Ref. Ares(2023)4658183 - 05/07/2023

MARIE CURIE ACTIONS

ELIXIRION

rEaLIzing healthcare 4.0 eXploIting the 6G netwoRk evolution

Marie Skłodowska-Curie Actions (MSCA)& Support to Experts

Doctoral Networks

HORIZON-MSCA-2022-DN-01

101120135 – ELIXIRION

ELIXIRION

WP1–High reliability and high capacity green 6G infrastructures for Healthcare 4.0

D1.1: SoA review on 6G healthcare infrastructures

Contractual Date of Delivery:	M12
Actual Date of Delivery:	22/11/2024
Responsible Beneficiary:	LU
Responsible Person:	Konstantinos Ntontin
Contact email:	kostantinos.ntontin@uni.lu
Contributing Beneficiaries:	LU, AUTH, MCS
Security:	Public
Nature:	R
Version:	4.0

PROPRIETARY RIGHTS STATEMENT

Document Information	
Version Date:	22/11/2024
Total Number of Pages:	43

List of Authors	
Organization	**Authors**
AUTH	Upaangana Das, Agapi Mesodiakaki, Amalia Miliou
LU	Alejandro Flores, Konstantinos Ntontin
MCS	Alexandros Alampanos, Fihmi Mousa, Andreas Schramm

Document History			
Revision	**Date**	**Modification**	**Contact Person**
0.1	09/09/2024	TOC Created	Konstantinos Ntontin
0.2	06/11/2024	First round of contributions	Upaangana Das, Alejandro Flores, Alexandros Alampanos
0.3	15/11/2024	Second round of contributions	Upaangana Das, Alejandro Flores, Alexandros Alampanos
0.4	22/11/2024	Checked by the coordinator's team	Agapi Mesodiakaki, Amalia Miliou

Table of Contents

Executive Summary 6
ABBREVIATIONS 7
1. INTRODUCTION 9
1.1 Background 9
1.1.1 Internet-of-Medical-Things 9
1.1.2 Non-terrestrial Networks 9
1.1.3 Flexible Functional Splits 9
1.1.4 X-haul Networks 10
1.2 Motivation 12
1.3 System Architecture 13
3. INTEGRATED TNs AND NTNs FOR UBIQUITOUS AND RELIABLE HEALTHCARE ACCESS 15
3.1 Literature review 15
3.1.1 Non-Terrestrial Networks 15
3.1.2 NTNs for the internet of medical things 16
3.1.3 Task offloading through LEO and GEO 17
3.1.4 Task offloading in integrated TN-NTN 18
4. INTEGRATED ACCESS AND X-HAUL NETWORK OPTIMIZATION FOR HIGH ENERGY- AND COST-EFFICIENT HEALTHCARE ACCESS 21
4.1 Literature Review 21
4.1.1 Offline-based SoA 21
4.1.2 Online-based SoA 26
5. CONTINUOUS LOW ENERGY COMPUTING (CLEC) TECHNIQUES ON SMART WEARABLE MULTI-SENSOR IOMT DEVICES TARGETING AT HIGH ENERGY EFFICIENCY AND PROLONGED BATTERY LIFETIME 29
5.1 Literature review 29
5.1.1 Introduction to Continuous Low Energy Computing (CLEC) Techniques 29
5.1.2 Approximate Computing (AC) and Near-Threshold Voltage (NTV) Computing 29
5.1.3 CLEC Applications in Edge and Fog Computing 30
5.1.4 Challenges and Limitations of CLEC 30
6. MATHEMATICAL TOOLS 32
6.1 Optimization Theory 32
6.2 AI-based Tools 33
6.3 Heuristics and greedy algorithms 34
6.4 Meta-heuristics 34
6.5 Stochastic Geometry 34

6.6 Game Theory 35
CONCLUSIONS 36
REFERENCES 37

Lists of Figures

Figure 1. Mobile access network. 9
Figure 2. System architecture. 13
Figure 3. Relative position of NTN nodes. 15

Lists of Tables

Table 1. Parameters of NTN nodes. 15
Table 2. Overview of key aspects on task offloading in NTNs literature. 20
Table 3. Bird-Eye view of Offline algorithms based SoA. 22
Table 4. Bird-Eye view of Online/Adaptive algorithm based SoA. 26
Table 5: Overview of Key Aspects in Energy-Efficient IoMT Architectures 31

Executive Summary

The aim of D1.1 is to provide a report of the literature review related to WP1: High reliability and high capacity green 6G infrastructures for Healthcare 4.0. Towards this, in Chapter 1 we first provide the related background that gives an overview of the current problems that existing networks face regarding the accommodation of healthcare related services. This smoothly leads to the rationale behind the consideration of the 3 areas of WP1, namely: i) the integration of terrestrial with non-terrestrial networks for ubiquitous and reliable healthcare access; ii) the joint access and X-haul network optimization targeting high energy- and cost-efficiency; and iii) the continuous low energy computing techniques on smart wearable multi-sensor IoMT devices targeting at high energy efficiency and prolonged battery lifetime. Finally, this chapter introduces the system architecture of WP1.

Subsequently, Chapters 2, 3, and 4 first provide an elaborate literature review on the aforementioned 3 areas, identifying the different domains and applications. Finally, Chapter 5 introduces the common mathematical tools to be used in the envisioned work, namely optimization theory, artificial intelligence, heuristics and greedy algorithms, meta-heuristics, stochastics geometry, and game theory.

ABBREVIATIONS

Abbreviations	Full Form
AC	Approximate Computing
ADMM	Alternating Direction Method of Multipliers
AOS	Atomic Orbital Search
BBU	Baseband Unit
BCD	Block Coordinate Descent
BS	Base Station
B5G	Beyond 5G
CCCP	Concave Convex Procedure
CLEC	Continuous Low Energy Computing
CU	Centralised Unit
C-RAN	Centralised Radio Access Network
CNC	Center Network Controller
CPU	Central Processing Unit
DPG	Deterministic Policy Gradient
DQN	Deep Q Networks
DRL	Deep Reinforcement Learning
DU	Distributed Unit
eMBB	Enhanced Mobile Broadband
GA	Genetic Algorithm
GEO	Geostationary Earth Orbit
GW	Gateway
H4.0	Healthcare 4.0
HAPS	High-Altitude Platform Station
IAB	Integrated Access and Backhaul
IoMT	Internet of Medical Things
IoT	Internet of Things
ITS	Intelligent Transportation System
KKT	Karush Kuhn Tucker
LEO	Low Earth-Orbit
LP	Linear Programming
MARDPG	Multi-Agent Recurrent Deterministic Policy Gradient
MDP	Markov Decision Process
MEC	Multi-Access Edge Computing
MEO	Medium Earth-Orbit
MILP	Mixed Integer Linear Programming
MIMO	Multiple-Input Multiple-Output

MINP	Mixed Integer Nonlinear Programming
mmWave	Millimeter-wave
NFV	Network function Virtualisation
NGSO	Non-Geostationary Orbit
NOMA	Non-Orthogonal Multiple Access
NP	Non-deterministic Polynomial
NTNs	Non-Terrestrial Networks
NTV	Near-threshold voltage
PPO	Proximal Policy Optimization
PSO	Particle Swarm Optimization
PXNs	Packet X-haul Networks
QoS	Quality of Service
RAN	Radio Access Network
RPM	Remote Patient Monitoring
RU	Radio Unit
RRH	Remote Radio Heads
SA	Simulated annealing
SAC	Soft Actor-Critic
SCA	Successive Convex Approximation
SG	Stochastic Geometry
SINR	Signal to Interference Noise Ratio
TN	Terrestrial Network
UAV	Unmanned Aerial Vehicle
URLLC	Ultra reliable low latency Communications
UTDs	User Terminal Devices
VNFs	Virtual Network Function
VSAT	Very Small Aperture Terminal

1. INTRODUCTION

1.1 Background

Healthcare 4.0 (H4.0) is a new revolution in healthcare, which foresees the interconnectivity of patients, healthcare providers and healthcare devices for improved monitoring, intervention and treatment for preventive and proactive care. To accomplish this, a level of interconnectivity must be guaranteed for all the stakeholders and devices. This interconnectivity must be provided with the required quality of service, while maintaining an efficient use of the resources. To that end, a number of technologies, which are detailed in the following, are expected to play a key role. These technologies include:

1.1.1 Internet-of-Medical-Things

An important element in H4.0 is the internet-of-medical-things (IoMT), which is an instance of internet of things (IoT), applied to devices used for healthcare. IoT is a paradigm consisting of a large number of devices, interconnected through the internet, or any other network. These devices usually have functionalities of sensors, actuators or coordinators and are commonly used for monitoring and automation. IoT devices are supposed to work autonomously as long as possible, for which battery management is an important issue. In the context of remote patient monitoring (RPM), some examples of IoMT devices include continuous glucose monitoring sensors [1], photoplethysmography sensors [2], portable electrocardiograms [3] and stress sensors [4].

1.1.2 Non-terrestrial Networks

IoMT monitoring devices need access to a network for offloading and processing the data it acquires. However, while 95% of the Earth's population has cellular coverage, less than 45% of the landmass of the Earth [5] and only 15% of its surface has such coverage [6]. In those areas, IoMT devices can only rely on non-terrestrial network (NTN) infrastructure to provide remote computing and routing services, complementing the terrestrial networks (TNs). NTNs may be comprised of unmanned aerial vehicles (UAVs), high-altitude platform stations (HAPSs), geostationary earth orbit (GEO) satellites, or non-geostationary orbit (NGSO) satellites, such as medium earth-orbit (MEO) satellites and low earth-orbit (LEO) satellites. Such a solution, apart from being the only possibility for dispatching essential medical data to healthcare providers, can even result in lower end-to-end latencies compared with the terrestrial fibre-only infrastructure [7].

1.1.3 Flexible Functional Splits

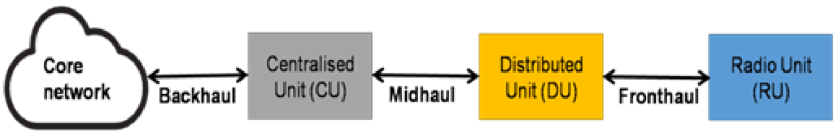

Figure 1. Mobile access network.

To meet the rigorous demands of modern healthcare applications in terms of **low latency, reliability, flexibility and efficiency**, X-haul networks and functional splits are to be leveraged. In telecommunications, **functional splits** refer to the separation of different processing tasks or "functions" within the network architecture, particularly in **radio access networks (RANs)**. This concept is central to 5G and beyond, where network components such as the radio unit (RU), distributed unit (DU), and centralized unit (CU) are separated and can be deployed flexibly across the network (see Figure 1).

Functional splits determine where and how specific processing tasks (e.g., signal encoding, data packet routing, and user control) occur within the RAN. Some commonly referenced splits include:

1) **High Layer Split (HLS)** - This is a split closer to the upper layers, often splitting control-plane processing from user-plane functions, which can be more centralized.
2) **Low Layer Split (LLS)** - This split occurs closer to the physical layer, so that more signal processing is centralized in the baseband units, allowing the RU to handle only the essential signal transmission tasks.

However, with rapid growth in mobile data traffic demand, adopting a fixed functional split is not a sustainable solution in the long term. Thus, given the rising mobile traffic demand and daily variations in traffic, it is crucial to have the flexibility to dynamically select the optimal functional split to effectively manage fronthaul bandwidth and baseband processing resources.

A **flexible functional split** allows operators to leverage the advantages of different splits by enabling them to adjust the split option according to their specific requirements. For instance, since various functional splits have distinct fronthaul bandwidth and latency demands, this flexibility allows operators to deliver tailored QoS settings for each service (e.g., low latency, high throughput). Additionally, it enables them to adapt to specific user densities and load demands across different geographic areas.

There are several splits considered in literature [8]. Among them, **Open Radio Access Network (ORAN)** alliance 3-layer split is the most popular one and is demonstrated in Figure 1. A lower layer split (LLS), which is close to the radio frequency (RF) refers to higher bandwidth on the fronthaul link and catering to very tight latency requirements. It also means the bandwidth scales with the increase in antenna ports. On the other hand, a higher layer split which is close to packet data convergence protocol (PDCP), means there is a drastic reduction in bandwidth and relaxation of latency leading to long distances between the CU and the RU. As far as scalability is concerned, higher layer splits scales with MIMO layers. The advantages of implementing a split architecture as outlined in 3GPP Release 14.0 are as follows:

1) **Flexible Hardware Implementation**: It supports scalable and cost-effective solutions, making it easier to meet deployment requirements.
2) **Coordination and Load Management**: By splitting the architecture between central and distributed units, it facilitates coordination for performance features, effective load management, and real-time performance optimization.
3) **Adaptability to Various Use Cases**: Configurable functional splits enable the system to adapt to different use cases, such as handling variable latency in transport networks.
4) **Virtualization**: Functional splits and disaggregation allow the use of diverse software solutions from different vendors for various purposes.

1.1.4 X-haul Networks

With the foresight of a large amount of IoMT-generated data, and heterogeneous access through TN and NTN nodes, H4.0 requires scalable network architectures. X-haul with functional splits enables **flexible scaling** for devices used in H4.0 by allocating resources dynamically, allowing networks to adapt to high device densities without compromising performance. For instance, a split architecture can easily expand computational resources in high-demand hospital areas while limiting resources in low-traffic zones, optimizing overall network load.

X-haul networks are referred to as either backhaul, midhaul or fronthaul (Figure 1).

- **Backhaul:** This is the network that connects the core network (e.g., data centers, the internet) to local distribution points, such as cell towers.
- **Fronthaul:** The link between the baseband units (BBUs) and remote radio heads (RRHs) in cellular networks, particularly in 5G.
- **Midhaul:** Midhaul connects the different elements of the RAN and can include links between small cells, hubs, and central base stations.

Further, to meet the required quality of service (QoS), integrated access backhaul (IAB) [9], proposed as a 5G New Radio fiber-alternative, can provide emerging data-hungry healthcare applications (e.g., hologram communication requiring up to a few Tbps [10]) with increased coverage via directional high-antenna-gain links, reliability via dual-connectivity-enabled IAB, and reduced congestion via flexible routing. Thus, IAB is envisioned to play a key role in 6G, alongside multi-GHz bands and NTN base stations, such as UAVs.

Finally, as we look at the future of IoMT, most modern devices still use conventional task offloading models, which severely hurt device lifetime due to continuous data transfer and service response [11]. Short battery lifetime, however, hinders the continuous vital signal monitoring by wearables, leading users to interrupt monitoring for the duration of charging it and even more, thus, ditching its use. Therefore, it is necessary to drastically improve the energy and autonomy of the wearables leveraging the computation capabilities of modern embedded systems.

1.1.5 Energy Efficiency and Autonomy in IoMT Wearables

The continuous monitoring of vital signs via IoMT wearables is becoming vital in modern healthcare, yet many current devices face limitations due to conventional task offloading models that significantly reduce battery life through constant data transfers for real-time health monitoring [12]. Such constraints on battery duration interrupt the monitoring of vital health metrics, as devices need frequent recharging, causing disruptions in data collection and potentially discouraging users from adopting these wearables long-term.

Addressing these energy challenges is crucial for sustainable health monitoring solutions, particularly as real-time data flow is critical in many healthcare applications. Recent advancements in embedded systems now allow for significant improvements in both energy efficiency and operational autonomy of IoMT devices. Techniques such as Continuous Low Energy Computing (CLEC) and Near-Threshold Voltage (NTV) scaling enable wearables to process data locally with reduced energy consumption, thus extending operational time without requiring frequent recharges [13], [14]. Specifically, CLEC achieves this by allowing approximate computations for non-essential tasks, effectively conserving energy for more critical functions. NTV scaling further supports these devices to function at minimal voltage levels, optimizing energy use while preserving monitoring capabilities.

These innovations enable IoMT wearables to perform continuous, real-time health monitoring independently, aligning with the goals of H4.0 for personalized, preventive care. By improving battery longevity and ensuring stable functionality even in resource-constrained settings, CLEC and NTV scaling collectively set the groundwork for more autonomous, sustainable wearables that can seamlessly integrate into patients' lives without frequent interruptions for charging [15], [16].

1.2 Motivation

Despite the potential advantages of NTNs, currently, there is no framework for the seamless integration of large-scale TN/NTNs consisting of LEO satellites, HAPSs, and UAVs that addresses their coordination and resource management, to meet the challenging H4.0 service requirements. ELIXIRION will provide a general framework for the distributed self-organization and resource optimization of wide-scale aerial/space NTNs comprising LEO satellites, HAPS, and UAVs that operate concurrently and are responsible to convey a large volume of IoMT generated data. Our goal is to provide dynamic routing algorithms that deal with the complexity and the dynamicity of mesh topology, caused by node mobility. To that end, convex optimization and AI/ML data-driven optimization will be leveraged together with offline data from analytical models when the collection of a large amount of online data is prohibitive [17]. In addition, NTN nodes will be also equipped with edge computing capabilities for alleviating the workload of TN edge nodes or in cases that the TN infrastructure is absent.

In regard to access, widespread IAB adoption introduces significant challenges, which have not been fully addressed yet in terms of topology control, interference coordination and capacity sharing, especially for time division duplex (TDD)-based mmWave transmissions and NTN base stations (BSs). Thus, development of scalable routing and scheduling algorithms is needed [18] for joint control of IAB/non-IAB links sharing network/compute resources. ELIXIRION formulates a joint access/X-haul problem, leveraging multi-GHz bands, computing capabilities at different nodes and incorporating both IAB/non-IAB links under a unified resource allocation scheme, so that all important network controls (e.g., TDD schedule, access/X-haul spectrum splitting, selected routes, virtual network function (VNF) placement) are jointly optimized dynamically, to exploit inherent spatiotemporal traffic fluctuations, diverse compute capabilities and BS availability.

Finally, despite advancements in IoMT technology, wearable devices are still limited by traditional power management models that require frequent recharging, constraining their ability to provide continuous, real-time health monitoring. Addressing these energy challenges is essential for sustaining patient monitoring in applications requiring uninterrupted data flow. In response, the ELIXIRION project integrates Continuous Low Energy Computing (CLEC) and Near-Threshold Voltage (NTV) scaling—innovations aimed at reducing energy demands and extending device autonomy. By leveraging Approximate Computing (AC), CLEC minimizes the energy required for IoMT data processing, allowing wearables to execute critical tasks locally and thereby conserve power—a key requirement for fully autonomous operation in real-time health monitoring environments [13], [15]

Additionally, NTV scaling enables ELIXIRION wearables to operate near their minimum voltage threshold, significantly reducing energy consumption without compromising monitoring capabilities. This approach not only prolongs battery life but also enhances data privacy by reducing dependency on external servers, thereby securing health data and optimizing real-time functionality. Collectively, CLEC and NTV techniques enable a self-sustained IoMT ecosystem that aligns with the Healthcare 4.0 vision for 24/7 monitoring, providing actionable health insights directly on wearable devices, and supporting personalized preventive medicine and proactive disease management [14], [16].

1.3 System Architecture

The envisioned system architecture for the ubiquitous interconnectivity of H4.0 can be seen in Figure 2.

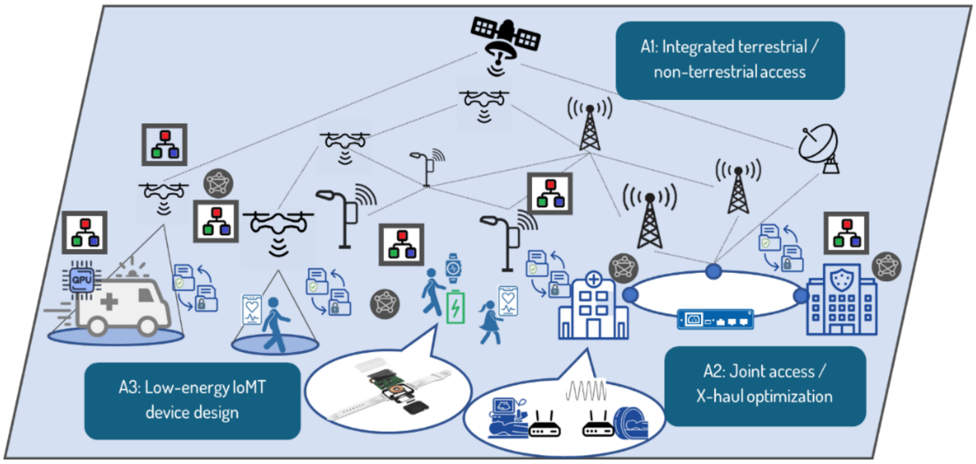

Figure 2. System architecture.

The system conveys the scenario of low energy IoMT devices accessing the network for data offloading directly or through second devices. The devices access the network through IAB or non-IAB links, making an efficient use of network resources. Devices in remote areas where TN infrastructure is not available, will make use of NTN nodes as relays to geographically separated gateways, or will aid with remote task computing. Under this framework, a continuous, reliable and efficient interconnectivity of H4.0 patients, healthcare providers and healthcare devices, is achieved.

The following elements are considered:

- Low energy IoMT devices, which monitor patients' health-related information and signals. Designed as to utilize its resources, and those of the network, efficiently.
- Aerial nodes, which provide immediate access to users and devices on the ground, which connect to the terrestrial infrastructure, as well as to other aerial or space nodes.
- Satellite nodes, with large coverage areas that can connect to aerial or ground nodes for information relating to healthcare facilities or to computing tasks. They connect directly to ground devices, or to intermediate RAN nodes, such as UAVs, HAPSs or ground BSs.
- Gateways, which connect directly to satellites through highly directive links for information relaying, and to the terrestrial infrastructure to effectively connect remote ground users.
- Ground BSs, that may not be available to the ground users but are accessible to the non-terrestrial network through ground gateways, or direct connections to aerial or satellite nodes.
- MEC servers, incorporated into the TN and NTN infrastructure for low-latency computing, and to alleviate the energy consumption of energy-constrained ground nodes.

- Fronthaul, midhaul and backhaul links between TN or NTN RAN nodes. The links can be non-IAB or IAB for efficient resource utilization across the network.

3. INTEGRATED TNs AND NTNs FOR UBIQUITOUS AND RELIABLE HEALTHCARE ACCESS

3.1 Literature review

3.1.1 Non-Terrestrial Networks

There has been increasing research interest in the integration of TNs and NTNs in the past years, particularly with the latest LEO constellation deployments. However, LEO satellites are not a new technology. Indeed, the first artificial Earth satellite was the LEO Sputnik 1, which was launched in low-earth orbit in 1957 [19], one of the first communications satellites, the LEO Telstar 1, was launched in 1962 [20] and the first LEO constellation, the first-generation Iridium constellation of 66 LEO satellites, completed its launch in 2002 [21]. The modern deployment of LEO satellites, which seek to provide global broadband coverage, was facilitated by advancements in miniaturization, a drop in cost of launch and an increase in demand on broadband access.

The launch of modern LEO constellations began in the late 2010s, with some of the modern LEO constellations deployments being the Starlink constellation which currently consists of over 7000 LEO satellites, the OneWeb constellation with 634 operational LEO satellites, the Iridium NEXT constellation with 80 operational LEO satellites, the Telesat Ligthspeed constellation with 3 LEO satellites in orbit and plans for over 1600, and the Kuiper constellation with 2 LEO satellites in orbit and plans of launching over 3000 in total.

Apart from LEO satellites, other NTN nodes can be GEO and MEO satellites, HAPSs and UAVs. Their common placement relative to ground level can be seen in Figure 3.

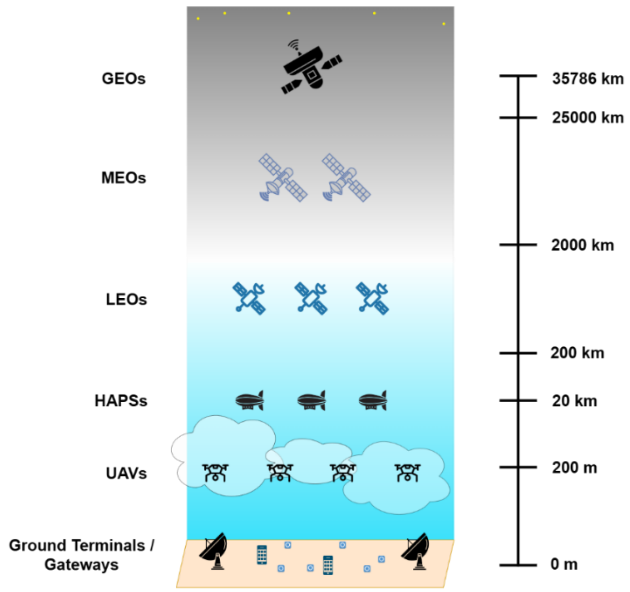

Figure 3. Relative position of NTN nodes.

A summary of the common altitudes, beam footprint size and mobility with respect to the surface of Earth of NTN nodes are given in Table 1.

Table 1. Parameters of NTN nodes.

Node	Altitude	Atmospheric Layer	Beam Footprint	Mobility

UAV	Hundreds of meters	Troposphere	Some kilometres	Fixed/Mobile
HAPS	Around 20 km	Stratosphere	5-200 km	Fixed/Mobile
LEO	200 – 2000 km	Thermosphere / Exosphere	100-1000 km	Mobile
MEO	2000 – 25000 km	Exosphere	100-1000 km	Mobile
GEO	35786 km	Exosphere	200-3500 km	Fixed

Satellites commonly have three types of communication links

- **Service link:** through which they connect to the ground terminals.
- **Feeder link:** through which they connect to the ground gateway (GW).
- **Inter-satellite link (ISL):** through which they connect to other satellites.

The frequency bands for integrated TN-NTN communications, considered by 3GPP [22] are

- **Ka-Band:** downlink 19.7 - 21.2 GHz, uplink 29.5 - 30 GHz. Intended for very small aperture terminals (VSAT) with directional antennas.
- **S-Band:** downlink 2170 - 2200 MHz, uplink 1980 - 2010 MHz. Intended for user equipment, such as handheld devices.

Depending on the processing done at a satellite, their payloads can be classified as

- **Transparent payload:** performs signal processing over the waveform at the lower layers, such as amplification, frequency conversion and filtering. It can work on analogue or digital domain, but it does not recover the sent symbols. These are the satellites currently in orbit
- **Regenerative payload:** a larger portion of the communication stack is present. It allows for the modulation/demodulation and coding/decoding of symbols, as well as the switching and routing of packets. It is not yet implemented in satellites.

The communication channels between ground terminals and satellites must account for atmospheric effects on the wireless signal. The main source of atmospheric impairments is the troposphere, where all the weather-related phenomena occur. The sources of attenuation in the troposphere are gas absorption [23], [24], cloud attenuation [25], scintillation [26], and most prominently, when present, rain attenuation [26], [27], [28], [29].

Further details on NTNs can be found in the surveys [30], [31] and [32]. Moreover, 3GPP standardization efforts on NTNs can be found on [22].

3.1.2 NTNs for the internet of medical things

IoMT devices allow for continuous monitoring of the bio-signals of the patient, which enables a more personalized healthcare. Particularly, the processing of the bio-signals can help detect adverse conditions within the patient, such as detecting fever from temperature sensors [33], cough from acoustic sensors [34], fatigue from heart rate [35], or shortness of breath from oxygen-based sensors [36]. Thus, IoMT devices not only acquire data for monitoring, but they also need to process it, either locally, at a linked smart-device or at a remote server. Such computing tasks generated by IoMT devices may have a maximum-allowed delay in terms of total processing time. IoMT devices, are also energy constrained, with limited battery life. Thus, remote processing of the tasks allows for an optimal use of resources, both local and remote.

Remote processing can be performed at remote cloud servers, or multi-access edge computing (MEC) servers:

- **Cloud servers:** commonly located at the core of the local or a remote network. Computing capabilities are vast, but a significant amount of latency is introduced, due to distance between the server and the end device, and the routing processing done by in-between devices.
- **MEC servers:** located closer to the end device. They have limited computing capabilities, but the latency introduced by the end-to-end transmission of the task is lowered. Also, security is improved by reducing the number of shared links used to transport the task.

The problem of resource allocation for task offloading considers a scenario, in which users generate tasks that can be computed locally or in a variety of remote nodes, where parameters of the system are leveraged to compute the tasks, such that some metric is optimized. Task offloading problems tend to consider two main metrics to minimize: the total energy spent by the system, and the total delay to compute the tasks in the system. There are also other metrics considered in the literature, such as service lease price, maximum delay, queue stability, load balancing, number of successfully computed tasks or combinations of the previous. The parameters of the system that are leveraged for task offloading, which we refer as optimization variables, can be, among others, transmission power of the nodes, offloading decision (where to compute), association decision, computing resource allocation, bandwidth, subchannel allocation, precoder and combiner design.

In this context, NTN nodes can act as computing servers or relays when IoMT devices intend to offload the processing of their tasks to a remote node, and no terrestrial infrastructure is available to access the network.

3.1.3 Task offloading through LEO and GEO

GEO satellites have a larger beam footprint, and thus, coverage. However, their high altitudes induce a considerable amount of delay and introduce substantial pathloss in their links to the surface of the Earth. Thus, in the context of task offloading, they are mainly considered for backhaul of satellite networks, or as cloud servers. For example, [37] considers the joint communication and computing resource allocation for satellite edge enabled IoT task offloading with collaborative computing across LEOs and a GEO as a relay for computing at a ground GW. The problem aims to minimize the total latency across tasks, and the optimization variables are the offloading decisions, per-beam scheduling, subchannel allocation and computing resources. The computing resource allocation is solved in closed form through Karush Kuhn Tucker (KKT) conditions, while the joint task offloading and channel allocation is solved through deep proximal policy optimization (PPO).

As an instance of a GEO cloud, [38] considers the task offloading from ground nodes to a MEC-enabled LEO network, or further to a GEO cloud, accounting for the mobility and coverage time of the LEOs. The problem minimizes the total delay of the tasks and energy consumption across the network, leveraging the offloading decision of the tasks. The solution is given in a distributed convex optimization algorithm based on alternating direction method of multipliers (ADMM) across LEOs.

Similarly, exemplifying a multi-layered NTN for task offloading, [39] considers the task offloading from remote intelligent transportation system (ITS) nodes to a multi-layered NTN comprising MEC CubeSats, MEC and scheduling LEOs or a cloud GEO. It minimizes the mean delay and computing service price by optimizing the offloading decisions, bandwidth allocation, computing allocation and cloud computing power. The offloading problem is solved by a cooperative multi-agent deep PPO scheme, while the other variables are iteratively solved through convex optimization techniques.

On the other hand, LEO satellites induce a lower amount of latency and pathloss to wireless communications from the surface of the earth, making communications between terrestrial nodes, UAVs or HAPSs to LEO satellites less stringent. The use of LEO satellites in task offloading has been the subject of research in the past years. For example, [40] considers the task offloading of ground IoT devices to a remote MEC server, using LEO satellites as relays. It considers a central ground node for coordination and information sharing. It minimizes the energy of IoT devices and LEOs and processing delay of the tasks, with the decision variables as the IoT-LEO bandwidth, the LEO-MEC bandwidth and the IoT-LEO transmit power. The problem is modeled as a partially observable Markov decision process (MDP). Each LEO, after receiving information from the ground info center, uses an actor-critic architecture to establish the bandwidths (in uplink and downlink) and user power allocations.

As an example of MEC-enabled LEOs, [41] considers the problem of task offloading to a MEC-enabled LEO network with peer-offloading capabilities. LEOs determine task admission from ground nodes and possible task offloading to other LEOs at each snapshot. It performs online minimization of the long-term task delay and energy consumption across LEOs over the task admission and offloading decisions utilizing Lyapunov optimization. It proposes a centralized approach and a simplified distributed approach.

As with the previous work, there is research on the long-term minimization of system parameters, such as in [42], which studies the long-term offloading decision of mobile users to a LEO satellite. The offloading is performed with consideration to the capacity of the task queues at users and at the LEO. The offloading can be performed locally or at the LEO, and non-processed packets remain for future processing. The solution considers queue long-term stability, and is handled by Lyapunov optimization and DRL, while the rest of the variables are optimized via block coordinate descent (BCD).

3.1.4 Task offloading in integrated TN-NTN

The use of satellites for task offloading allows for global coverage and ubiquitous availability of computing resources. Nonetheless, the communication channels to the LEO induce relative high latency and pathloss, which is an issue particularly in coordination. LEOs also have limited resources on board. For these reasons, the use of other aerial nodes, such as UAVs and HAPSs, proves useful in task offloading in NTNs, particularly for delay-constrained tasks. Some examples of the use of aerial and satellite nodes in task offloading are included in [43], which considers the task offloading framework involving civil aircrafts and LEOs as edge servers and ground stations as cloud servers with partial processing. The multi-objective problem minimizes the weighted sum of energy consumption and end-to-end delay, where the optimization variables are the association strategy, the amount of processed data, the computing resources, the offloading ratio and user transmit power. The solution is found by utilizing generalized benders decomposition to solve the primal problem for the transmit power, and the master problem over the rest of the variables, by means of successive convex approximation (SCA).

Contemplating the use of UAVs as an access node before LEOs, the work in [44] considers the task offloading from ground nodes to cooperative MEC-enabled UAVs or further to a MEC-enabled LEO. It minimizes the total task latency while accounting for maximum energy constraints, over the offloading decision variables from

IoT-UAV, from UAV-UAV and UAV-LEO. The problem is solved through a block successive upper-bound minimization algorithm.

As an example of the use of a HAPS as coordinator, [45] considers the task offloading from ground industrial IoT nodes to a set of collaborative MEC-enabled UAVs coordinated by a MEC-enabled HAPS. It maximizes the number of satisfactory connections defined as the ones that meet a delay threshold, by optimizing over the power allocation, subcarrier allocation, computing resources and UAV trajectory design. The solution is given by a multi-agent recurrent deterministic policy gradient (MARDPG) architecture, with centralized training and decentralized execution scheme. On the other hand, aerial nodes can be MEC servers for tasks generated by satellites, such as in [46], which considers the offloading of tasks generated by LEOs, to aircrafts, other LEOs or the cloud. The queue-aware problem minimizes the long-term total task delay by the offloading and scheduling decisions. The problem is solved via Lyapunov optimization.

Moreover, given the limited resources in NTN nodes, a common approach as well, is to consider integrated TN-NTNs in which ground MEC or cloud servers are also available for processing, and the NTN nodes serve as MEC servers or as relays for ground computing. TN nodes may also be considered as relays towards NTN nodes. For example, [47] considers the task offloading through a BS-LEO-core path, or through a LEO-core path. It uses a centralized controller for decision making, which minimizes the total task delay by optimizing the path selection and bandwidth allocation. The paper models the problem as a two-timescale hybrid MDP with a large time-scale agent that handles the offloading and allocations to LEOs and a short-timescale agent which handles the offloading and allocations to the BSs. The agents are modelled as a hybrid deep PPO solution with action preprocessing to accommodate continuous and discrete actions.

As another example, which employs partial offloading, [48] studies the multi-layer offloading decision of user terminals directly to a LEO and/or subsequent offloading to a ground GW. Partial offloading is assumed, where a fractional amount of the task is computed at each layer, having this fraction as a degree of freedom. The problem is solved by optimizing the transmit precoding of multiple-input multiple-output (MIMO) signals, and the computing resource allocation at each layer.

If the access to the TN-NTN nodes is not orthogonal, interference will be experienced, such as described in [49], which considers the offloading of tasks generated by ground nodes, to MEC-enabled LEOs, or further relayed to ground cloud. UE-LEO access is based on non-orthogonal multiple access (NOMA), but the LEO-GW backhaul is orthogonal in frequency. The solution minimizes the total energy consumption of the system by leveraging the user association, power allocation, task scheduling (partial offloading) and bandwidth assignment. It is solved by BCD over each set of variables iteratively using convex optimization tools.

Even if the LEO links induce considerable latency, they may constitute a preferable path to the cloud, rather than the TN, such as in [50], which considers the resource allocation for partial task offloading at local ground MEC, at cloud through ground (wireless) backhaul, or at cloud through ultra-dense LEO formation backhaul. It considers average task generation and processing rate by the ground terminals and the servers respectively. It minimizes the total energy spent in the system with optimization variables as the task partition, cell association, user transmit power, MEC computing resources and cloud offloading decision. The offloading and association variables are solved with a multilayer perceptron (MLP), while the rest of the variables are jointly optimized through SCA.

Lastly, considering different timescales, the deployment of specific services in MEC-enabled LEOs may also be leveraged, as in [51], which considers the task offloading and service deployment in a cooperative LEO network, for tasks generated by ground users. The system considers multi-hop offloading to MEC-enabled LEO, or to a ground cloud server. The problem aims to improve load balancing and minimize energy consumption of the nodes and packet loss rate. In the small timescale, it optimizes over the task scheduling using Lyapunov-based transformations and applying deep reinforcement learning through a soft actor-critic (SAC) architecture. In the large timescale, it optimizes over the deployment of service replicas in the MEC servers, solved through the atomic orbital search (AOS) metaheuristic.

While the literature considers many layers in TN-NTN integration for resource allocation in task offloading, the widespread nature of TN-NTN networks has not been completely addressed in terms of scalability. Centralized solutions are not adequate for networks that span the Earth, and even distributed solutions that place the algorithm execution at nodes that have to transverse high propagation delay links between the ground and space, introduce delay that is not adequate for some applications.

An overview of key aspects of the works described can be found in Table 2.

Table 2. Overview of key aspects on task offloading in NTNs literature.

Ref.	Nodes					Opt. Metric			Solution			Computing		
	UAV	HAPS	LEO	MEO GEO	TN	Latency	Energy	Other	Opt.	AI	Heur.	Local	MEC	Cloud
[37]			x	x	x	x			x	x			x	x
[38]			x	x		x	x		x				x	x
[39]			x	x		x		x	x	x			x	x
[40]			x		x	x	x			x			x	
[41]			x			x	x		x			x	x	
[42]			x				x		x	x		x	x	
[43]		x	x		x	x	x		x			x	x	x
[44]	x		x			x			x			x	x	
[45]	x	x						x		x			x	
[46]	x	x	x		x	x		x	x			x	x	x
[47]			x		x	x				x			x	
[48]	x		x		x		x		x				x	
[49]			x		x		x		x			x	x	x
[50]			x		x		x		x	x			x	x
[51]			x		x		x	x	x	x	x		x	x

4. INTEGRATED ACCESS AND X-HAUL NETWORK OPTIMIZATION FOR HIGH ENERGY- AND COST-EFFICIENT HEALTHCARE ACCESS

4.1 Literature Review

In telecommunications, "integrated access" refers to the integration and optimization of different network access technologies (e.g., 5G and 6G) to ensure seamless and reliable connectivity. The goal is to allow users to move across different access networks without interruption and with minimal energy consumption. Optimizing joint access and X-haul provides reliable, low-latency, high-bandwidth communication, with a focus on reducing energy consumption and cost. The X-haul network refers to either backhaul [52] [53], midhaul or fronthaul [54].

As mobile access networks evolve into denser configurations (from 1G to 5G and 6G), they face challenges related to interference management, network design, and efficient resource utilization. A promising solution to these issues is joint operation, which enhances network flexibility, effectiveness, and resource efficiency [55]. This approach integrates access and X-haul networks, enabling resource pooling, interdependence, and improved cooperation for more efficient network management.

Healthcare 4.0 integrates technologies such as Internet of Medical Things (IoMT), AI and Machine learning, Big Data, Multi-Access Edge Computing and Cloud computing into healthcare systems. It focuses on creating intelligent, interconnected systems that improve patient care, hospital operations, and health outcomes. **Joint access and X-haul network optimization in Healthcare 4.0** is essential to handle the growing complexity, scale, and real-time demands of modern healthcare systems while keeping energy consumption and operational costs low. By optimizing the network holistically, healthcare providers can deliver better quality services with improved efficiency, security, and reliability, which is crucial for patient care and technological advancement in healthcare. It can ensure that the network infrastructure can support the demands of healthcare applications, which are often mission-critical, without incurring excessive energy costs or operational expenses.

Figure 1 above demonstrates the **Open Radio Access Network (ORAN)** alliance 3-layer split, which is the most popular one among several other functional splits [8]. Radio Unit (RU), Distributed Unit (DU), and Centralized Unit (CU) are key architectural concepts in the ORAN framework. The decision on how to split New Radio (NR) functions within the architecture is influenced by factors such as the deployment scenario, network constraints, and the desired software services. In [56], focus is on enhancing the resiliency and efficiency of 5G X-haul networks through reliable functional splitting. X-haul networks used in 5G, integrates wired and wireless technologies under a unified framework for flexible and software-defined management.

In this section, a thorough study into 5G/6G access and X-haul networks is provided. The algorithms used in networks can be classified as either online or offline, primarily based on the stage of the network in which they operate. **Offline algorithms** are typically applied during the network deployment phase, where they rely on pre-existing data to support tasks such as network planning and optimization. In contrast, **online algorithms** function during the network's operational phase, dynamically adapting to real-time changes like user mobility, sudden link failures, or temporary blockages. So, based on the nature of the algorithm, the following subsections present the state-of-the-art (SoA) on it.

4.1.1 Offline-based SoA

This sub-section presents a comprehensive study on offline algorithms-based network deployment and optimization which leverages pre-existing data to support essential tasks in network planning. These algorithms operate without real-time inputs, relying instead on historical data, traffic patterns, and network usage forecasts gathered through measurements, simulations, or statistical analysis.

Table 3. Bird-Eye view of Offline algorithms based SoA.

Ref.	Cost Efficiency	Energy Efficiency	Resource Allocation					Flexible Functional Split
			Spectrum	Power	User Association	Traffic Routing	Computation	
[18]			x		x	x		
[57]		x	x	x	x	x	x	x
[58]		x	x		x	x	x	
[59]		x	x	x	x			
[60]					x	x		
[61]	x	x		x	x	x		
[62]		x	x	x	x			
[63]	x	x				x		x
[64]	x					x	x	x
[65]	x					x		
[66]		x		x	x			
[67]		x				x	x	
[68]		x		x	x		x	
[69]	x				x	x	x	x
[70]			x		x			

As shown in Table 3, the papers [57], [58], [59], [61], [62], [63], [67] discuss methods for optimizing **energy consumption** in networks.

The authors in paper [57] address the rising energy consumption in industrial wireless sensor networks (IWSNs) as the world progresses toward beyond 5G (B5G) and 6G networks. The paper mainly focuses on improving energy efficiency through a multi-agent system (MAS) combined with distributed artificial intelligence (DAI). The method clusters sensor nodes, predicts their main node locations using backpropagation and convolutional neural networks (CNN), and allocates resources efficiently. It offers a new approach focusing on efficient resource allocation and clustering methods. However, scalability (particularly in handling large networks or highly dynamic environments with low latency requirements) and environment adaptability to real-world conditions like changing satellite link quality, mobility, may limit the performance in actual deployment of their proposed work. Also, optimizing node positioning standards for different IoT applications is one of the research gaps that needs to be explored.

Paper [58] also focuses on addressing the challenges of optimizing energy efficiency while managing computational and network resources in B5G mobile networks. The paper is based on the placement of Virtual Network Functions (VNFs) and the allocation of computational and communication resources, with the goal of minimizing power consumption and ensuring service requirements. It presents Network function Virtualisation (NFV) emphasizing its ability to support diverse services with high flexibility and configurability.

VNFs need to be placed optimally within the network to meet requirements like throughput and delay while minimizing power usage. The work jointly addresses VNF placement, traffic routing, and user association while explicitly accounting for the Access Network (AN).

Similarly, paper [67] also explores the energy-efficient placement of VNFs in NFV-enabled telecom networks. It addresses the challenge of minimizing energy consumption while ensuring efficient traffic routing. The study presents an Integer Linear Programming (ILP) model to optimize VNF placement and traffic routing. Given the NP-hard nature of the problem, the authors propose a polynomial algorithm using the Markov approximation technique to find near-optimal solutions. Simulation results show that the proposed approach reduces energy consumption by up to 14.84% compared to previous methods, making it highly effective for telecom networks. The research emphasizes balancing server energy consumption and communication resources to enhance overall network efficiency. However, the proposed solution involves an ILP model, which is computationally intensive, especially for large-scale networks. While a Markov approximation is used, it may still be challenging to apply in real-time large-scale network scenarios. Although energy efficiency is the paper's primary focus, the algorithm does not account for potential trade-offs between routing energy costs and other operational costs and reliability.

Unlike [57] [58] [67], paper [59] focuses on the radio sub-system, which is the primary energy consumer in IoT devices. The authors propose a mixed-integer nonlinear programming (MINP) problem to jointly optimize power allocation, sub-channel allocation, user selection, and the activation of remote radio units (RRUs) based on channel conditions. The proposed work encounters practical challenges related to computational complexity and it prioritizes energy efficiency, potentially at the expense of other key metrics like data rate, latency, and network reliability. The authors are exploring as an extension of this work, the case where each IoT device can allocate energy resources with multi-objective heuristics strategies.

Authors in paper [61] proposes an optimization framework for UAV deployment and resource allocation, targeting energy and cost-efficient management of access and backhaul networks. It includes ground user (GU) association, route selection, and power allocation, utilizing an enhanced k-means clustering for Aerial Base Stations (ABS) deployment and Lagrange methods for power allocation. A minimum circle algorithm optimizes Aerial Relay (AR) placement for backhaul, showing a 19% performance gain over existing schemes with large packet sizes and limited bandwidth. However, the study assumes ideal conditions like reliable backhaul connectivity and line-of-sight (LoS), which may not be applicable in urban areas. Additionally, it focuses on cost and energy without accounting for latency or user quality of experience (QoE), posing potential scalability challenges in larger networks.

In [69] [63], authors address the challenge of jointly optimizing DU and CU placements and routing in 5G Packet X-haul Networks (PXNs), essential for meeting latency and throughput needs for services like eMBB and URLLC. The paper models this as a "Latency-Aware DU/CU Placement and Flow Routing" (LADCPR) problem using Mixed-Integer Linear Programming (MILP) to reduce latency across fronthaul, midhaul, and backhaul flows. Due to scalability limits in large networks, two MILP reformulations and heuristic algorithms (DFCMB and DF+CMB) are proposed to balance accuracy and computational feasibility. While the MILP model provides precise solutions, heuristic methods perform well in large networks, though they may sacrifice some optimality. However, the paper does not focus on the optimization of PXNs under varying traffic loads.

As tabulated in Table 3 the papers [58] [18] [61] [63] [64] [65] discuss methods for **optimizing cost and energy consumption** in different networks. Papers [58] [63] [69] [61] [18], aim to achieve both cost and energy efficiency and they use different kind of resource allocation techniques to optimize various aspects of the access network, while paper [64] focus more on minimizing costs while adhering to latency requirements.

In [18], the paper focuses on addressing challenges in optimizing network topology and routing for Integrated Access and Backhaul (IAB) networks, especially in the context of 5G networks. The paper develops genetic algorithm (GA)-based methods for placing IAB nodes and distributing non-IAB backhaul links. It also evaluates how routing affects bypassing blockages in millimeter-wave (mmWave) communications. Whereas, in [64], authors explore optimization methods for DU placement and flow routing in 5G Radio Access Networks (RAN). They tackle the Packet X-haul network design, a network architecture that allows efficient transport of data flows between various elements of 5G RAN, such as RUs, DUs, and CUs. A significant challenge in 5G X-haul networks is handling dynamic latency due to packet buffering at switches, especially in packet-switched networks. The paper uses a worst-case latency model that excludes non-affecting co-routed flows, providing tighter latency estimations. However, this work does not take latency constraints into account. Also, deployment planning in paper [18] may be affected by, e.g., the availability of non-IAB backhaul connection in specific areas.

Paper [60], investigates the problem of VNF placement in 6G Non-Terrestrial Networks (NTNs), focusing on reducing delay and improving resource utilization in resource-constrained environments like satellite networks. It presents a solution that formulates VNFs placement as a weighted graph matching problem, aiming to optimize resource utilization while satisfying the delay requirements of Service Function Chain (SFC) requests. The authors tackle the Packet X-haul network design, a network architecture that allows efficient transport of data flows between various elements of 5G RAN, such as RU, DU and CUs. The proposed heuristic (MaSCA-RA) performs well in large-scale networks where traditional Mixed Integer Programming (MIP) models struggle. The methodology developed offers a practical solution for real-world X-haul networks. Although the proposed methods improve resource utilization, they lack a focus on minimizing the energy cost of computational processes, particularly for the LP-based and Hungarian algorithms, which can be resource-intensive. Developing low-power and energy-efficient algorithms for VNF placement in constrained NTN environments would address this gap.

Paper [62], aims to optimize the usage of IAB technology for providing efficient access and backhaul in aerial networks, a problem largely unexplored in literature. The study examines IAB-assisted UAV networks, proposing solutions to mitigate the backhaul bottleneck in aerial networks. A sum-rate maximization problem is formulated to optimize the frequency spectrum allocation between aerial and terrestrial networks. The proposed solution involves beamforming and spectrum allocation using non-convex optimization techniques. A hybrid beamforming strategy is proposed to spatially separate access and backhaul links. The paper evaluates the effectiveness of IAB technology in aerial networks using simulation results. The results demonstrate that the IAB framework improves both sum-rate and energy efficiency compared to non-IAB setups. However, it is important to integrate terrestrial and non-terrestrial networks in more realistic scenarios based on aerial mobile-IAB taking into account factors such as imperfect channel conditions, UE mobility, and the effects of higher altitudes. The authors aim to extend this work by leveraging machine learning approaches that can help address the challenges posed by this complex system model.

In correspondence to our research interest on cost efficient networks, the authors in paper [65] seek to minimize Total Cost of Ownership (TCO), which includes both Capital Expenditure (Capex) and Operational Expenditure (Opex), for deploying optical fronthaul networks in dense and sparse scenarios in 5G/6G systems.

They propose an Integer Linear Program (ILP) to find the optimal placement of fronthaul components such as RRHs, Power Splitters, and BBU pools. The proposed algorithms were tested in both dense and sparse deployment scenarios (5×5 km² for dense and 20×20 km² for sparse). The paper shows that using a 1:8 PON split ratio is the most cost-efficient option in both dense and sparse areas, saving up to 13% in the sparse scenario and 8% in the dense scenario compared to other split ratios like 1:4 and 1:16. However, this work does not address wavelength and bandwidth assignment techniques, which could allow more than one operator to use the same network.

Paper [71] presents an innovative solution for joint QoS and energy-efficient resource allocation in 5G network slices. By combining Virtual Backbones and Cognitive Cycles, it ensures that the network operates efficiently with minimal energy wastage while satisfying the QoS requirements of diverse traffic types. The proposed framework integrates two core components:

- Virtual Backbone (VB) for Resource Allocation: A virtual backbone is used to form optimized routing paths for resource allocation. The VB is formed using nodes with high fitness values based on multiple objectives, such as energy efficiency, latency, and resource utilization. The VB allows dynamic reconfiguration of routing paths to adapt to changing network conditions.
- Cognitive Cycles (CC) for Energy Efficiency: Cognitive Cycles are employed to periodically monitor network states, evaluate traffic loads, and dynamically adjust the configuration of the network slices. This enables energy savings by shutting down or deactivating underutilized nodes and paths while maintaining QoS guarantees.

Simulation results show that the proposed method outperforms existing approaches by up to 23% in energy savings while maintaining QoS guarantees, demonstrating its effectiveness in 5G environments.

Paper [66] explores the challenges of resource allocation in dense mmWave multiple access networks, particularly focusing on scenarios where both wireless access and backhaul links share the same mmWave spectrum. The paper formulates the problem of joint access and relay-assisted backhaul resource allocation as a Stackelberg game. In this model, the network (represented by the macro base station and access points) serves as the leader, while smart IoT devices act as followers. The game aims to maximize the network's energy efficiency and data rates while handling interference and relay challenges. The authors introduce backward induction, a technique used to solve the Stackelberg game, where the leader makes decisions first, and the followers react based on the leader's actions. Simulation results show that their proposed scheme achieves a better balance between the access data rate and backhaul data rate for each access point. While the study aims to improve energy efficiency, it highlights the complexity and computational overhead associated with its proposed Stackelberg game model, suggesting a need for lower-complexity alternatives. Additionally, balancing the trade-off between energy efficiency and data rate remains unresolved in dense networks.

Paper [68] provides a detailed analysis and solution to the problem of efficiently managing computation offloading and resource allocation in a multi-access edge computing (MEC) environment. Their primary objective is to minimize the energy consumption of user terminal devices (UTDs) while ensuring that the computation tasks meet strict delay requirements, since UTDs are limited in computational power and energy. The authors present the optimization problem as a MINP problem. The objective function seeks to minimize the energy consumption of UTDs, constrained by offloading decisions, channel selection, power allocation, and resource allocation. The system model comprises multiple UTDs and MEC servers. Each UTD has a computational task that can either be processed locally or offloaded to a MEC server. The study shows that increasing the number of available communication channels or MEC servers reduces the energy consumption of UTDs and increases the number of tasks offloaded. The solution is scalable and performs well even as the number of users or servers increases.

The paper [70] addresses uplink transmission challenges in systems integrating LEO satellites with terrestrial cellular networks, with LEO satellites providing backhaul for BSs. The study proposes centralized and decentralized algorithms to optimize user association, sub-channel assignment, bandwidth, and power control, minimizing transmission time despite LEO satellites' mobility. The centralized algorithm uses convex approximation and compressed sensing for accuracy, while the decentralized approach reduces computational load by distributing tasks across local controllers. Both methods outperform baseline algorithms in efficiency, though the decentralized model achieves greater scalability. The authors in this paper highlight the research gap in addressing the challenges of handover and scalability. Particularly, the handover between BSs and LEO sats, which consumes time for connection establishment and reduces system performance, should be minimized. In terms of scalability, the controller link limitations at the Center Network Controller (CNC) could lead to overload when scaling up the system.

Authors in [69] delve into how network slicing isolates multiple services (such as eMBB and URLLC) within a shared network infrastructure while ensuring service isolation and fulfilling service requirements like latency and throughput. The problem involves placing DU and CU and routing the fronthaul and midhaul data flows between RU and the central processing entities. The authors mainly aim to minimize the number of processing nodes (PP) required to fulfill all services, subject to constraints on latency, processing capacity, and bandwidth. Given the computational complexity of solving large MILP problems, the paper proposes the price-and-branch algorithm (PBA), which enables scalable optimization in packet-switched networks where MILP becomes computationally infeasible as network size grows. The algorithm accounts for worst-case buffering delays and ensures QoS requirements for both eMBB and URLLC. The paper demonstrates that the proposed PBA can solve the network slice optimization problem more efficiently than a standard MILP approach.

Papers [60], [62], [65] **optimally allocate computational resources** across devices, which are resource-constrained (in terms of CPU, memory, and bandwidth) [60], optimizing throughput and balancing resource use between access and backhaul functions [62], while [65] deals with resource allocation related to infrastructure placement and traffic routing, aiming to minimize costs in the network and [69] deals with allocating resources in network slicing scenario within 5G X-haul transport networks. The paper [68] focuses on optimizing computation resource allocation and power management in a multi-server environment, with some attention to user association between UTDs and MEC servers.

4.1.2 Online-based SoA

This sub-section presents a comprehensive study on online based algorithms which are used during the operational phase of a network, as they enable dynamic adaptation to real-time changes in the network. Unlike offline algorithms that rely on pre-existing data and are applied in the planning stages, online algorithms operate on live data, making rapid, responsive decisions to maintain network quality and performance as conditions fluctuate.

Table 4. Bird-Eye view of Online/Adaptive algorithm based SoA.

Ref.	Cost Efficiency	Energy Efficiency	Resource Allocation					Flexible Functional Split
			Spectrum	Power	User Association	Traffic Routing	Computational	
[72]		x		x	x	x		

[73]		x	x		x	x	x	
[74]			x	x	x	x		
[75]			x	x		x		
[71]		x		x	x	x	x	
[76]			x	x	x		x	x

As tabulated inTable 4, the papers [72] [73] [71], discuss methods for optimizing energy efficiency using different resource allocation techniques in different networks.

The paper [72] discusses methods for optimizing resource allocation and path selection in multi-hop networks in a decentralized manner. It introduces multi-hop networks which are commonly employed in communication systems, where data is relayed through several intermediate nodes before arriving at its destination. Optimizing the performance of these networks requires efficient resource allocation, such as bandwidth and power, along with effective path selection to determine the best nodes for data transmission. In this paper, without any central authority control process, each node makes decisions based on its local information. Individual nodes make decisions about resource allocation and path selection based on local observations and limited communication with neighboring nodes. This approach reduces the overhead and complexity associated with centralized control, while still aiming for globally optimal performance. However, the proposed design in the paper does not accommodate delay sensitive applications which could be done by including the number of hops to donors in the decision-making process.

The paper [73] tackles resource allocation in next-generation networks integrating terrestrial and satellite components. It focuses on real-time user association, traffic routing, and x-Network Function (xNF) placement in a 6G integrated terrestrial and satellite network (TN-SN) to maximize energy efficiency and user acceptance. The proposed TERA algorithm solves this joint problem in two steps:

1) User Association and Traffic Routing: TERA constructs a weighted graph, selecting the shortest power-efficient path that meets capacity and delay constraints. If no path is available, the request is blocked.
2) xNF Placement: TERA places xNFs along the path based on node ranking (centrality, computational capacity, CPU load). If no node meets the constraints, it revisits the routing step. TERA effectively balances energy efficiency and complexity for real-time applications.

The proposed TERA algorithm, while efficient, relies on heuristic methods that only approximate the optimal solution. As a result, it achieves about 85% of optimal energy efficiency but does not guarantee an exact optimal solution, which may be a limitation in scenarios requiring precise optimization.

The authors in [71] present a joint approach to QoS and energy-efficient resource allocation and scheduling in 5G network slicing. They propose a novel framework using machine learning algorithms to classify traffic based on QoS attributes, allowing for better traffic prioritization in network slices. The study emphasizes the importance of optimizing traffic management and resource allocation for both QoS and energy efficiency. The framework incorporates a virtual backbone and cognitive cycles (CC) for dynamic resource allocation and energy-efficient scheduling, enhancing network reconfiguration and reducing energy consumption. Simulations show that the method improves energy savings by up to 23% compared to existing approaches while maintaining QoS, demonstrating its effectiveness in 5G environments. While the proposed solution improves energy savings by up to 23%, the paper does not fully explore cost trade-offs associated with resource allocation and scheduling across slices. There is a gap in evaluating operational costs tied to energy savings, such as balancing computational load between high and low-power components or implementing cost-efficient AI models for real-time scheduling.

Unlike paper [71], authors in paper [74] aim to explore the feasibility and benefits of applying IAB in dynamic aerial-terrestrial networks, particularly in enhancing coverage probability. To address the frequent handover issue, the paper proposes a mobility-adaptable IAB scheme. The proposed IAB architecture improves network performance compared to traditional terrestrial-only networks, especially when the mobility of Aerial Base Stations is managed effectively. Although the paper provides a solution for UAV backhaul connections to ground BSs, it does not thoroughly investigate backhaul load balancing or capacity optimization across multiple backhaul routes. Also, the paper's IAB scheme does not explicitly consider cost or energy consumption and power constraints of UAVs when optimizing their deployment and operation. The papers [71], [74] and [75] utilize online algorithms to support decision-making during network operation.

The paper [76] jointly optimizes communication (resource blocks and power) and computing (virtual machine allocation) resources, along with user association and BBU-RRH mapping, to minimize network delay. In a two-tier C-RAN architecture with RRHs as radio stations and BBUs for processing, user requests pass through RRHs to BBUs over a fronthaul link. Two queues—one for RRH transmission and one for BBU processing—manage requests to minimize response time. The paper simulates a network environment where multiple RRHs are deployed under a macro-cell base station. The study analyzes bandwidth utilization, user density, and SINR thresholds, comparing the proposed method to centralized (FCFS) and auction-based RB allocation. Results show reduced response time, improved efficiency, and fair resource distribution, Jain's fairness index.

The adaptive algorithm in paper [74] as discussed above, continuously adjusts to real-time changes in UAV positions, user demands, and network conditions, using stochastic modelling and optimization techniques to ensure optimal coverage and backhaul performance.

Paper [75] focuses on the IAB concept, which integrates access and backhaul functionalities to improve network efficiency, especially at mmWave frequencies. While the paper presents a system design framework for out-band IAB networks, it primarily emphasizes resource allocation strategies (centralized vs. distributed) and the challenges associated with propagation loss**,** multi-hop backhaul**,** and interference management, along with an insightful geometric analysis to optimize system design. If the IAB nodes are densely deployed to expand the coverage of IAB network, it is expected that power consumption will significantly increase due to the circuit power consumed by a large number of IAB nodes. Hence it is necessary to evaluate the energy efficiency of IAB network, however, it remains to be further explored by the authors. The Centralized based Resource Allocation introduced in the paper is an offline algorithm, as it relies on full Channel State Information (CSI) and traffic data from all nodes in the network, which are collected and processed at a central entity (IAB donor) and the Distributed based Resource Allocation is an online algorithm, since it allows each IAB node to make independent resource allocation decisions based on local CSI. Here, the decisions are made dynamically as new data (CSI) becomes available at each node.

Paper [71] utilizes a mix of online and offline algorithms: Cognitive Cycles continuously monitor network conditions and adjust resource allocation dynamically, forming the core of the online approach and ML-based classification which are trained offline and then applied online for traffic prioritization and Virtual Backbone Formation, forms the core of the offline approach, making the algorithm an adaptive one.

The joint design of access and X-haul networks presents new opportunities in the development of 5G/6G networks. This approach enables optimal utilization of valuable resources, enhancing the cost, energy, and resource efficiency of 5G/6G networks.

5. CONTINUOUS LOW ENERGY COMPUTING (CLEC) TECHNIQUES ON SMART WEARABLE MULTI-SENSOR IOMT DEVICES TARGETING AT HIGH ENERGY EFFICIENCY AND PROLONGED BATTERY LIFETIME

5.1 Literature review

5.1.1 Introduction to Continuous Low Energy Computing (CLEC) Techniques

Continuous Low Energy Computing (CLEC) refers to a set of strategies aimed at optimizing energy efficiency in Internet of Medical Things (IoMT) devices, particularly wearables where battery life is critical for sustained functionality [15]. CLEC focuses on minimizing energy consumption by managing computational resources, data transmission, and energy-intensive operations, making it ideal for devices designed for continuous data collection and real-time monitoring. Traditional approaches, often relying on frequent data transmissions to centralized servers, quickly deplete battery life, limiting device longevity and performance [15], [77]. CLEC, however, promotes local data processing and selective data transmission, reducing cloud dependency and achieving significant energy savings without compromising monitoring quality [13].

Through the use of local processing capabilities, IoMT wearables can maintain operation even in environments with unreliable connectivity, enhancing device reliability and enabling real-time management of medical emergencies [15]. This approach improves device efficiency, allowing the storage and processing of critical data without excessive battery drain, as opposed to traditional cloud-dependent methods [13], [78]. Moreover, CLEC-based systems enhance user privacy by minimizing the data sent to external servers, a considerable advantage in healthcare environments where data security is important.

By implementing CLEC, IoMT devices can extend battery life, improve operational efficiency, and achieve greater autonomy, especially important in healthcare applications where continuous monitoring is essential for timely intervention in critical situations. Additionally, CLEC-based wearables can function reliably even in remote areas with limited connectivity, making them suitable for medical practice in low-power settings [77], [78].

5.1.2 Approximate Computing (AC) and Near-Threshold Voltage (NTV) Computing

Approximate Computing (AC) is a key CLEC technique that enables IoMT devices to perform less accurate calculations for non-critical tasks, achieving energy savings without significant impact on the functionality of health monitoring wearables. AC is particularly effective in applications such as signal filtering and preliminary data analysis, where only approximate accuracy is required [16]. By prioritizing calculations, IoMT devices can focus on critical data, such as heart rate monitoring, conserving energy from non-essential functions [79].

In contrast, Near-Threshold Voltage (NTV) Computing reduces the operating voltage of electronic components in IoMT devices to levels just above the functional threshold, significantly lowering energy consumption. This method is ideal for applications requiring long-term operation with minimal power usage, enabling IoMT devices to run for extended periods without recharging [80]. NTV Computing is well-suited for wearables that provide continuous monitoring, as it minimizes energy drain without impacting performance [14].

The combination of AC and NTV enhances the flexibility and resilience of IoMT devices, enabling them to achieve maximum performance with minimal energy consumption. By incorporating AC for adaptive data analysis and NTV to minimize power in critical sensors, IoMT devices can effectively address the demands of continuous health monitoring, particularly in remote or low-power environments [16], [14], [79], [81].

5.1.3 CLEC Applications in Edge and Fog Computing

Integrating CLEC with Edge and Fog Computing architectures boosts energy efficiency by enabling extensive local data processing on IoMT devices, thereby reducing the need for constant cloud communication [77]. Through this localized processing, IoMT wearables can store and process essential data without draining their batteries through continuous transmission, a crucial advantage in healthcare applications [82].

Moreover, fog computing techniques enable distributed data processing, providing faster response times in critical situations and reducing latency in real-time data analysis. This improvement in response time is particularly essential for healthcare applications where rapid detection and alerting of vital sign changes is critical [82]. Additionally, fog computing enhances wearable device resilience in environments with limited connectivity, allowing them to function effectively without continuous cloud reliance [77].

CLEC integration into edge and fog architectures also reduces the overall energy requirements, enabling simultaneous data processing across multiple devices. This feature proves particularly beneficial in remote monitoring applications, where IoMT devices must operate for extended periods without access to centralized servers or continuous connectivity [77], [82].

5.1.4 Challenges and Limitations of CLEC

Despite the advantages of CLEC, certain challenges and limitations exist. Approximate Computing can compromise data accuracy, potentially affecting the quality of health data in critical applications [80]. Ensuring reliable monitoring requires careful calibration of AC settings to balance energy savings with desired accuracy levels for dependable data analysis.

NTV Computing is sensitive to environmental factors such as temperature changes, which can impact the stability and performance of IoMT devices. Temperature fluctuations affect device reliability, necessitating the development of control and stabilization strategies to ensure the consistent performance of NTV technology in varying conditions [81].

Addressing these limitations requires ongoing research to improve CLEC stability and accuracy. Advanced energy management systems and stability controls, with mechanisms to detect and adjust for environmental variables, can enhance wearable performance across diverse and demanding environments [80], [81].

CLEC techniques, including Approximate Computing and Near-Threshold Voltage Computing, offer innovative approaches to energy management in IoMT devices, supporting the development of sustainable and efficient solutions for health monitoring. By focusing on practical applications and addressing the limitations of these techniques, CLEC opens up new possibilities for more resilient, long-lasting IoMT devices that meet modern healthcare demands [83].

Future advancements may incorporate artificial intelligence (AI) into wearables, enhancing adaptive response through algorithms that reduce energy consumption. Such developments are expected to extend IoMT device autonomy and improve user experience, enabling comprehensive monitoring with minimal maintenance needs [12].

The integration of these advanced technologies has the potential to make IoMT wearables self-sustaining and resilient, providing long-term monitoring even in remote environments with limited energy and connectivity resources [83]. The literature review has been summarized in Table 5.

Table 5: Overview of Key Aspects in Energy-Efficient IoMT Architectures

Ref.	Energy-Efficient Framework	IoMT Architecture	Edge Computing	Privacy in IoMT	Hardware Efficiency	AC	NTV	Challenges in CLEC	Integration with Fog Computing	Autonomous IoMT Devices
[12]										x
[13]		x	x						x	x
[14]	x		x							
[15]	x	x								
[16]		x								
[77]	x	x	x						x	x
[78]		x	x	x						x
[79]	x				x	x	x	x		
[80]					x					
[81]					x		x	x		
[82]			x						x	x
[83]		x								
[84]			x	x						x

6. MATHEMATICAL TOOLS

In wireless communications, mathematical tools for analysis of performance evaluation of a network or communication system can be broadly categorized into optimization theory, metaheuristics, heuristics, greedy algorithms, with AI-based tools and stochastic geometry also playing crucial roles in the network analysis. Each of these categories involves specific methods and techniques, which are used to address various challenges, such as resource allocation, network configuration and network performance optimization, e.g. in terms of cost/energy efficiency.

6.1 Optimization Theory

Optimization theory focuses on finding the best solution to a problem from a set of feasible solutions, subject to constraints. It is widely used for problems based on resource allocation, power control, and scheduling.

An optimization problem in standard form is written as [85].

$$\begin{array}{ll} \min_{x} & f_0(x) \\ \text{subject to} & Ax = b \\ & f_i(x) \leq b_i \qquad \forall i = 1, \dots, M \end{array}$$

where x is the variable to be optimized over, $f_0(x)$ is the objective function evaluated at x, the first constraint is an equality constraint, and the second constraint are M inequality constraints over the optimization variable x. The set of all points x that meet the equality and inequality constraints is called the feasible set denoted as $\mathcal{X}$, and the points contained in it are called feasible points.

To solve an optimization problem is to find a point that minimizes the objective function, while meeting the equality and inequality constraints (a feasible point). In other words, the objective is to find a point x^* that is contained in $\mathcal{X}$, in which the objective function finds its minimum value.

Depending on the structure of the objective function and the constraints, an optimization problem can be classified, among others, as:

- **Linear Programming (LP)**: Involves optimizing a linear objective function subject to linear equality and inequality constraints. Example: Minimizing power consumption in a network subject to throughput constraints.
- **Nonlinear Programming (NLP)**: It is similar to LP but the objective function or constraints are nonlinear. Example: Maximizing spectral efficiency where power and interference relationships are nonlinear.
- **Mixed Integer Linear Programming (MILP)**: Involves both integer and continuous variables. Commonly used for user association and resource allocation in networks. Example: Optimizing traffic routing in 6G networks where binary variables decide if a user is connected to a base station.
- **Convex Optimization**: Focuses on problems where the objective function and constraints are convex, meaning local minima are also global minima. Example: Widely used for problems like power allocation or energy-efficient scheduling.
- **Dynamic Programming**: A method used for solving problems by breaking them into smaller sub-problems, typically applied in resource management for multi-hop wireless networks.

Some optimization-based tools include:

- Convex relaxations.
- Block coordinate descent (BCD).
- Alternating direction method of multipliers (ADMM) [86].
- Lagrangean decomposition.
- Successive convex approximation (SCA).
- Concave convex procedure (CCCP) [87].
- Lyapunov optimization.

6.2 AI-based Tools

Artificial intelligence (AI) and **machine learning (ML)** techniques are used for tasks like resource allocation, fault detection, and network optimization. Specifically, **reinforcement learning**, a type of AI where agents learn optimal policies through interaction with the environment, is particularly effective for dynamic resource management and adaptive decision-making in network optimization, allowing systems to improve performance based on observed network behavior.There are various reinforcement learning solutions used to solve the resource allocation problem, such as proximal policy optimization (PPO), deterministic policy gradient (DPG), deep Q networks (DQN), soft actor-critic (SAC), among others. The term deep reinforcement learning (DRL) is used when referring to reinforcement learning solutions utilizing deep neural networks.

Reinforcement learning is a subcategory of machine learning which models the system as a Markov decision process (MDP), which is characterized by an agent, interacting with the environment through actions it takes, and an environment that presents itself to the agent in the form of a state or an observation. The dynamics of an MDP are summarized by the following loop

1) The agent observes the state of the environment.
2) The agent takes action based on their observation.
3) The environment reacts to the action of the agent.
4) The environment sends a reward back to the agent.
5) The environment presents a state (that may be a different one) to the agent.

The set of all possible states the environment can present to the agent is called the state space. The set of all possible actions the agent can take, which may depend on the state, is called the action space. The objective of the agent is to build an action-taking policy that maximizes its long-term reward. Depending on the structure of the state and action space, the policy can take the form of a table of state-action pairs, of a probability distribution, or of a black box (e.g. neural networks). More details on reinforcement learning are available in [88].

Data-driven optimization builds upon large datasets gathered from IoMT devices to enhance decision-making in real time. This method relies on machine learning models to predict energy consumption patterns and adjust system parameters accordingly [80], [78].

6.3 Heuristics and greedy algorithms

One of the complexities of solving a non-convex optimization problem, particularly with integer constraints, is the need to search through the whole feasible set to find the solution (not an issue for convex optimization problems). Deep Reinforcement Learning (DRL) solutions attempt to parametrize and characterize the state-space to be able to take the optimal action (in the long-term) at each step. However, these solutions often require extensive training.

As an alternative, **heuristics** are problem-specific methods or rules of thumb used to quickly find satisfactory (but not necessarily optimal) solutions. They are often used in real-time systems where speed is more important than finding the absolute best solution. These tools iteratively improve the current solution by exploring neighboring solutions. They also use predefined rules to make decisions.

Similarly, **greedy algorithms** work by reformulating the optimization problem as a sequence, or series of disjointed decisions, but making the local, or current best decision at each step. This does not guarantee optimality, as local optimality from a given time step or entity is not, in general, conducive to global optimality. Nevertheless, although greedy algorithms are not guaranteed to reach global optimality, when properly designed and implemented, they can be simple, low-complexity algorithms that reach stationary points and do not require extensive computation or training.

Predictive Modeling utilizes historical data to forecast energy needs, allowing IoMT systems to adjust their power usage proactively [80], [82], [84], [89]. Stochastic Modeling, on the other hand, incorporates probability-based methods to address uncertainties in power demand, ensuring stable and efficient IoMT operation under unpredictable conditions [83], [90].

6.4 Meta-heuristics

Metaheuristics are higher-level iterative procedures designed to generate or select heuristics to find near-optimal solutions for complex optimization problems. They try to balance the exploration of the set of solutions, while exploiting the current best estimated solution as reinforcement learning methods, but without strong guarantees in terms of optimality or convergence.

They are used when exact optimization techniques are computationally infeasible due to the complexity or size of the problem. Some common examples of metaheuristics include:

- Genetic algorithms (GA), inspired by the process of natural selection.
- Particle swarm optimization (PSO), inspired by the social behavior of animals.
- Simulated annealing (SA), inspired by the concept of annealing in metallurgy.
- Atomic Orbital Search (AOS), inspired by the configurations of electrons around the nucleus.

6.5 Stochastic Geometry

Stochastic Geometry (SG) provides mathematical tools to model and analyze the random nature of wireless networks, such as the distribution of nodes (base stations, users) and signal propagation. This analytical tool is used in designing sustainable and high-performance wireless networks, especially as network environments grow increasingly dense and complex.

- SG models, especially **Poisson Point Processes (PPP)**, are employed to statistically represent the random spatial distribution of network elements. This allows efficient resource allocation by predicting Signal to Interference Noise Ratio (SINR) distributions, which helps manage power and bandwidth across various network configurations [91].
- SG offers insights into optimal configurations in complex, multi-layered networks, such as multi-tier and heterogeneous networks. By analyzing spatial dependencies between network components, SG helps manage resource allocation dynamically, aiding in the configuration of small cells, macro cells, and relays to improve network coverage and load balancing [91].
- SG frameworks enable capacity optimization through probabilistic modeling, which helps manage interference and optimize power use, enhancing energy efficiency across the network. In this way, SG provides a structure to evaluate coverage probabilities, throughput, and latency, balancing network performance with cost and energy constraints.

6.6 Game Theory

Some optimization problems imply the interplay between different entities in the system. For example, various nodes may compete for access to a given subchannel, where a bad choice may induce a large amount of interference; the presence of tasks from other nodes in a link for offloading reduces the quality of the link for other tasks; if many tasks are chosen to be computed in a given node, the share of the computation resources of that node will be low at each task. Such problems can be modeled by game theory.

In **game theory**, a game in normal form is characterized by a

- A set of players.
- A set of available actions per player.
- A payoff function.

The purpose for each player is to choose actions such that its payoff is maximized, in games that might be cooperative or non-cooperative. Important concepts of game theory are:

- **Nash equilibrium**, which is a state in which no player is incentivized to change its current strategy (effectively a stable point).
- **Pareto optimal** point, in which no player can greedily change its strategy without negatively affecting the rest of the players.

Game theory is a mathematical formulation through which the problem can be modelled. To solve the games, different algorithms can be used, including heuristics, greedy algorithms multi-agent reinforcement learning, among others.

CONCLUSIONS

This deliverable has provided an overview of the literature works related to the 3 technological areas of WP1, namely: i) the integration of terrestrial with non-terrestrial networks for ubiquitous and reliable healthcare access; ii) the joint access and X-haul network optimization targeting high energy- and cost-efficiency; and iii) the continuous low energy computing techniques on smart wearable multi-sensor IoMT devices targeting at high energy efficiency and prolonged battery lifetime, and of potential areas of investigation related to these. Furthermore, the mathematical tools that will be used in the investigations have been detailed.

D1.1 constitutes the basis for which the subsequent deliverable of WP1, D1.2: Novel infrastructure enhancements for 6G TNs/NTNs towards healthcare delivery, will rely on as a collection of state-of-the-art works that the technical innovations of the DCs will be compared against.

REFERENCES

[1] B. M. Ahmed, M. E. Ali, M. M. Masud and M. Naznin, "Recent trends and techniques of blood glucose level prediction for diabetes control," *Smart Health,* vol. 32, p. 100457, 2024.

[2] S. A. Lubitz, A. Z. Faranesh, C. Selvaggi, S. J. Atlas, D. D. McManus, D. E. Singer, S. Pagoto, M. V. McConnell, A. Pantelopoulos and A. S. Foulkes, "Detection of Atrial Fibrillation in a Large Population Using Wearable Devices: The Fitbit Heart Study," *Circulation,* vol. 146, pp. 1415-1424, 2022.

[3] W. Qi and H. Su, "A Cybertwin Based Multimodal Network for ECG Patterns Monitoring Using Deep Learning," *IEEE Transactions on Industrial Informatics,* vol. 18, pp. 6663-6670, 2022.

[4] L. Rachakonda, S. P. Mohanty, E. Kougianos and P. Sundaravadivel, "Stress-Lysis: A DNN-Integrated Edge Device for Stress Level Detection in the IoMT," *IEEE Transactions on Consumer Electronics,* vol. 65, pp. 474-483, 2019.

[5] GSMA Intelligence, "The Mobile Economy 2024," 2024. [Online]. Available: https://www.gsma.com/mobileeconomy/. [Accessed 16 10 2024].

[6] World Economic Forum, "What is a satellite-connected smartphone – and how can it help bridge the digital divide?," October 2022. [Online]. Available: https://www.weforum.org/agenda/2022/10/satellite-connected-smartphones-digital-divide/. [Accessed 16 10 2024].

[7] A. U. Chaudhry and H. Yanikomeroglu, "When to Crossover From Earth to Space for Lower Latency Data Communications?," *IEEE Transactions on Aerospace and Electronic Systems,* vol. 58, pp. 3962-3978, 2022.

[8] P.-H. K. a. A. Mourad, "Millimeter wave for 5G mobile fronthaul and backhaul," in *2017 European Conference on Networks and Communications (EuCNC),* Oulu, 2017.

[9] R. Henrik, E. Jonas, E. Anders and Ö. Christer, "Integrated access and backhaul – a new type of wireless backhaul in 5G," 2020.

[10] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland and F. Tufvesson, "6G Wireless Systems: Vision, Requirements, Challenges, Insights, and Opportunities," *Proceedings of the IEEE,* vol. 109, pp. 1166-1199, 2021.

[11] Y. Li, J. Huang, Q. Sun, T. Sun and S. Wang, "Cognitive Service Architecture for 6G Core Network," *IEEE Transactions on Industrial Informatics,* vol. 17, pp. 7193-7203, 2021.

[12] P. Mercati and G. Bhat, "Self-Sustainable Wearable and Internet of Things (IoT) Devices for Health Monitoring: Opportunities and Challenges," *IEEE Design & Test,* 2024.

[13] D. K. J. Rajendiran, B. V. Ramasamy, J. T. T, K. S. Palanisamy and V. Chandramohan, "An Effective and Efficient Computation Architecture for Edge Computing Devices on IoMT-Based Deep Belief Networks," in *Edge Computing - Architecture and Applications for Smart Cities*, IntechOpen, 2024.

[14] B. U. Demirel, I. A. Bayoumy and M. A. Al Faruque, "Energy-Efficient Real-Time Heart Monitoring on Edge-Fog-Cloud Internet of Medical Things," *IEEE Internet of Things Journal,* vol. 9, no. 14, 2021.

[15] T. Saba, K. Haseeb, I. Ahmed and A. Rehman, "Secure and energy-efficient framework using Internet of Medical Things for e-healthcare," *Journal of Infection and Public Health,* vol. 13, no. 10, pp. 1567-1575, 2020.

[16] Z. Ashfaq, A. Rafay, R. Mumtaz, S. M. H. Zaidi, H. Saleem, S. A. R. Zaidi, S. Mumtaz and A. Haque, "A review of enabling technologies for Internet of Medical Things (IoMT) Ecosystem," *Ain Shams Engineering Journal,* vol. 13, no. 4, 2022.

[17] A. Zappone, M. Di Renzo and M. Debbah, "Wireless Networks Design in the Era of Deep Learning: Model-Based, AI-Based, or Both?," *IEEE Transactions on Communications,* vol. 67, pp. 7331-7376, 2019.

[18] C. Madapatha, B. Makki, A. Muhammad, E. Dahlman, M.-S. Alouini and T. Svensson, "On Topology Optimization and Routing in Integrated Access and Backhaul Networks: A Genetic Algorithm-Based Approach," *IEEE Open Journal of the Communications Society,* 2021.

[19] NASA, "Sputnik 1," [Online]. Available: https://nssdc.gsfc.nasa.gov/nmc/spacecraft/display.action?id=1957-001B. [Accessed 17 10 2024].

[20] NASA, "Telstar 1," [Online]. Available: https://nssdc.gsfc.nasa.gov/nmc/spacecraft/display.action?id=1962-029A. [Accessed 17 10 2024].

[21] W. Graham, *Iridium NEXT-5 Mission Successfully Launched by SpaceX's Falcon 9,* 2018.

[22] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on New Radio (NR) to support non-terrestrial networks (Release 15)," 2020.

[23] ITU-R, "Attenuation by atmospheric gases and related effects," 2022.

[24] ITU-R, "Model parameters for the physical-statistical wideband model inRecommendation ITU-R P.681," 2017.

[25] ITU-R, "Attenuation due to clouds and fog," 2023.

[26] ITU-R, "Propagation data and prediction methods required for the design of Earth-space telecommunication systems," 2023.

[27] ITU-R, "Characteristics of precipitation for propagation modelling," 2017.

[28] ITU-R, "Specific attenuation model for rain for use in prediction methods," 2005.

[29] ITU-R, "Rain height model for prediction methods," 2013.

[30] J. Liu, Y. Shi, Z. M. Fadlullah and N. Kato, "Space-Air-Ground Integrated Network: A Survey," *IEEE Communications Surveys & Tutorials,* vol. 20, pp. 2714-2741, 2018.

[31] M. M. Azari, S. Solanki, S. Chatzinotas, O. Kodheli, H. Sallouha, A. Colpaert, J. F. Mendoza Montoya, S. Pollin, A. Haqiqatnejad, A. Mostaani, E. Lagunas and B. Ottersten, "Evolution of Non-Terrestrial Networks From 5G to 6G: A Survey," *IEEE Communications Surveys & Tutorials,* vol. 24, pp. 2633-2672, 2022.

[32] O. B. Yahia, Z. Garroussi, O. Bélanger, B. Sansò, J.-F. Frigon, S. Martel, A. Lesage-Landry and G. K. Kurt, "Evolution of High-Throughput Satellite Systems: A Vision of Programmable Regenerative Payload," *IEEE Communications Surveys & Tutorials,* pp. 1-1, 2024.

[33] J. Medina, M. Espinilla, Á. L. García-Fernández and L. Martínez, "Intelligent multi-dose medication controller for fever: From wearable devices to remote dispensers," *Computers & Electrical Engineering,* vol. 65, pp. 400-412, 2018.

[34] Y. Umayahara, Z. Soh, K. Sekikawa, T. Kawae, A. Otsuka and T. Tsuji, "A Mobile Cough Strength Evaluation Device Using Cough Sounds," *Sensors,* vol. 18, 2018.

[35] D. Ichwana, R. Z. Ikhlas and S. Ekariani, "Heart Rate Monitoring System During Physical Exercise for Fatigue Warning Using Non-invasive Wearable Sensor," in *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*, 2018.

[36] A. Gaidhani, K. S. Moon, Y. Ozturk, S. Q. Lee and W. Youm, "Extraction and Analysis of Respiratory Motion Using Wearable Inertial Sensor System during Trunk Motion," *Sensors,* vol. 17, 2017.

[37] G. Cui, P. Duan, L. Xu and W. Wang, "Latency Optimization for Hybrid GEO–LEO Satellite-Assisted IoT Networks," *IEEE Internet of Things Journal,* vol. 10, pp. 6286-6297, 2023.

[38] J. Zhou, Q. Yang, L. Zhao, H. Dai and F. Xiao, "Mobility-Aware Computation Offloading in Satellite Edge Computing Networks," *IEEE Transactions on Mobile Computing,* vol. 23, pp. 9135-9149, 2024.

[39] S. S. Hassan, Y. M. Park, Y. K. Tun, W. Saad, Z. Han and C. S. Hong, "Satellite-Based ITS Data Offloading & Computation in 6G Networks: A Cooperative Multi-Agent Proximal Policy Optimization DRL With Attention Approach," *IEEE Transactions on Mobile Computing,* vol. 23, pp. 4956-4974, 2024.

[40] Y. Lyu, Z. Liu, R. Fan, C. Zhan, H. Hu and J. An, "Optimal Computation Offloading in Collaborative LEO-IoT Enabled MEC: A Multiagent Deep Reinforcement Learning Approach," *IEEE Transactions on Green Communications and Networking,* vol. 7, pp. 996-1011, 2023.

[41] X. Zhang, J. Liu, R. Zhang, Y. Huang, J. Tong, N. Xin, L. Liu and Z. Xiong, "Energy-Efficient Computation Peer Offloading in Satellite Edge Computing Networks," *IEEE Transactions on Mobile Computing,* vol. 23, pp. 3077-3091, 2024.

[42] Q. Tang, Z. Fei, B. Li, H. Yu, Q. Cui, J. Zhang and Z. Han, "Stochastic Computation Offloading for LEO Satellite Edge Computing Networks: A Learning-Based Approach," *IEEE Internet of Things Journal,* vol. 11, pp. 5638-5652, 2024.

[43] Q. Chen, W. Meng, T. Q. S. Quek and S. Chen, "Multi-Tier Hybrid Offloading for Computation-Aware IoT Applications in Civil Aircraft-Augmented SAGIN," *IEEE Journal on Selected Areas in Communications,* vol. 41, pp. 399-417, 2023.

[44] Y. K. Tun, K. T. Kim, L. Zou, Z. Han, G. Dán and C. S. Hong, "Collaborative Computing Services at Ground, Air, and Space: An Optimization Approach," *IEEE Transactions on Vehicular Technology,* vol. 73, pp. 1491-1496, 2024.

[45] A. Mohammadisarab, A. Nouruzi, A. Khalili, N. Mokari, B. A. Arand and E. A. Jorswieck, "Resilient Disaster Relief in Industrial IoT: UAV Trajectory Design and Resource Allocation in 6G Non-Terrestrial Networks," *IEEE Open Journal of the Communications Society,* vol. 5, pp. 1827-1845, 2024.

[46] Y. Liu, L. Jiang, Q. Qi, K. Xie and S. Xie, "Online Computation Offloading for Collaborative Space/Aerial-Aided Edge Computing Toward 6G System," *IEEE Transactions on Vehicular Technology,* vol. 73, pp. 2495-2505, 2024.

[47] D. Han, Q. Ye, H. Peng, W. Wu, H. Wu, W. Liao and X. Shen, "Two-Timescale Learning-Based Task Offloading for Remote IoT in Integrated Satellite–Terrestrial Networks," *IEEE Internet of Things Journal,* vol. 10, pp. 10131-10145, 2023.

[48] C. Ding, J.-B. Wang, Y. Chen, H. Yu, M. Cheng, M. Lin and J. Wang, "Satellite-Terrestrial Assisted Multi-Tier Computing Networks With MIMO Precoding and Computation Optimization," *IEEE Transactions on Wireless Communications,* vol. 23, pp. 3763-3779, 2024.

[49] Y. Zhang, H. Zhang, K. Sun, J. Huo, N. Wang and V. C. M. Leung, "Partial Computation Offloading in Satellite-Based Three-Tier Cloud-Edge Integration Networks," *IEEE Transactions on Wireless Communications,* vol. 23, pp. 836-847, 2024.

[50] Z. Tang, K. Yu, G. Yang, L. X. Cai and H. Zhou, "New Bridge to Cloud: An Ultra-Dense LEO Assisted Green Computation Offloading Approach," *IEEE Transactions on Green Communications and Networking,* vol. 7, pp. 552-564, 2023.

[51] Q. Tang, R. Xie, Z. Fang, T. Huang, T. Chen, R. Zhang and F. R. Yu, "Joint Service Deployment and Task Scheduling for Satellite Edge Computing: A Two-Timescale Hierarchical Approach," *IEEE Journal on Selected Areas in Communications,* vol. 42, pp. 1063-1079, 2024.

[52] M. Jaber, M. A. Imran, R. Tafazolli and A. Tukmanov, "5G Backhaul Challenges and Emerging Research Directions: A Survey," *IEEE Access,* 2016.

[53] T. A. C. a. P. F. Sharma, "Review of optical and wireless backhaul networks and emerging trends of next generation 5G and 6G technologies.," *2. Sharma, Teena, Abdellah Chehri, and Paul Fortier. "ReviTransactions on Emerging Telecommunications Technologies* , 2021.

[54] A. Fayad, T. Cinkler and J. Rak, "Toward 6G Optical Fronthaul: A Survey on Enabling Technologies and Research Perspectives," *Toward 6G OptIEEE Communications Surveys & Tutorials,* 2024.

[55] R. I. Rony, A. Jain, E. Lopez-Aguilera, E. Garcia-Villegas and I. Demirkol, "Joint access-backhaul perspective on mobility management in 5G networks," in *2017 IEEE Conference on Standards for Communications and Networking (CSCN),* Helsinki, Finland, 2017.

[56] B. e. a. Masood Khorsandi, "Dedicated and shared solutions for reliable function splitting in 5G X-haul," *Photonic Network Communications,* 2020.

[57] A. Mukherjee, P. Goswami, M. A. Khan, L. Manman, L. Yang and P. Pillai, "Energy-Efficient Resource Allocation Strategy in Massive IoT for Industrial 6G Applications," *IEEE Internet of Things Journal,* 2021.

[58] A. M. G. K. a. N. P. M. Gatzianas, "Energy-efficient Joint Computational and Network Resource Planning in Beyond 5G Networks," in *2021 IEEE Global Communications Conference (GLOBECOM)*, Madrid, Spain, 2021 .

[59] J. Ansere, M. Kamal, I. Khan and M. Aman, "Dynamic Resource Optimization for Energy-Efficient 6G-IoT Ecosystems.," *Sensors ,* 2023.

[60] Y. Yue, X. Tang, W. Yang, X. Zhang, Z. Zhang and C. Gao, "Delay-aware and Resource-efficient VNF placement in 6G Non-Terrestrial Networks," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, Glasgow, United Kingdom, 2023.

[61] R. Chai, L. He, R. Sun and Q. Chen, "System Cost-Oriented UAV Deployment for Integrated Access and Backhaul Networks," *IEEE Transactions on Vehicular Technology,* 2024.

[62] R. G. Alavicheh, S. M. Razavizadeh and H. Yanikomeroglu, "Integrated Access and Backhaul (IAB) in Low Altitude Platforms," *IEEE Open Journal of the Communications Society,* 2024.

[63] M. Klinkowski, "Optimized Planning of DU/CU Placement and Flow Routing in 5G Packet Xhaul Networks," *IEEE Transactions on Network and Service Management,* 2023.

[64] D. Mroziński, M. Klinkowski and K. Walkowiak, "Cost-Aware DU Placement and Flow Routing in 5G Packet xHaul Networks," *IEEE Access,* 2023.

[65] A. Fayad, T. Cinkler, J. Rak and M. Jha, "Design of Cost-Efficient Optical Fronthaul for 5G/6G Networks: An Optimization Perspective," *Sensors,* 2022.

[66] J. Gui, L. Yin, X. Deng and L. Cai, "Joint Access and Relay-Assisted Backhaul Resource Allocation for Dense mmWave Multiple Access Networks," *IEEE Transactions on Vehicular Technology,* 2024.

[67] Z. Xu, X. Zhang, S. Yu and J. Zhang, "Energy-Efficient Virtual Network Function Placement in Telecom Networks," in *2018 IEEE International Conference on Communications (ICC)*, Kansas City, MO, USA, 2018.

[68] H. Li, H. Xu, C. Zhou, X. Lü and Z. Han, "Joint Optimization Strategy of Computation Offloading and Resource Allocation in Multi-Access Edge Computing Environment," *IEEE Transactions on Vehicular Technology,* 2020.

[69] M. Klinkowski, "A Price-and-Branch Algorithm for Network Slice Optimization in Packet-Switched Xhaul Access Networks," *Applied Sciences,* 2024.

[70] H. Nguyen-Kha, V. N. Ha, E. Lagunas, S. Chatzinotas and J. Grotz, "Joint Two-tier User Association and Resource Management for Integrated Satellite-Terrestrial Networks," *IEEE Transactions on Wireless Communications,* 2024.

[71] S. S. M. a. C. H. Saibharath, "Joint QoS and energy-efficient resource allocation and scheduling in 5G Network Slicing," *Computer Communications ,* 2023.

[72] H. Alghafari and M. S. Haghighi, "Decentralized joint resource allocation and path selection in multi-hop integrated access backhaul 5G networks," *Computer Networks,* 2022.

[73] A. Mesodiakaki, M. Gatzianas, C. Bratsoudis, G. Kalfas, C. Vagionas and R. Maximidis, "Online Energy-Efficient Resource Allocation in Integrated Terrestrial and Satellite 6G Networks," in *ICC 2024 - IEEE International Conference on Communications*, Denver, CO, USA, 2024.

[74] M. Sheng, Y. Zhang, J. Liu, Z. Xie, T. Q. S. Quek and J. Li, "Enabling Integrated Access and Backhaul in Dynamic Aerial-Terrestrial Networks for Coverage Enhancement," *IEEE Transactions on Wireless Communications,* 2024.

[75] J.-H. Kwon, B. Lim and Y.-C. Ko, "Resource Allocation and System Design of Out-Band Based Integrated Access and Backhaul Network at mmWave Band," *IEEE Transactions on Vehicular Technology,* 2022.

[76] L. Ferdouse, A. Anpalagan and S. Erkucuk, "Joint Communication and Computing Resource Allocation in 5G Cloud Radio Access Networks," *IEEE Transactions on Vehicular Technology,* 2019.

[77] S. Tahir, S. T. Bakhsh, M. Abulkhair and M. O. Alassafi, "An energy-efficient fog-to-cloud Internet of Medical Things architecture," *International Journal of Distributed Sensor Networks,* 2019.

[78] R. Gupta and J. Gupta, "Privacy and convergence analysis for the internet of medical things using massive MIMO," *e-Prime - Advances in Electrical Engineering, Electronics and Energy,* vol. 8, 2024.

[79] K. Tamilselvan and P. Thangaraj, "A novel intelligent energy-efficient and dynamic frequency scaling for multi-core embedded architectures in an IoT environment," *Microprocessors and Microsystems,* vol. 72, 2020.

[80] S. Akter, K. Khalil and M. Bayoumi, "Hardware Security in the Internet of Things: A Survey," in *IEEE 36th International Conference*, 2023.

[81] R. Dreslinski, Near-threshold computing: From single core to many-core energy-efficient architectures, ProQuest, 2011.

[82] I. Ben Dhaou, M. Ebrahimi, M. Ben Ammar, G. Bouattour and O. Kanoun, "Edge Devices for Internet of Medical Things: Technologies, Techniques, and Implementation," *Electronics,* 2021.

[83] T. Abbas, A. H. Khan, K. Kanwal, A. Daud, M. Irfan, A. Bukhari and R. Alharbey, "IoMT-Based Healthcare Systems: A Review," *Computer Systems Science and Engineering,* 2024.

[84] Y. Han, "A Privacy Preserving Federated Learning System for IoT Devices Using Blockchain and Optimization," *Journal of Computer and Communications,* vol. 12, no. 9, 2024.

[85] S. Boyd and L. Vandenberghe, Convex optimization, Cambridge university press, 2004.

[86] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, 2011.

[87] A. L. Yuille and A. Rangarajan, "The Concave-Convex Procedure (CCCP)," in *Advances in Neural Information Processing Systems*, 2001.

[88] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, Cambridge, MA: A Bradford Book, 2018.

[89] S. B. Baker, W. Xiang and I. Atkinson, " Internet of Things for Smart Healthcare: Technologies, Challenges, and Opportunities," *IEEE Access,* vol. 5, 2017.

[90] M. Karatas, L. Eriskin, M. Deveci, D. Pamucar and H. Garg, "Big Data for Healthcare Industry 4.0: Applications, Challenges and Future Perspectives," *Expert Systems with Applications,* vol. 200, 2022.

[91] Y. Hmamouche, M. Benjillali, S. Saoudi, H. Yanikomeroglu and M. D. Renzo, "New Trends in Stochastic Geometry for Wireless Networks: A Tutorial and Survey," *Proceedings of the IEEE,* vol. 109, pp. 1200-1252, 2021.