ or $1+\varepsilon$ respectively. Thus, the minimum of the clipped and unclipped objectives be a lower bound or pessimistic objective on the unclipped one.

PPO is usually regarded as a fast and efficient method for online, on-policy reinforcement algorithm [40].

4. DRL-Based BLCOO for Inter-Band CA Energy Saving

In this chapter, we first formulate the operational CA employed LTE-A network energy consumption problem as a sequential decision making problem in infinite horizon. Then we propose to use DRL methods for the solution. We apply Q-learning based DQN and Actor-Critic framework based PPO RL algorithms to train our BLCOO energy saving algorithm and find optimal policies. Evaluate energy saving performances for the two cases and come up with numerical results for next analysis.

4.1. System Model

Our system model is based on urban LTE-A network with indoor users. Inter-band CA enabled urban macro LTE-A network environment with a set of $\mathbb{N}$ serving CCs and bandwidths $\{CC_1, CC_2, \ldots, CC_{\mathbb{N}}\}$ (in MHz) with a capacity to allocate a maximum of 100MHz aggregated transmission bandwidth per user for a realistic network is considered for this work. Seven CA enabled tri-sector eNBs with ISD of $3 \times \mathcal{R}$ and each sector with multiple hexagonal cells. Each cell is configured with either of LTE standard carrier in $\mathbb{N}$ different bands.

Moreover, $\aleph$ number of random static users per sector with CA capability are uniformly distributed to capture spatial impacts of serving network on received signal qualities into account. The six eNBs in the "Honey Comb" layout is to designate the single-tier interference for target eNB (eNB_1) connected users. For a 4G/LTE downlink based on OFDMA and frequency reuse factor unity, the main interference is ICI due to collision of same PRBs allocated simultaneously to users by neighbor eNBs.

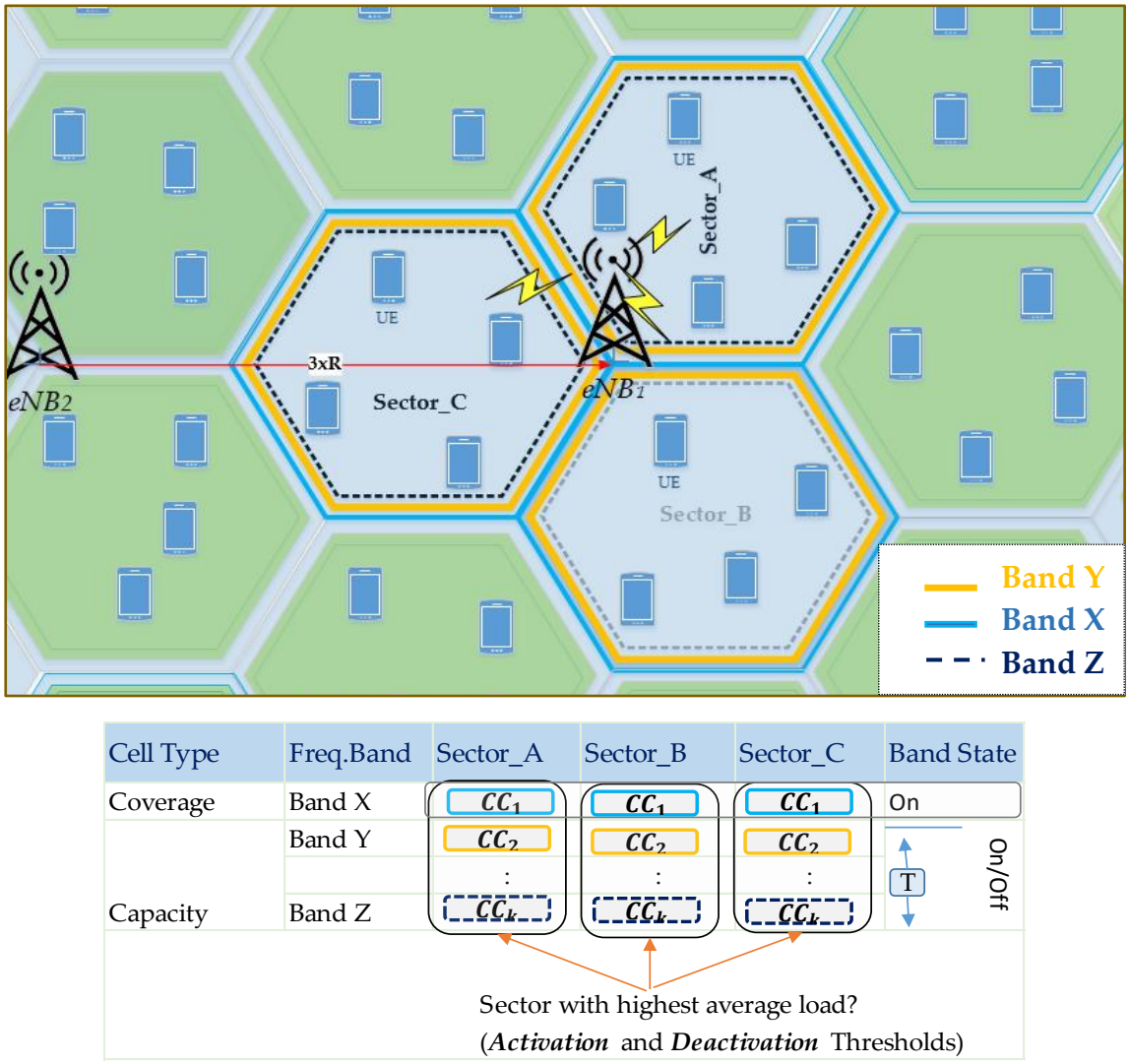

Figure 4.1: Network deployment layout

Moreover, the radio topology is based on deploying an RF unit for each LTE-A cell or CC added. The power consumption of RF units varies for multiple reasons, including the manufacture, working band, version and operating technology. Hence, addressing configuration-specific RF unit power consumption trends should be a critical task to bring a better energy reduction solution.

4.2. RF Power Measurement and Modeling

Adopting previous studies on the linear dependency of an active RF unit dynamic power consumption to traffic load [7], [41] and to PRBs usage [12], we built our own configuration specific power models.

The traffic dependent power consumption of remote radio units (RRUs) which we are also referring them radio frequency (RF) units or modules, operating in 4G mode ($P^{4G}{}_{RRU}$) is also studied in [42] asserting the linear dependency with the number of utilized PRBs (l^{rb}).

$$P^{4G}{}_{RRU} = \alpha_0 + \alpha_1(l^{rb}),\ l^{rb} \in [0,\ 100] \qquad (4.1)$$

Where, the first component refers the idle state power consumption in zero traffic plus active PA while the second component denotes the transmission power consumption.

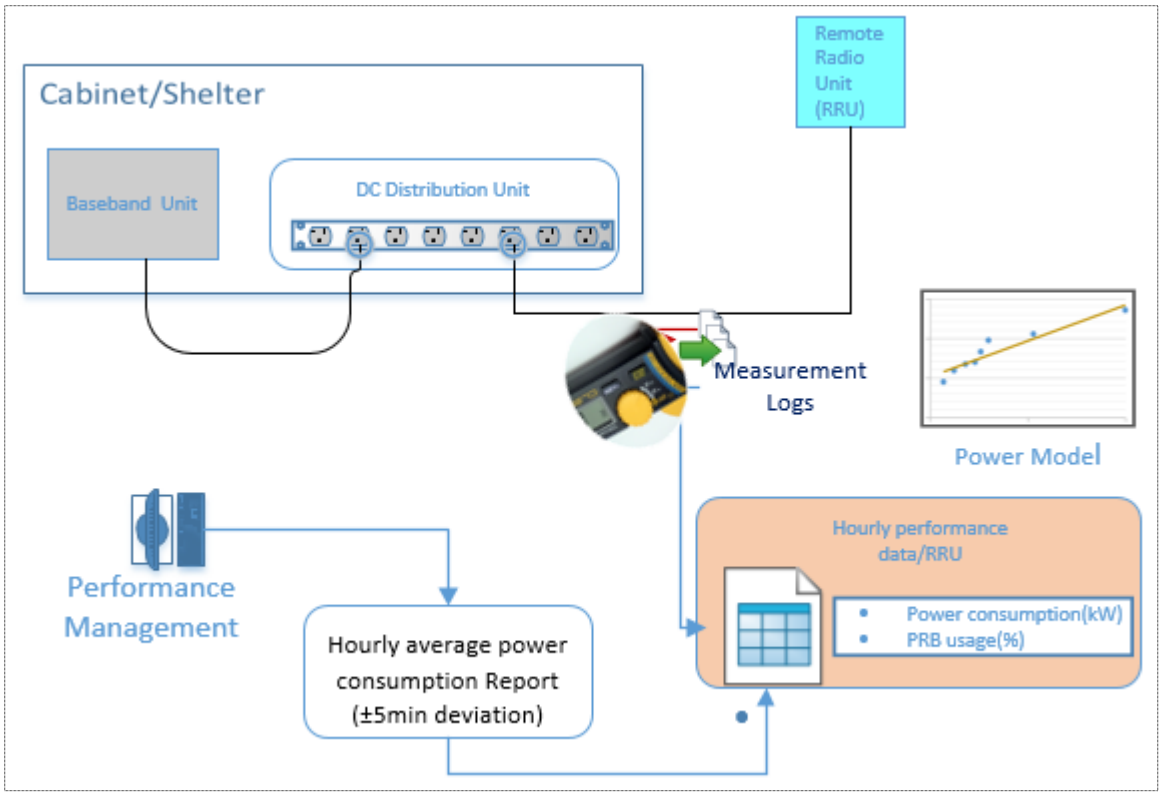

Figure 4.2: RF power measurement setup and modeling

The operational data for energy consumption performance of RF modules (kWh) is the average of 12 samples reported at 5 minutes intervals. Reported values delayed by a maximum of 5 minutes introduce some in accuracy or deviations on hourly performances.

To narrow such deviations, we have conducted on-site measurements by increasing sampling resolution time. A random sample per 2-minutes interval or 30 samples tally for five consecutive hours is incorporated. We have ensured the time deviation between

our records time reference and the network management system is less than a minute and almost no impact for our 2-minutes sampling intervals.

Moreover, PRB utilization performance, which is also highly correlated to the traffic volume as presented in our analysis for a one-week daily performance revealing 99.8% correlation, is used for the corresponding dynamic power consumption dataset preparation.

A simple linear regression implemented to model power consumption for RRUs operating on different bands and specific configurations. R-square is used in evaluating the accuracy of the model.

4.3. MDP Problem Formulation

Our optimal number of CCs selection based on the expected load and current states of eNB cells for state changes and save energy at low load conditions, is formulated as an MDP problem with unknown transition probabilities. An MDP problem is defined with tuples of $\{S, A, \rho, R\}$.

Where:

- S -is the set of eNB states perceived by the energy saving controller
- A -is the set of eNB triggering actions implemented by eNB
- $\rho(s'|s,a)$ -is the state transition probability function, which gives the probability of transitioning from state s to state s' after taking action a
- $R(s,a)$ -is the reward function, which gives the reward received after taking action (a) in state (s)

All relevant terms are explain as follows.

A) **State** –Our state is the observation from the eNB at each time step and for the decision making or action selection for optimal set of serving bands. We design it as a vector consisting of sectors hourly cumulative load and the current activation status of CCs. Thus, the set of information or metrics $(\mho)$ required for decision at each time step(t):

$$\mho = \{ m,\ l_{s1}^{t}, l_{s2}^{t}, l_{s3}^{t} \} \quad (4.2)$$

Where: $m \in \{0,1,\dots,N-1\}$

Instead, current decisions shall to have the expected load information. To this end, we use weighted moving average where the most recent past is higher weighted than old ones. In addition, we determined the five most recent historical data of the base station to influence our expected load. Different window sizes can be selected. The expected load for time $(t+1)$ is denoted by $\left(l_s^{t+1}\right)$, and WMA of window size 5 is used in this work.

For N number of bands values ranges my scale differently. We better encoded them as binary vectors. Below is one hot encoded binary mapping for eNB enabled CA with a maximum of three bands. A maximum 8 serving bands can be represented by three bits. Current states of all cells and serving bands encoded as binaries. Three binary-bits can denote all our band states.

Table 4.1: One –hot encoded state vector for N value 3

Band State (m)	*Number of Active Bands (N)*	*One-hot Encoding (m)*	*Available PRBs /Sector*
0	1	[1 0 0]	[50, 0, 0]
1	2	[1 1 0]	[50, 100, 0]
2	3	[1 1 1]	[50, 100, 100]

B) **Action** -Each decision in selecting the optimal number of CCs based on the observation vector.

C) **Reward Design** -Each possible action for the state and transition probabilities is assigned a value of rewards or cost. Our reward (negative cost) function is defined as a decreasing function of energy consumption and an increasing function of target throughput, a QoS measure. A third metrics CC usage denotes the percentage of the ratio of the average number of serving bands in our optimized network to the total available ones is included. Low CC usage indicates higher reduction in user devices multiple CCs monitoring power. Thus, the three reward functions are formulated as follows.

a. Optimal number of CCs reward function (r_{cc}^t): CC usage metric denotes the percentage of the ratio of the average number of serving bands in our optimized network (N_{opt}^t) to the total available (N) is included. Low CC usage indicates higher monitoring power reduction for user devices.

$$r_{cc}^t = 1 - \frac{N_{opt}^t}{N} \tag{4.3}$$

b. Power reward function (γ_p^t):

$$\gamma_p^t = \frac{P_{RRU,\ Total}^t}{P_{RRU,\,Coverage\ band}^t} \tag{4.4}$$

Where:

- $P_{RRU,\ Total}^t$ –power summed up from all active RRUs including coverage and capacity cells.
- $P_{RRU,\,Coverage\ band}^t$ –refers the sum of minimum power consumption from coverage or always-on RRUs. By minimizing the normalized power reward value means maximizing network energy saving.

The maximum power reward is designed to be at all capacity cells off state while coverage cells are active but zero traffic.

c. CA throughput maximization reward function (r_{thp}^t): Maximizing CA users' average throughput (Tp_{av}^t) normalized to the target threshold throughput (Tp_{target}^t) avoids greedy power saving during high traffic loads. In prioritizing capacity over power saving, additional precondition set by penalizing power saving when average throughput is less than target.

$$r_{thp}^t = \begin{cases} \frac{Tp_{av}^t}{Tp_{target}^t}, & Tp_{av}^t \geq Tp_{target}^t \\ \log(\frac{Tp_{av}^t}{Tp_{target}^t}), & Otherwise \end{cases} \tag{4.5}$$

The use of log(.) function avoids discontinuity of r_{thp}^t values with a zero penalty for the achievable rate (Tp_{av}^t) equals to our target throughput (Tp_{target}^t) and negative increasing values or penalties as achievable rate is degrading below target.

To this end, the immediate reward value (r_{total}^{t}) at each execution time step is a weighted sum of rewards computed from the above three functions.

$$r_{total}^{t} = \alpha r_{p}^{t} + \beta r_{thp}^{t} + \zeta r_{cc}^{t} \quad (4.6)$$

Where, weights α, β and ζ can be implementation specific values.
The ultimate goal is to minimize energy consumption cost during light loads, guarantee traffic capacity demand and minimize the number of always-active serving bands.

In order to activate and deactivate band for all sectors, the optimal decision selection observes the sector highest load state. This is achieved by collecting experiences via trial and error of self-learning algorithms to find the optimal policy ($\pi^{*}(s|a)$).

To solve MDP problem for a network with uncertain state transitions or hard to determine transition probabilities, RL is now adays a preferred approach for its trial-and-error self-learning performance [36]. The RL agent iteratively interacts with the environment to collect more experiences and maximize reward until convergence.

4.4. Network Simulation Environment

In using self-learning approaches to solve problems of MDP type, the agent need to interact with the environment and collect experiences by trial-and-error before deployment. However, its random actions can have irreversible catastrophes if applied on realistic network.

To avoid such risks, general-purpose discrete-event-based simulators or customized simulation environments [38] is feasible for the control agent to interact with by sending random action and to get observation for the next optimal action selection. Higher overheads and environment complexity are challenges when using general-purpose environments.

Building customized environments replicating specific scenarios are better suited to solve specific problems. This enables to customize all aspects of the simulation

environment with low overhead, to experiment with different scenarios and address specific use cases.

Simulation tools NS3 and 5G discrete time simulator have been used in [15] and [14] respectively to generate data input and build environment. Such approaches lack to mimic the realistic network traffic pattern. Whereas in [13] problem specific customized simulator is used to replicate real network environment.

In this work, realistic data from a cellular operator is used as an input to build our custom simulator. Urban macro LTE-A network employing inter-band CA composed of target eNB with randomly distributed users and six neighboring eNBs as a source of ICI has been built for interaction with the RL agent i.e. the optimal set of CCs controller.

Moreover, configuration parameters from a realistic network ISD, always full-load scenario CA serving cells configuration carriers and bands, and others as summarized in *Table 4.2* have been adopted to replicate the real network scenario. The maximum bandwidth each eNB can allocate to a user is 50 MHz or 150 PRBs of which 25% of the total PRBs are allocated for control channels signaling overheads for this work.

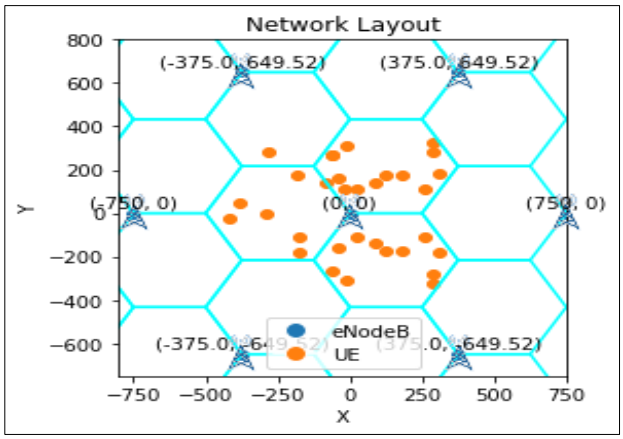

Figure 4.3: Network user layout

In this simulation setup, each user is assumed to have CA capability and can exploit all CA benefits thus each connected user performs multiple SINR measurements, mapped to CQI values and reported for all aggregated bands. Each reported CQI is mapped to modulation schemes and coding rates according to 3GPP standards where the highest

supported modulation scheme is upgraded from 64QAM in R10 to 256QAM in R12. Based on our assessment, it is already implemented in ethio telecom operational network. DL PRBs are allocated from each active CC in a round robin fashion. Of course, the presence of CA non-capable LTE-A users in the actual environment does not affect our generalization of users' capability for this simulation setup. Non-CA users can be perceived as CA capable users but experiencing bad radio conditions from some of their radio links or CCs. We have used full buffer traffic type which is more convenient to static user scenarios.

Thus, based on the triggering information from the controlling agent at each iteration time step, the environment updates traffic load from dataset, activate/deactivate cells, evaluate new RF power consumption based on new load, compute user level SINR and aggregated throughput. Traffic load split and resource scheduling are not specified by 3GPP and implementation specific operations at MAC layer. Thus, at each interaction time step and based on the action triggered by the energy saving RL agent, our custom environment executes or considers the following tasks:

a. ***Load Split:*** When a CC is deactivated at low traffic load, the user traffic is to be redistributed among active CCs/ bands. For CA scenario with all CCs equal bandwidths and balanced loads, the offloaded traffic can be shared equally. However, this is highly relaxed and ideal. To consider for flexible bandwidths and load imbalance among CCs in the realistic case, the split ratio is designed to be based on their available PRBs and pre-offloading loads.
b. ***SCC Activation/Deactivation:*** Once a capacity cell is switched off, no more signal measurement reports from users to upper layers. Thus, the MAC CE messages sent to UEs for SCC deactivation followed by RRC signaling for SCC removal. The deactivated LTE band/carrier will remain out of CA set of serving bands and not available for schedulers for the whole ***T*** time step period. To the reverse, for a cell switched on, SCC addition and activation messages sent by upper layers. In this aspect, no changes on the existing resource allocation and

scheduling mechanisms except the number of available CCs varying for optimal case.

c. ***Evaluate Metrics:*** Each metrics are computed based on the action event. User throughput(TP^t), number of CCs saved, RF power consumption are computed either to respond a real number value reward (r^t) to the agent at training phase, or to evaluate our proposed energy saving performances at inference phase.

Table 4.2: Network simulation parameters

Parameters	**Value**
Network Scenario	Urban Macro
CA Type / Number of CCs	Inter-Band FDD / 3
LTE Carrier Frequency	800 MHz, 1800MHz, 2600 MHz
Bandwidth	10 MHz, 20 MHz, 20 MHz
Subcarrier Spacing	15 KHz
eNB TX Power	40 dBm, 43 dBm, 46 dBm
Cell Radius	250 m
UE TX Power	23 dBm
Number of UEs/Mobility	30/ Static (UE Random distribution)
Height (BS/UE)	35/1.5 (meters)
BS Antenna Gain/ Pattern(Horizontal)	18 dBi $A_H(\theta) = -\min[12\left(\frac{\theta}{\theta_{3dB}}\right)^2,\ A_m],$ $\theta_{3dB} = 65\ degrees,\ A_m = 20\ dB$
UE Antenna Gain	0 dBi
Propagation Model	COST-231 Hata
Resource Scheduler	Round Robin
Penetration loss/Noise Figure	20 dB / 7 dB
Channel Model	AWGN
Thermal Noise PSD	-174 dBm/Hz

4.5. DRL-Based BLCOO Energy Saving

Our RL based energy saving setup general network architecture is built as follows. The $\widehat{\boldsymbol{SNIR}}^{\boldsymbol{t}}$ is SNIR values matrix ($N_u \times N$) at a time t representing SINRs measured for N number of serving CCs by N_u number of users.

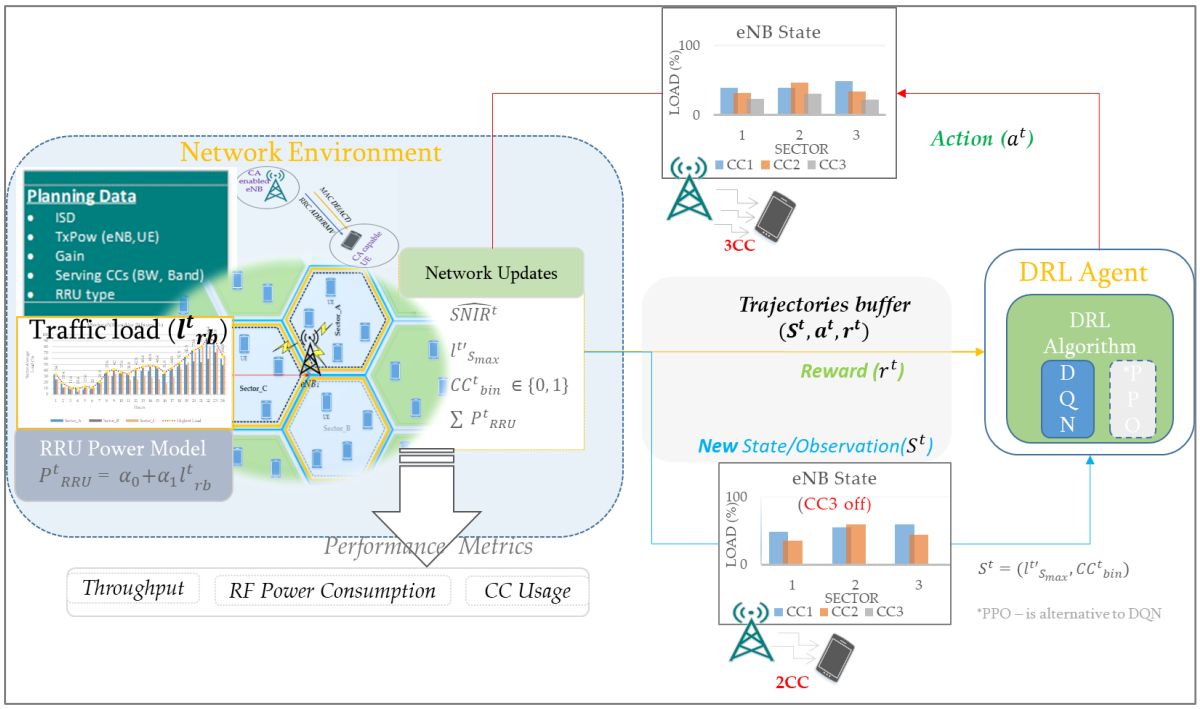

Figure 4.4: DRL based BLCOO energy saving architecture

The expected load ($l^{t\prime}_{S_{max}}$) is computed as a weighted moving average (WMA) of past experiences, where the most recent history is weighted higher and oldest one takes the least contribution in determining the expected value.

$$l^{t\prime}_{S_{max}} = f(\, l^{t-(w-1)}_{S_{max}}, \ldots, l^{t-1}_{S_{max}}, l^{t}_{S_{max}}\,) \tag{4.7}$$

Where: $f(.)$ computes WMA and w denotes for sliding window size.

The maximum load among the three sectors at a time step t is $l^t_{S_{max}}$ is computed as:

$$l^t_{S_{max}} = \max(l^t_{S_1}, l^t_{S_2}, l^t_{S_3}) \tag{4.8}$$

Our network environment, discussed in next section, is built based on configuration data and performance data of real base station as input. Each time triggering signal from the RL agent, it updates the new set of serving bands, power consumption, per user and sum throughputs.

In solving our MDP problem, different RL algorithms are adapted depending on discrete and continuous state scenarios.

4.5.1. RL Algorithms Selection and Training

To solve our energy saving problem as an MDP problem with uncertain conditions, different RL algorithms can be chosen but the problem type matters. Q-learning is known to be the basic RL algorithm which works by iteratively updating a table of Q-values, which represent the expected reward for taking each action in each state. However, Q-learning has limitations for problems with a large number of states and actions, and limitation for unseen conditions.

To overcome these limitations, we used DQN, an extension of Q-learning algorithm that uses a deep neural network to approximate the Q-values. DQN has been shown to be effective for solving problems with a large number of states and actions, and for problems with uncertain conditions. It is a powerful and versatile RL algorithm that has been shown to be effective for a variety of problems. We also chose DQN because it is relatively easy to implement and can be used with a variety of environments.

We chose Bayesian optimization to tune the hyperparameters of the DQN algorithm. Bayesian optimization is a powerful and efficient technique for tuning hyperparameters by iteratively evaluating the algorithm with different hyperparameter settings.

In our performance evaluation using Deep Q-learning, to adapt its limitation for network traffic load continuous values, we have designed loads as discrete values in range of 0 to 100. The DQN target network weights updated after the Q network of batch size 64. The diversities in the samples of our eNB states at each training step improved the exploration of different actions.

We trained the DQN algorithm using our custom environment interaction. We trained the algorithm until it converged, which means that the Q-values no longer changed significantly. We then evaluated the performance of the DQN algorithm by plotting the learning curves.

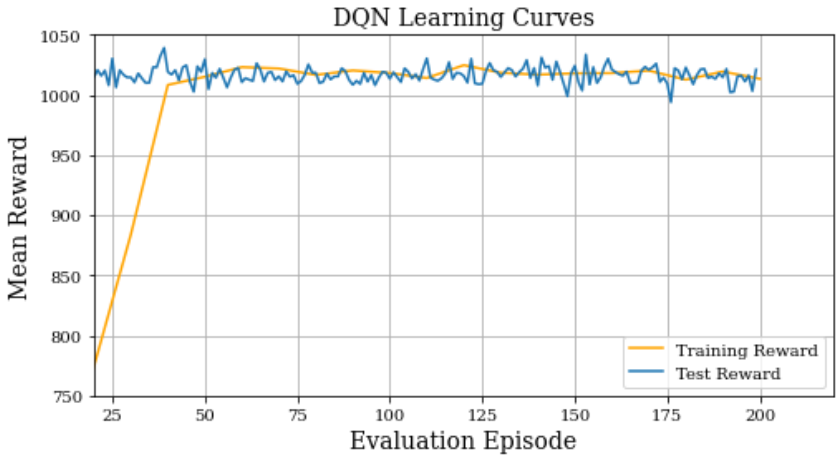

Figure 4.5: DQN training performance

The learning curves showed that the DQN algorithm was able to learn a policy that maximizes the reward within 200 training episodes. We used the learned policy to control the energy saving system.

Second approach is to design our network load states as continuous from 0.0 to 1.00 infinite number of states, which was not supported by the previous RL algorithm. The two Actor-Critic RL framework based algorithms PPO and DDPG supports continuous values of the eNB loads. However, DDPG is applicable for continuous actions whereas our environment has discrete actions and continuous states. In this aspect, to choose PPO is the convenient one.

PPO is Actor-Critic RL framework based algorithm for better learning performance. Actor-Critic uses two deep learning models, one called Actor model and the other called Critic model. The Actor model performs the task of learning what action to be selected under a particular observation of the environment (i.e., the control policy). When the action selected by the Actor model is performed, the agent gets a reward from the environment. This reward is taken in by the Critic model.

The role of the Critic model is to learn to evaluate if the action taken by the Actor model led the environment to be in a better state or not, and its feedback is used to the Actor model optimization. It outputs a real number indicating a rating of the action taken in the previous state. By comparing this rating, the agent can compare its current policy

with a new policy and decide how it improves the Actor model to take better actions. we set up Actor–Critic framework into the agent working on an eNB and implement the PPO algorithm into the framework. Our Actor network designed to have inputs layers equivalent to the size of our observation or state from the network. Determining the number of hidden layers and neurons is part of our hyper parameter tuning task. The output layer is to match the number of random actions the actor to choose from.

Our Critic network shares the same number of input and hidden layers except its number of output neurons is 1, which is a feedback scalar value to the Actor. This network is used for smooth training updates by the actor, so unlike the actor, it will not be used at inference stages.

Xavier initialization is applied for a random initialization of weights and biases of both Actor and Critic Multi-Layer Perceptron (MLP) dense neural networks.

The PPO agent samples data through interaction with the given environment and optimizes its objective function using a Stochastic Gradient Descent (SGD) optimization algorithm. PPO hyper parameters such as clip ratio, discount factor, learning rate are tested for different values.

We have used Adam optimizer for optimal weights and biases. Adam is a popular choice for optimizing the parameters of deep learning models. In addition to that, the drastic policy change is mitigated by the clip function.

Next task is to train proposed learning algorithms and evaluate metrics at the inference stage using the trained models. The learning performance converges to its maximum reward within the first 100 training episodes.

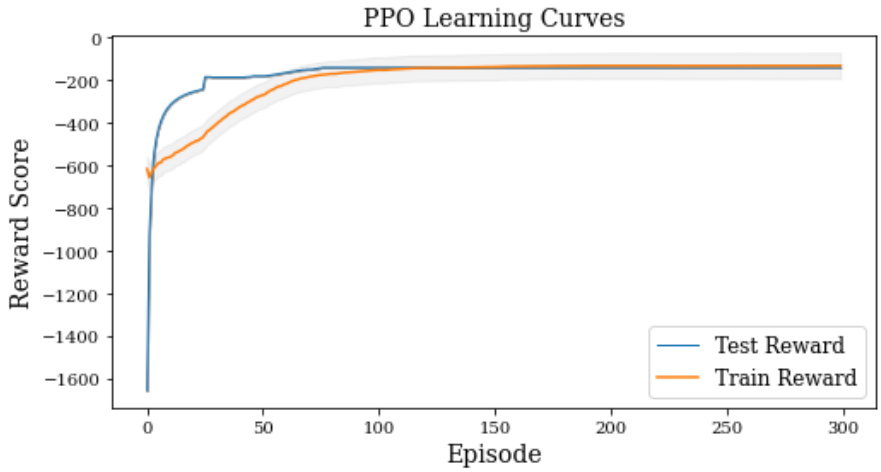

Figure 4.6: PPO training performance

The goal of the model training is to obtain an optimal policy. The optimal policy can predict the right activation and deactivation actions as accurate as possible for the optimal set of CA serving bands in CA_3A-7A-20A with bandwidths (in *MHz*) 20/20/10 licensed operator, given the current traffic loads and activation status of eNB cells as an input vector.

4.5.2. Simulation Parameters and Assumptions

All parameters and assumptions in our simulation setup are listed in *Table 4.3*. Additionally, the tuned hyperparameters for our trained DQN and PPO models are provided. The performance of each proposed solution is compared to the power consumption of the existing operational network, which serves as a baseline for evaluating the effectiveness of the proposed solutions.

Table 4.3: DRL hyper parameters

RL Algorithm	***Hyper-Parameters and Values***					***Optimizer***
	Learning rate	*Clip ratio*	*Discount factor*	*Batch/Buffer size*	*Exploration rate*	
DQN	0.003	-	0.95	64/72000 Bytes	0.07	BayesianOpt.
PPO	0.001	0.2	0.98	-	0.2	Adam

5. Results and Discussion

5.1. RF Power Consumption Model Based Analysis

To begin with, R-squared (R^2) is a metric used to evaluate goodness-of-fit of the model in linear regression, numerical value between 0% - 100%. Usually, the larger the R^2, the better the regression model fits your observations. In this aspect, R^2 value of 85% and above taken for our evaluation shows our models are almost capable to reproduce the realistic power consumption.

In our power consumption modeling, we did not observe any significant changes when incorporating site measurements. This lack of change could accredited to two factors: firstly, the inclusion of more performance data samples minimized the inaccuracies, particularly the 5-minute deviations, in the performance reports. Secondly, the consecutive network traffic performances exhibit a higher degree of correlation, which may also contribute to the absence of notable differences.

The subsequent plots illustrate the RF power consumptions operating at different loads and frequency bands

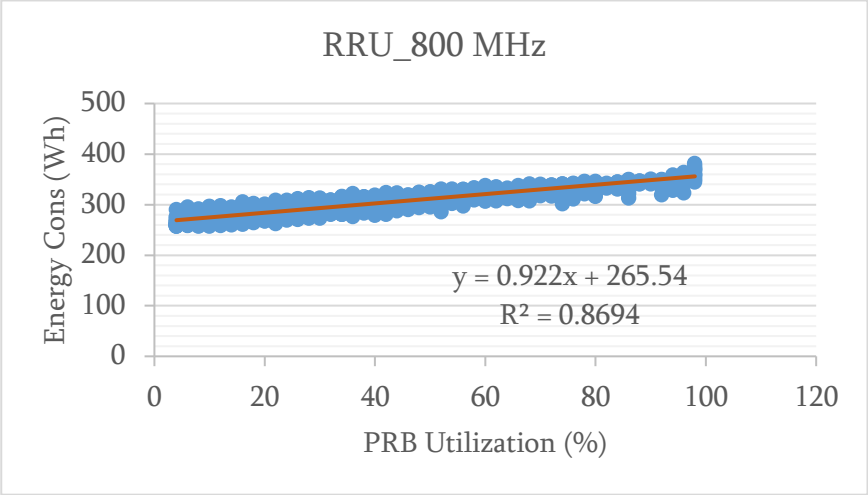

Figure 5.1: RRU power model (800 MHz)

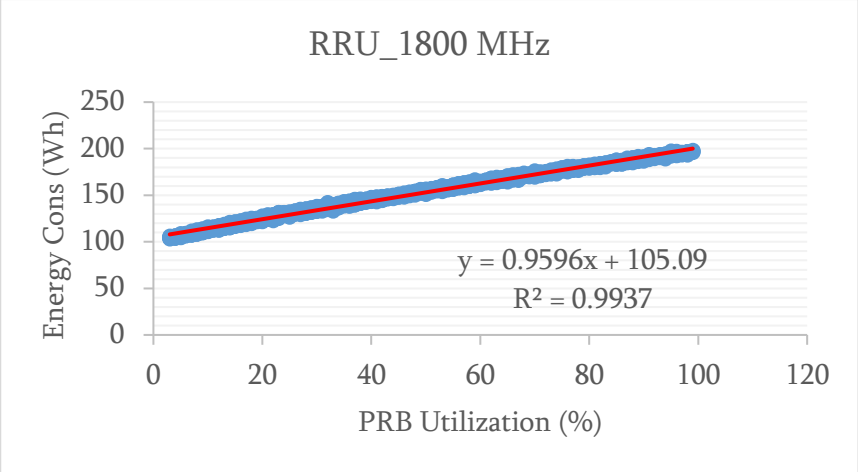

Figure 5.2: RRU power model (1800 MHz)

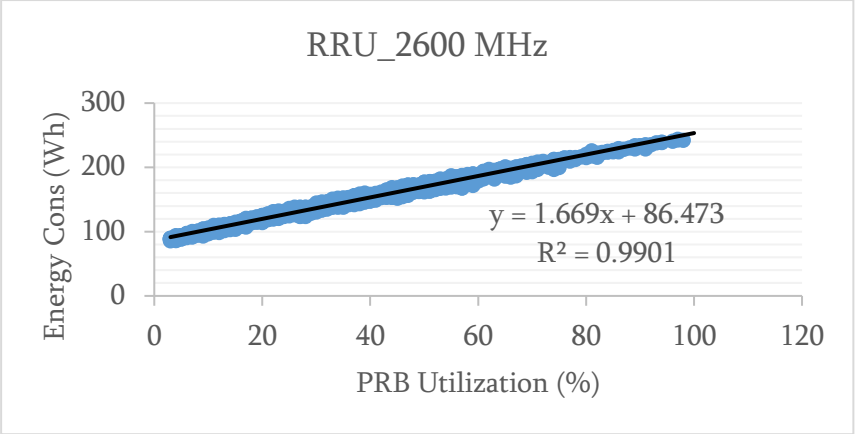

Figure 5.3: RRU power model (2600 MHz)

In the 800 MHz band, the variation in RF power consumption with PRB utilization is relatively less steep, which can be attributed to a couple of factors. Firstly, lower frequency bands necessitate lower transmission power. Secondly, the 10 MHz carrier in this band consumes less power compared to the 20 MHz carriers in the other two serving bands. Here are the power consumption models summarized for all RF units in target.

Table 5.1: RRU power consumption models and datasheet ratings

Model/Type	***Power Consumption Model*** $l^{rb} \in [0, 100]$	R^2	***Maximum Power Consumption (Watt)***
RRU_A 2600MHz	$P^{4G}{}_{RRU\_2600} = 86.47 + 1.67(l^{rb})$	*0.99*	*358*
RRU_B 1800MHz	$P^{4G}{}_{RRU\_1800} = 105 + 0.96(l^{rb})$	*0.99*	*270*
RRU_C 800MHz	$P^{4G}{}_{RRU\_800} \sim 265.54 + 0.92(l^{rb})$	*0.87*	*510*

From the power consumption trends of those RF units working in different bands, the per PRB utilization (l^{rb}) power consumption weighting coefficients showed increment with the working frequency. Thus, deactivating the 2.6 GHz band would save more power than deactivating the 1800 MHz band. Therefore, prioritizing the deactivation order from the highest frequency band (2.6 GHz) to the lowest would be the most energy-efficient approach. This means that deactivating 2.6 GHz band preferred to 1800 MHz band, as long as the same number of resource blocks can be provisioned. Of course, the latter band is more resilient to environmental impairments for a better capacity as well.

At no user traffic, active PAs dominantly contributes to the static power part. Hence, in deactivating each band, the RF component power consumption becomes insignificant.

5.2. RL Based BLCOO Energy Saving Result Analysis

The results presented in this section are obtained using a custom network simulator replicating eNB. For our case, an inter-band CA enabled configuration management and performance management realistic data with a hourly resolution time is used. The downlink PRB utilization for its high correlation with the downlink traffic volume is selected to capture the dynamic load of target network. Randomly distributed static users of 10 per sector or 30 per base station capable of exploiting the CA benefits are uniformly dropped in an interference limited network environment. The network to replicate the real environment is composed of a hexagonal deployment layout, a target eNB with six neighboring eNBs as sources of ICI. Hence, users' SINRs and respective throughputs are affected by ICI of neighbor multi-cell eNBs, and updated at each training steps following the activation/deactivation actions from our RL agent.

In this subsection, both for trained DQN and PPO RL algorithms energy saving performances results are presented and discussed thoroughly. In our evaluation the reward function weight values for α, β and ζ are -3.0, 1.0 and 1.5 respectively. In this

regard, the reward function value will be zero when all SCells are switched off and average throughput is equals target.

5.2.1. DQN-Based Energy Saving Result

Our trained DQN agents for different average users' throughputs have been applied for a 24 hours energy saving performances.

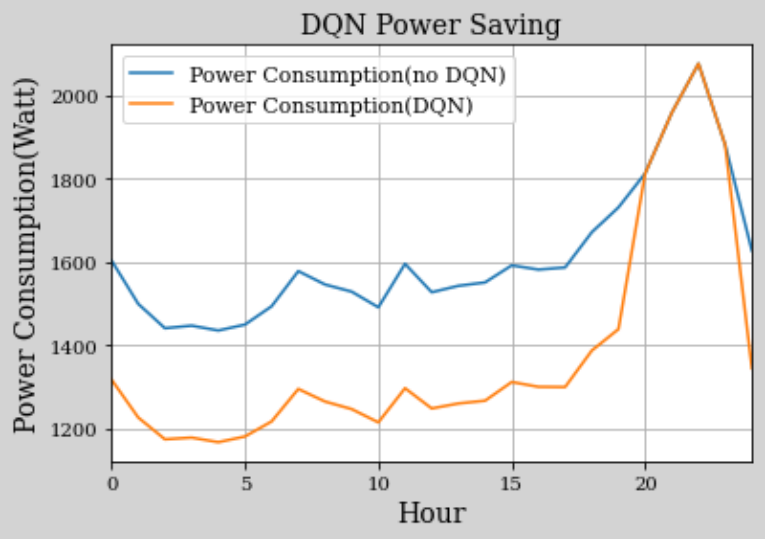

Figure 5.4: DQN 24-hour energy saving result

The result depicted in *Figure 5.4* is based on average throughput of 3 Mbps as a minimum requirement. This evaluation is done for the eNB with highest sector cumulative hourly load ranging from 8% to 85%. This daily performance can represent typical scenarios in normal operation.

The per hour energy saving shows variations in power reduction. The highest power saving is observed 18.71% on 3rd hour (3:00 AM) and the lowest is 0% or no saving for three consecutive high load states. The power consumption reduction is emanated from deactivation of cells and lowering the number of serving CCs to two. The daily average energy saving performance was 14.6%. Besides, the average number of serving CCs evaluated 72.0%. This indicates the eNB in optimal operation was using lower number of serving bands for several hours of the day. In such operation, CA users' devices due to multiple CCs monitoring power consumption would be reduced by 28.0%. In

achieving the above energy saving, we have compared per sector achievable rates or sum rates of optimal and non-optimal cases.

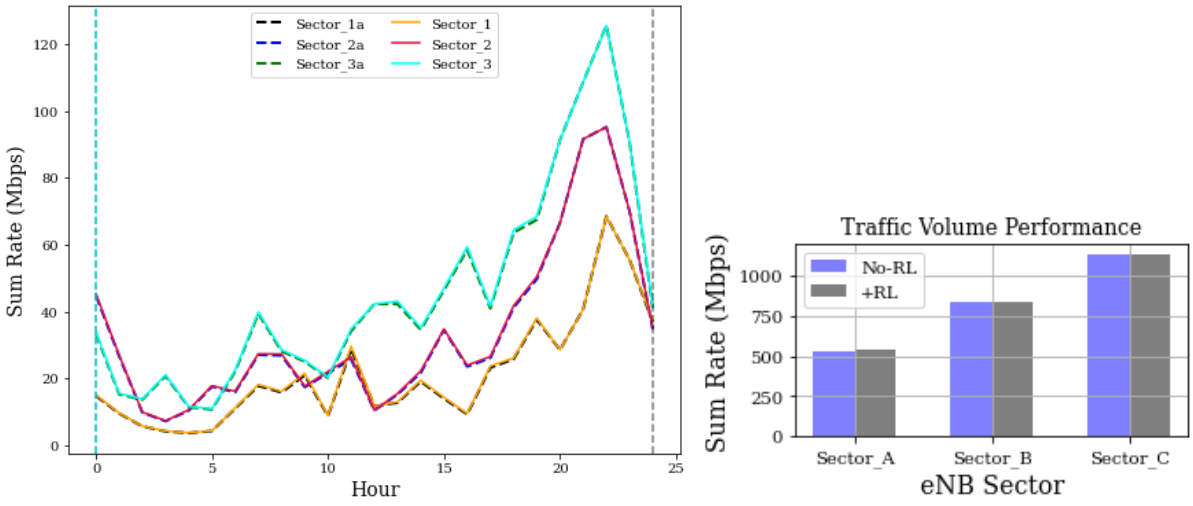

Figure 5.5: Per sector sum rate

In the figure, *Figure 5.5,* dashed lines denoting non-optimized achievable rates and solid lines for achievable rates with energy saving. As one can closely notice, the optimized capacity showed a bit increment except for high-load hours. The overall result affirms the per hour capacity demand is ensured.

5.2.2. PPO Based Energy Saving Result

The previous DQN based energy saving was limited to discrete values. We extend our work to PPO based energy saving which can support continuous values of network load performances.

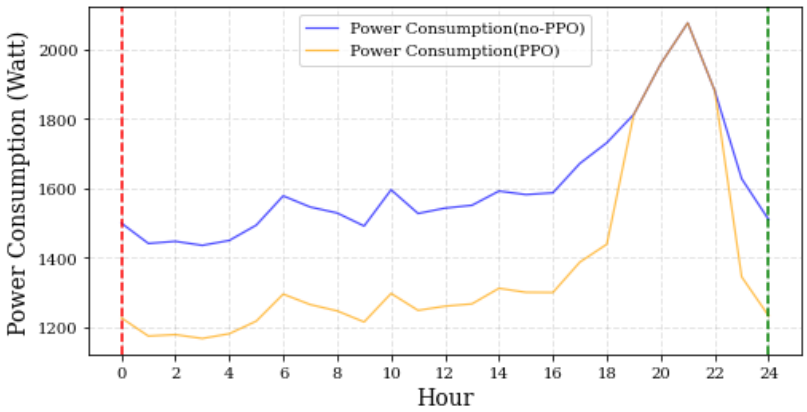

Figure 5.6: PPO based energy saving result

From the simulation results in *Figure 5.6*, the same performances to DQN based energy saving is achieved.

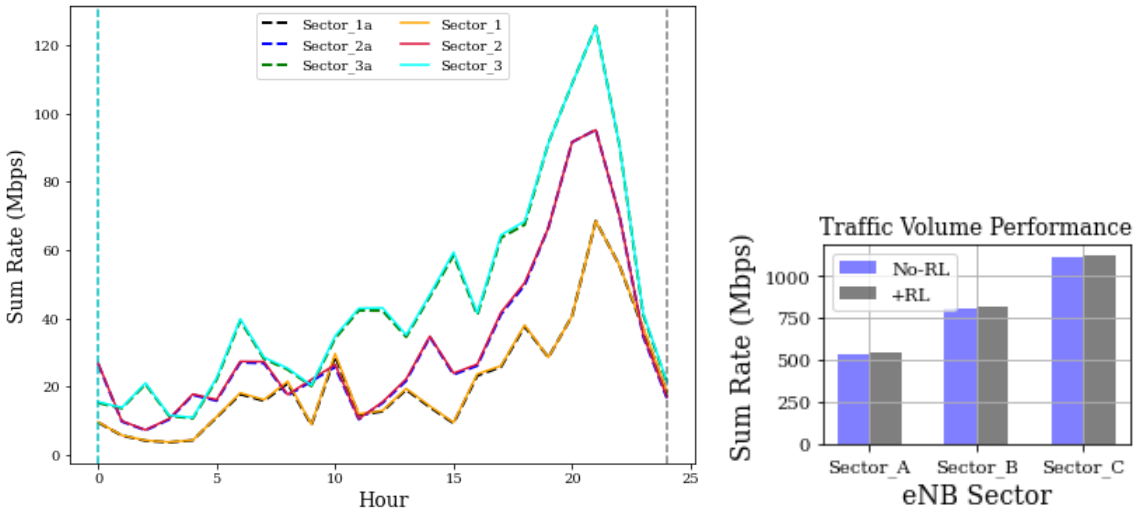

Figure 5.7: Normalized sum rate with and without BLCOO power saving

Lastly, one of our objectives is to ensure capacity satisfying the data traffic demand while performing power saving optimization task. As depicted in *Figure 5.7*, the sum rate which is the sum of achievable rates of each CCs per sector simulated based on the hourly PRB load information and SINR values calculated from all 30 CA users served by the base station. Within the simulated environment and actual utilized resources, the achievable throughputs with and without activation/deactivation are represented by solid and broken lines respectively. Taking for a single sector, the number of CCs is less than 3, the maximum configured, leaves load shared ratio and resources allocated from the remaining active carriers increases. That is defined in our load split strategy. Upon exhibiting for close proximity of plot curves to assure QoS impact, it is also noteworthy that in sum rate line plots, values for optimal case is a bit higher.

This performance deviation highlights deactivating higher band carriers led to allocate more fading resilient spectral resources and achieve better data rate. Moreover, the load balancing strategy implemented by the operator to allocate more carrier channel resources from the scarce lower band primary carriers at higher loads is also another aspect. Because it is from the operational traffic load experiences that the RL agent ultimately learns.

Once we have designed our MDP problem to fit those RL algorithms i.e. discrete states for DQN and continuous states for PPO, both algorithms performed equally for our energy saving. However, in DQN training, the testing curve showed higher standard deviations or less smooth compared to PPO. This is not exhibited in PPO due to our clip factor parameter to control large deviations. In addition, DQN demands a memory buffer for batch samples and we set as buffer-size hyperparameter. In this aspect DQN need more memory size than PPO. The total number of samples required for DQN are almost doubled to PPO case.

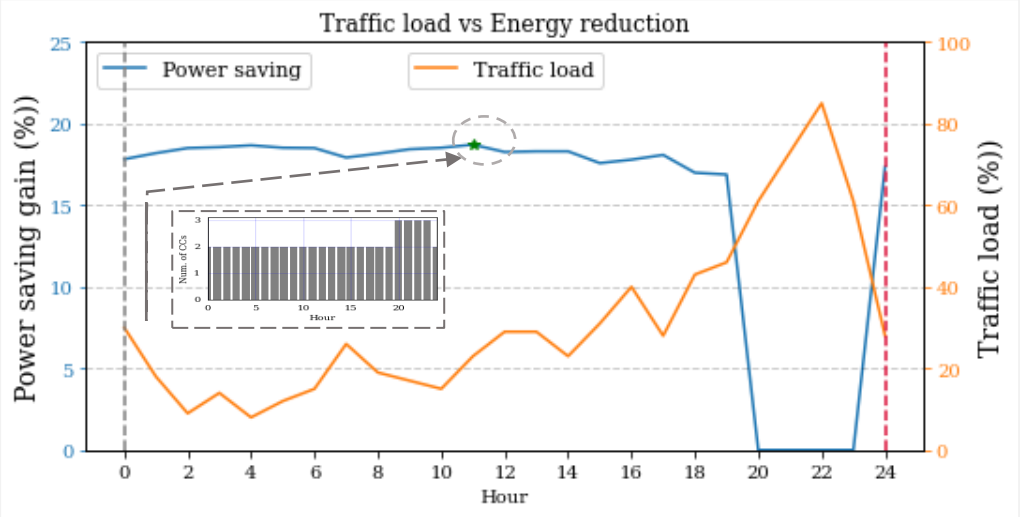

Figure 5.8: Energy saving vs Traffic load

As the daily traffic load traverses from 8% (low traffic) to 85% (peak-hour) and depicted in orange color, the energy saving performance (blue color plot) with maximum value of 18.71% (*) drops to 16.88% at mid-traffic and 0% at peak loads. It is equivalent to a maximum hourly saving of 298.53 Watt-hour (*) and a total energy saving of 5.88 kWh for 24 hours evaluation period. The energy reduction mainly emanated from the deactivation of Band 7 but also from more traffic offloading to lower bands. While satisfying their PRBs demands, this in turn leaves 3CC CA users in 2CC states and hence saves UE channel monitoring power consumption by one-third throughout the action implementation period.

We further evaluated our BLCOO energy saving solution for different traffic patterns of same eNB taking different days and starting hours.

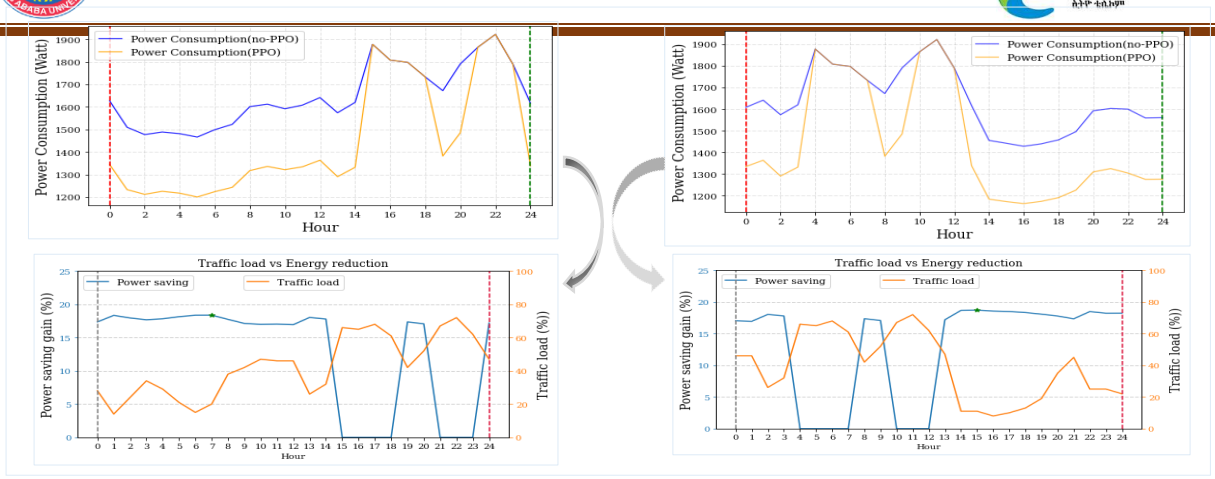

Figure 5.9: Energy saving evaluation for different traffic trends

The evaluations done for Day-2 (left) and Day-3 (right) and different starting times of 00:00 AM and 11:00 AM i.e. for different traffic load patterns of the same eNB. Based on the results, the 24-hour or daily average energy saving was lower (energy saving 12.14% and CC usage 76%) as expected for days with more traffic capacity demands. However, in all cases those energy saving action decision hours show our algorithm is performing as intended and robust to traffic load patterns.

The below table, *Table 5.2*, summarizes the overall results from one of our evaluations representing the typical daily traffic performance in optimal planning.

Table 5.2: Energy saving performances summary

Performances / **Metrics**		**Non-optimal /Baseline**	**Optimal**	**Improvements**
CC usage		100%	72%	28%
Daily RF energy consumption		40.25 kWh	34.38 kwh	14.6%
Hourly energy consumption with highest saving gain		1595.83 Wh	1297.3 Wh	18.71%
Sum rate (Avg. over 24 hrs.)	*Sector_A*	19.8 Mbps	20.03 Mbps	+ 1.14%
	Sector_B	31.09 Mbps	31.36 Mbps	+ 0.87%
	Sector_C	42.12 Mbps	42.52 Mbps	+ 0.95%

The sum rate deviations came from allocating more PRBs from lower band LTE carriers when higher band cells were switched-off. The hourly traffic volume and capacity demand is thus ensured by controlling set serving of CCs for sufficient PRBs.

As a note to the end of this section, the operational network power consumption performance is our baseline. However, it is already using energy saving techniques such as DRX and hibernation at no traffic. Thus, the achieved performance is on top of those existing solutions.

6. Conclusion and Future Directions

Based on our analysis on operator data for urban LTE-A mobile traffic hourly performance, the operator needs to employ CA to serve the peak-hour traffic demand. Those realized 3CC CA operating hours after we optimized reveal that the operator using the global common practice is forced to upgrade the target eNB's set of CA serving carriers to 3, or QoS would be impacted at peak hours.

Extending our surveys on power consumption of RF units from the same vendor but operating at different technologies have shown that the fixed power is higher in legacy networks. In the case of 4G, the transmission-dependent power extends to 87%, much higher than the fixed power. This indicates promising improvements in hardware designs by manufacturers to reduce power consumption. This could be an indicator that equipment manufacturers are striving to bring better alternatives to the competitive telco market to meet: (i) Operators' cost reduction expectations from future low power consumption designs. (ii) The global initiative to carbon reduction 45% target by 2030.

But, more works yet for operation phase soft energy saving alternatives.

However, optimizing spectral resources and energy consumption requires trade-off solutions that are not achievable through hardware design alone. Hence, to devise operational energy saving alternatives and AI (DRL) for a well-informed decision is irreplaceable.

The daily average of 14.6% and including a maximum 18.71% energy saving, while ensuring network capacity demand, for a planning optimal eNB reveals our approach to propose BLCOO, define the optimization problem MDP and solve with DRL is practical.

Moreover, relying on operator data from realistic network deployments also makes our multi-objective energy-saving solution more practical. The operator can thus save more power at the expense of capacity, or vice versa, depending on its specific demands. And

due to different manufacturers and deployment scenarios, such solution avoids generalization.

Another observation is that using discrete approximates of traffic loads as our samples (DQN) versus using the actual continuous values (PPO) to our RL controllers exhibited comparable energy saving performance. This indicates RL agents in both scenarios still acquired sufficient samples to determine optimal CCs and increases flexibility in choosing algorithms. However, due to large memory storage demand for increasing replay buffer size in DQN and also its limitation to finite state-action space, PPO is recommended in advance.

With the implementation of BLCOO, capacity can also be increased. From a capacity perspective, operators can redistribute underutilized and deactivated capacity bands, equivalent to a 20 MHz bandwidth LTE carrier, to an eNB operating at its peak load. This approach is more feasible for operator under discussion with two acknowledged busy hours, specifically from 10:00 to 11:00 and 21:00 to 22:00. Thus, operators can minimize expenses for soft license resources such as carrier power, carrier bandwidth, RRC connected users, and related features.

Another observation is that provisioning RF units and frequency bands in multi-technologies scenario denies energy saving expected from PAs shutdown. Because, decision for one technology affects the other.

Lastly, with the new enhancements and evolutions of mobile technologies, CA remains a key capacity-enhancing feature, especially with the inclusion of higher bands, including millimeter Waves. Unlike in macro BS, per BS power consumption in millimeter Waves is lower, but their low serving ranges require more BS per area. Therefore, we recommend future research on per BS cluster-based energy-saving adaptive solutions for centralized Baseband Units, along with RF densification, as edge computing gains higher market attraction.

References

[1] GSMA, "The Mobile Economy 2022 - GSMA," Feb. 2022.

[2] M. Agiwal, H. Kwon, S. Park, and H. Jin, "A Survey on 4G-5G Dual Connectivity: Road to 5G Implementation," *IEEE Access*, vol. 9, pp. 16193–16210, 2021, doi: 10.1109/ACCESS.2021.3052462.

[3] R. Antonioli *et al.*, "Dual Connectivity for LTE-NR Cellular Networks: Challenges and Open Issues," *Journal of Communication and Information Systems*, vol. 33, no. 1, pp. 282–294, 2018, doi: 10.14209/jcis.2018.28.

[4] ethio telecom, "Mobile Network Traffic and KPI Analysis," Addis Ababa, 2023.

[5] A. Kassie, A. Advisor, Dr.-E. Y. Wondie, and A. Ababa, "Users Throughput Enhancement of LTE Network using Carrier Aggregation for Addis Ababa," 2019.

[6] S. B. A. and K. A. P. A. James D.Gadze, "Real Time Traffic Base Station Power Consumption Model for Telcos in Ghana", Accessed: Sep. 02, 2022. [Online]. Available: https://www.researchgate.net/publication/309618577_Real_Time_Traffic_Base_Station_Power_Consumption_Model_for_Telcos_in_Ghana

[7] A. Mourato, D. Duarte, I. Pinto, and P. Vieira, "A Novel and Realistic Power Consumption Model for Multi-Technology Radio Networks." doi: 10.23919/URSIRSB.2018.8486764.

[8] H. Chen Dongxu, "5G Power: Creating a green grid that slashes costs, emissions & energy use," Jul. 2020. https://www.huawei.com/en/technology-insights/publications/huawei-tech/89/5g-power-green-grid-slashes-costs-emissions-energy-use (accessed Sep. 24, 2022).

[9] ethio telecom, "Power Consumption and Monitoring," Addis Ababa, 2023.

[10] P. Chang and G. Miao, "Optimal Operation of Base Stations With Deep Sleep and Discontinuous Transmission," *IEEE Trans Veh Technol*, vol. 67, no. 11, pp. 11113–11126, Nov. 2018, doi: 10.1109/TVT.2018.2869668.

[11] K. Kanwal, G. A. Safdar, M. Ur-Rehman, and X. Yang, "Energy Management in LTE Networks," *IEEE Access*, vol. 5, pp. 4264–4284, 2017, doi: 10.1109/ACCESS.2017.2688584.

[12] S. Wu, Y. Wang, and L. Bai, "Deep Convolutional Neural Network Assisted Reinforcement Learning Based Mobile Network Power Saving," *IEEE Access*, vol. 8, pp. 93671–93681, 2020, doi: 10.1109/ACCESS.2020.2995057.

[13] M. Choi *et al.*, "Cell On/Off Parameter Optimization for Saving Energy via Reinforcement Learning," in *2021 IEEE Globecom Workshops, GC Wkshps 2021 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/GCWkshps52748.2021.9682160.

[14] J. S. Pujol-Roigl, S. Wu, Y. Wang, M. Choi, and I. Park, "Deep Reinforcement Learning for cell on/off energy saving on Wireless Networks," in *2021 IEEE Global Communications Conference, GLOBECOM 2021 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/GLOBECOM46510.2021.9685279.

[15] M. Elsayed *et al.*, "Reinforcement Learning Based Energy-Efficient Component Carrier Activation-Deactivation in 5G," in *2021 IEEE Global Communications Conference, GLOBECOM 2021 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/GLOBECOM46510.2021.9685223.

[16] N. Ludant, N. Bui, A. Garcia Armada, and J. Widmer, "Data-driven performance evaluation of carrier aggregation in LTE-Advanced," in *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, IEEE, Oct. 2017, pp. 1–6. doi: 10.1109/PIMRC.2017.8292590.

[17] C.-T. Tung, Y.-L. Chung, and Z. Tsai, *An efficient power-saving downlink transmission scheme in OFDM-based multiple component carrier systems*. 2012.

[18] Ericsson, "Ericsson Mobility Report June 2022," 2022.

[19] B. Barakat and K. Arshad, "Energy efficient carrier aggregation for LTE-Advanced," in *2015 IEEE 8th GCC Conference & Exhibition*, IEEE, Feb. 2015, pp. 1–5. doi: 10.1109/IEEEGCC.2015.7060066.

[20] S. Sasikumar and J. Jayakumari, "Spectral Efficiency-Energy Efficiency Tradeoff Analysis for a Carrier Aggregated 5G NR Based System," 2020, pp. 45–55. doi: 10.1007/978-981-15-3992-3_5.

[21] J. Joao Bazzo, R. de Melo Pires, A. Javier Ortega, N. Portela Salehi, and P. Thiago Marreiros Santos, "UE current consumption on carrier aggregation in LTE-A systems," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, IEEE, Oct. 2020, pp. 600–602. doi: 10.1109/ICTC49870.2020.9289502.

[22] X. Hu, W. Xu, and Y. Tian, "Robust Clock Timing Recovery for Performance Improvement in Digital Self-Synchronous OFDM Receivers," *Wirel Pers Commun*, vol. 95, no. 2, pp. 683–701, Jul. 2017, doi: 10.1007/s11277-016-3792-9.

[23] 3GPP TS 36.211 version 12.9.0, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation (3GPP TS 36.211 version 12.9.0 Release 12)," 2017.

[24] S. Parkvall, A. Furuskar, and E. Dahlman, "Evolution of LTE toward IMT-advanced," *IEEE Communications Magazine*, vol. 49, no. 2, pp. 84–91, Feb. 2011, doi: 10.1109/MCOM.2011.5706315.

[25] "Requirements related to technical performance for IMT-Advanced radio interface(s)."

[26] E. Dahlman, S. Parkvall, and J. Sköld, "Carrier Aggregation," in *4g, LTE Evolution and the Road to 5G*, Elsevier, 2016, pp. 309–330. doi: 10.1016/B978-0-12-804575-6.00012-1.

[27] TSGR, "TS 136 300 - V10.2.0 - LTE; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (3GPP TS 36.300 version 10.2.0 Release 10)," 2011. [Online]. Available: http://portal.etsi.org/chaircor/ETSI_support.asp

[28] TSGR, "TS 136 321 - V10.0.0 - LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification (3GPP TS 36.321

version 10.0.0 Release 10),” 2011. [Online]. Available: http://portal.etsi.org/chaircor/ETSI_support.asp

[29] C. Bouras, G. Diles, V. Kokkinos, K. Kontodimas, and A. Papazois, “A Simulation Framework for Evaluating Interference Mitigation Techniques in Heterogeneous Cellular Environments,” *Wirel Pers Commun*, vol. 77, no. 2, pp. 1213–1237, Jul. 2014, doi: 10.1007/s11277-013-1562-5.

[30] R. Falkenberg, B. Sliwa, and C. Wietfeld, “Rushing Full Speed with LTE-Advanced Is Economical - A Power Consumption Analysis,” in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*, IEEE, Jun. 2017, pp. 1–7. doi: 10.1109/VTCSpring.2017.8108515.

[31] G. Yuan, X. Zhang, W. Wang, and Y. Yang, “Carrier aggregation for LTE-advanced mobile communication systems,” *IEEE Communications Magazine*, vol. 48, no. 2, pp. 88–93, Feb. 2010, doi: 10.1109/MCOM.2010.5402669.

[32] Q. Wu, X. Chen, Z. Zhou, L. Chen, and J. Zhang, “Deep Reinforcement Learning With Spatio-Temporal Traffic Forecasting for Data-Driven Base Station Sleep Control,” *IEEE/ACM Transactions on Networking*, vol. 29, no. 2, pp. 935–948, Apr. 2021, doi: 10.1109/TNET.2021.3053771.

[33] Y. L. Chung, “Energy-Efficient Transmissions for Green Base Stations with a Novel Carrier Activation Algorithm: A System-Level Perspective,” *IEEE Syst J*, vol. 9, no. 4, pp. 1252–1263, Dec. 2015, doi: 10.1109/JSYST.2014.2331030.

[34] T. Saraiva, D. Duarte, I. Pinto, and P. Vieira, “An Improved BBU/RRU Energy Consumption Predictor for 4G and Legacy Mobile Networks using Mixed Statistical Models,” in *2020 International Conference on Computing, Networking and Communications, ICNC 2020*, Institute of Electrical and Electronics Engineers Inc., Feb. 2020, pp. 320–325. doi: 10.1109/ICNC47757.2020.9049673.

[35] Ericcson.com, “Ericsson Mobility Report; Mobile network traffic Q1 2023.” https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/mobile-traffic-update (accessed Jul. 12, 2023).

[36] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction Second edition, in progress."

[37] H. M. Taylor and S. Karlin, "An Introduction to Stochastic Modeling, Third Edition."

[38] A. M. Metelli, "Configurable Environments in Reinforcement Learning: An Overview," 2022, pp. 101–113. doi: 10.1007/978-3-030-85918-3_9.

[39] H.-K. Lim, J.-B. Kim, J.-S. Heo, and Y.-H. Han, "Federated Reinforcement Learning for Training Control Policies on Multiple IoT Devices," *Sensors*, vol. 20, no. 5, p. 1359, Mar. 2020, doi: 10.3390/s20051359.

[40] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," Jul. 2017.

[41] Teshale Belete, " Energy assessment and optimization in second generation wireless access network, the case of ethio-telecom.," 2018, Accessed: Sep. 02, 2022. [Online]. Available: Teshale, Belete. Energy assessment and optimization in second generation wireless access network, the case of ethio-telecom. Diss. Addis Ababa University, 2018

[42] T. Saraiva, D. Duarte, I. Pinto, and P. Vieira, "An Improved BBU/RRU Energy Consumption Predictor for 4G and Legacy Mobile Networks using Mixed Statistical Models," in *2020 International Conference on Computing, Networking and Communications, ICNC 2020*, Institute of Electrical and Electronics Engineers Inc., Feb. 2020, pp. 320–325. doi: 10.1109/ICNC47757.2020.9049673.

APPENDIX-I

Traffic-Aware Band-Level Cells On-Off for Energy Saving in LTE-Advanced Networks with Inter-Band Carrier Aggregation

Yilma Melaku* and Dereje Hailemariam**
School of Electrical and Computer Engineering
Addis Ababa Institute of Technology, Addis Ababa University
Addis Ababa, Ethiopia
E-mail: yilmamlk@gmail.com* and dereje.hailemariam@aait.edu.et**

***Abstract*—In this research, a novel traffic load adaptive band-level cells on/off (BLCOO) approach for inter-band carrier aggregation (CA) in LTE-A networks have been proposed to save RF power. BLCOO optimizes the number of serving component carriers (CCs) during off-peak hours based on network traffic load. The energy-saving problem is formulated as a Markov decision process (MDP) with uncertain network conditions. Deep reinforcement learning algorithms are trained to solve the MDP problem and achieve significant RF power savings, up to 18.71%, while maintaining the required quality of service.**

Keywords—DQN, MDP, Custom Network Environment, PPO

I. INTRODUCTION

Carrier Aggregation (CA) is one of the transmission capacity enhancing features added to LTE-Advanced and beyond networks, and to overcome the ever-increasing smart phone users and high-rate data demands. The statically provisioed fixed set of serving CCs to support peak-hour traffic demand requires more number of and always-on RF units. These RF chains which in turn relies on low efficient PAs with high peak-to-average power ratio (PAPR) in orthoonal frequency diviosion multiple access (OFDMA) radio access scheme also furtherly affect efficiency of PAs which are high dynamic power consuming RF components [1]. In our survey on ethio telecom, CA employed mobile network operator (MNO) in Addis Ababa city, LTE-A network PS traffic share to other technologies is increasing and accounts to 83% as of January, 2023. In the same survey, the CA supporting RF units dynamic to static power consumption ratio ranges to 87% compared to 30.96% for legacy network counterparts. Hence, the static power contribution is much improved compared to RF units operating on legacy technologies. This reveals promising hardware design improvements by manufactures along with emerging technologies, whereas more works traffic-adaptive energy saving solutions but in uncertain traffic load state transitions is expected.

Authors in [2], evaluated CA impact both for energy saving and energy wasting conditions depending on the boost rate on demand. In [3], the authors evaluated the impact of CA for different releases of CA-capable user terminals, though improvements but still 15% higher in CA. Thus, with the statically provisioned CA resources contradicting for the dynamically varying mobile data hourly traffic load, in parallel to the high CA capacity gain demanded at peak-hours, the power wastage impact of CA overweighing its capacity gain at low-traffic hours should also be considered.

This research is aimed at saving the power consumed by multiple RF chains deployed to support the multiple CCs and bands in inter-band CA scenario at low traffic load times or when CA capacity gain is not or partly not mandatory. Determining the optimal set of serving CCs to save power at low traffic hours while guaranteeing the traffic capacity demanded and for the dynamic traffic load states of an inter-band CA enabled, tri-sector and multi-CCs eNB from a realistic network is defined as a sequential decision-making or MDP problem.

However, with increasing number of state-action spaces and metrics to consider, the use of tabular updates or mathematical models for uncertain network conditions is not practical. Authors in [4-6] have formulated energy saving under constraint as MDP problem and used DRL algorithms to tackle such probabilistic network state transitions.

In this aspect, a new traffic-aware band-level cells on-off (BLCOO) energy saving solution is proposed. It is based on determining the optimal set of serving CCs in an evolved node B (eNB), depending on the hourly traffic load. The energy-saving controller is trained using deep reinforcement learning (DRL) algorithms. Deep Q-Networks (DQN) which is limited to finite number of state-action spaces and Proximal Policy Optimization (PPO) that can support continuous states to adapt the actual traffic load performances, have been adopted.

In solving our multi-objective optimization problem under a target achievable rate as a constraint, the following are contributions as an outcome this thesis work.

- New dataset for traffic load and RF units power consumption performances
- RF power consumption models for RF units operating on different frequency bands
- Custom network simulation environment for RL agent interaction
- DRL trained BLCOO energy saving solution

II. SYSTEM MODEL

Our system model is based on urban LTE-A network with indoor users. Inter-band CA employed urban macro LTE-A network environment with a set of $\mathbb{N}$ serving CCs and bandwidths $\{CC_1, CC_2, \ldots, CC_{\mathbb{N}}\}$ (in MHz) with a capacity to allocate a maximum of 100 MHz aggregated transmission bandwidth per user for a realistic network is considered for this work. Seven CA enabled tri-sector eNBs with ISD of $3 \times \mathcal{R}$ and each sector with multiple hexagonal cells of serving ranges $\mathcal{R}$ each. Each cell is configured with either of LTE standard carrier in $\mathbb{N}$ different bands.

Moreover, $\aleph$ number of random static users per sector with CA capability are uniformly distributed to capture spatial impacts of serving network on received signal qualities into account. The six eNBs in the "Honey Comb" layout is to designate the single-tier interference for target eNB (eNB_1) connected users. For a 4G/LTE downlink based on OFDMA and frequency reuse factor unity, the main interference is inter-cell interference (ICI) due to collision of same PRBs allocated simultaneously to users by neighbor eNBs.

Moreover, the radio topology is based on deploying an RF unit for each LTE-A cell or CC added. The power consumption of RF units varies for multiple reasons, including the manufacture, working band, version and operating technology. Hence, addressing configuration-specific RF unit power consumption trends is essential to bring a more accurate energy reduction solution.

A. RF Power Measurement and Modeling

Adopting previous studies on the linear dependency of an active RF unit dynamic power consumption to traffic load [7], we built our own configuration specific power models.

The traffic dependent power consumption of remote radio units (RRUs) which we are also referring them as RF units or modules, operating in 4G mode (${P^{4G}}_{RRU}$) is also studied in [7] asserting the linear dependency with the number of utilized PRBs (l^{rb}).

$$ {P^{4G}}_{RRU} = \alpha_0 + \alpha_1(l^{rb}),\ l^{rb} \in [0,\ 100] \qquad (1) $$

Where, the first component refers the idle state power consumption in zero traffic plus active PA while the second component denotes the transmission power consumption. The operational data for energy consumption performance of RF modules (kWh) is the average of 12 samples reported at 5 minutes intervals. Reported values delayed by a maximum of 5 minutes introduce some in accuracy or deviations on hourly performances.

To narrow such deviations, we have conducted on-site measurements by increasing sampling resolution time. A random sample per 2-minutes interval or 30 samples tally for five consecutive hours is incorporated. We have ensured the time deviation between our records time reference and the network management system is less than a minute and almost no impact for our 2-minutes sampling intervals.

Moreover, PRB utilization performance, which is also 99.8% correlated to the traffic volume as per our analysis for a one-week daily performance, is used for the corresponding dynamic power consumption dataset preparation.

A simple linear regression implemented to model power consumption for RRUs operating on different bands and specific configurations. R-square is used in evaluating the accuracy of the model.

TABLE I. TABLE RRU POWER CONSUMPTION MODELS

Model/Type	*Power Consumption Model* $l^{rb} \in [0, 100]$	R^2
RRU_A 2600MHz	${P^{4G}}_{RRU\_2600} = 86.47 + 1.67(l^{rb})$	*0.99*
RRU_B 1800MHz	${P^{4G}}_{RRU\_1800} = 105 + 0.96(l^{rb})$	*0.99*
RRU_C 800MHz	${P^{4G}}_{RRU\_800} \sim 265.54 + 0.92(l^{rb})$	*0.87*

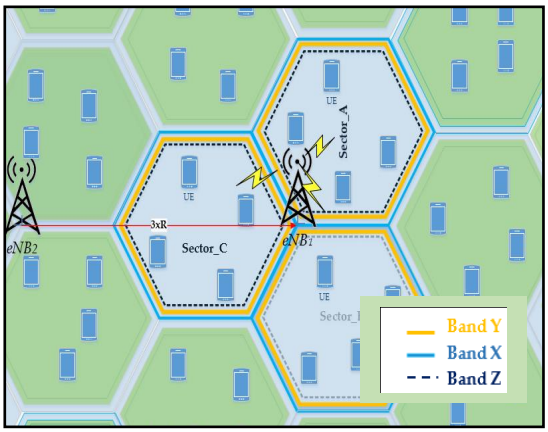

Fig. 1. Network deployment layout

III. MDP PROBLEM FORMULATION

Our optimal number of CCs selection based on the expected load and current states of eNB cells for state changes and save energy at low load conditions, is formulated as an MDP problem with unknown transition probabilities.

An MDP problem is defined with tuples of $\{S,A,\rho,R\}$, where:

- S - is the set of eNB states perceived by the energy saving controller
- A - is the set of eNB triggering actions implemented by eNB
- $\rho(s'|s,a)$ - is the state transition probability function, which gives the probability of transition from state s to state s' after taking action a
- $R(s,a)$ - is the reward function, which gives the reward received after taking action a in state s.

All relevant terms are explained as follows.

1) State: Our state is the observation from the eNB at each time step for the decision making or action selection for optimal set of serving bands. We design it

as a vector consisting of sectors hourly cumulative load and the current activation status of CCs. Thus, the set of information or metrics ($\mho$) required for decision at each time step(t):

$$\mho = \{ m,\ l_{s1}^t, l_{s2}^t, l_{s3}^t \} \quad (2)$$

Where: $m \in \{0,1,\dots,N-1\}$

Instead, current decisions shall to have the expected load information. To this end, we use weighted moving average where the most recent past is higher weighted than old ones. In addition, we determined the five most recent historical data of the base station to influence our expected load. Different window sizes can be selected.

Expected load for time $(t+1)$ is denoted by $\left(l_s^{t+1}\right)$, and WMA of window size 5 is used in this work. All the bands' possible states have been one –hot encoded in 3 binary bits.

2) Action: Each decision in selecting the optimal number of CCs based on the observation vector. The action space includes all three possible actions representing active bands to be coverage band only, including one or two capacity cell serving bands.

3) Reward Design: Each possible action for the state and transition probabilities is assigned a value of rewards or cost. Our reward (negative cost) function is defined as a decreasing function of energy consumption and an increasing function of target throughput, a QoS measure. A third metrics CC usage denotes the percentage of the ratio of the average number of serving bands in our optimized network to the total available ones is included. Lower number of CCs usage indicates higher power reduction in CA-capable user devices due to multiple CCs monitoring. Thus, the three reward functions are formulated as follows.

a) Optimal number of CCs reward function (r_{cc}^t): CC usage metric denotes the percentage of the ratio of the average number of serving bands in our optimized network (N_{opt}^t) to the total available (N) is included. Low CC usage indicates higher monitoring power reduction for user devices.

$$r_{cc}^t = 1 - \frac{N_{opt}^t}{N} \quad (1)$$

b) Power reward function (γ_p^t):

$$\gamma_p^t = \frac{P_{RRU,\ Total}^t}{P_{RRU,\ Coverage\ band}^t} \quad (2)$$

Where:

The $P_{RRU,\ Total}^t$ –power summed up from all active RRUs including coverage and capacity cells.

$P_{RRU,\ Coverage\ band}^t$ –refers the sum of minimum power consumption from coverage or always-on RRUs. By minimizing the normalized power reward value means maximizing network energy saving.

The maximum power reward is designed to be at all capacity cells off state while coverage cells are active but zero traffic.

c) CA throughput maximization reward function (r_{thp}^t): Maximizing CA users' average throughput (Tp_{av}^t) normalized to the target threshold throughput (Tp_{target}^t) avoids greedy power saving during high traffic loads. In prioritizing capacity over power saving, additional precondition set by penalizing power saving when average throughput is less than target.

$$r_{thp}^t = \begin{cases} \dfrac{Tp_{av}^t}{Tp_{target}^t}, & Tp_{av}^t \geq Tp_{target}^t \\ \log\left(\dfrac{Tp_{av}^t}{Tp_{target}^t}\right), & Otherwise \end{cases} \quad (3)$$

The use of $\log(.)$ function avoids discontinuity of r_{thp}^t values with a zero penalty for the achievable rate (Tp_{av}^t) equals to our target throughput (Tp_{target}^t) and increasing negative values or penalties when achievable rate is degrading below target.

To this end, the immediate reward value (r_{total}^t) at each execution time step is a weighted sum of rewards computed from the above three functions.

$$r_{total}^t = \alpha r_p^t + \beta r_{thp}^t + \zeta r_{cc}^t \quad (6)$$

Where, weights α, β and ζ can be implementation specific values.

The ultimate goal is to minimize energy consumption cost during light loads, guarantee traffic capacity demand and minimize the number of always-active serving bands.

In order to activate and deactivate band for all sectors, the optimal decision selection observes the sector highest load state. This is achieved by collecting experiences via trial and error of self-learning algorithms to find the optimal policy ($\pi^*(s|a)$).

To solve MDP problem for a network with uncertain state transitions or hard to mathematically model or determine transition probabilities, RL is now adays a preferred approach for its trial-and-error self-learning performance [8]. The RL agent iteratively interacts with the environment to collect more experiences and maximize reward until convergence.

A. Network Simulation Environment

In using self-learning approaches to solve problems of MDP type, the agent need to interact with the environment and collect experiences by trial-and-error before deployment. However, its random actions can have irreversible catastrophes if applied on realistic network.

To avoid such risks, general-purpose discrete-event-based simulators or customized simulation environments [9] is feasible for the control agent to interact with by sending random action and to get observation for the next optimal action selection. Higher overheads and environment complexity are challenges when using general-purpose environments.

Building customized environments replicating specific scenarios are better suited to solve specific problems. This enables to customize all aspects of the simulation environment with low overhead, to experiment with different scenarios and address specific use cases.

Simulation tools NS3 and 5G discrete time simulator have been used in [5] and [6] respectively to generate data input and build environment. Such approaches lack to mimic the realistic network traffic pattern. Whereas in [4] problem

specific customized simulator is used to replicate real network environment.

In this work, realistic data from a cellular operator is used as an input to build our custom simulator. Urban macro BS LTE-Advanced network employing inter-band CA composed of target eNB with randomly distributed users and six neighboring eNBs as a source of ICI is built for interaction with the RL agent controlling the optimal set of CCs sufficing the dynamic DL traffic demand.

Moreover, configuration parameters from a realistic network ISD, always full-load scenario CA serving cells configuration carriers and bands, and others as summarized in TABLE II, have been considered to replicate the real network scenario. The maximum bandwidth each eNB can allocate to a user is 50 MHz or 150 PRBs of which 25% of the total PRBs have been considered for control channels signaling overheads.

In this simulation setup, each user is assumed to have CA capability and can exploit all CA benefits thus each connected user performs multiple SINR measurements, mapped to CQI values and reported for all aggregated bands. Each reported CQI is mapped to modulation schemes and coding rates according to 3GPP standards where the highest supported modulation scheme is upgraded from 64QAM in Release 10 to 256QAM in R12. Based on our assessment, it is already implemented in ethio telecom operational network. DL PRBs are allocated from each active CC in a round robin fashion. Of course, the presence of CA non-capable LTE-A users in the actual environment does not affect our generalization of users' capability for this simulation setup. Non-CA users can be perceived as CA capable users but experiencing bad radio conditions (SINR value below -5 dBm) from some of their radio links or CCs and hence no PRB will be allocated for this link. We have used full-buffer traffic type which is more convenient to static user scenarios.

Thus, based on the triggering information from the controlling agent at each iteration time step, the environment updates traffic load from dataset, activate/deactivate cells, evaluate new RF power consumption based on new load, compute user level SINR and aggregated throughput. Traffic load split and resource scheduling are not specified by 3GPP and implementation specific operations at MAC layer. Thus, at each interaction time step and based on the action triggered by the energy saving RL agent, our custom environment executes or considers the following tasks.

1) Load split: When a CC is deactivated at low traffic load, the user traffic is to be redistributed among active CCs/ bands. For CA scenario with all CCs equal bandwidths and balanced loads, the offloaded traffic can be shared equally. However, this is highly relaxed and ideal. To consider for flexible bandwidths and load imbalance among CCs in the realistic case, the split ratio is designed to be based on their available PRBs and pre-offloading loads.

2) SCC Activation/Deactivation: Once a capacity cell is switched off, no more signal measurement reports from users to upper layers. Thus, the MAC CE messages sent to UEs for SCC deactivation followed by RRC signaling for SCC removal. The deactivated LTE band/carrier will remain out of CA set of serving bands and not available for schedulers for the whole T time step period. In this work, the value of T is one hour. To the reverse, for a cell switched on, SCC addition and activation messages sent by upper layers. In this aspect, no changes on the existing resource allocation and scheduling mechanisms except the number of available CCs varying for optimal case.

3) Evaluate Metrics: Each metrics are computed based on the action event. User throughput(TP^t), number of CCs saved, RF power consumption are computed either to respond a real number value as an immediate reward (r^t) to the agent at training phase, or to evaluate our proposed energy saving performances at inference phase.

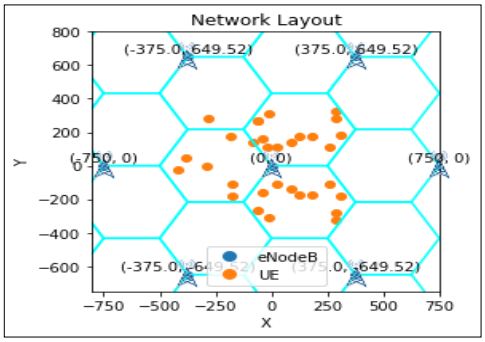

Fig.2. User distribution layout

TABLE II. NETWORK SIMULATION PARAMETERS

Parameters	Value
Network Scenario	Urban Macro
CA Type / Number of CCs	Inter-Band FDD / 3
LTE Carrier Frequency	800 MHz, 1800MHz, 2600 MHz
Bandwidth	10 MHz, 20 MHz, 20 MHz
Subcarrier Spacing	15 KHz
eNB TX Power	40 dBm, 43 dBm, 46 dBm
Cell Radius	250 m
UE TX Power	23 dBm
Number of UEs/Mobility	30/ Static (UE Random distribution)
Height (BS/UE)	35/1.5 (meters)
BS Antenna Gain/ Pattern	18 dBi $A(\theta) = -\min[12\left(\frac{\theta}{\theta_{3dB}}\right)^2, A_m]$, $\theta_{3dB} = 65\ degrees,\ A_m = 20\ dB$
UE Antenna Gain	0 dBi
Propagation Model	COST-231 Hata
Resource Scheduler	Round Robin
Penetration/Fading (losses)	20 dB / 7 dB
Channel Model	AWGN
Noise PSD	-174 dBm/Hz

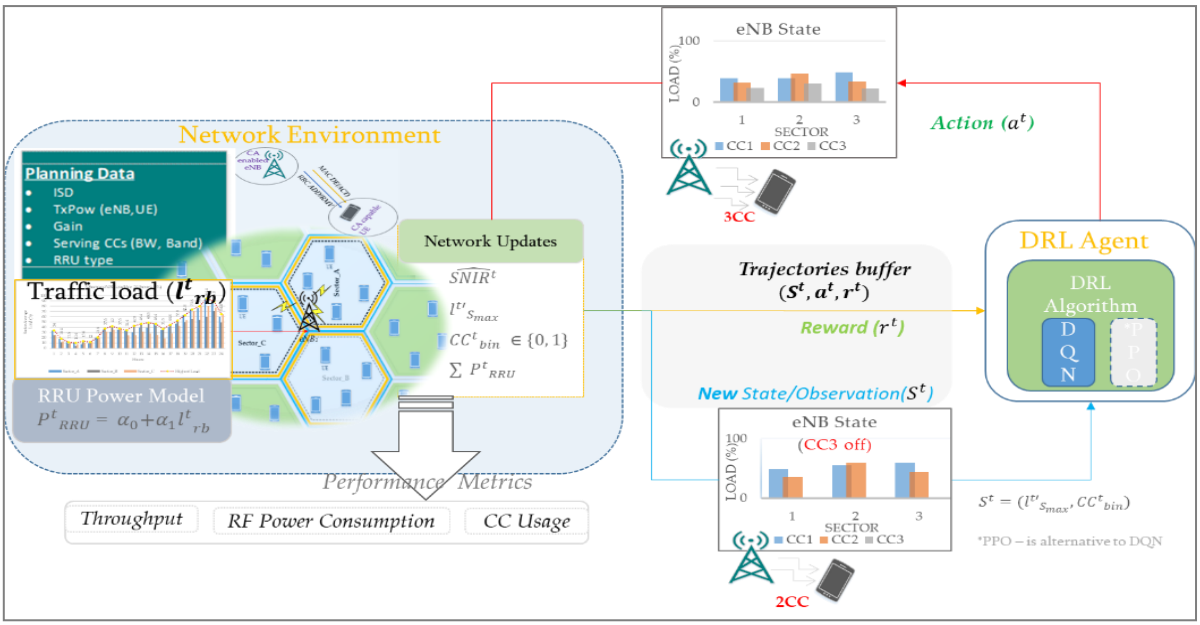

Fig. 3. DRL based BLCOO energy saving architecture

C. DRL-Based BLCOO Energy Saving

Our RL based energy saving setup general network architecture is built as in "Fig.3". The $\widehat{SNIR}^t$ value matrix representing SINRs measured for N number of serving CCs by N_u number of users.

The expected load ($l^{t'}{}_{S_{max}}$) is computed as a weighted moving average (WMA) of past experiences, where the most recent history is weighted higher and oldest one takes the least contribution in determining the expected value.

$$l^{t'}{}_{S_{max}} = f(\, l^{t-(w-1)}{}_{S_{max}}, \dots, l^{t-1}{}_{S_{max}}, l^{t}{}_{S_{max}}) \qquad (7)$$

Where: $f(.)$ denotes WMA and w is for sliding window size.

The maximum load among the three sectors at a time step t is $l^t{}_{S_{max}}$ is computed as:

$$l^t{}_{S_{max}} = max(l^t{}_{S_1}, l^t{}_{S_2}, l^t{}_{S_3}) \qquad (8)$$

Our network environment, discussed in next section, is built based on configuration data and performance data of real base station as input. Each time triggering signal from the RL agent, it updates the new set of serving bands, power consumption, per user and sum throughputs.

In solving our MDP problem, different RL algorithms are adapted depending on discrete and continuous state scenarios.

D. RL Algorithms Selection and Training

To solve our energy saving problem as an MDP problem with uncertain conditions i.e. to determine optimal number of serving CCs for dynamically changing traffic loads evaluated as PRB utilization which in turn is 98% correlated to traffic volume, different RL algorithms can be chosen but the problem type matters. Q-learning is known to be the basic RL supports continuous values of the eNB loads. However, DDPG is applicable for continuous actions whereas our environment has discrete actions and continuous states. In this aspect, we got PPO is the convenient one for the target problem.

algorithm which works by iteratively updating a table of Q-values, which represent the expected reward for taking each action in each state. However, Q-learning has limitations for problems with a large number of states and actions, and for unseen conditions.

To overcome these limitations, we used a deep Q-network (DQN), an extension of Q-learning algorithm that uses a deep neural network to approximate the Q-values. DQN has been shown to be effective for solving problems with a large number of states and actions, and for problems with uncertain conditions. It is a powerful and versatile RL algorithm that has been shown to be effective for a variety of problems. We also chose DQN because it is relatively easy to implement and can be used with a variety of environments. Bayesian optimization is applied for tuning hyperparameters by iteratively evaluating the algorithm with different hyperparameter settings.

DQN being limited to discrete states ($\boldsymbol{s}$) values, in contrast to network traffic load continuous values in actual scenarios, we have designed traffic load values as discrete values in range of 0 to 100. The DQN target network weights updated after the Q network of batch size 64. Explorations for different actions ($\boldsymbol{\alpha}$) is improved with diversified samples of our eNB states at each training step.

We trained the DQN algorithm using our custom environment interaction until it converges, which means that the Q-values no longer changed significantly. It is done by iterating over a hourly PRB usage of 500 samples and 200 episodes.

Second approach is to design our network load states as continuous from 0.0 to 1.00 infinite number of states, which was not supported by the previous, DQN, algorithm. The two Actor-Critic RL framework based algorithms PPO and DDPG

Actor-Critic uses two deep learning models, one called Actor model and the other called Critic model. The Actor model performs the task of learning what action to be selected under a particular observation of the environment (i.e., the control policy). When the action selected by the Actor model, triggers the network environment to execute and respond

feedback as a reward based on our objective function, the Critic model takes the reward to compute Q-value and feedback the Actor network for its previous action selection.

Thus, the role of the Critic model is to learn to evaluate if the action taken by the Actor model led the environment to be in a better state or not, and its feedback is used to the Actor model optimization. It outputs a real number indicating a rating of the action taken in the previous state. By comparing this rating, the agent can compare its current policy with a new policy and decide how it improves the Actor model to take better actions. we set up Actor–Critic framework into the agent working on an eNB and implement the PPO algorithm into the framework. Our Actor network designed to have inputs layers equivalent to the size of our observation or state from the network. Determining the number of hidden layers and neurons is part of our hyper parameter tuning task. The output layer is to match the number of random actions the actor to choose from.

Our Critic network shares the same number of input and hidden layers except its number of output neurons is 1, which is a feedback scalar value to the Actor. This network is used for smooth training updates by the actor, so unlike the actor, it will not be used at our inferencing stages.

Xavier initialization is applied for a random initialization of weights and biases of both Actor and Critic Multi-Layer Perceptron (MLP) dense neural networks.

The PPO agent samples data through interaction with the given environment and optimizes its objective function using a Stochastic Gradient Descent (SGD) optimization algorithm. PPO hyper parameters such as clip ratio, discount factor, learning rate tested for different values.

We have used Adam optimizer for optimal weights and biases. Adam is a popular choice for optimizing the parameters of deep learning models. In addition to that, the drastic policy change is mitigated by the clip function.

Next task is to train proposed learning algorithms and evaluate metrics at the inference stage using the trained models. The learning performance converges to its maximum reward within the first 100 training episodes.

The ultimate goal of the model training is to obtain an optimal policy i.e. $\boldsymbol{\pi}^{*}: \boldsymbol{s} \rightarrow \boldsymbol{\alpha}$. The optimal policy can thus predict the right activation and deactivation actions as accurate as possible for the optimal set of CA serving bands in CA_3A-7A-20A scenario with bandwidths 20/20/10 (in MHz) licensed operator, given the current traffic loads and activation status of all eNB cells as an input vector.

IV. SIMULATION AND NUMERICAL RESULTS

The results presented in this section are obtained using a custom network simulator replicating eNB. For our case, an inter-band CA enabled configuration management and performance management realistic data with a hourly resolution time is used. The downlink PRB utilization for its high correlation with the downlink traffic volume is selected to capture the dynamic load of target network. Randomly distributed static users of 10 per sector or 30 per base station capable of exploiting the CA benefits are uniformly dropped in an interference limited network environment. The network to replicate the real environment is composed of a hexagonal deployment layout, a target eNB with six neighboring eNBs as sources of ICI. Hence, users' SINRs and respective throughputs are affected by inter cell interference (ICI) of neighbor multi-cell eNBs, and updated at each training steps following the activation/deactivation actions from our RL agent.

In this subsection, both for trained DQN and PPO RL algorithms energy saving performances results are presented and discussed thoroughly. In our evaluation the reward function weight values for α, β and ζ are -3.0, 1.0 and 1.5 respectively. In this regard, the reward function value will be zero when all SCells are switched on and average throughput is below target.

In this subsection, both for trained DQN and PPO RL algorithms energy saving performances results are presented and discussed thoroughly. In our evaluation the reward function weight values for α, β and ζ are -3.0, 1.0 and 1.5 respectively. In this regard, the reward function value will be zero when all SCells are switched on and average throughput is below target.

Our trained DQN agents for different average users' throughputs have been applied for a 24 hours energy saving performances.

The DQN based energy saving was limited to discrete values. We extend our work to PPO based energy saving which can support continuous values of network load performances.

From the simulation results in "Fig.5"., the same performances to DQN based energy saving is achieved.

The result depicted in "Fig.7" is based on average throughput of 3 Mbps as a minimum requirement. This evaluation is done for the eNB with highest sector cumulative hourly load ranging from 8% to 85%. This daily performance can represent typical scenarios in normal operation.

TABLE III. DRL HYPER-PARAMETERS

DRL Algorithm	**Hyper-parameters and values**					**Optimizer**
	Learning rate	*Clip ratio*	*Discount factor*	*Batch/Buffer size*	*Exploration rate*	
DQN	0.003	-	0.95	64/72000 Bytes	0.07	BayesianOpt.
PPO	0.001	0.2	0.98	-	0.2	Adam

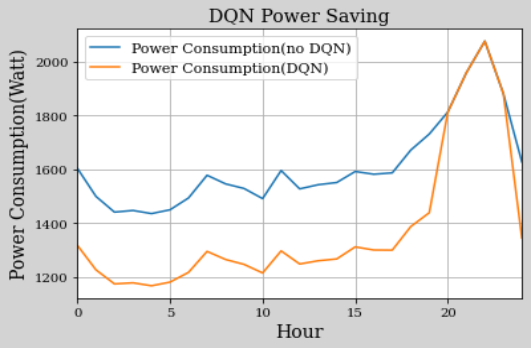

Fig.4.PPO based energy saving result

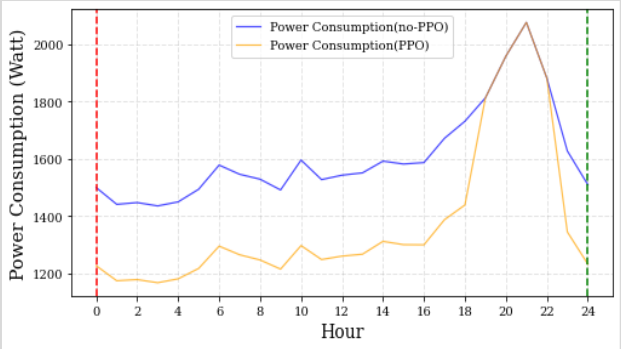

Fig.5. DQN 24-hour energy saving result

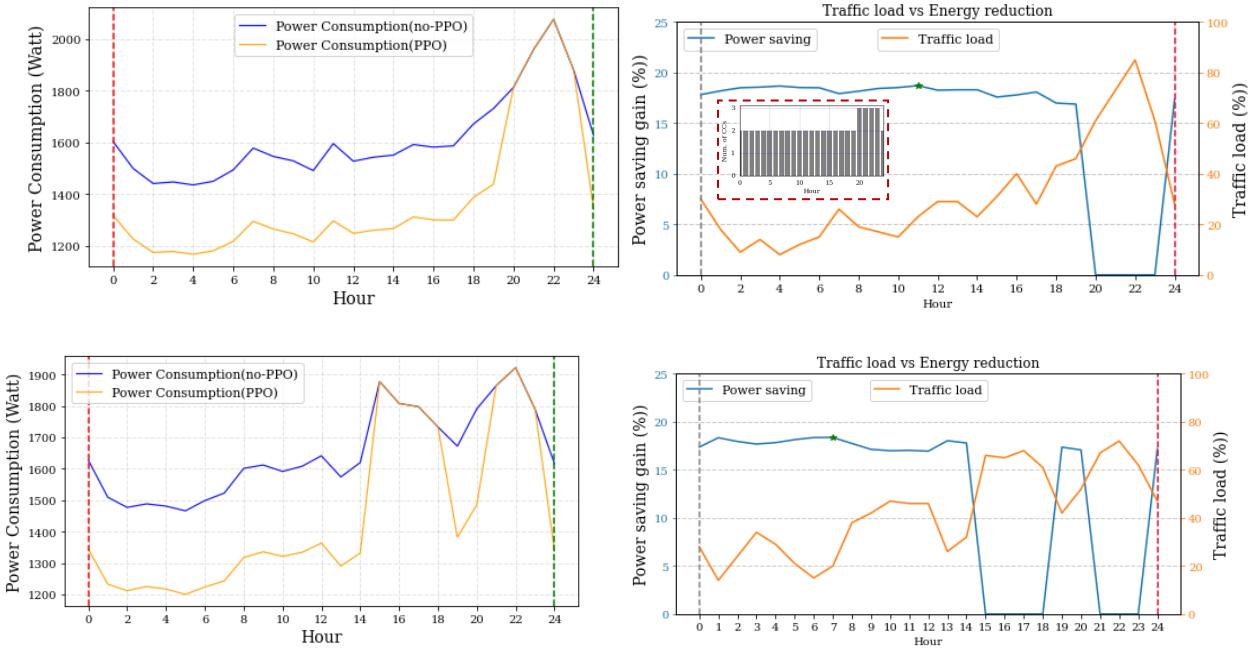

Fig.6. Energy saving performance evaluation results for different

TABLE I. ENERGY SAVING RESULTS SUMMARY

Performances / Metrics		Non-optimal/ Baseline	Optimal	Improvements
CC usage		100%	72%	28%
Daily RF energy consumption		40.25 kWh	34.38 kwh	14.6%
Hourly energy consumption with highest saving gain		1595.83 Wh	1297.3 Wh	18.71%
Sum rate (Avg. over 24 hrs.)	*Sector A*	19.8 Mbps	20.03 Mbps	+ 1.14%
	Sector B	31.09 Mbps	31.36 Mbps	+ 0.87%
	Sector C	42.12 Mbps	42.52 Mbps	+ 0.95%

The per hour energy saving shows variations in power reduction. The highest power saving is observed 18.67% on 3rd hour (3:00 AM) and the lowest is 0% or no saving for three consecutive high load states. The power consumption reduction is emanated from deactivation of cells and lowering the number of serving CCs to two. The daily average energy saving performance was 14.6%. Besides, the average number of serving CCs evaluated 72.0%. This indicates the eNB in optimal operation was using lower number of serving bands for several hours of the day. In such operation, CA users devices multiple CCs monitoring power consumption would be reduced by 28.0%. In achieving the above energy saving, we have compared per sector achievable rates or sum rates of optimal and non-optimal cases.

As the daily traffic load traverses from 8% (low traffic) to 85% (peak hour) and depicted in orange color, the energy saving performance (blue color plot) with maximum value of 18.71% drops to 16.88% at mid-traffic and 0% at peak loads.

As a note to the end of this section, the operational network power consumption performance is our baseline. However, it is already using energy saving techniques such as DRX and hibernation at no traffic. Thus, the achieved performance is on top of those existing solutions.

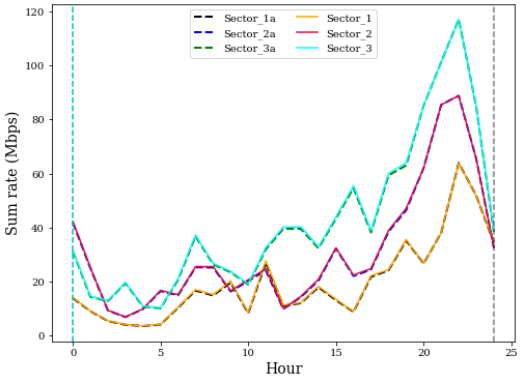

Fig.7. Per sector sum rates before and after energy optimized

V. CONCLUSION

The lower static power consumption ratio in 4G RF units compared to the legacy ones, and vice versa for traffic load dependent power, revealing promising hardware design improvements by manufacturers. On the other hand, more energy saving is expected from load adaptive operational saving solutions.

There showed times of 100% CCs demanding hours even after we optimized indicating the implemented CA by ethio telecom is mandatory to serve peak-hour loads. On the other hand, the energy saving hours of up to 18.71% and achieved at 2/3rd of CCs with minimal QoS impact ensures the approach we followed to model our energy saving problem as MDP followed by the use of AI or DRL trained BLCOO is applicable. Furthermore, same performances achieved both in DQN & PPO increases flexibility in choosing RL algorithms. However, the additional fixed memory spaces for replay buffer in DQN that can also increase for higher iteration steps and its limitation to handle continuous state values, PPO would be a preference with its additional clip ratio parameter to avoid large updates.

The proposed DRL-based BLCOO energy saving operational solution can be implemented with less complexity to trigger the auto sleep switching time i.e. DormancyTimer of RF units in the existing network. With such an hour lasting decision implementation period, idle resources can be redistributed to other eNB operating at its peak-load hence maximizes capacity and also saves soft licenses costs.

REFERENCES

[1] C.-T. Tung, Y.-L. Chung, and Z. Tsai, *An efficient power-saving downlink transmission scheme in OFDM-based multiple component carrier systems*. 2012.

[2] R. Falkenberg, B. Sliwa, and C. Wietfeld, “Rushing Full Speed with LTE-Advanced Is Economical - A Power Consumption Analysis,” in 2017 IEEE 85th Vehicular Technology Conference (VTC Spring), IEEE, Jun. 2017, pp. 1–7. doi: 10.1109/VTCSpring.2017.8108515.

[3] J. Joao Bazzo, R. de Melo Pires, A. Javier Ortega, N. Portela Salehi, and P. Thiago Marreiros Santos, “UE current consumption on carrier aggregation in LTE-A systems,” in 2020 Traffic-Aware Band-Level Cells On Off for Energy Saving in LTE-Advanced Networks with Inter-Band Carrier Aggregation 73 International Conference on Information and Communication Technology Convergence (ICTC), IEEE, Oct. 2020, pp. 600–602. doi: 10.1109/ICTC49870.2020.9289502.

[4] M. Choi et al., “Cell On/Off Parameter Optimization for Saving Energy via Reinforcement Learning,” in 2021 IEEE Globecom Workshops, GC Wkshps 2021 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/GCWkshps52748.2021.9682160.

[5] J. S. Pujol-Roigl, S. Wu, Y. Wang, M. Choi, and I. Park, “Deep Reinforcement Learning for cell on/off energy saving on Wireless Networks,” in 2021 IEEE Global CommunicationsConference, GLOBECOM 2021 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/GLOBECOM46510.2021.9685279.

[6] M. Elsayed et al., “Reinforcement Learning Based Energy-Efficient Component Carrier Activation-Deactivation in 5G,” in 2021 IEEE Global Communications Conference, GLOBECOM 2021 - Proceedings, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/GLOBECOM46510.2021.9685223.

[7] A. Mourato, D. Duarte, I. Pinto, and P. Vieira, “A Novel and Realistic Power Consumption Model for Multi-Technology Radio Networks.” doi: 10.23919/URSIRSB.2018.8486764.

[8] H. M. Taylor and S. Karlin, “An Introduction to Stochastic Modeling, Third Edition.”

[9] A. M. Metelli, “Configurable Environments in Reinforcement Learning: An Overview,” 2022, pp. 101–113. doi: 10.1007/978-3-030-85918-3_9.