

Ultra-Dense Mobile Networks: Optimal Design and Communications Strategies

Emanuele Gruppi

A thesis submitted for the degree of Doctor of Philosophy
to
University College London

Communication and Information Systems Research Group
Department of Electronic and Electrical Engineering
University College London
December 2020

Statement of Originality

I, Emanuele Gruppi, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This thesis conducts an extensive analysis within the mobile telecommunications sub-field of the ultra-dense mobile networks, in which a massive deployment of network's pieces of equipment is assumed. Future cache-enabled mobile networks are expected to meet most of the generated content demands directly at the edge, where each node has the availability to proactively store a set of contents in a local memory. This thesis makes several important contributions. The research being presented in this thesis proposes new analytical expressions to modeling the performance associated to the network's edge. Base-stations' idling technologies are also investigated to temporarily turn off some network nodes, saving energy and, in some circumstances, improving the overall performance by contributing less interference at the network's edge. On the other hand, making use of fewer base-stations however reduces the amount of available resources at the network's edge. A trade-off is investigated, which balances among interference saturation and available resources to increase the average user's quality of experience. In this work, we treat the edge node density as a variable of the problem. This greatly increases the difficulty of obtaining analytical expressions, but also offers a direct access for optimizing the users' average performance and network's energy consumptions. An energy-focused performance metric is subsequently proposed, with the intention to highlight an interesting duality within the same network's tier, which can transition from a better efficient to a more performing state, according to the energy expenses from the operators. Nonetheless, under an ultra-dense scenario, line-of-sight wireless links between the user and the nodes become more likely. The introduction of a main component of the multi-path propagated copies of a signal involves analytical complications. A feasible approximation is proposed and validated through a set of computer simulations. The scalability of the proposed technique allows to generalise existing results in the literature.

Impact Statement

This thesis exhibits novel contributions to the design of communication strategies for future mobile networks. Driven by the expected substantial increase in the number of low-range access points, new obstacles arise when it comes to accommodate the foreseen increased demand in network performance. This research work aims to highlight the importance of considering a limit to the network nodes densification, which can be detrimental to the overall performance. The analytical findings in this thesis allow the reader to quantify the negative effect to performance and energy consumption brought by over-densification, and provide arbitrary close-to-optimal strategies to increase the overall quality of user experience.

The contributions of this work include a set of more encompassing network metrics, with the derived conclusions and observations supported by numerical validations. The industrial collaboration with Toshiba Europe Ltd has considerably driven the attention of this work to the precise demands for higher performance and efficiency of future mobile network. This thesis has produced results that bridged the research work with the more complicated set up in real-world problems.

The improved analytical derivations being presented in this thesis are expected to attract the attention of wireless telecommunication researchers who aims to tackle better communication strategies as a result of the optimal management of densely deployed network access points. Most of the presented research work in this thesis has been peer-reviewed and subsequently published to international journals of the sector.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to Prof. Kai-Kit Wong for his wise supervision which greatly contributed toward this important outcome. With his strong background and natural attitude in understanding complex theoretic problems, he has always been capable of anticipating the impractical direction I could have taken and patiently guided me toward the realization of effective outcomes. To the same extent, I would like to explicitly thank Dr. Woon Hau Chin and Dr. Mohammud Zubeir Bocus from Toshiba Europe Ltd, for the provided broad exposure to the industrial demands for future mobile networks and for having given me the opportunity to pursue a doctoral degree by means of an industrial sponsorship.

I would also like to thank my closest friends I met during my Ph.D. studies, who made the whole process smoother than it could have ever been.

Finally, this dissertation is dedicated to my family and girlfriend, for the love they have been showing to me during the whole time of this long process.

Contents

Abstract	16
1 Introduction	17
1.1 Motivation	18
1.1.1 Main Contributions	20
1.2 Thesis Outline	22
1.2.1 A Coherent Direction of the Work	23
1.3 List of Publications	24
2 Introductory Material	25
2.1 Stochastic Geometry Tools	26
2.1.1 Operations over Point Processes	29
2.1.2 A Look at Slivnyak’s Theorem	31
2.2 Analytical Standings for the SCDP	33
2.2.1 Analytical Derivation of the SCDP	35
2.3 General Model Design	37
2.3.1 Small-scale Channel Fading Models	39
2.3.2 Content Caching Schemes	42
2.3.3 Cooperation Gain Vs. Diversity Gain	45
3 Edge Node Density and Content Caching Strategy for Arbitrary Spatial User Density	47
3.1 Introduction and Problem Statement	47
3.1.1 Literature Review	48
3.2 Network Model and Design	52

3.2.1	UE Non-Homogeneity	54
3.2.2	Content Delivery	56
3.3	Performance Metric	58
3.3.1	The Scaled Probabilistic Content Caching Model	61
3.3.2	User-Load with K Cooperating Caching Nodes	62
3.3.3	A Zero Truncated Poisson Distribution	66
3.3.4	The Objective Function	67
3.4	The Proposed Approach	71
3.4.1	The Problem and Subproblems	71
3.4.2	Backtracking Line Search based Optimisation	72
3.4.3	Performance Trade-off	75
3.5	Simulation Results	77
3.5.1	SCDP vs User Target Bit Rate	78
3.5.2	Network Energy Consumption	81
3.5.3	Optimality for the Proposed Method	82
3.6	Conclusions	83
4	Fundamental Network Energy Operating Point	85
4.1	Introduction and Problem Statement	85
4.1.1	Literature Review	85
4.2	The Model	87
4.3	The Proposed Performance Metric	89
4.3.1	The Employed Numerical Solution	93
4.3.2	Network Energy Dualism	94
4.3.3	Extension to Cache-Enabled HetNet	95
4.4	Numerical Results and Analysis	95
4.4.1	Sub-optimal Network Tier Densities Vs. ρ	96
4.4.2	Energy Expenses Measures	97
4.4.3	Latency	98
4.4.4	Comparison among EE Approaches	99
4.5	Conclusions	101

5	On LOS Contribution to UDN	102
5.1	Introduction and Problem Statement	102
5.1.1	Literature Review	102
5.2	Network Model and Content Delivery	104
5.2.1	The LOS Thinning Function	107
5.3	The Analytical Findings	108
5.3.1	Essential Tools for Γ_f^S and Γ_f^M Approximations	108
5.3.2	The Approximations	111
5.3.3	Network Trade-off	116
5.3.4	Extension to Double-Slope Path-Loss Model	117
5.4	Numerical Results	118
5.4.1	Validation against λ^S	118
5.4.2	Validation against Rician k -factor	119
5.4.3	Effective Bandwidth and User Density	120
5.4.4	Maximum Number of Cooperating Nodes	121
5.4.5	The Proposed Approximation against σ	122
5.5	Conclusions	123
6	Conclusions	124
6.1	Summary of the Thesis Outcomes	124
6.1.1	Overview of Parameters' Choice for Numerical Simulations	126
6.2	Further Developments	128
	Appendices	132
A		133
A.1	Proof of Theorem.7	133
B		137
B.1	Proof of (4.4)	137
B.2	Proof of Theorem 8	140
B.3	Proof of (4.10)	144
B.4	Theorem 8 applied to $\mathcal{G}^{\text{SCDP}}$	145
B.5	Proof of Theorem 9	147

B.6	Proof of Monotonic Decreasing Trend for $\lambda_{\Psi_1}(\bar{r})$ and $\lambda_{\Psi_2}(\bar{r})$	150
B.7	Proof of Monotonic Decreasing Trend for $\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r})$	152
C		153
C.1	Decoupling Interfering Terms	153
References		154

List of Figures

1.1	Multi-tier HetNet network.	19
2.1	Block diagram digital communication.	34
2.2	Multi-path signal propagation.	38
2.3	Content caching without replacement with $F = 10$ and $M = 3$. In this example, the cache contents are 1 and 3.	43
2.4	Comparison between <i>sampling with replacement</i> and <i>sampling without replacement</i> cache filling for various content caching schemes.	44
3.1	The bin-based discretisation of the space, where the sets $\phi_{n,f}$, $\bar{\phi}_{n,-f}$ and $\tilde{\phi}_i$ will be defined in Section 3.2.2.	55
3.2	The average SCDP from a UCL and the random free-of-bin results for various wordbook contents.	60
3.3	Original and scaled probabilities for the most and least popular contents with $F = 10$ and $M = 3$	62
3.4	A visual representation of the sets ζ , $\bar{\zeta}$, $\tilde{\zeta}$ for a wordbook and cache size of $F = 8$ and $M = 3$ and with user-load $\xi_k = 2$	65
3.5	Visual feedback on the $\Pr(\xi_k = m)$ distribution for the most popular content, with $F = 20$, $M = 5$, and various values of λ_0^U	66
3.6	Match between empirical measure of \mathcal{G} from (3.23) and derived analytical result from Theorem 7.	69
3.7	Resulting Jensen's gap between the target SCDP and the proposed lower-bound.	70
3.8	A user intensity function λ_n^U for the generating PPP.	78
3.9	The SCDP results.	79
3.10	The probability hit-cache results.	79

3.11	The probabilistic content caching policy by the proposed model for various ρ/B . The x-axis shows the index of the UCL while the y-axis corresponds to the content index.	80
3.12	The SBS density $\{\lambda_n^S\}$ by the proposed method.	81
3.13	Average total network power consumption for various work-load values ρ/B	82
3.14	Comparison of SCDP results for different methods in a 3×3 network.	83
4.1	Visual representation of the network model, with the interfering set \mathcal{I} and $\bar{\mathcal{I}}$, closest user-node distance \bar{r} and CSA of radius R	87
4.2	Block diagram of BS transceiver.	88
4.3	Theoretic and empirical measurements of the cdf for the random shortest user-node distance.	89
4.4	Sub-optimal network tiers' densities and associated SCDP values.	97
4.5	Energy expenses for λ^* and $\bar{\lambda}^*$ within the CSA.	98
4.6	Averaged latency time for λ^* and $\bar{\lambda}^*$	99
4.7	Comparison among network tier's EE.	100
5.1	Visual representation of the network model, with interfering power indicated in red and desired power from the cooperating set K_f in black.	105
5.2	Target and approximated SCDP against λ^S	119
5.3	Target and approximated SCDP against Rician k -factor.	120
5.4	Target and approximated SCDP against λ^U	121
5.5	Target and approximated SCDP against maximum number of cooperating edge nodes.	122
6.1	Users' activity urban area of Milan.	129
6.2	The multi-agent reinforcement learning architecture, where the agent network is composed by a combination of multilayer perceptron, gated recurrent networks and a probabilistic ϵ -greedy decision making process.	131

List of Tables

3.1	Comparison of content caching schemes.	50
3.2	The network parameters.	77
4.1	The network and femto-cell energy consumption parameters.	96
5.1	Resume of key variables.	110
5.2	The network parameters.	118

Abbreviations

5G	5 th Generation
UDN	ultra-dense network
FDMA	frequency division multiple access
LOS	line-of-sight
NLOS	non line-of-sight
SINR	signal-to-interference-plus-noise ratio
SCDP	success content delivery probability
JT	joint transmission
CoMP	coordinated multipoint
CSA	content searching area
UCL	user candidate location
BS	base station
SBS	small base station
MBS	macro base station
PPP	Poisson point process
EE	energy efficiency
HetNet	heterogeneous network
CoMP	coordinated multi point
3GPP	3rd Generation Partnership Project
QoS	quality of service

AWGN	additive white Gaussian noise
UE	user equipment
MPC	most popular content
UC	uniform content
MCP	maximum conversion point
SISO	single-input single-output
CSIR	channel state information at the receiver
CSIT	channel state information at the transmitter
pdf	probability density function
cdf	cumulative density function
pgfl	probability-generating functional
pmf	probability mass function

List of Symbols and Operators

λ^S	intensity function of small base stations point process
λ^U	intensity function of users point process
λ^M	intensity function of macro base stations point process
$\phi^S \triangleq \phi(\lambda^S)$	Poisson point process for SBS
$\phi^U \triangleq \phi(\lambda^U)$	Poisson point process for SBS
$\phi^M \triangleq \phi(\lambda^M)$	Poisson point process for SBS
p_f	local content caching probability for content f
\hat{p}_f	content popularity for content f
R	radius circular CSA
M	local cache size
F	wordbook size
v	skewness factor for content popularity
B	available bandwidth for contents' transmission
r	user-node link distance
α	path-loss coefficient
h	Rayleigh faded channel coefficient
h'	Rician faded channel coefficient
ξ	random user-load at single network's node

Ξ	maximum random user-load experienced by a set of network's node
ρ	network's target bit-rate
γ	generic signal-to-noise ratio
W	noise thermal power
$\mathbb{E}[\cdot]$	expected value operator
$\mathcal{L}(\cdot)$	Laplace functional
$\text{Pr}(\cdot)$	probability of an event

Chapter 1

Introduction

Future mobile communications networks will face an unprecedented growth of content requests generation, with the global mobile data traffic increasing nearly eightfold between 2015 and 2020, reaching 30.6 exabytes per month by 2020 [1]. The 5th Generation (5G) of mobile networks are therefore required to meet a variety of key performance indicators such as capacity, latency, energy efficiency (EE) and will consequently adopt several new technologies to deliver enhanced user experience [2–5]. Network densification has been indicated as a key strategy to overcome the data traffic increase, with promises of enhanced network capacity and spectrum reuse [6]. This technique is able to shorten the user-to-node distances, unlocking the benefits of low-powered and low-latency wireless communications [7, 8]. In addition, diffuse network components can directly meet the users' requests, if provided with a local memory storage, alleviating the traffic demand at the backhaul. Introduced from the latest 3rd Generation Partnership Project (3GPP) Long Term Evolution Advanced [9], distributed content caching has demonstrated great quality of service (QoS) improvements when applied to a radio access network [10–13]. Furthermore, the introduction of a heterogeneous network (HetNet) allows users to be multiplexed in different domains, but also unlocks content caching at different layers of the network [14–16].

To reap the benefits from these techniques, success content delivery probability (SCDP) has paved the way to gain new insights in wireless mobile networks. SCDP is a QoS performance metric which allows to capture the health of a wireless link

by addressing the event that the channel transmitting rate at the receiver is higher than a minimum target bit-rate.

In this first chapter, the fundamentals of the thesis work are presented. In Section 1.1, the reasons which motivate the directions of our efforts are discussed with the list of contributions to the state-of-the-art reported. The structure of the work is illustrated in Section 1.2 and the list of our contributions to the scientific literature is given in Section 1.3.

1.1 Motivation

Normally, a content request is forwarded to a base station (BS) which connects to the clouds to fetch the required data by means of a physical wired link, named backhaul. During peaks of traffic demand, the burden at the backhaul wired links represents a bottleneck to the users' QoS [17–19]. Cache-enabled HetNets serve as a promising solution to the backhaul burden obstacle, by differentiating distinct layers based on the size of the coverage area of the cells. A representative example of a multi-tier HetNet structure is depicted in Fig. 1.1. The intuition to potentially meet a good portion of content requests directly at the network's edge not only serves as an alleviating solution to the backhaul burden, but also unlocks short distance content transmissions, encouraging low-powered transmissions and frequency reuse. The strategy of massive network's densification of cache-enabled cells has promised to bridge the network capacity gap with the 1000-fold increase in traffic demand, expected to occur in the 5G mobile networks [20]. Therefore, on top of the HetNet architectures, the deployment of an ultra-dense network (UDN) as layers in a HetNet, has widely attracted the attention of both research and operators due to its promising increase in network capacity [21–23]. However, the massive deployment of network's pieces of equipment introduces some difficulties which can be summarized as: dramatic increase in network's energy consumptions; interferences from serving a massive number of content requests escalate the users' competition for content retrieving at the edge; complex derivations of QoS metrics due to the combination of line-of-sight (LOS) and non line-of-sight (NLOS) wireless links at the user side as higher densities allow direct LOS wireless links to occur more frequently.

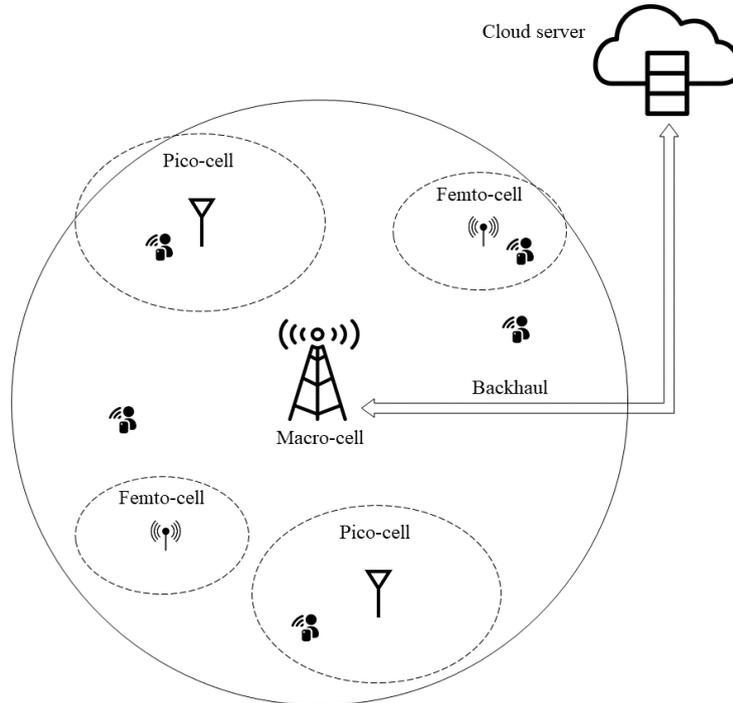


Figure 1.1: Multi-tier HetNet network.

Current literature has widely investigated optimal or sub-optimal network policies with the intention to maximize a utility function that resembles the average user performance. Typically, the problem of augmenting users' QoS is presented by independently optimizing either the probabilistic content caching strategy or the network densities. Moreover, the detrimental effects of a local over-densification of pieces of equipment (*i.e.*, edge node) to neighbouring content transmissions remain an open challenge due to its intrinsic analytical difficulties. However, due to the non-trivial analytical derivations, the random number of caching nodes was often considered as a conditioning factor for the evaluation of a network performance metric. After the optimization is performed, the random quantity is commonly picked deterministically, with its effects empirically studied. Being able to arbitrarily choose the operating density of a network's tier is becoming possible by means of idling technologies, widely predicted to be employed in future 5G mobile networks [24–27]. Idling technologies have been investigated as beneficial as a mat-

ter of network's energy consumptions. However, the potential benefit to the users QoS deriving from idling a subset of the active nodes remains an open question.

Massive network densification unlocks frequency reuse as the size of deployed content source cell decreases, *i.e.*, pico and femto cells. The same frequency bands can thus be utilised at a local scale given the lower power levels associated to smaller cells, whose interfering power is less destructive for neighbouring transmissions. However, given the foreseen massive increase in content requests generation, a full characterisation of the bandwidth usage for multicasting content dissemination is needed for reliable SCDP estimate. It is a common practice in the literature to assume light user-load conditions. However, under UDN conditions, especially when coordinated multi point (CoMP) techniques are engaged, this assumption is no longer valid.

A direct consequence of employing UDNs for future mobile networks is the combination of LOS and NLOS wireless links in the environment. This follows as consequence of having more nodes within the same portion of network's space such that the probability of a direct wireless link is increased. The implied difficulty of introducing possible LOS links lies in the different small-scale channel distribution which does not have closed-form solutions and thus hinder any insights being reviewed.

1.1.1 Main Contributions

Contributions of this work shed more light on the limitations and reveal the true potentials of UDN. The main contributions are listed and summarized as follows:

- **Joint SCDP maximization under constraints of probabilistic content caching policy and density of active network's nodes** (Chapter 3): This is achieved by proposing a Jensen's lower bound of the target SCDP measure, which allows to average out the random number of edge nodes and considers the whole maximization to be performed on two distinct sets of decision variables.
- **Proposal of a localised SCDP maximization approach** (Chapter 3): This is obtained through a spatially binned interpretation of the network

area, such that the center of each bin determines the geographical coordinates of a representative user in the area covered by the bin. A global SCDP is then maximized as a collection of the experienced metric by the set of representative users. The proposed approach allows to deploy an arbitrary user density. It is demonstrated, by means of a numerical analysis, that the representative location works as a useful description for the averaged SCDP perceived by a randomly located user in the network.

- **Full characterisation of bandwidth usage for content multicasting** (Chapter 3): By means of a rigorous representation of the user-load forwarded to a small base station (SBS), the whole available bandwidth at the edge node is divided into homogeneous portions of frequency bands to be allocated to distinct users for simultaneous contents' transmission.
- **Improved analytical derivation for network's EE** (Chapter 4): Considering the expected value of the random ratio of the SCDP over energy consumption within the typical user's content searching area (CSA), an improved definition of network efficiency is derived. Numerical evaluations validate the proposed approach against the state-of-the-art as a matter of efficiency metrics.
- **Improved access to network's insights for LOS wireless links** (Chapter 5): When the power contribution from multiple nodes is accounted, a combination of LOS and NLOS wireless links is observed at the receiver, thus complicating analytical performance metric derivations. A closed-form approximation is proposed and, by means of numerical analysis, it is shown to be an appropriate estimation of the target SCDP.
- **Extension method for full-NLOS models to account for LOS contribution** (Chapter 5): By means of a feasible correctness scaling factor, Rayleigh channel fading models can be extended by applying the proposed approach. The suggested correcting factor for Rayleigh mobile networks is shown to be a tight approximation of the targeted SCDP and to greatly simplify the analytical derivation of the performance metric.

1.2 Thesis Outline

The remainder of the thesis work is organized as follows.

Chapter 2 provides an overview of the basic concepts thoroughly employed during the thesis. Elements from stochastic geometry, probability theory and information theory are discussed in Section 2.1 with a representative derivation of the SCDP been derived in Section 2.2. The network structural choices and fundamental modeling techniques are described in Section 2.3, with a more specific network design description being discussed at the begin of each chapter.

Chapter 3 analyses the problem of finding local sets of optimal probabilistic content caching policies and caching node density by maximizing a lower-bound of the target SCDP. In Section 3.2, the specific model design choices are discussed with employed binned network approach introduced. Section 3.3 delineates the proposed lower-bound of the target SCDP, with the focus on the utilized scaled probabilistic content caching model, the rigorous definition for user-load and with the utilized zero-truncated Poisson distribution for the set of cooperating nodes illustrated. The gradient-based solution is examined in Section 3.4. The resulting numerical outcomes are discussed in Section 3.5 with the final conclusions been reported in Section 3.6.

Chapter 4 explores an improved definition of network's EE, highlighting the existence of a fundamental energetic dualism in HetNets. This metric is then maximized against the edge node density, to find the network's optimum energy conversion. In Section 4.2 the specific network design choices are indicated. Section 4.3 covers the analytical findings for the upgraded definition of network EE with the employed numerical search of its global maximum described. In Section 4.4 the retrieved numerical sub-optimal edge node density is analysed, with an extensive investigation conducted over a wide set of network's measures. The chapter is concluded in Section 4.5 with the resume of the final understanding of the problem.

Chapter 5 examines a SCDP approximation when LOS wireless links and cooperating content transmitting strategies are accounted. The model specifications are discussed in Section 5.2, while the analytical fundamentals of the approximated SCDP are elaborated in Section 3.4.3. Section 5.4 shows the numerical results and the validation of the proposed approximation is examined. Section 5.5 concludes

the chapter with the final considerations.

Chapter 6 concludes the work of this thesis, presenting a summary of the achievements and introducing to the possible future developments of the presented work.

1.2.1 A Coherent Direction of the Work

The contributions presented in this thesis constitute a coherent direction of the conducted research work. To face the unprecedented growth in content requests generation, improved models which can capture the effects on the network metrics of higher densities are needed. Cooperative transmission techniques, channel power gain models, user-load quantification and EE are some of the measures that are greatly affected by higher density levels. The work presented in the technical Chapters 3, 4 and 5 provide more representative mathematical tools to better quantify the enhanced performance of UDNs.

More specifically, the efforts in Chapter 3 to find global optimal strategies which maximise the SCDP can be further improved when contributions from Chapter 5 and Chapter 4 are integrated. The assumptions over the nature of the channel power gains in Chapter 3 permit to ease the derivation of the performance metric, which then allow to acknowledge substantial improvements in performance as a result of an optimisation. Unfortunately, the complexity introduced by accounting the distribution of the channel power gain for LOS links does not allow to have a direct access to key parameters. More efforts are required to include the contributions of LOS wireless links, which are likely to occur in UDNs due to the involved high densities that shorten the user-node distances. In Chapter 5, this problem is addressed by proposing a more accessible mathematical derivation of the channel power gain with combined LOS and NLOS wireless links. A novel interpretation of the target power levels is proposed and its integration to Chapter 3 is presented in more details in Section 5.3.2. In conclusion, Chapter 5 aims to improve the quality of the model from Chapter 3 while keeping the complexity as low as possible.

The contribution discussed in Chapter 4 addresses the problem of efficiency in UDNs. The intuition behind this work is that a performance focused metric which numerically quantifies the QoS does not suffice to have a more comprehensive

knowledge of the network optimal strategies. The results from Chapter 4 allow to extend our understanding of the optimal behaviours of UDNs under both a performance and efficiency perspective. The improved EE metric can directly be applied to the framework presented in Chapter 3, and the combined knowledge allows to derive local optimal strategies of the network for its most efficient and most performing states.

1.3 List of Publications

The work being conducted in this thesis has produced three major contributions, two of which have been published in two technical IEEE journals and one is being prepared for major submission. The list of publications is given below.

- **Journal Publications:**

- Emanuele Gruppi, Kai-Kit Wong, Mohammud Zubeir Bocus, and Woon Hau Chin. "Ultra Dense Edge Caching Networks with Arbitrary User Spatial Density", *IEEE Transactions on Wireless Communications*. IEEE, 2020. Volume, 19. Issue, 7. Pages, 4363-4377.
- Emanuele Gruppi, Kai-Kit Wong, and Woon Hau Chin. "On LOS Contribution to Ultra-Dense Network", *IEEE Access*. IEEE, 2020. Volume, 8. Pages, 100288-100297.

- **Submission in preparation:**

- Emanuele Gruppi, Kai-Kit Wong, and Woon Hau Chin. "Fundamental Duality in Heterogeneous Networks"

Chapter 2

Introductory Material

Random spatial models are widely employed to represent the randomness in both number and locations of nodes and users for mobile network applications. Stochastic geometry allows to step from a random spatial displacement of elements to closed-forms (or semi-closed-forms) of the network's performance metrics. The analytical derived results are essential to average out the desired performance metrics from the influence of a collection of randomly displaced nodes. Hence, there is a set of network performance metrics (such as signal-to-interference-plus-noise ratio (SINR), channel capacity or SCDP) that can be derived as a direct consequence of the application of stochastic geometry to random spatial models. On the other hand, stochastic geometry provides us with the tool to statistically determine the necessary quantities to analyze or model the network. Both users and edge nodes are suitable for being accounted as random Poisson point process (PPP), defined on a Euclidean space \mathbb{R}^d [28–30]. Stochastic geometry has therefore become a useful mathematical tool over decades to approximate random spatial patterns in wireless applications. With a combined use of notions from point processes theory, probability theory and information theory, researchers can provide answers to questions as

- How to determine the average desired power and interference from a random set of transmitters randomly placed across the network?
- How do we address link-distance distribution over all the possible realization of a random collection of points?

In the following outline the focus on two-dimensional Euclidean spaces \mathbb{R}^2 is considered. However, the reported notions can be extended to either one or three dimensional Euclidean spaces.

This chapter provides an overview of the analytical tools from stochastic geometry. Section 2.1 covers the basics of this applied probability theory. In Section 2.2, the introduced notions have been utilised to analytically derive a representative closed-form for the SCDP. In Section 2.3 the mobile network model employed in this work is described, with more detailed versions specifically described at the beginning of each chapter in detail.

2.1 Stochastic Geometry Tools

The concepts of intensity measure Λ and intensity function λ are first introduced, to define the general PPP with its main properties. Making use of the *random measure formalism*, a point process can be described by the counting measure of all its points in a set (or finite space) $B \in \mathbb{R}^d$. The number of points in B writes as $N(B)$ and corresponds to an integer random variable which follows a Poisson distribution and whose mean is defined as the intensity measure $\Lambda = \mathbb{E}[N(B)]$. A PPP ϕ can be fully characterised by an intensity function λ which might or might not depend on the spatial coordinates. In the first case, we refer to it as a non-homogeneous PPP, and in the latter case a homogeneous PPP. Intensity measure $\Lambda(B)$ and intensity function $\lambda(x)$ are strictly linked. One might say the intensity measure is the integral of the intensity function over a compact Euclidean space, or the expected number of points over the same area. On the other hand, the intensity function $\lambda(x)$ is the expected number of points per unit of area dx . The following equivalences formalise the relation between intensity measure and intensity function as

$$\Lambda(B) = \mathbb{E} [|\phi(\lambda | B)|] = \mathbb{E} [N(B)] = \int_B \lambda(x) dx.$$

A general PPP, defined over the Euclidean space \mathbb{R}^d , is such that

1. Over the compact set $B \in \mathbb{R}^d$, $N(B)$ has a Poisson distribution such that

the random number of points writes as $\mathbb{E}[N(B)] = \int_B \lambda(x) dx$ and has a distribution

$$\Pr(N(B) = n) = \frac{\left(\int_B \lambda(x) dx\right)^n}{n!} e^{-\int_B \lambda(x) dx}.$$

2. If $N(B_1), N(B_2), \dots, N(B_K)$ are computed over disjoint sets, then they are also independent random variables.

When the intensity function is not space-dependent, we have $\int_B \lambda(x) dx = \lambda|B|$ and the PPP is said to be spatially homogeneous. If we try to investigate the average number of points over an infinite area, the result will always be infinite regardless of the density function. Therefore, compact sets are generally in use.

The following properties on point processes have been employed in this thesis.

Property 1 (Translation). *If $\phi = \{x_1, \dots, x_K\}$ defines the point process, then a translation of $x \in \mathbb{R}^d$ defines the process $\phi_x = \{x_1 + x, \dots, x_k + x\}$.*

Property 2 (Stationarity). *A point process is said to be stationary on \mathbb{R}^d if its intensity function λ is translation-invariant. As a general example, if the intensity function of ϕ is λ , the same intensity function holds for the translated point process ϕ_x .*

Property 3 (Isotropy). *A point process is isotropic on \mathbb{R}^d if its intensity function is rotationally invariant with respect to a rotation over the origin.*

Property 4 (Motion-Invariance). *A point process is motion-invariant on \mathbb{R}^d if it is stationary and isotropic.*

The following holds from the formal definition of a PPP.

Property 5 (Independence). *The number of points in the Euclidean ball $\mathcal{B}(x, \epsilon)$ is independent on the number of points in any region outside this ball $\forall \epsilon > 0$.*

The independent thinning of a PPP is a common transformation which removes a subset of points of an initial PPP based on a probabilistic rule. When the removal event is independent for all points, we refer to it as an independent thinning. The following theorem is known to hold for stationary PPP.

Theorem 1 (Independent thinning of a PPP). *Let a thinning function $g : \mathbb{R}^d \rightarrow [0, 1]$ be applied to a stationary PPP such that each point is independently deleted with probability $1 - g(x)$. The result of the thinning procedure generates a non-homogeneous PPP whose intensity function writes as $\lambda g(x)$.*

The independent thinning transformation of an initial PPP is particularly useful when it is necessary to discriminate among LOS and NLOS wireless links. The following applied probability concepts are reported.

Void Probability The probability that no elements of a homogeneous PPP fall into a ball of radius R is called void or null Probability. This can be interpreted as the distance over which no elements are found in a ball centered at $(0, 0)$ of radius R . This result stems from the probability of having no points in a PPP within a generic R as

$$\Pr(N(\pi R^2) = 0) = e^{-\lambda \pi R^2}.$$

From the void probability, the notion of nearest neighbour distance directly follows.

Nearest-Neighbour Distance Distribution Function The distance from a point $x \in \phi$ to its closest element of ϕ is called nearest-neighbour distance. The probability density function (pdf) of this random quantity is a direct result of the application of the void probability. The probability of the event ϕ to have no elements given a distance r from a point centered at $(0, 0)$ writes as

$$\Pr(R > r) = e^{-\lambda \pi r^2}.$$

The cumulative density function (cdf) can be directly obtained as

$$\Pr(R \leq r) = F_R(r) = 1 - e^{-\lambda \pi r^2},$$

and therefore the pdf of the closest element of ϕ from the origin $(0, 0)$ is

$$f_R(r) = \frac{\partial F_R(r)}{\partial r} = 2\pi \lambda r e^{-\lambda \pi r^2}.$$

2.1.1 Operations over Point Processes

The importance of being able to perform sums and products over the elements of a point process has important implications for the analysis of a mobile network. As a matter of fact, the interference perceived at a random spatial location is equivalent to a sum of the single interfering contributions from the modeling interfering point process. It is essential to have a full distributional characterisation of the interference. However, this is possible for limited cases and point process functionals have still an important role when it is necessary to analyse the expected value of product of functions calculated at the elements of a point process.

One of the underlying results of point process functionals is the mean of the sum of functions defined over a point process, envisioned by Campbell in 1909 [31]. The following theorem, also known as Campbell-Hardy theorem, states this important result.

Theorem 2 (Campbell's theorem for functions summed over a point process). *Let ϕ be a point process defined on \mathbb{R}^d and f a measurable function such that $f : \mathbb{R}^d \rightarrow \mathbb{R}$. The random sum*

$$S = \sum_{x \in \phi} f(x),$$

stands as a random variable whose mean writes as

$$\mathbb{E}[S] = \int_{\mathbb{R}^d} f(x)\lambda(x) dx.$$

From Theorem 2, the direct consequence of the Campbell's theorem applied over a stationary point process permits the expression

$$\mathbb{E}[S] = \lambda \int_{\mathbb{R}^d} f(x) dx.$$

2.1.1.1 The moment-generating function of a sum over PPP

The derivation of the moment-generating function for a PPP stems directly from the application of Campbell's theorem.

Theorem 3 (Campbell's theorem for moment-generating functions for PPPs). *Let ϕ be a homogeneous PPP defined on \mathbb{R}^d and f a measurable function such that $f : \mathbb{R}^d \rightarrow \mathbb{R}$. When the random sum*

$$S = \sum_{x \in \phi} f(x),$$

is absolutely convergent, i.e. $\int_{\mathbb{R}^d} \min(|f(x)|, 1) dx < \infty$, then the moment-generating function of S is given by

$$\mathbb{E} [e^{tS}] = \exp \left(\lambda \int_{\mathbb{R}^d} (e^{tf(x)} - 1) dx \right).$$

The same result can potentially be extended to non-homogeneous PPP, when the absolute convergence condition holds

$$\int_{\mathbb{R}^d} \min(|f(x)|, 1) \lambda(x) dx < \infty,$$

then the moment-generating function for S writes as

$$\mathbb{E} [e^{tS[f]}] = \exp \left(\int_{\mathbb{R}^d} (e^{tf(x)} - 1) \lambda(x) dx \right).$$

2.1.1.2 The probability-generating and Laplace functionals for PPP

Theorem 4 (probability-generating functional (pgfl) for a PPP). *Let ϕ be a homogeneous PPP defined on \mathbb{R}^d with intensity measure Λ and a function such that $v : \mathbb{R}^d \rightarrow \mathbb{R}$ is a measurable function. Then the pgfl of ϕ is the following result*

$$\mathcal{P}(v) = \mathbb{E} \left(\prod_{x \in \phi} v(x) \right) = \exp \left(- \int_{\mathbb{R}^d} [1 - v(x)] \Lambda(dx) \right),$$

where, for a homogeneous PPP, simplifies as

$$\mathcal{P}(v) = \exp \left(-\lambda \int_{\mathbb{R}^d} [1 - v(x)] dx \right).$$

From the definition of pgfl for a PPP, the definition of Laplace functional di-

rectly follows by noting that the relationship that links the pgfl and the Laplace functionals over the same PPP is

$$\mathcal{L}(v) \triangleq \mathcal{P}[e^{-u}],$$

or, equivalently

$$\mathcal{P}(v) \triangleq \mathcal{L}[-\log(v)].$$

Hence, the following theorem of a Laplacian over a PPP stands.

Theorem 5 (Laplace functional for a PPP). *Let ϕ be a homogeneous PPP defined on \mathbb{R}^d with intensity measure Λ and a function such that $u : \mathbb{R}^d \rightarrow \mathbb{R}$ is a measurable function. Then the Laplace functional for ϕ is the following outcome*

$$\mathcal{L}(v) = \mathbb{E}[e^{-S}] = \exp\left(-\int_{\mathbb{R}^d} [1 - e^{-f(x)}] \Lambda(dx)\right),$$

where, for a homogeneous PPP, simplifies as

$$\mathcal{L}(v) = \exp\left(-\lambda \int_{\mathbb{R}^d} [1 - e^{-f(x)}] dx\right).$$

The Laplace functional can be directly seen to be retrieved from the moment-generating function over a PPP for $t = -1$. That is,

$$\mathbb{E}[e^S] \triangleq \mathbb{E}[e^{-S}].$$

The Laplace functional and pgfl can be directly employed to average out the interfering contributions of a PPP as it will be discussed for a representative example in Section 2.2.1.

2.1.2 A Look at Slivnyak's Theorem

Slivnyak's theorem represents one of the most known and significant results available to researchers to investigate wireless networks modelled as a homogeneous PPP. In general, it allows to condition the distribution of a PPP to have a point at the origin of an Euclidean space without changing the distribution itself. The

user at the origin of the Euclidean space is called typical user and it is representative to all the users of the PPP. This theorem hugely simplifies the derivation of performance metrics.

Slivnyak's theorem stems directly from the notion of Palm probability which can be synthesised using the following definition.

Definition 1 (Palm probability). *The Palm probability of a point process ϕ is the probability of an event given that the point process contains an element at some location. Defining Y as an event (or property) of a point process ϕ , the Palm probability writes as*

$$\Pr(\phi \in Y | x \in \phi) \triangleq P_x(Y).$$

For stationary processes, the following holds

$$\Pr(\phi \in Y | x \in \phi) \triangleq \Pr(\phi_x \in Y | 0 \in \phi) \triangleq P_0(Y),$$

with the translated ϕ_x containing the origin of the Euclidean space. From the notion of Palm probability, its variant such that a specific point is not included in ϕ stands as the following reduced Palm probability.

Definition 2 (Reduced Palm probability). *The reduced Palm probability of a point process ϕ is the conditioned probability of an event Y to occur given that a point x is not included in the distribution of ϕ . Namely, we have*

$$\Pr(\phi \setminus \{x\} \in Y | x \in \phi) \triangleq P_x^!(Y).$$

The independence property for PPP, *i.e.* **Property 5**, suggests that conditioning the point process on x does not change the distribution of the PPP. Slivnyak's theorem formalizes this intuition.

Theorem 6 (Slivnyak's theorem). *If the intensity function of a PPP is such that $\lambda(0) > 0$, then*

$$\Pr(\phi \in Y | 0 \in \phi) = \Pr(\phi \cup \{0\} \in Y)$$

or, more compactly

$$P(Y) \equiv P_0^!(Y).$$

When the typical user is considered, the origin of the \mathbb{R}^d Euclidean space is included in the modeling point process of the users. As a PPP can be considered to spatially represent the random spatial locations of the users in a mobile network, the stationary property allows the translation, $\phi \triangleq \phi_x$ such that $0 \in \phi$. Slivnyak's theorem states that we can condition the users PPP on having an element at the origin of the Euclidean space, without influencing the distribution of the process. The formalisation of this statement holds as

$$P_0^!(Y) \triangleq \Pr(\phi \setminus \{0\} \in Y \mid 0 \in \phi) \triangleq P(Y),$$

where P indicates the unconditional distribution of ϕ . The implications of this on the analysis of mobile networks are huge. The typical user, or the element conditioned at the origin of the Euclidean space, can be regarded as the average receiver of the whole network such that the metrics can be computed at its location in order to obtain an averaged result over the whole set of users. The following results yield as a consequence of the fundamentals discussed in this section.

2.2 Analytical Standings for the SCDP

In this work, the SCDP, also acknowledged as coverage probability, is considered as the primary QoS metric to capture the health of the network transmitting conditions. Unlike the Shannon channel capacity, which gives a measure in bits per second of the additive white Gaussian noise (AWGN) affected user-to-node wireless link, the SCDP is capable of providing a measure over the user's ability of correctly decoding the transmitted message. The numerical value of the SCDP is mainly influenced by the interference pattern, the experienced nature of the wireless links and the available bandwidth for the content transmission. Research being conducted on the matter has widely made use of stochastic geometric principles of network design which allows elegant closed forms of the metric.

The maximum transmission rate of a channel, or Shannon capacity, is a QoS metric which provides an upper-bound to the maximum mutual information between the input and output of a channel affected by noise. Namely, the mutual information between two random variables is the maximum amount of informa-

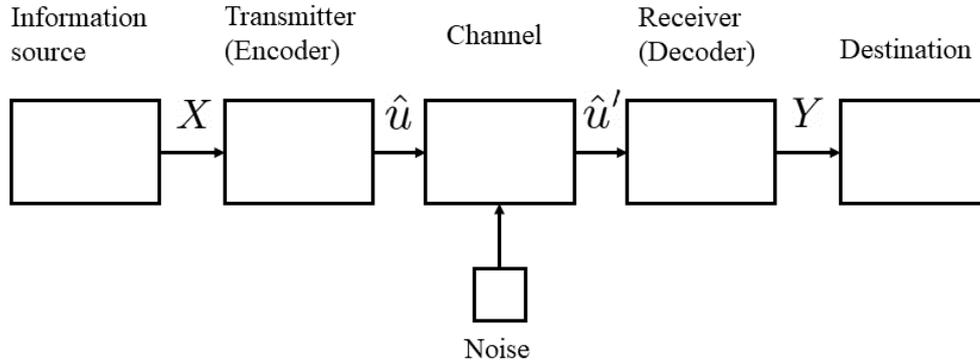


Figure 2.1: Block diagram digital communication.

tion we can get from X after having observed the output Y . When applied to an AWGN channel, the capacity can be formalised as

$$C^{AWGN} = \sup I(X; Y),$$

which gives the maximum error-free data rate a channel can support.

From Fig. 2.1, it can be seen how this notion of mutual information reflects the action of decoding a received message after being passed to a channel such that X has been affected by some kind of perturbation, *e.g.* noise. The set of random variables that define the noise in a wireless channel is particularly high such that the resulting perceived noise by the receiver is a combination of a huge number of unknown random variables. The central limit theorem of probability theory suggests that the summation of many random processes results in a zero-mean Gaussian distributed random variable $n \sim \mathcal{N}(0, N)$, where N is the variance of the process. Firstly theorised by Claude Shannon in 1948 [32], the maximum achievable transmission rate of a channel allows an elegant closed form when the channel is affected by AWGN [33], giving

$$C^{AWGN} = B \log_2(1 + \gamma) \text{ [bps]},$$

where B indicates the effective bandwidth for the content transmission and γ is the experienced SINR. This QoS metric has intrinsic fairness issues when, for instance, the maximization of the channel rate is performed over a set of wordbook contents.

Lack of fairness might result in the most popular contents being encouraged to experience better transmitting conditions while least popular contents can suffer from a lack of fairness as most of the network resources are allocated to a limited set of contents. The same issue can be encountered when trying to maximize the channel capacity for a set of users. In this case, those users that experience the best channel conditions are advantaged at the expenses of other users. To overcome this problem, the SCDP, also indicated as coverage probability, lies as a fairer metric which accounts for the whole set of contents or users within the network. The formal definition of SCDP is

$$\text{SCDP} = \Pr \left(B \log_2 (1 + \gamma) > \rho \mid \gamma \right), \quad (2.1)$$

which is conditioned on the SINR distribution.

2.2.1 Analytical Derivation of the SCDP

When Rayleigh channel fading envelope is considered, closed-form results of SCDP can be attained, given the standard exponential distribution of the channel power gain which writes as $|h|^2 \sim \exp(1)$ (this result will be analytically derived in Section 2.3.1.1). Namely, given the generic definition of SINR, with considered transmitting power terms $P^S = 1$ for simplicity, we have

$$\gamma = \frac{P}{I + W} = \frac{|h|^2 r^{-\alpha}}{\sum_{i \in \mathcal{I}} |h_i|^2 r_i^{-\alpha} + W},$$

where P , I and W indicate the received, the interfering and the noise thermal power terms respectively, \mathcal{I} is the set of interfering network nodes, r is the user-node link-distance and $|h|^2$ represents the channel power gain.

It can be easily computed, using the SCDP definition in (2.1), that

$$\begin{aligned}
\text{SCDP} &= \Pr \left(\log_2 (1 + \gamma) > \frac{\rho}{B} \middle| \gamma \right) \\
&= \Pr \left(\gamma > 2^{\frac{\rho}{B}} - 1 \middle| \gamma \right) \\
&= \Pr \left(|h|^2 > \underbrace{r^\alpha (2^{\frac{\rho}{B}} - 1)}_{\tilde{\rho}} (I + W) \middle| |h|^2, r, I \right) \\
&= \mathbb{E}_{\tilde{\rho}} \left[\int_{\tilde{\rho}}^{\infty} e^{-x} dx \right] \\
&= \mathbb{E}_{\tilde{\rho}} [e^{-\tilde{\rho}}].
\end{aligned} \tag{2.2}$$

The derivation (2.2) can be interpreted as

$$\text{SCDP} = \mathbb{E}_{I, |h|^2, r} \left[e^{-s(I+W)} \Big|_{s=r^\alpha (2^{\frac{\rho}{B}} - 1)} \right] = \mathcal{L}_I(s) \Big|_{s=r^\alpha (2^{\frac{\rho}{B}} - 1)} e^{-r^\alpha (2^{\frac{\rho}{B}} - 1)W}.$$

Hence, the SCDP metric is shown to be a function of the Laplace functional for the interfering point process using the tools described in Section 2.1.1. In the following, the derivation of the Laplacian term is provided:

$$\begin{aligned}
\mathcal{L}_I(s) &= \mathbb{E}_{I, |h|^2, r} \left[\exp \left(-s \sum_{i=0}^I |h_i|^2 r_i^{-\alpha} \right) \right] \stackrel{a}{=} \mathbb{E}_I \left[\prod_{i=0}^I \mathbb{E}_{|h|^2, r} \left[\exp \left(-s |h_i|^2 r_i^{-\alpha} \right) \right] \right] \\
&\stackrel{b}{=} \exp \left(-\lambda \iint_{x, y \in \mathcal{D}} \int_0^\infty \left[1 - e^{-sa(x^2+y^2)^{-\alpha/2}} \right] e^{-a} da dx dy \right) \\
&\stackrel{c}{=} \exp \left(-2\pi\lambda \iint_{x, y \in \mathcal{D}} \left[1 - \frac{1}{1 + s(x^2 + y^2)^{-\alpha/2}} \right] dx dy \right) \\
&\stackrel{d}{=} \exp \left(-2\pi\lambda \int_0^R \left[1 - \frac{1}{1 + sr^{-\alpha}} \right] r dr \right) \\
&= \exp \left(-2\pi\lambda \int_0^R \left[\frac{sr^{-\alpha}}{1 + sr^{-\alpha}} \right] r dr \right),
\end{aligned} \tag{2.3}$$

where (a) comes from the independence property for a PPP, (b) derives from utilizing the pgfl of the interfering PPP discussed in Section 2.1.1, (c) derives from averaging out the Rayleigh channel fading coefficient and (d) is obtained from a

change from Cartesian to polar coordinates. The derivation (2.3) is representative for the use of element from stochastic geometry employed to infer analytical forms of the averaged network SCDP. However, specific cases which derive from design choices of the model are indicated in each chapter.

2.3 General Model Design

Downlink single-input single-output (SISO) HetNets, which comprise of different layers differentiated by their composing cells' radii size, are considered. More specifically, a macro base station (MBS) layer provides signal ubiquity over greater portion of space while a denser SBS tier is responsible for locally delivering high-rate transmissions to the users. Random spatial displacement of both users and network's nodes are defined in a 2D space \mathbb{R}^2 which describes the entire network space. The users PPP is indicated as ϕ^U with intensity function λ^U . Similarly, the PPPs for SBS and MBS nodes are respectively ϕ^S and ϕ^M with their associated intensity functions λ^S and λ^M . If not otherwise stated, we consider homogeneous PPPs over the whole network space \mathbb{R}^d , with their intensity functions not spatially dependent. Resorting the result of Slivnyak's theorem in Section 2.1.2, the typical user or typical receiver is placed at the origin of the \mathbb{R}^2 Euclidean network space and is able to perform its content search over a local region centered at $(0, 0)$, referred to as CSA.

Time is split into frames, within which each user equipment (UE) generates its content request. The required contents are all considered to have unit length and be drawn from a wordbook \mathcal{F} with size $|\mathcal{F}| = F$ and the set of probabilities $\hat{p}_f \in [0, 1]$ indicates the popularity of the single f -th content. These content popularity coefficients are drawn from a zipf-like distribution, with the skewness factor ν such that $\sum_{f=1}^F \hat{p}_f = 1$. This specific distribution of the content popularity has been widely investigated and validated to be highly representative for empirical observations [34–37]. Without loss of generality, we consider the set of content popularity coefficients to be ordered as a decreasing sequence such that $\hat{p}_1 > \hat{p}_2 > \dots > \hat{p}_F$, with the first content corresponding to the most popular request. Content popularity usually varies much slower than the density for content requests generation and in particular, in this thesis, is assumed static. Furthermore, in

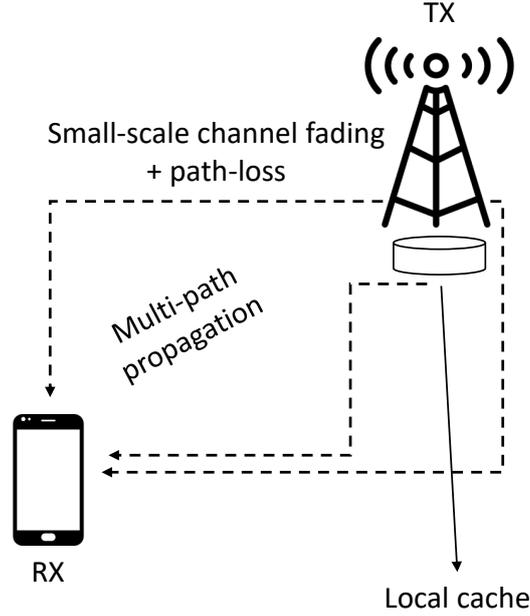


Figure 2.2: Multi-path signal propagation.

this work, it is assumed that both UE density λ^U and content popularity $\hat{\mathbf{p}} = \{\hat{p}_1, \dots, \hat{p}_F\}$ are known a-priori.

If not stated otherwise, each network node is equipped with a cache of size M which can store up to M distinct contents. All the pieces of equipment are allowed to operate on the same frequency bands with a total bandwidth indicated by B [Hz]. When a node is required to simultaneously perform multiple content transmissions, a frequency division multiple access (FDMA) scheme is adopted to equally split the bandwidth into a number of distinct content requests for delivery. The user-load ξ_k is defined as the random number of distinct content requests forwarded to the same k -th edge node. When CoMP techniques, such as joint transmission (JT), are in place, the experienced user-load for content transmission is indicated as $\Xi = \max\{\xi_1, \dots, \xi_{K_f}\}$ and is driven by the maximum user-load perceived by the cooperating set of nodes.

The path loss is modeled by a factor $r^{-\alpha}$, where r denotes the user to node distance and α is the path loss exponent. Typical path loss exponent values for urban areas range within $\alpha \in [2, 4]$ (see [38]). A simplified path-loss model is used, where r_k is the distance between the k -th edge node and the reference user such

that $r_k^{-\alpha} = \max(d_0, r_k^{-\alpha})$ where $d_0 = 1$ [m] is the reference distance.

In the following of this chapter, the set of employed channel fading models, the probabilistic content caching model and the performance gains from considering a CoMP technique are presented.

2.3.1 Small-scale Channel Fading Models

In wireless environments, transmitted signals may be subjected to multiple scatterings before arriving to the receiver, see Fig. 2.2. This causes random fluctuations in the received signal and this phenomenon is regarded as channel fading. In this section, two of the main models for channel fading are described.

2.3.1.1 Rayleigh Fading

This model best describes the phenomenon of having multi-path signals bouncing off buildings and obstacles before reaching the destination, without a direct LOS path. Hence, h is employed in this work to indicate the random fading channel coefficient, which is assumed to follow a circularly symmetric zero-mean unit-variance complex Gaussian distribution $h \sim \mathcal{CN}(0, 1)$. This is the direct result of the application of the *central limit theorem*. When the number of scattered multi-path components is sufficiently large, the channel impulse response can be interpreted as a Gaussian process, independently on the distribution of each scattered component. The envelope of the channel response is Rayleigh distributed, as it becomes the magnitude of the sum of the independent and identically distributed real and imaginary part of a Gaussian process. Given two independent standard Gaussian random variables $X \sim \mathcal{N}(0, 1/2)$ and $Y \sim \mathcal{N}(0, 1/2)$, the magnitude of the Rayleigh faded channel coefficient is given by

$$|h| = |h_I + jh_Q| = \sqrt{X^2 + Y^2},$$

such that $|h| \sim \text{Rayleigh}(\sqrt{1/2})$. The distribution of the channel power gain, *i.e.*, $|h|^2$, is obtained by initially computing its cdf $F_{|h|^2}(y)$ which is then derived to

obtain the pdf $f_{|h|^2}(y)$. Hence, given $|h| \sim \text{Rayleigh}(\sigma)$, it writes

$$\begin{aligned} F_{|h|^2}(y) &= \Pr(|h|^2 \leq y) = \Pr(-\sqrt{y} \leq |h| \leq \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx \stackrel{a}{=} \int_0^{\sqrt{y}} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx = 1 - e^{-\frac{y}{2\sigma^2}}, \end{aligned}$$

where (a) follows from the definition of the Rayleigh distribution

$$f_{|h|}(x) = \begin{cases} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0. \end{cases}$$

The derivative of $F_{|h|^2}(y)$ gives the pdf of $|h|^2$ as

$$f_{|h|^2}(y) = \frac{\partial F_{|h|^2}(y)}{\partial y} = \frac{1}{2\sigma^2} e^{-y/(2\sigma^2)},$$

from which, re-inserting $\sigma^2 = 1/2$ back to the obtained distribution, it follows

$$f_{|h|^2}(y) = e^{-y},$$

that can easily be recognised as the standard exponential pdf.

2.3.1.2 Rician Fading

The Rayleigh fading model is less accurate when it comes to the UDN setup in which LOS paths tend to exist. When a LOS link is experienced, then a dominant component of the multi-path signal affects the channel power fading experienced by the UE and so does the performance metric. According to the 3GPP standard, the probability of having a LOS link depends on specific cities' architectures [39].

In this case, the dominant LOS component affects the Gaussian process modeling the channel medium, whose envelope follows a Rician distribution. Specifically, the magnitude of the channel fading coefficient for modeling LOS wireless links writes as

$$|h'| = |h'_I + jh'_Q|,$$

where

$$h'_I \sim \mathcal{N} \left(\underbrace{\sqrt{\frac{k}{2(k+1)}}}_{\mu_I}, \underbrace{\sqrt{\frac{1}{2(k+1)}}}_{\sigma_I} \right)$$

and

$$h'_Q \sim \mathcal{N} \left(\underbrace{\sqrt{\frac{k}{2(k+1)}}}_{\mu_Q}, \underbrace{\sqrt{\frac{1}{2(k+1)}}}_{\sigma_Q} \right).$$

The power ratio between the LOS component over the scattered components is indicated as the k -factor and stands as an important parameter of this model. The channel fading coefficient $|h'|$ can be expressed as a function of two independent standard Gaussian random variables $X \sim \mathcal{N}(0, 1)$ and $Y \sim \mathcal{N}(0, 1)$

$$|h'| = |h'_I + jh'_Q| = \sqrt{(h'_I)^2 + (h'_Q)^2} = \sqrt{(\sigma_I X + \mu_I)^2 + (\sigma_Q Y + \mu_Q)^2}, \quad (2.4)$$

such that, given $\mu = \mu_I = \mu_Q$ and $\sigma = \sigma_I = \sigma_Q$, the magnitude of the channel coefficient is distributed as $|h'| \sim \text{Ricean}(\mu, \sigma)$. In other words, the magnitude of two circularly-symmetric independent Gaussian random variables is Ricean distributed, whose pdf writes as

$$\begin{aligned} f_{|h'|}(x) &= \frac{x}{\sigma^2} e^{-\frac{(x^2 + \mu^2)}{2\sigma^2}} I_0 \left(\frac{x\mu}{\sigma^2} \right) \\ &\stackrel{a}{=} 2x(k+1) e^{-x^2(k+1)-k} I_0 \left(2x\sqrt{k(k+1)} \right), \end{aligned}$$

where in (a) the values of μ and σ are introduced and I_0 is the zero-order modified Bessel function of the first kind, which does not hold a closed-form. The associated power gain, *i.e.* $|h'|^2$, writes as a noncentral Chi-squared random variable [40]. This result can also be observed by considering the squared channel magnitude from (2.4), such that the channel power gain writes as

$$|h'|^2 = (h'_I)^2 + (h'_Q)^2 = (\sigma_I X + \mu_I)^2 + (\sigma_Q Y + \mu_Q)^2,$$

which stands as the sum of squared Gaussian random variables with nonzero mean and non unitary variance. The channel power gain distribution follows from the definition of the pdf of a noncentral Chi-squared distribution with 2 degree of freedom, which writes as

$$\begin{aligned} f_{|h'|^2}(x) &= \frac{1}{2\sigma^2} \left(\frac{x}{2\mu^2} \right)^{\frac{1}{2}} e^{-\frac{x+2\mu^2}{2\sigma^2}} I_1 \left(\frac{\sqrt{2x}\mu}{\sigma^2} \right) \\ &= (k+1) \left(\frac{x(k+1)}{k} \right)^{\frac{1}{2}} e^{-x(k+1)-k} I_1 \left(2(k+1) \sqrt{x \frac{k}{(k+1)}} \right), \end{aligned}$$

where a first-order modified Bessel function of the first kind, which does not hold a closed-form, is involved. Despite this channel model is more suitable to capture the effects of LOS components at the receiver, it also involves a more difficult analytical expression, which greatly complicates the network's performance metric derivations.

2.3.2 Content Caching Schemes

In this section, a detailed discussion over the main content caching schemes, adopted cache filling procedure and sampling techniques is presented. The content caching schemes can be grouped in three main categories.

- **most popular content (MPC)** caching: the memory space is entirely dedicated to the most popular contents.
- **uniform content (UC)** caching: the contents in the wordbook \mathcal{F} , or, in some circumstances, a subset of \mathcal{F} , are cached following a uniform distribution.
- **probabilistic content caching** caching: it makes use of a derived set of probabilities, *i.e.* $\{p_f\} \forall f \in \mathcal{F}$, which indicate the likelihood of storing a content at each node's cache.

Given the knowledge of the content popularity, the MPC lies as the simplest caching scheme among the indicated, as it does not require a sampling technique

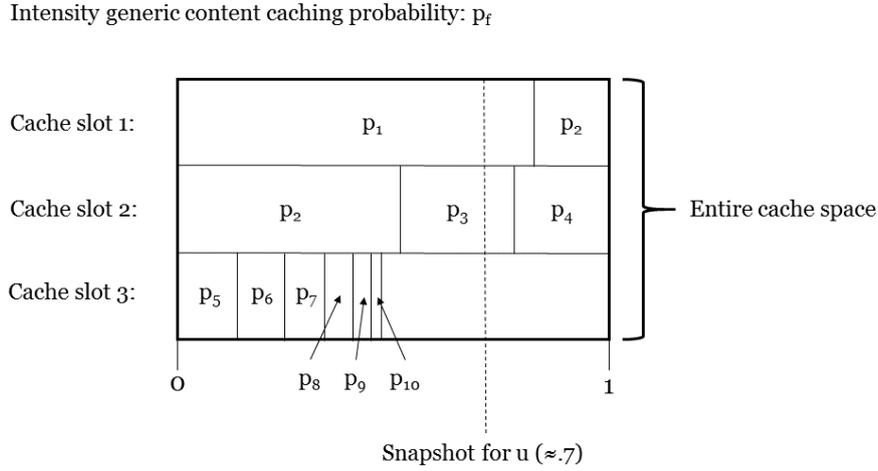


Figure 2.3: Content caching without replacement with $F = 10$ and $M = 3$. In this example, the cache contents are 1 and 3.

to create instances of the cache. Differently, when the UC scheme or a probabilistic content caching strategy is adopted, a set of instructions to accurately fill the cache slots is required. Authors in [41] extended the standard method for generating samples of Bernoulli distributed random variables to content caching in mobile wireless networks. Each edge node has a cache of size M which can store up to M distinct contents. The probability of the f -th content being cached at each cache-enabled edge node is denoted as $p_f \in [0, 1]$. Our caching model is regulated by the constraint $\sum_{f=1}^F p_f \leq M$. At each access node, the whole cache space $M \in \mathbb{N}_+$ is split into unit-lengthed chunks, and subsequently filled by the intensity of each probability p_f . Fig. 2.3 shows such setup when $M = 3$, where $\sum_{f=1}^F p_f < M$. To create a snapshot of the cache, a uniform random variable is drawn, i.e., $u \sim \mathcal{U} \in [0, 1]$ and the intersection between u and the cached probabilities determines the set of stored contents. This procedure is widely used to generate instances of the node's cache that follow the required content caching probabilities, i.e. $\{p_f\}$. Note that, in UC schemes, we set $p_f = 1/F, \forall f \in \mathcal{F}$.

The illustrated cache filling procedure is a *sampling without replacement* technique. More precisely, when a content $f \in \mathcal{F}$ is selected to be part of the local cache, it can not be considered to be a valid element when filling the successive slots. The difference between *sampling with replacement* and *sampling without*

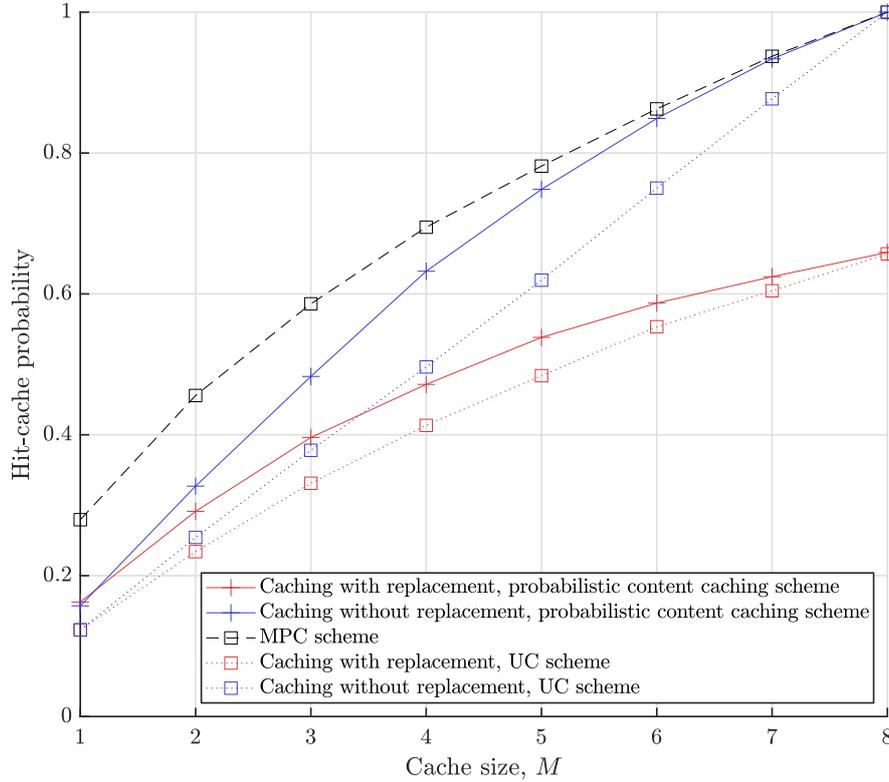


Figure 2.4: Comparison between *sampling with replacement* and *sampling without replacement* cache filling for various content caching schemes.

replacement is the presence of duplicates in the sampled set. The act of storing multiple copies of the same content within the same node reduces the capacity of the cache by a factor which depends by the number of copies. With a slightly more rigorous approach, given M the number of available slots for content caching and $\tilde{M} \in [0, M]$ which indicates the number of copies at the same node, the effective storing capacity reduces to $M - \tilde{M}$ contents. A comparative study between *sampling with replacement* and *sampling without replacement* is addressed in Figure 2.4. To numerically target difference between these two sampling techniques, a number of 20^4 of cache instances and content requests is generated and the hit-cache ratio is computed for the presented caching schemes against different cache sizes M for a wordbook of size $F = 8$. The representative employed probabilistic content caching scheme and content request generation process are generated by means of a zipf-distribution with skewness factor $\nu = 0.7$. This analysis allows to

have a visual feedback over the loss, in terms of hit-cache ratio, which incurs by storing multiple copies within the same memory space.

As indicated by the results of the numerical simulation in Fig. 2.4, caching with replacement strategies have a detrimental effect to the hit-cache ratio, due to the reduced capacity of the local memory space. However, it is important to point out that, having more copies of the same element within the same node's cache does not contribute to an increase of the SINR, thus the SCDP. Nonetheless, this same aspect is worth a deeper discussion when copies of the same content are cached at different nodes of the network. This is be addressed in more details in the following section.

2.3.3 Cooperation Gain Vs. Diversity Gain

When cache-enabled nodes are in use, it is useful to make use of the concept of cooperation gain and diversity gain, which we discuss in the following.

Cooperation gain benefits from the increase in the overall transmit power to compensate the path loss by promoting the use of CoMP transmission from more than one SBSs. This encourages the same contents to be cached at the SBSs to perform cooperative transmission but reduces the content diversity available in a particular CSA. Cooperative transmission could increase the level of interference perceived by neighbouring users. On the other hand, diversity gain comes about when different contents are cached at the SBSs within a CSA so that it is more likely for a user to find the requested content within the CSA to reduce the miss-cache probability.

There is a trade-off between cooperation gain and diversity gain given by the limited content storing resources available at the cache-enabled nodes. A more UC caching strategy would place more distinct contents at the edge to benefit from the content diversity gain whereas a more biased content caching strategy would place more popular contents to exploit cooperative transmission.

The discussion presented in this section can be extended to account for the effect of cache capacity reduction when *sampling with replacement* cache filling techniques are adopted. It is important to make a clear distinction between copies which are cached at distinct network nodes from the act of caching multiple copies

within the same device. While the first case allows to combine the transmitting power from distinct sources, thus enhancing the cooperating gain of the transmission, the second case reduces the content diversity gain of the network, while not contributing to the cooperating gain simultaneously. This further motivates the choice of adopting *sampling without replacement* over their to their counterpart.

Chapter 3

Edge Node Density and Content Caching Strategy for Arbitrary Spatial User Density

3.1 Introduction and Problem Statement

To alleviate the burden at the backhaul link, content caching has emerged as an attractive solution which also brings the benefits of shorter distance between content servers and users, greatly reducing latency and energy consumption. This approach has often been investigated in the setting of a multi-tier HetNet, see e.g., [42–46], which makes it easy to distinguish among different kinds of SBS by their radii. Meanwhile UDN is becoming reality as traffic lights, lamp posts, or drones can act as access nodes serving at the network edge. Having denser cells not only shortens the communication distance of each cell but also encourages spectrum reuse to increase capacity per unit area. The network edge however becomes the main source of interference as a single content transmission has to compete against a massive number of different requests to be successful. An information-theoretic approach that fully characterises the asymptotic limits for edge node deployment remains an open problem.

3.1.1 Literature Review

Recent years have witnessed considerable efforts on content caching in HetNets. In [41], an optimal content caching policy to maximise the diversity gain under different coverage models was provided focusing on the evaluation of the SCDP as the performance metric. In [47], a probabilistic caching model has been proposed for homogeneously distributed cache-enabled edge nodes. The authors provided a closed-form expression to optimise the SCDP over the content caching probabilities under the noise-limited case while a lower bound was used for interference-limited scenarios. Edge node cooperation was not considered and users were assigned to nodes to which they experience the best channel quality. In contrast, [48] extended the probabilistic content caching scheme to a multicasting network of cache-enabled BSs. In [49, 50], content caching with CoMP to improve the SCDP was investigated. However, the user-load at the BSs for simultaneous content deliveries was ignored and an arbitrarily chosen number of cooperating edge nodes was considered to examine the SCDP. By basing the analysis on a fixed set of cooperating content providers, the ability to properly consider the network dynamics and observe the effects of edge node density is however obscured. In [51], the authors provided both a lower-bound and approximation for the SCDP for a set of cooperating nodes, with the intention to investigate a trade-off between cooperation gain and content diversity gain. In particular it is shown that by increasing of the number of cooperating nodes, an initial benefit for the SCDP can be experienced. This trend is reversed to favour the diversity gain when the number of cooperating nodes is sufficiently large. This motivates our efforts in considering the edge node density as a decision variable of our maximisation problem.

A common assumption of these studies is that homogeneous point processes for modelling users spatial distribution were considered. Cooperative transmission and content caching have been widely studied under homogeneity conditions for the information sources deployment but the case for spatially dependent densities is less understood [52–54]. User’s spatial distribution normally does not follow a homogeneous point process, and is often spatially dependent due to manmade structures such as buildings and roads, and social events. In [55], the authors considered cooperative content transmission from non-homogeneously and identi-

cally distributed sources. However, the study was limited to Thomas cluster point processes and was not able to cope with any arbitrary geometric models. This motivates our work on edge caching networks that copes with any arbitrary user spatial distributions.

UDNs using cooperative transmission and content caching are the key enabler for content-centric mobile communications but as the number of content requests increases, interference becomes a serious bottleneck. Recent research in [22, 23, 56–58] investigated the impact of UE and BS density in the form of interference for UDNs. Specifically, the authors in [22] considered the use of higher frequency bands and network densification to improve the UE rates. They also showed the possibility of saving energy and reducing interference by idling some edge nodes in a UDN. Then [23] illustrated the increase in network capacity by having a denser SBS implementation. In [56], the authors accounted for the backhaul limitation to address the user’s outage probability with homogeneously deployed small cells, providing insights over the enhanced SCDF when adjusting the access nodes’ density given a fixed content caching placement strategy. Most recently, in [57], queuing theory was employed to model the movement of UEs and evaluate the throughput and EE performance. Though spatially dependent UE density was considered, homogeneous deployment of SBS was assumed within the same hot-spot. Cooperation was also not considered, and spatially dependent downlink interference was not analysed. In other words, each user perceives interference only from a homogeneous PPP of BSs belonging to the same cluster.

More specifically, [22] conducted an extensive investigation on the effect on the performance of a UDN under three degrees of freedom of exploitation of higher frequency bands, multi-antenna techniques and network’s equipment densification. The authors particularly highlight the benefits of network densification to reach up to 18x on average, with peaks up to 48x at the cell edge. The deployment of diffused network equipment allows shorter user-node distance, triggering low-power content transmissions and frequency reuse, greatly improving users’ experience. However, as the content sources of information move toward the network edge, more interference will be produced closer to the users.

CoMP content delivery strategies are also considered as key technologies to enhance the performance [59]. Non coherent JTs allow to increase the received

Reference	Description of the adopted content caching scheme
[41]	Probabilistic content caching scheme at single tier of BSs.
[47]	Probabilistic content caching scheme at single tier of caching helpers.
[48]	Probabilistic content caching scheme at single tier of BSs.
[49]	Combined MPC and probabilistic content caching schemes at single tier of SBSs.
[50]	Probabilistic content caching scheme at SBSs and no caching scheme at the MBS tier.
[51]	Probabilistic content caching scheme at single tier of caching helpers.
[55]	Probabilistic content caching scheme to assists device-to-device service.
[56]	MPC caching scheme at a single tier of SBSs.
[23, 57]	No content caching scheme is adopted.

Table 3.1: Comparison of content caching schemes.

power at the user side by combining the desired power from multiple sources with limited overhead signalling among the cooperating nodes [50]. CoMP is therefore a method to achieve much better performance if compared to the non cooperative case [60–63]. However, a close-form analysis of the SCDP has not been found for cooperative content transmissions.

Different from the previous work, our objective is to design a cache-enabled UDN that can cope with any arbitrary non-homogeneous UE density on a global scale by locally adapting the cache-enabled SBS density and the content caching strategy for enhancing the SCDP. Our model is that cache nodes can be turned on or off to adjust the SBS density for an optimal trade-off between coverage and interference. In [52], a model for the generation of non-homogeneous user distributions was considered, by means of a quantized representation of the continuous space wherein the density of users is locally constant. A similar network binning approach is proposed in this chapter, with the intention of being able to characterise each bin by means of an edge node density and local content caching policy. In the literature, a random number of caching nodes was often considered as a conditioning factor for the evaluation of a network performance metric. After the optimisation is performed, the random quantity is commonly picked deterministically, with its effects empirically studied.

In this chapter, we show that it is important to average the SCDP over the

random set of cooperating nodes to achieve better performance under multiple aspects. Our analysis is different from the literature where the number of cooperating nodes is assumed fixed which fails to account for most of the network information.

The contributions of our work are summarised as follows:

1. We consider a network model that discretises the coverage area into a finite number of bins where the SBS density and the content caching probabilities are optimised to cooperatively deliver contents to a spatially non-homogeneous content request generation.
2. We derive an SCDP lower-bound which can be adopted as a metric for performance maximisation. The SBS edge-node density and content caching probability are optimised via the SCDP lower-bound.
3. We analytically study the statistics of the user-load to account for a random set of cooperating nodes, which plays a major role in the achievable SCDP. The insight also allows us to tailor the derivation to suit different situations such as single node transmission, non-homogeneous user density, probabilistic caching model, and so on.
4. We adopt a steepest ascent algorithm to jointly optimise the SBS density and content caching probabilities over the entire network space according to local content popularity and user density, based on the SCDP lower bound and the gradients. Simulation results demonstrate that the proposed algorithm achieves significant SCDP performance gain compared to conventional approaches.
5. Different from previous studies, our results are not conditioned on the number of cooperative edge caching nodes. Instead, by averaging out the random number of cooperative edge caching nodes from the performance metric, we obtain a solution that depends on the SBS spatial density. This allows us to optimise the decision variables valid over all possible realisations of caching nodes.

6. We show that our solution of content caching probabilities and SBS density is spatially adaptive to the user density, highlighting the need to look for locally optimised solutions to improve the SCDP performance.

The remainder of this chapter is organised as follows. Section 3.2 presents the specific design choices to encompass the proposed SCDP maximisation. The chosen performance metric and its derivation are provided in Section 3.3. Then Section 3.4 gives details of the proposed solution to the problem. Numerical results are provided in Section 3.5. Finally, we conclude the chapter in Section 3.6.

3.2 Network Model and Design

A downlink ultra-dense small-cell wireless network¹ with short-range low-power cache-enabled SBS nodes is the subject of our investigation, which comprises of equipments, commonly referred to as fog access points² SBSs or simply caching nodes. The locations of SBSs are modeled as the atoms of a homogeneous PPP Φ_S with intensity function $\bar{\lambda}^S$. The SBS intensity function is considered to be upper-bounded by $\bar{\lambda}^S$, such that the decision variable can be modeled by the network to chase its optimal value by idling some of its nodes. Each edge node is equipped with a local cache and is capable of simultaneously performing up to M distinct content transmissions to UEs, when a hit-cache is experienced.

Those requests that experience a hit-cache with the SBS-tier within the CSA, can directly be processed by the network edge. A more rigorous definition of geometric properties of the CSA will be given in Section 3.2.1. In case a request cannot be met within the CSA, it can be either forwarded to the MBS-tier to be processed by means of backhaul resources, or fetched by closer SBSs across the network. These cases are however at the cost of extra power consumption and latency. The presented work focuses on the hit-cached cases, and it is assumed

¹In HetNets, MBSs are present to provide coverage with the aid of SBSs, forming a multi-tier structure. In this chapter, the inclusion of MBS is omitted in our problem formulation for simplicity but some discussion will be provided in Section 3.3.4.2 to extend our work to HetNets with MBSs.

²By the term ‘fog’, we indicate a network architecture that adopts near-user edge devices to carry out a significant amount of storage and communication.

that the interference pattern from across the network is to be tailored to respond to the local UE density.

Non coherent cooperative transmission of the same content from multiple caching nodes is adopted to exploit the cooperation gain. With this approach, the message is delivered without prior phase-alignment and little signalling between the cooperating edge nodes is required to work out the amount of bandwidth to perform the transmission [50]. A multicasting scheme for content dissemination is employed to perform multipoint-to-multipoint transmissions over the usable bandwidth. A FDMA scheme is adopted by each caching node to equally split the bandwidth into the number of distinct content requests (which is referred to as the user-load) for delivery. The user-load at an arbitrary node depends on (i) the probabilistic caching model, (ii) content popularity, (iii) UE density and (iv) SBS caching node density. This will be studied in the subsequent sections.

In this work, the network nodes' cache is considered to be updated over non traffic peaks of time and a backhaul which connects the network nodes is not admitted at this stage. The introduction of a wired link to directly link the network nodes to the cloud permits to replace obsolete instances of a node's cache across the entire network with low-latency. This allows to keep the adopted strategy optimal over time. However, the focus of this work lies on the study of a punctual moment of time. The presence of a backhaul link serving the network nodes also allows to resolve missed-cache cases by fetching the request from higher tiers of the network at the cost of extra energy and burden at the backhaul. Consequently, to target these expenses is crucial when designing the whole model as they represent a major factor which influences the derived optimal strategy. Future research efforts are needed in order to generate a adequate model which can precisely capture the expenses from the introduction of the backhaul.

The edge nodes are to be switched between an active or idle state to achieve certain SBS density, which is optimised according to the content popularity and the UE density, to balance between cooperation and diversity gain, energy consumption as well as interference. Hence, it is allowed the network to adjust the local densities of caching nodes to best accommodate the user requests. The optimisation operates to determine the optimal spatial displacement of active edge nodes from an initial homogeneous density of SBS such that $\lambda_n^S \leq \bar{\lambda}^S$. As a re-

sult of the optimisation, it is permitted the optimal set of $\{\lambda_n^S\}$ to generate a non-homogeneous PPP over the considered network space. At the same time, probabilistic proactive caching is performed at the active SBSs for further improving the SCDP. To derive an achievable lower-bound of the target SCDP, we condition our performance metric on the existence of at least one edge nodes that have cached the required content. Further details will be provided in Section 3.3.3.

3.2.1 UE Non-Homogeneity

To best cater for the non-homogeneous nature of UE distribution, we partition the entire space of network coverage into square-shaped bins. The center of each bin is regarded as a representative user for that bin, with its coordinates (x_n, y_n) . We refer to this location as a user candidate location (UCL), a reference for all the users within the same bin in terms of user density dependent parameters such as SINR and user-load at the set of cooperating edge nodes. It is noted that although hexagonal shapes are usually adopted for tessellation to mimic circular coverage of radio signals, the use of square-shaped bins is chosen for simplicity. In addition, we define the CSA for a UCL as the squared space of side $2d$ over which a content is requested. It is assumed that the nodes within the CSA are the possible content providers for the reference UCL whereas the nodes outside, if active, cause interference. The CSA also serves to provide the boundary for cooperative transmission.³

For a given UCL, (x_n, y_n) , the average number of UEs within its CSA, \mathcal{D}_n , is given by

$$U_n = \int \int_{(x,y) \in \mathcal{D}_n} \lambda^U(x,y) dx dy. \quad (3.1)$$

Therefore, given the area of each bin, denoted by, $Area(\mathcal{D}_n) = 4d^2$, we can use

$$\lambda_n^U = \left(\frac{U_n}{Area(\mathcal{D}_n)} \right) = \frac{U_n}{4d^2} \quad (3.2)$$

³In [52], an analogous network binning was used to emulate a non-homogeneous point process of UEs. Under this approach, local probabilistic caching model, caching node density, content popularity and all the statistical parameters involved are the same for all the UEs within the CSA of a reference user.

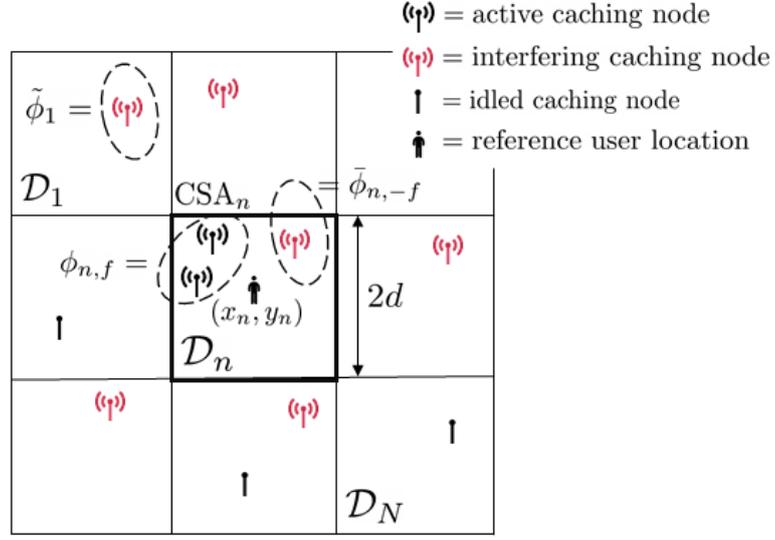


Figure 3.1: The bin-based discretisation of the space, where the sets $\phi_{n,f}$, $\bar{\phi}_{n,-f}$ and $\tilde{\phi}_i$ will be defined in Section 3.2.2.

to approximate any arbitrary 2D continuous UE density function at the n -th network bin.

In doing so, the discretisation of the entire coverage area into bins can be easily applied to compute any global performance metric of the entire network. Based on the discretised network space, the PPPs for the SBSs Φ_S and UEs Φ_U are modeled as a homogeneous PPP within each network bin.

The grid-based discretisation of the coverage space with a representative n -th UCL and CSA is shown in Fig. 3.1. Focusing on the highlighted UCL, we see that the edge nodes belonging to other squares are not included in the content transmission but considered as the source of interference, if active. This shows that it may be desirable to move some edge nodes to an idle state to reduce interference. In this work, a CSA corresponds to a single network bin for simplicity sake.

A numerical validation of the proposed binned network approach has been carried out in Section 3.3, as the employed performance metric is analytically introduced and compared with its numerical obtained averaged value.

3.2.2 Content Delivery

When a content request is initiated from a UCL, the request is extended over its CSA. Those edge nodes inside the CSA that will experience a hit-cache are in charge of processing the request. Note that the number of cooperating nodes is a random variable whose probability mass function (pmf) follows a Poisson distribution. To keep the complexity low at the SBSs, non-coherent JT will be considered from the edge nodes. This cooperation technique has been widely used in HetNets due to its low complexity and limited communication overhead [50]. We therefore consider channel state information at the receiver (CSIR) to be available for coherent detection of cooperative content delivery while the channel state information at the transmitter (CSIT) is not exploited. Since we are dealing with a random number of cooperating edge nodes, managing the estimation of the channel knowledge for a random number of cooperative transmitters would be highly complex and less practical. Our results can be interpreted as a performance lower-bound for coherent cooperative transmission. Assuming a user at the UCL (x_n, y_n) asking for content f and having unit-power transmissions at the edge nodes, the received signal power for that user can be expressed as

$$P_{n,f} = \left| \sum_{k \in \phi_{n,f}} h_{n,k} r_{n,k}^{-\frac{\alpha}{2}} \right|^2 + \sum_{\bar{k} \in \bar{\phi}_{n,-f}} |h_{n,\bar{k}}|^2 r_{n,\bar{k}}^{-\alpha} + \sum_{i \neq n}^N \omega_i \sum_{\bar{k} \in \tilde{\phi}_i} |h_{i,\bar{k}}|^2 r_{i,\bar{k}}^{-\alpha} + W, \quad (3.3)$$

where $\phi_{n,f}$ denotes the set of cooperating edge nodes serving the requested content, $\bar{\phi}_{n,-f}$ denotes the set of interfering edge nodes transmitting different contents within the same CSA, and $\tilde{\phi}_n \triangleq \phi_{n,f} \cup \bar{\phi}_{n,-f}$. Thus, $\tilde{\phi}_i$ for $i \neq n$ corresponds to the set of all the interfering edge nodes outside the CSA of interest. Note that the interference originated from a given edge node outside the CSA is present only if that edge node is active and there is at least one UE within its CSA. As a consequence, we scale the interference power from the i -th CSA by the probability $\omega_i = 1 - e^{-\lambda_i^U 4d^2}$, whose derivation is a direct application of the void probability applied to the users PPP for the i -th CSA. This allows to model the interference pattern to adapt to the spatially dependent user density. For simplicity, all the edge nodes that have not cached the required content within the CSA are

considered as interferers.

Although the network is initially given as a homogeneous PPP of SBSs with intensity function $\bar{\lambda}^S$, we aim to adapt the regional intensity function for each UCL, i.e., to have λ_n^S for UCL (x_n, y_n) , so that the network performance can be maximised. For the PPPs $\phi_{n,f}$, $\bar{\phi}_{n,-f}$, $\tilde{\phi}_i (i \neq n)$, we have the following intensity measures:⁴

$$\Lambda(\phi_{n,f}) = \mathbb{E} \left[\left| \phi \left(\lambda_n^S p_{n,f} 4d^2 \right) \right| \right] \equiv \mathbb{E} \left[\left| \phi \left(\mu_{n,f} \right) \right| \right], \quad (3.4a)$$

$$\Lambda(\bar{\phi}_{n,-f}) = \mathbb{E} \left[\left| \bar{\phi} \left(\lambda_n^S (1 - p_{n,f}) 4d^2 \right) \right| \right] \equiv \mathbb{E} \left[\left| \bar{\phi} \left(\bar{\mu}_{n,f} \right) \right| \right], \quad (3.4b)$$

$$\Lambda(\tilde{\phi}_i) = \mathbb{E} \left[\left| \tilde{\phi} \left(\lambda_i^S 4d^2 \right) \right| \right] \equiv \mathbb{E} \left[\left| \tilde{\phi} \left(\tilde{\mu}_i \right) \right| \right], \text{ for } i \neq n. \quad (3.4c)$$

From (3.4c), note that the interference from across the network is not influenced by its local probabilistic content caching model but only by the cardinality of the sets $\tilde{\phi}_i$. This design choice has been taken to simplify the subsequent derivation of the SCDP. More specifically, outer CSA nodes can potentially fetch a content request in case a missed-cache occurs. While the transmitting scheme is synchronised within the same CSA to avoid concurrent transmissions of hit and missed caches, it is not between CSAs. This results in the interfering pattern from the outer CSAs, which is not driven by the regional probabilistic caching policy. From (3.3), we can write the SINR for the reference user at UCL n as

$$\gamma_{n,f} = \frac{\left| \sum_{k \in \phi_{n,f}} h_{n,k} r_{n,k}^{-\frac{\alpha}{2}} \right|^2}{\underbrace{\sum_{\bar{k} \in \bar{\phi}_{n,-f}} |h_{n,\bar{k}}|^2 r_{n,\bar{k}}^{-\alpha} + \sum_{i \neq n} \omega_i \sum_{\bar{k} \in \tilde{\phi}_i} |h_{i,\bar{k}}|^2 r_{i,\bar{k}}^{-\alpha} + W}_{\text{Interference } \mathcal{I}_{n,f}}}}. \quad (3.5)$$

A successful content delivery is deemed to occur if a target rate for transmission ρ is achieved. That is,

$$\mathcal{E}_{n,f} \triangleq \left\{ \frac{B}{\Xi_{n,f}} \log_2(1 + \gamma_{n,f}) \geq \rho \right\} = \left\{ \gamma_{n,f} \geq 2^{\frac{\rho \Xi_{n,f}}{B}} - 1 \right\}, \quad (3.6)$$

⁴The expected values are made on stochastic point processes. Therefore, the mean is made over the random location of the atoms within the area over which the process is defined and over the number of atoms of the process.

in which B denotes the whole available bandwidth, and $\Xi_{n,f}$ indicates the perceived load at the set of cooperating edge nodes.

It is possible that the achievable rate far exceeds the target rate. In this case, the interference caused by those edge nodes can be reduced by turning off some nodes while the target rate is still met. The benefit is twofold as both interference and power consumption can be reduced. This is one of the intuitions of this work that attempts to adapt the spatial intensity of the SBSs by selectively idling some edge nodes.

3.3 Performance Metric

Considering the event (3.6), we can express the SCDP conditioned on a large number of network parameters as

$$\Pr(\mathcal{E}_{n,f}) = \Pr \left(\frac{\left| \sum_{k \in \phi_{n,f}} h_{n,k} r_{n,k}^{-\frac{\alpha}{2}} \right|^2}{\sum_{\bar{k} \in \bar{\phi}_{n,-f}} |h_{n,\bar{k}}|^2 r_{n,\bar{k}}^{-\alpha} + \sum_{i \neq n}^N \omega_i \sum_{\tilde{k} \in \tilde{\phi}_i} |h_{n,\tilde{k}}|^2 r_{n,\tilde{k}}^{-\alpha} + W} \geq 2^{\frac{\rho \Xi_{n,f}}{B}} - 1 \mid \begin{array}{l} \{r_{n,k}\}_{\forall k}, \{h_{n,k}\}_{\forall k}, \\ \phi_{n,f}, \bar{\phi}_{n,-f}, \\ \{\tilde{\phi}_i\}_{i \neq n}, \Xi_{n,f} \end{array} \right). \quad (3.7)$$

To better model the SCDP metric, we first focus on the desired signal power term assuming $K_{n,f} \triangleq |\phi_{n,f}|$ cooperating edge nodes. Thus, we have

$$Z \triangleq \left| \sum_{k=1}^{K_{n,f}} h_{n,k} r_{n,k}^{-\frac{\alpha}{2}} \right|^2 = |X + iY|^2 = X^2 + Y^2, \quad (3.8)$$

where X and Y each follow $\mathcal{N}(0, \sigma_{n,f}^2 \equiv \frac{1}{2} \sum_{k=1}^{K_{n,f}} r_{n,k}^{-\alpha})$. It can be observed that

$$Z \sim \sigma_{n,f}^2 \mathcal{X}_2^2 \sim 2\sigma_{n,f}^2 \exp(1), \quad (3.9)$$

where \mathcal{X}_2^2 refers to the Chi-squared distribution with 2 degrees of freedom, and $\exp(1)$ is the standard exponential distribution. As such, the probability of the

event $\mathcal{E}_{n,f}$ to occur is

$$\begin{aligned} \Pr(\mathcal{E}_{n,f}) &= \Pr \left(\bar{Z} \geq \frac{1}{\sigma_{n,f}^2} \underbrace{\frac{(\mathcal{I}_{n,f} + W)}{2} (2^{\frac{\rho \Xi_{n,f}}{B}} - 1)}_{\tilde{\rho}} \middle| \begin{array}{l} \mathcal{I}_{n,f}, \\ \sigma_{n,f}^2, \\ \Xi_{n,f} \end{array} \right) \\ &= \mathbb{E} \left[\int_{\frac{\tilde{\rho}}{\sigma_{n,f}^2}}^{\infty} e^{-x} dx \middle| \tilde{\rho}, \sigma_{n,f}^2 \right] = \mathbb{E} \left[e^{-\frac{\tilde{\rho}}{\sigma_{n,f}^2}} \middle| \tilde{\rho}, \sigma_{n,f}^2 \right], \end{aligned} \quad (3.10)$$

where \bar{Z} corresponds to the standard exponentially distributed random variable. The SCDP of a single UCL can be found by averaging the expression (3.10) over the indicated conditioning variables and combining the results for all contents $f \in \mathcal{F}$. Unfortunately, no closed-form expression can be obtained for the averaged result. For this reason, we resort to Jensen's inequality⁵ that leads to

$$\begin{aligned} \text{SCDP}_{n,f} &= \mathbb{E}_{\mathcal{I}_{n,f}, \sigma_{n,f}^2, \Xi_{n,f}} \left[\exp \left(-\frac{(2^{\frac{\rho \Xi_{n,f}}{B}} - 1) (\mathcal{I}_{n,f} + W)}{2 \sigma_{n,f}^2} \right) \right] \\ &\geq \exp \left(-\mathbb{E}_{\mathcal{I}_{n,f}, \sigma_{n,f}^2, \Xi_{n,f}} \left[\frac{(2^{\frac{\rho \Xi_{n,f}}{B}} - 1) (\mathcal{I}_{n,f} + W)}{2 \sigma_{n,f}^2} \right] \right). \end{aligned} \quad (3.11)$$

The validity of Jensen's inequality for $\text{SCDP}_{n,f}$ follows from the convexity of the exponential function with negative argument. As the random ratio

$$\left\{ \frac{(2^{\frac{\rho \Xi_{n,f}}{B}} - 1) (\mathcal{I}_{n,f} + W)}{2 \sigma_{n,f}^2} \right\} > 0,$$

Jensen's inequality holds.

The accuracy of the binned-network approximation from Section 3.2.1 can now be tested. Monte-carlo numerical results are obtained by averaging the target metric from (3.11) for all the randomly placed users across the central bin in a 7×7 bins network, while the same quantity is computed for the representative user placed

⁵Jensen's inequality is a well known result in the fields of mathematics and probability theory such that, given a random variable X and a convex function f , then $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$

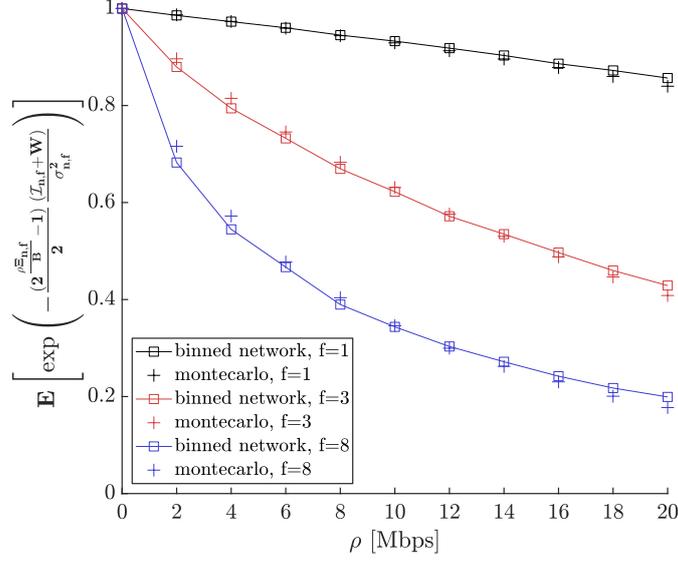


Figure 3.2: The average SCDP from a UCL and the random free-of-bin results for various wordbook contents.

at the center of the bin. In Fig. 3.2, the results are obtained for a representative $\{\hat{p}_f\}$ and $\{p_f\}$ retrieved from a zipf distribution with skewness factor $v = 0.7$, with $F = 8$, $M = 3$, an available bandwidth of $B = 40$ [MHz] and with two homogeneous PPPs for UE and SBS of density respectively $\lambda_n^U \approx 0.0153$, $\forall n \in \mathcal{N}$ and $\lambda_n^S \approx 0.0893$, $\forall n \in \mathcal{N}$. The considered values for user and caching node density adopted in Fig. 3.2 are chosen to be consistent with the guidelines for UDNs indicated in [7]. More precisely, higher values for the elements of the set $\{\lambda_n^S\}$ improve the cooperation and diversity gain across the network but also contribute to generate more interference at the network's edge. Differently, the values of the element in the set $\{\lambda_n^U\}$ govern the content request generation, controlling the competition among distinct content requests. The selected simulation parameters are selected to be consistent with the intensities employed during the future optimisation. The choice of a reduced available bandwidth for the simulation in Fig. 3.2 has been made to best represent the illustrated SCDP values. The analysis is conducted on single contents from the wordbook. As a consequence of that, the choice over the employed bandwidth works to best represent the quantities, allowing the most popular content, *i.e.* $f = 1$, to be delivered with an high probability of success and the least popular content, *i.e.* $f = 8$, to experience lower probability of success

over the considered spectrum of values for ρ . In this way, the binned network space approach can be validated for a wider range of SCDP values to resemble different qualities of content transmissions. The obtained numerical results indicate the binned network approach is suitable to represent the average randomly displaced users' SCDP as the same metric evaluated at the UCL. The notion of representative user is therefore validated.

Before we evaluate the lower bound of SCDP for the f -th content given by Jensen's inequality in (3.11), we look at several important parameters of our model.

3.3.1 The Scaled Probabilistic Content Caching Model

When a given content f is being considered, it is assumed that a hit-cache has occurred. The knowledge of the realisation of this event conditions the content caching probabilities at the set of cooperating edge nodes. Given that the generic f request has been cached, and the constraint over the cache size $\sum_{f=1}^F p_f \leq M$, the set of content caching probabilities becomes

$$p_{f'|f} = \begin{cases} p_{f'} \times \left(\frac{\sum_{\tilde{f}=1}^F p_{\tilde{f}}^{-\eta-1}}{\sum_{\tilde{f}=1}^F p_{\tilde{f}}^{-\eta} p_{\tilde{f}}} \right) & \text{if } f' \neq f, \\ 1 & \text{if } f' = f, \end{cases} \quad (3.12)$$

where η denotes the number of contents with caching probability of one. The scaling factor, introduced in (3.12), is important to correct the caching probabilities when content f is considered cached and the whole set of content caching probabilities changes accordingly. This adopted scaled version of the set of content caching probabilities is necessary to correctly evaluate the user-load and hence the effective bandwidth used by the set of cooperating nodes. Note that we have omitted the location index n as the discussion is the same regardless of the UCL. In the sequel, unless otherwise stated, the reference UCL 0 will be considered. In Fig. 3.3, the validity of (3.12) is proven starting from an initial representative set of content caching probabilities $\{p_{\tilde{f}}\}$ with skewness factor $\nu = 0.7$. The original set of caching probabilities is reported against the results from (3.12). The obtained monte-carlo results follow by producing random instances of the cache from the original set of caching probabilities $\{p_{\tilde{f}}\}$. When f is found to be cached, the

instance of the cache is stored and the resulting distribution of the caching probabilities is obtained for 5×10^4 executions. The conditioned set of probabilities deriving from (3.12) is shown to precisely follow the monte-carlo results. Also, it can be seen that when a specific content is considered to be locally stored, its content caching probability stands as one, as expected.

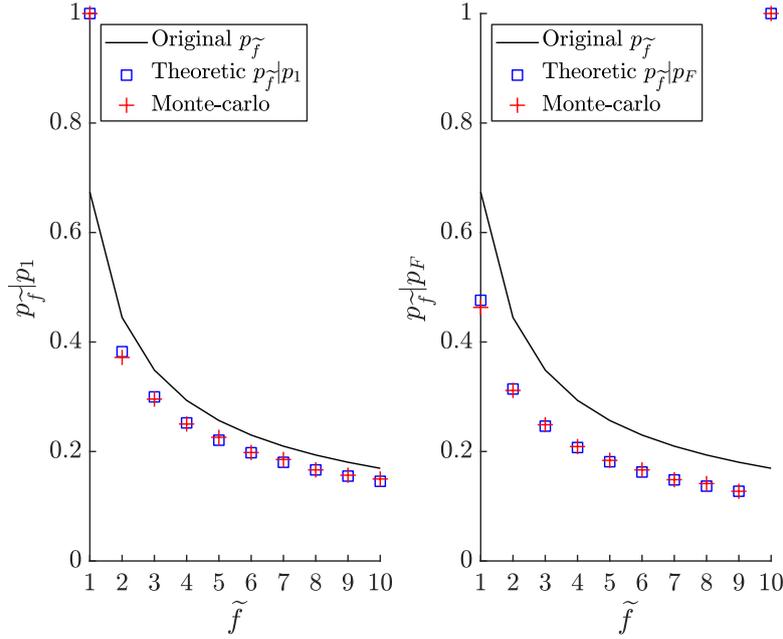


Figure 3.3: Original and scaled probabilities for the most and least popular contents with $F = 10$ and $M = 3$.

3.3.2 User-Load with K Cooperating Caching Nodes

In our model, it is assumed that the set of cooperating nodes has access to operate over the whole available bandwidth B when delivering content f . However, the effective bandwidth usage for the single content transmission depends on the user-load experienced by the whole set of edge nodes that cooperate to perform the transmission. If we now define the user-load for a generic active edge node k as $\xi_k \in \{1, 2, \dots, M\}$ which represents the number of distinct simultaneous contents to be delivered by that node, then the amount of bandwidth that can be used to deliver a single content for that node would be given by B/ξ_k . If there are

multiple edge nodes cooperatively delivering the same content, according to our JT approach, all the cooperative edge nodes will need to occupy a common portion of bandwidth to deliver the content to the UE. A centralized approach is considered and we assume that resource allocation is suitably done to ensure that the same portion of bandwidth is used by all the cooperating edge nodes for delivering the same content. Specifically, let us say that \mathcal{C} is the set of some cooperating edge nodes of interest. Then for $k \in \mathcal{C}$, those cooperating nodes should use

$$\frac{B}{\max_{k \in \mathcal{C}} \xi_k} \equiv \frac{B}{\Xi} \text{ bandwidth} \quad (3.13)$$

to deliver the same requested content. Note that the subscripts n and f have been dropped here for conciseness. Clearly, Ξ is a random variable and we can work out

$$\begin{aligned} \Pr(\Xi = m) &= \Pr(\Xi \leq m) - \Pr(\Xi \leq m - 1) \\ &= F_{\xi_k}(m)^K - F_{\xi_k}(m - 1)^K, \end{aligned} \quad (3.14)$$

where $F_{\xi_k}(m)$ stands for the cdf of the user-load for a single generic edge node ξ_k . Also, it is known that for discrete random variables, we can write

$$F_{\xi_k}(M) = \sum_{m=1}^M \Pr(\xi_k = m). \quad (3.15)$$

As such, if we can obtain the pmf of ξ_k , then we will be able to derive the pmf of the user-load for a set of collaborating edge nodes Ξ . As our discussion is always based on the condition that the f -th content is cached, ξ_k is certainly at least one. To derive $\Pr(\xi_k = m)$, we first define the sets of indices for hit-cache contents as ζ , missed-cache contents as $\bar{\zeta}$ and not-cached contents as $\tilde{\zeta}$. Considering a generic $\xi_k = m$ with $m > 1$, it means that $m - 1$ contents from the set $\mathcal{F} \setminus f$ contribute to the user-load. These contents belong to the set ζ and the total number of possible index combinations belonging to this set is

$$\binom{|\mathcal{F} \setminus f|}{m - 1} = \binom{F - 1}{m - 1}. \quad (3.16)$$

Given an instance of ζ , we can write that the total number of possible index combinations for $\bar{\zeta}$ stands as

$$\binom{|\mathcal{F} \setminus \{f, \zeta\}|}{M - m} = \binom{F - m}{M - m}, \quad (3.17)$$

with the remaining elements $\mathcal{F} \setminus \{f, \zeta, \bar{\zeta}\}$ that define the indices in $\tilde{\zeta}$. Note that when computing the pmf of ξ_k for $m = M$ (i.e., all the contents in the cache contribute to the user-load), the corresponding set $\bar{\zeta}$ would be empty. Similarly, when computing the pmf for $m = 1$ (i.e., no contents in the cache except f contribute to the user-load), we would consider an empty ζ . According to the set the content belongs to, for a generic $\tilde{f} \in \mathcal{F} \setminus f$, we have, at UCL 0, the probabilities:

$$\begin{aligned} \text{hit-cache} &\rightarrow \left(1 - e^{\lambda_0^U \hat{p}_{\tilde{f}} 4d^2}\right) p_{\tilde{f}|f} \\ \text{missed-cache} &\rightarrow e^{\lambda_0^U \hat{p}_{\tilde{f}} 4d^2} p_{\tilde{f}|f} \\ \text{not-cache} &\rightarrow 1 - p_{\tilde{f}|f} \end{aligned} \quad (3.18)$$

where $p_{\tilde{f}|f}$ is defined in (3.12), and \hat{p}_f denotes the global content popularity for content f which is known a priori and is not location dependent. Clearly, $\sum_{f=1}^F \hat{p}_f = 1$. A visual representation of an example with $m = 2$ is provided in Fig. 3.4. As we can see, the shaded box is the sure-cached content f . Moreover, the cache line separates the cached contents from the not-cached contents while the user-load line is used to distinguish the hit-cached contents from the missed-cached contents. In this example, one content experiences a hit-cache $|\zeta| = m - 1 = 1$, one experiences a missed-cache $|\bar{\zeta}| = M - m = 1$ while the remaining contents are not cached $|\tilde{\zeta}| = F - M = 5$. The probability of occurrence of each possible index combination can be found by

$$\underbrace{\left(1 - e^{\lambda_0^U \hat{p}_i 4d^2}\right) p_{i|f}}_{i \in \zeta} \times \underbrace{e^{\lambda_0^U \hat{p}_j 4d^2} p_{j|f}}_{j \in \bar{\zeta}} \times \underbrace{\prod_{k \in \tilde{\zeta}} (1 - p_{k|f})}_{k \in \tilde{\zeta}}. \quad (3.19)$$

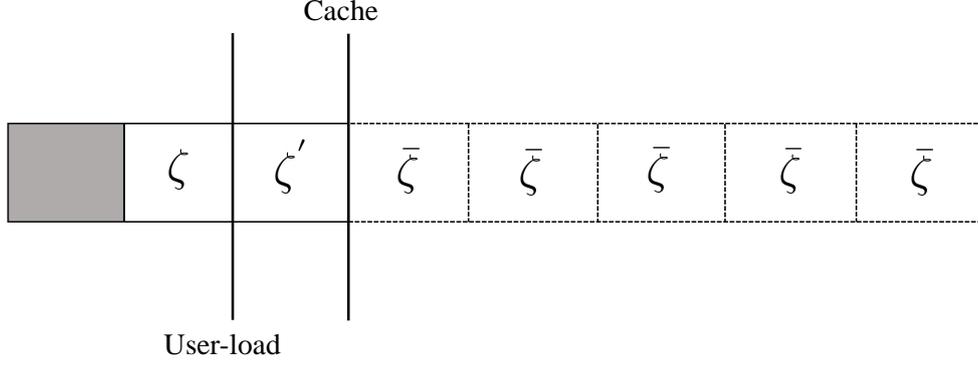


Figure 3.4: A visual representation of the sets ζ , $\bar{\zeta}$, $\tilde{\zeta}$ for a wordbook and cache size of $F = 8$ and $M = 3$ and with user-load $\xi_k = 2$.

Summing up all the contribution of the combinations will give $\Pr(\xi_k = 2)$ for the example in Fig. 3.4. As a result, we can obtain the generic pmf of ξ_k as

$$\Pr(\xi_k = m) = \sum_{c(m)} \prod_{i \in \zeta_c} \left(1 - e^{-\lambda_0^U \hat{p}_i 4d^2}\right) p_{i|f} \sum_{g(c(m))} \prod_{j \in \bar{\zeta}_g} e^{-\lambda_0^U \hat{p}_j 4d^2} p_{j|f} \prod_{k \in \tilde{\zeta}_g} (1 - p_{k|f}), \quad (3.20)$$

where $c(m)$ specifies a combination of hit-cached indices such that $\xi_k = m$, $g(c(m))$ indicates a combination of indices as a function of $c(m)$ and the sets ζ_c , $\bar{\zeta}_g$ and $\tilde{\zeta}_g$ are defined as before except they are now specific to a given combination, either $c(m)$ for ζ or $g(c(m))$ for $\bar{\zeta}_g$ and $\tilde{\zeta}_g$. By replacing (3.20) into (3.15), the pmf of the user-load at the set of cooperating edge nodes can be found from (3.14).

Note that because of the way we define $p_{f'}$ and $p_{f'|f}$ in the probabilistic caching model, the probability (3.20) may have a scaling issue, but this can be easily fixed by

$$\Pr(\xi_k = m) \leftarrow \frac{\Pr(\xi_k = m)}{\sum_{i=1}^M \Pr(\xi_k = i)}. \quad (3.21)$$

In Fig. 3.5, the theoretic pmf of $\Pr(\xi_k = m)$ in (3.21) is reported against the monte-carlo numerical validation. Our analytical derivation is shown to match with the empirical averaged results for a variety of values of λ_0^U , showing the accuracy of the theoretic approach for different user densities.

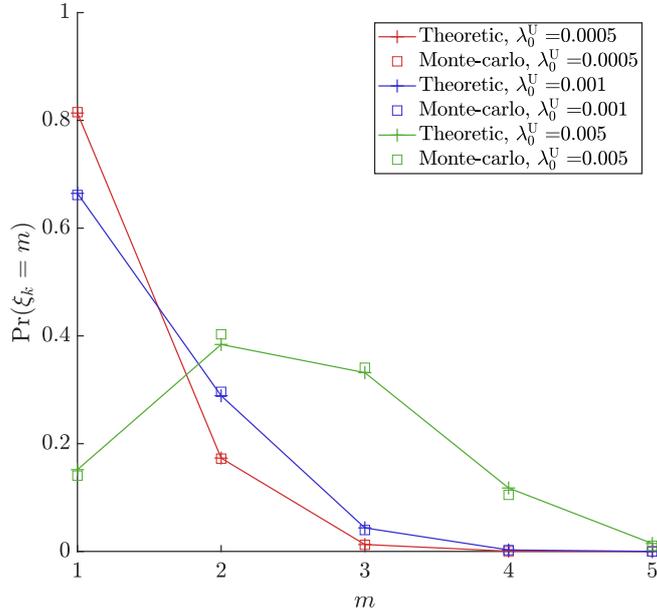


Figure 3.5: Visual feedback on the $\Pr(\xi_k = m)$ distribution for the most popular content, with $F = 20$, $M = 5$, and various values of λ_0^U .

3.3.3 A Zero Truncated Poisson Distribution

As described, the number of edge nodes follows a Poisson distribution across the network. As a consequence, there exists a probability of not having any caching nodes at the network edge to perform the required transmission, and in this case the content will have to be fetched from the nearest MBS over the backhaul link. For a more complete analysis, this would mean that the latency as well as the power consumption for the MBS will need to be accounted for. Our objective is however on the benefits of using the cache-enabled edge nodes in terms of the SCDP. Thus, we focus on the lower bound of SCDP in (3.11) where the probability is conditioned on the fact that there is at least one edge nodes (i.e., $K \geq 1$) able to provide the required content, successfully or not.

Because the case $K = 0$ is not valid in our analysis, we adopt the zero truncated Poisson distribution when accounting for the number of cooperating caching nodes, thus removing the case $K = 0$. Hence, given μ as the mean of the general Poisson

distribution $f(k; \mu)$, the zero truncated pmf, $g(k; \mu)$, can be expressed as

$$g(k; \mu) = \Pr(K = k | K > 0) = \frac{f(k; \mu)}{1 - f(0, \mu)} = \frac{\mu^k}{k!(e^\mu - 1)}, \quad (3.22)$$

where $\mu_{n,f} = p_{n,f} \lambda_n^S 4d^2$ if the n -th CSA for content f is considered. The zero truncated pmf will be useful when retrieving the unconditioned lower bound of the SCDP in (3.11).

3.3.4 The Objective Function

In order to come up with a global performance metric that can capture all the essential parameters of the caching network, we define $\mathcal{G}(\boldsymbol{\lambda}^S, \mathbf{p})$ as the metric, based on the lower bound (3.11) and averaged over the random variables, given by

$$\mathcal{G}(\boldsymbol{\lambda}^S, \mathbf{p}) = \sum_{n=1}^N \lambda_n^U \sum_{f=1}^F \hat{p}_f \exp \left(-\mathbb{E} \begin{pmatrix} \mathcal{I}_{n,f}, \{r_{n,k}\}_{k \in \phi_{n,f}}, \\ K_{n,f}, \Xi_{n,f} \end{pmatrix} \left[\frac{1}{2} \left(2^{\frac{\rho \Xi_{n,f}}{B}} - 1 \right) \frac{\mathcal{I}_{n,f} + W}{\frac{\sum_{k=1}^{K_{n,f}} r_{n,k}^{-\alpha}}{2}} \right] \right), \quad (3.23)$$

where $\boldsymbol{\lambda}^S \triangleq \{\lambda_1^S, \lambda_2^S, \dots, \lambda_N^S\}$ and $\mathbf{p} \triangleq \{p_{n,f}\}_{\forall n,f}$ are the network parameters to be optimised for maximising the function \mathcal{G} . Note that the variables $\{\mathcal{I}_{n,f}\}$, $\{K_{n,f}\}$, $\{\Xi_{n,f}\}$, depend on the choices of $\boldsymbol{\lambda}^S$ and \mathbf{p} . The index n in (3.23) indicates over which UCL the SCDP is computed.

Theorem 7. *The global performance metric, $\mathcal{G}(\boldsymbol{\lambda}^S, \mathbf{p})$ in (3.23), permits the expression*

$$\mathcal{G}(\boldsymbol{\lambda}^S, \mathbf{p}) = \sum_{n=1}^N \lambda_n^U \sum_{f=1}^F \hat{p}_f \exp \left(-\varphi_{n,f}^D \varphi_{n,f}^I \right), \quad (3.24)$$

where

$$\varphi_{n,f}^D(\lambda_n^S, \mathbf{p}_n) = \frac{1}{e^{\mu_{n,f}} - 1} \int_0^\infty \sum_{m=1}^M (2^{\frac{\rho m}{B}} - 1) \left(e^{\mu_{n,f} F_{\xi_{n,f}}(m) J_n(t)} - e^{\mu_{n,f} F_{\xi_{n,f}}(m-1) J_n(t)} \right) dt, \quad (3.25)$$

and

$$\varphi_{n,f}^I(\lambda_n^S, p_{n,f}) = \bar{\mu}_{n,f} \bar{J}(n) + \sum_{i \neq n}^N \omega_i \tilde{\mu}_i \tilde{J}(i) + W, \quad (3.26)$$

in which the functions $\mu_{n,f}$, $\bar{\mu}_{n,f}$ and $\tilde{\mu}_i$ have been defined earlier in (3.4), $\mathbf{p}_n \triangleq (p_{n,1}, \dots, p_{n,F})$, $F_{\xi_{n,f}}(m)$ is given by

$$\frac{\partial F_{\xi_{n,f}}(m)}{\partial p_{n,f}} = \quad (3.27)$$

$$\sum_{\bar{m}=1}^m \frac{\partial}{\partial p_{n,f}} \left[\sum_{c(\bar{m})} \prod_{i \in \zeta_c} (1 - e^{\lambda_n^U \hat{p}_i 4d^2}) p_{n,i|\bar{f}} \sum_{g(c(m))} \prod_{j \in \zeta_g} e^{\lambda_n^U \hat{p}_j 4d^2} p_{n,j|\bar{f}} \prod_{k \in \zeta_g} (1 - p_{n,k|\bar{f}}) \right] \quad (3.28)$$

with the indices n and f re-inserted to the equation, and

$$J_n(t) \triangleq \frac{1}{4d^2} \left(\iint_{\mathcal{D}_n \setminus \mathcal{B}_0} e^{-\frac{t}{(x^2+y^2)^{\alpha/2}}} dx dy + \pi e^{-t} \right), \quad (3.29)$$

where \mathcal{B}_0 is the circle of unit radius centered at the n -th UCL.

Proof. See Appendix A.1. □

3.3.4.1 Numerical Validations for $\mathcal{G}(\lambda^S, \mathbf{p})$

Having analytically defined $\mathcal{G}(\lambda^S, \mathbf{p})$ in Theorem 7, numerical evaluations can now be undertaken with the intention to both validate the obtained metric in (3.24) and the gap with the target SCDP from Jensen's inequality in (3.11). To obtain the following analysis, homogeneous intensity functions for both UEs and SBSs as $\lambda_n^U \approx 0.0153$, $\forall n \in \mathcal{N}$ and $\lambda_n^S \approx 0.0893$, $\forall n \in \mathcal{N}$ are considered over a 7×7 binned network, for a wordbook and cache size of respectively $F = 8$ and $M = 3$, a representative content popularity retrieved as a zipf distribution of

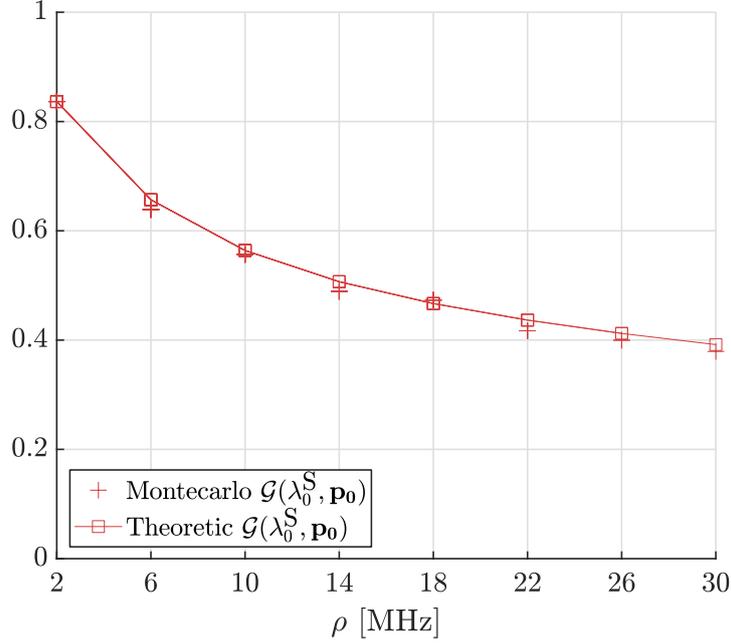


Figure 3.6: Match between empirical measure of \mathcal{G} from (3.23) and derived analytical result from Theorem 7.

skewness factor $v = 0.7$ and UC probabilistic caching policy. Results from Fig. 3.6 show the exact match of our analytical derivation from Theorem 7 and the monte-carlo outcomes obtained from (3.23). Furthermore, in Fig. 3.7, the gap between the target SCDP and the proposed metric (3.24) is inspected against the decision variables λ_0^S and $p_{0,1}$ for the representative central network bin. The reported numerical values highlight that the Jensen's gap turns out to be higher when either the SBS density is low or the caching probability for the content under investigation is low. Conversely, the gap is greatly reduced when the content is more accessible at the network edge, namely when more copies of the content are disseminated and when high density of caching node is deployed. It can be concluded that when UDN are accounted, the two measures are tighter as the contents are more available at the edge, thus making the application of Jensen's inequality suitable for the purpose of this study.

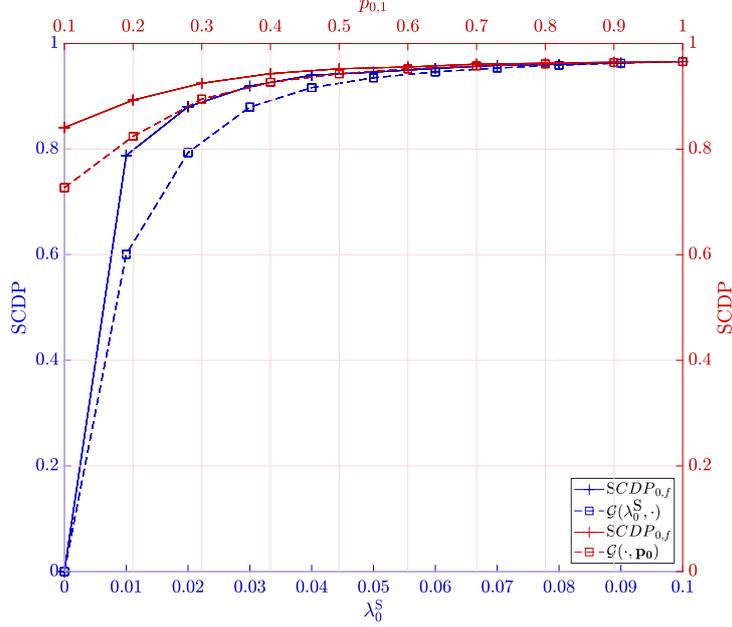


Figure 3.7: Resulting Jensen's gap between the target SCDP and the proposed lower-bound.

3.3.4.2 Extension with MBS Sharing the Same Frequency Bands

To consider the presence of MBSs, it is necessary to add an independent interfering term φ^M to $\varphi_{n,f}^I$ in the argument of the exponential function in (3.24). The derivation of φ^M can be easily done following the steps in Appendix A.1 as

$$\begin{aligned} \varphi^M &= \mathbb{E}_{r_b^M, h_b^M, \phi_b^M} \left[\sum_{b=1}^{N_M} |h_b^M|^2 (r_b^M)^{-\alpha} \right] = \mathbb{E}_{r^M, h^M, N_M} \left[N_M |h^M|^2 (r^M)^{-\alpha} \right] \\ &= \frac{\mu^M}{\mathcal{D}} \left(\iint_{\mathcal{D} \setminus \mathcal{B}_0} (x^2 + y^2)^{-\alpha/2} dx dy + \pi \right), \end{aligned} \quad (3.30)$$

where μ^M , r^M and h^M are, respectively, the average number of MBSs, the random link-distance and the channel fading coefficient. Note that the indices for location (i.e., n) are no longer needed because the same density for the MBSs is considered over the entire coverage area.

3.4 The Proposed Approach

In this section, the joint maximisation is defined, and the employed techniques are illustrated. The difficulties of the chosen problem are also discussed with final observations in terms of performance trade-offs retrievable from this approach highlighted in Section 3.4.3.

3.4.1 The Problem and Subproblems

Our objective is to maximise the global metric (3.23) (and hence (3.24)) by adapting the SBS density and the content caching probabilities for all the UCLs. That is,

$$(\mathbb{P}_0) : \quad \underset{\lambda^S, \mathbf{p}}{\text{maximise}} \quad \mathcal{G}(\lambda^S, \mathbf{p}) \quad (3.31a)$$

$$\text{subject to} \quad \sum_{f=1}^F p_{n,f} \leq M \quad (3.31b)$$

$$0 \leq p_{n,f} \leq 1 \quad (3.31c)$$

$$0 \leq \lambda_n^S \leq \bar{\lambda}^S, \quad (3.31d)$$

where $\bar{\lambda}^S$ denotes the upper limit of the caching node density. Note that while the initial upper-bound on the SBS intensity function is homogeneously considered, it can potentially be adapted as a non-homogeneous upper-bound as $\bar{\lambda}_n^S$, with no changes to be made in our method. This allows operators to better mimic the existing initial SBS distribution and eventually investigate the benefit from introducing edge nodes.

The problem (\mathbb{P}_0) needs some interpretation. We observe that the linear constraints for (\mathbb{P}_0) are jointly independent. The problem can be decoupled as a combination of subproblems which can be solved via an iterative algorithm. Therefore, we decompose (\mathbb{P}_0) in (4.9) into $N + 1$ subproblems, where N is the total number of distinct network bins. In particular, (\mathbb{P}_0) can be solved by repeatedly finding the solutions to N problems for the local optimum content caching probability (one for each n), and the solution for the SBS density optimisation problem, in an

iterative fashion. We refer to the two subproblems as (\mathbb{P}_1) and (\mathbb{P}_2) , given by

$$(\mathbb{P}_1) : \quad \underset{\mathbf{p}_n}{\text{maximise}} \quad \mathcal{G}_n(\boldsymbol{\lambda}^S, \mathbf{p}) \equiv \lambda_n^U \sum_{f=1}^F \hat{p}_f \exp(-\varphi_{n,f}^D \varphi_{n,f}^I) \quad (3.32a)$$

$$\text{subject to} \quad \sum_{f=1}^F p_{n,f} \leq M \quad (3.32b)$$

$$0 \leq p_{n,f} \leq 1, \quad (3.32c)$$

and

$$(\mathbb{P}_2) : \quad \underset{\boldsymbol{\lambda}^S}{\text{maximise}} \quad \mathcal{G}(\boldsymbol{\lambda}^S, \mathbf{p}) \quad \text{subject to} \quad 0 \leq \lambda_n^S \leq \bar{\lambda}^S. \quad (3.33)$$

Although Theorem 7 gives an expression to evaluate $\mathcal{G}(\boldsymbol{\lambda}^S, \mathbf{p})$, a closed-form expression is not possible, and a steepest ascent gradient based method is used when searching for the maximisers. Also, note that the constraints for both (\mathbb{P}_1) and (\mathbb{P}_2) are convex sets. Ideally, it would be necessary to prove that the involved objective functions are concave so that the search of a local maxima would lead to the global maxima. For both (\mathbb{P}_1) and (\mathbb{P}_2) we are dealing with two continuous optimisation of differentiable functions over a convex set. However, the study of the concavity of $\mathcal{G}_n(\boldsymbol{\lambda}^S, \mathbf{p})$ and $\mathcal{G}(\boldsymbol{\lambda}^S, \mathbf{p})$ is rather arduous. We target a stationary point for the problem (\mathbb{P}_0) by means of a diminishing stepsize gradient-based maximisation. The obtained results will thus be compared with the numerically found optimal solutions, to validate the proposed method.

To solve (\mathbb{P}_0) , the two subproblems (\mathbb{P}_1) and (\mathbb{P}_2) allow the gradients to be determined with respect to the decision variables $\{\lambda_n^S\}$ and $\{p_{n,f}\}$. We propose to solve the two subproblems separately and iteratively to provide the joint solution. In particular, (\mathbb{P}_1) addresses the probabilistic caching problem while (\mathbb{P}_2) deals with the effects of caching node density at a global scale. The pseudocode of the proposed algorithm is given as Algorithm 1.

3.4.2 Backtracking Line Search based Optimisation

Backtracking optimisation is a gradient-based inexact line search method which is particularly suitable for the exploration of stationary points and the avoidance of

Algorithm 1 Alternating optimization for solving (\mathbb{P}_0)

```

1: initialize the iteration index  $t = 1$ 
2: initialize  $\tau =$  some large number
3: initialize  $\mathbf{p}_n^{(t)}$  from a uniform content caching probability
4: initialize  $\boldsymbol{\lambda}^{\text{S},(t)} = \bar{\boldsymbol{\lambda}}^{\text{S}}$ 
5: initialize  $\delta_p^{\text{init}}, \delta_\lambda^{\text{init}}, \beta, \epsilon_p, \epsilon_\lambda, \kappa$ 
6: while  $\tau >$  some small threshold do
7:   for  $n = 1$  to  $N$  do
8:      $\delta_p \leftarrow \delta_p^{\text{init}}$ 
9:     while  $\delta_p \geq \epsilon_p$  do
10:      compute  $\mathbf{g}_p = \frac{\nabla \mathcal{G}_n}{\|\nabla \mathcal{G}_n\|}$ 
11:      if  $\mathcal{G}_n(\mathbf{p}_n^{(t)} + \delta_p \mathbf{g}_p) \geq \mathcal{G}_n(\mathbf{p}_n^{(t)}) + \delta_p \kappa \mathbf{g}_p^T \nabla \mathcal{G}_n$  then
12:         $\mathbf{p}_n^{(t+1)} \leftarrow \mathbf{p}_n^{(t)} + \delta_p \mathbf{g}_p$ 
13:      else
14:         $\delta_p \leftarrow \beta \delta_p$ 
15:      end if
16:    end while
17:  end for
18:   $\delta_\lambda \leftarrow \delta_\lambda^{\text{init}}$ 
19:  while  $\delta_\lambda \geq \epsilon_\lambda$  do
20:    compute  $\mathbf{g}_\lambda = \frac{\nabla \mathcal{G}}{\|\nabla \mathcal{G}\|}$ 
21:    if  $\mathcal{G}(\boldsymbol{\lambda}^{(t)} + \delta_\lambda \mathbf{g}_\lambda) \geq \mathcal{G}(\boldsymbol{\lambda}^{(t)}) + \delta_\lambda \kappa \mathbf{g}_\lambda^T \nabla \mathcal{G}$  then
22:       $\boldsymbol{\lambda}^{\text{S},(t+1)} \leftarrow \boldsymbol{\lambda}^{\text{S},(t)} + \delta_\lambda \mathbf{g}_\lambda$ 
23:    else
24:       $\delta_\lambda \leftarrow \beta \delta_\lambda$ 
25:    end if
26:  end while
27:  update  $\tau = \max\{\|\boldsymbol{\lambda}^{\text{S},(t+1)} - \boldsymbol{\lambda}^{\text{S},(t)}\|, \|\mathbf{p}_n^{(t+1)} - \mathbf{p}_n^{(t)}\|\}$ 
28:   $t = t + 1$ 
29: end while

```

saddle points⁶ of an objective function. This line search approach moves along the dimensions computed by the gradients of the objective function, with a tailored step-size found by starting from an initial large value, which is then iteratively decreased until a satisfying condition is met, *i.e.* Armijo-Goldstein condition. This is the general idea behind inexact line search algorithms, which represent a cheap option in terms of computational burden and avoid divergence of the solution. On the other hand, exact line search algorithms are quite cumbersome in terms of calculations, as they aim to find a local optimum, or an approximation of it, of the function along the directions. Typically, this class of algorithms are used to have a faster convergence in terms of number of iterations needed to find the solution. However, the search of the exact step size represents a significant burden which greatly increases the complexity of this class of algorithms. As a consequence of that, inexact line search methods are more indicated to perform the optimisation in this work. The proposed backtracking line search is among the simplest line search methods to implement, with strong guaranteed convergence properties to the solution.

As seen in Algorithm 1, the backtracking line search with Armijo-Goldstein condition [64,65] is employed when solving (\mathbb{P}_1) or (\mathbb{P}_2) . In the search, the objective at each iteration is to find a step size δ which satisfies the following Armijo-Goldstein condition at the t -th iteration

$$f(\mathbf{x}^{(t)} + \delta \mathbf{g}) \geq f(\mathbf{x}^{(t)}) + \delta \times \kappa \times \underbrace{\mathbf{g}^T \nabla f(\mathbf{x}_k^{(t)})}_{\text{local slope along direction } \mathbf{g}}, \quad (3.34)$$

where the superscript (t) is the iteration index, \mathbf{g} is a unit vector computed in the direction where a local increase occurs and $\kappa = 10^{-4}$ is the control parameter which ensures the increment to be at least a fraction κ of the Taylor approximation of f at \mathbf{x} . In addition, $f(\cdot)$ and $\nabla f(\cdot)$ correspond to the objective function and its gradient, respectively.

The initial step-sizes for the probability and density maximisation problems are chosen to be $\delta_p^{\text{init}} = 0.1$ and $\delta_\lambda^{\text{init}} = \frac{\bar{\lambda}^S}{4}$, respectively. Note that $\bar{\lambda}^S$ denotes

⁶A saddle is a point on the surface of a function where the gradients are zero but which is not a local extremum.

the maximum SBS density of the network. The search terminates if a sufficiently small step-size is reached. In our simulations, we set the stopping thresholds to be $\epsilon_p = \frac{\delta_p^{\text{init}}}{20}$ and $\epsilon_\lambda = \frac{\delta_\lambda^{\text{init}}}{20}$. Also, at the t -th iteration, if the Armijo-Goldstein condition is not met, the step-size δ will be reduced by a factor $\beta = 0.8$; otherwise, the optimising variables will be updated by $\mathbf{x}^{(t+1)} = \mathbf{x}^{(t)} + \delta \mathbf{g}$.

To carry out the steepest ascent algorithm, we also need the expression for the gradient $\nabla f(\mathbf{x}_k)$.

For the subproblem (\mathbb{P}_1) , we need to know $\frac{\partial \mathcal{G}_n}{\partial p_{n,f}}$, which after some lengthy derivations gives the expression (3.39) (see page 69), where

$$\frac{\partial \mu_{n,f}}{\partial p_{n,f}} = \lambda_n^S 4d^2, \quad (3.35)$$

$$\frac{\partial \bar{\mu}_{n,f}}{\partial p_{n,f}} = -\lambda_n^S 4d^2, \quad (3.36)$$

and $\frac{\partial F_{\xi_{n,\tilde{f}}}(m)}{\partial p_{n,f}}$ is given by (3.27) in which $p_{n,i|\tilde{f}}$ has been defined in (3.12) with the index n re-insterted in the expression.

Similarly, the gradient $\nabla f(\mathbf{x}_k)$ for the subproblem (\mathbb{P}_2) over the n -th UCL, i.e., $\frac{\partial \mathcal{G}}{\partial \lambda_n^S}$, writes as (3.40) (see page 69), where

$$\frac{\partial \mu_{n,f}}{\partial \lambda_n^S} = p_{n,f} 4d^2, \quad (3.37)$$

$$\frac{\partial \bar{\mu}_{n,f}}{\partial \lambda_n^S} = (1 - p_{n,f}) 4d^2. \quad (3.38)$$

3.4.3 Performance Trade-off

There is a performance trade-off achievable by controlling the local intensity of edge nodes density. If the number of edge caching nodes storing multiple copies of the same content is increased, then the cooperation gain is increased to enhance the SCDP. However, having more cooperative edge nodes increases the experienced user-load, resulting in less bandwidth which can be exploited for the content transmission. At the same time, when SBS density is too high, it might cause too much interference outside the CSA. The optimisation aims to strike a good balance by finding the edge caching node density to maximise the overall SCDP.

$$\begin{aligned}
\frac{\partial \mathcal{G}_n}{\partial p_{n,f}} &= \lambda_n^U \hat{p}_f \exp(-\varphi_{n,f}^D \varphi_{n,f}^I) \left[- \left(- \frac{\varphi_{n,f}^I \frac{\partial \mu_{n,f}}{\partial p_{n,f}} e^{\mu_{n,f}}}{(e^{\mu_{n,f}} - 1)^2} \int_0^\infty \sum_{m=1}^M (2^{\frac{m}{B}} - 1) \left(e^{\mu_{n,f} F_{\xi_{n,f}}(m) J_n(t)} - e^{\mu_{n,f} F_{\xi_{n,f}}(m-1) J_n(t)} \right) dt + \right. \right. \\
&\quad \left. \frac{\varphi_{n,f}^I}{e^{\mu_{n,f}} - 1} \int_0^\infty \sum_{m=1}^M (2^{\frac{m}{B}} - 1) \left[\begin{aligned} &J_n(t) \left(F_{\xi_{n,f}}(m) \frac{\partial \mu_{n,f}}{\partial p_{n,f}} + \mu_{n,f} \frac{\partial F_{\xi_{n,f}}(m)}{\partial p_{n,f}} \right) e^{\mu_{n,f} F_{\xi_{n,f}}(m) J_n(t)} \\ &- J_n(t) \left(F_{\xi_{n,f}}(m-1) \frac{\partial \mu_{n,f}}{\partial p_{n,f}} + \mu_{n,f} \frac{\partial F_{\xi_{n,f}}(m-1)}{\partial p_{n,f}} \right) e^{\mu_{n,f} F_{\xi_{n,f}}(m-1) J_n(t)} \end{aligned} \right] dt \right. \\
&\quad \left. + \frac{\varphi_{n,f}^D}{4d^2} \frac{\partial \bar{\mu}_{n,f}}{\partial p_{n,f}} \left(\iint_{\mathcal{D}_n \setminus \mathcal{B}_0} (x^2 + y^2)^{-\alpha/2} dx dy + \pi \right) \right] - \lambda_n^U \sum_{\tilde{f} \neq f} \hat{p}_{\tilde{f}} \exp(-\varphi_{n,f}^D \varphi_{n,\tilde{f}}^I) \left[\frac{\varphi_{n,\tilde{f}}^I}{(e^{\mu_{n,\tilde{f}}} - 1)} \times \right. \\
&\quad \left. \int_0^\infty \sum_{m=1}^M (2^{\frac{m}{B}} - 1) \mu_{n,\tilde{f}} J_n(t) \left(\frac{\partial F_{\xi_{n,f}}(m)}{\partial p_{n,f}} e^{\mu_{n,\tilde{f}} F_{\xi_{n,\tilde{f}}}(m) J_n(t)} - \frac{\partial F_{\xi_{n,\tilde{f}}}(m-1)}{\partial p_{n,f}} e^{\mu_{n,\tilde{f}} F_{\xi_{n,\tilde{f}}}(m-1) J_n(t)} \right) dt \right] \quad (3.39)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathcal{G}}{\partial \lambda_n^S} &= \lambda_n^U \sum_{f=1}^F \hat{p}_f \exp(-\varphi_{n,f}^D \varphi_{n,f}^I) \times \left[-\varphi_{n,f}^I \left(- \frac{\frac{\partial \mu_{n,f}}{\partial \lambda_n^S} e^{\mu_{n,f}}}{(e^{\mu_{n,f}} - 1)^2} \int_0^\infty \sum_{m=1}^M (2^{\frac{m}{B}} - 1) \left(e^{\mu_{n,f} F_{\xi_{n,f}}(m) J_n(t)} - e^{\mu_{n,f} F_{\xi_{n,f}}(m-1) J_n(t)} \right) dt \right. \right. \\
&\quad \left. \left. + \frac{1}{e^{\mu_{n,f}} - 1} \int_0^\infty \sum_{m=1}^M (2^{\frac{m}{B}} - 1) \left[\begin{aligned} &J_n(t) F_{\xi_{n,f}}(m) \frac{\partial \mu_{n,f}}{\partial \lambda_n^S} e^{\mu_{n,f} F_{\xi_{n,f}}(m) J_n(t)} \\ &- J_n(t) F_{\xi_{n,f}}(m-1) \frac{\partial \mu_{n,f}}{\partial \lambda_n^S} e^{\mu_{n,f} F_{\xi_{n,f}}(m-1) J_n(t)} \end{aligned} \right] dt \right) \right. \\
&\quad \left. - \frac{\varphi_{n,f}^D}{4d^2} \frac{\partial \bar{\mu}_{n,f}}{\partial \lambda_n^S} \left(\iint_{\mathcal{D}_n \setminus \mathcal{B}_0} (x^2 + y^2)^{-\alpha/2} dx dy + \pi \right) \right] - \sum_{\tilde{n} \neq n} \lambda_{\tilde{n}}^U \sum_{f=1}^F \hat{p}_f \left[\exp(-\varphi_{\tilde{n},f}^D \varphi_{\tilde{n},f}^I) \left(\varphi_{\tilde{n},f}^D \omega_n \iint_{\mathcal{D}_n} (x^2 + y^2)^{-\alpha/2} dx dy \right) \right] \quad (3.40)
\end{aligned}$$

Variable	Value	Description
$\lambda^S(x, y)$	≈ 0.0893 [unit/ m^2]	Initial caching node density
$\bar{\lambda}^S$	≈ 0.0893 [unit/ m^2]	Upper caching node density limit
#UE	300	Total number of users
-	140×140 [m^2]	Total network space
d	10 [m]	Half side length of each CSA
W	-174 [dBm]	Thermal noise power
F	8	Wordbook size
M	3	Cache size
v	0.7	Skewness factor content popularity
α	3	Path-loss coefficient
B	100 [MHz]	Bandwidth
ρ	$\{2, \dots, 30\}$ [Mbps]	Target bit-rate
P^S	1	Transmitting power at the edge node

Table 3.2: The network parameters.

3.5 Simulation Results

In this section, the simulation results to evaluate the performance of the proposed algorithm that jointly optimises the spatial cache node density and the content caching probability are reported. Table 3.2 provides the values of the network parameters used in the simulations, if not stated otherwise.

The following baselines are also considered and compared with the proposed algorithm:

1. MPC caching policy with $\lambda^S \approx 0.0893$. This can be used together with any SBS density.
2. UC caching policy with $\lambda^S \approx 0.0893$. This can also be used with any SBS density.
3. Content caching optimisation in [48] with fixed caching node density of $\lambda^S \approx 0.0893$.

In Fig. 3.8, an example of the employed arbitrary user density over a 1D projection of the 2D network space is shown, with $\min_{n \in \mathcal{N}} \lambda_n^U \approx 0.0037$ and $\max_{n \in \mathcal{N}} \lambda_n^U \approx 0.0350$. User density is considered zero outside the 140×140 [m^2]

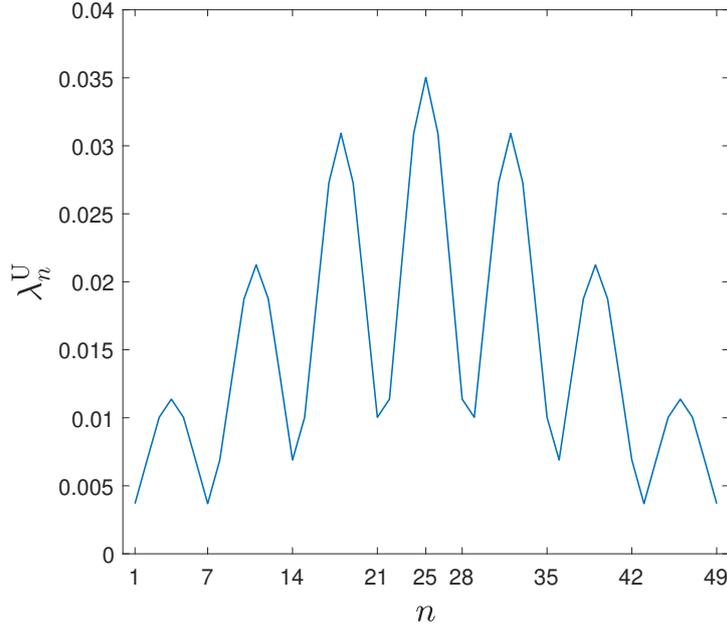


Figure 3.8: A user intensity function λ_n^U for the generating PPP.

network space. When zero UE density is experienced at some location, the corresponding ω_i coefficient in (3.26) would be zero and no contribution to the interference is given. Note that the bin model is used only for our optimisation to compute the edge node density and content caching probabilities but the SCDP results in the figures were obtained using Monte-Carlo simulations without the bin model restriction. Similarly, the following results will consider the case of no cooperating nodes, i.e., $K = 0$, avoided during the optimisation of (\mathbb{P}_0) , as a zero contribution to the reported SCDP.

3.5.1 SCDP vs User Target Bit Rate

Fig. 3.9 provides the SCDP results for the proposed algorithm and baselines against the network work-load (i.e., spectral efficiency usage) ρ/B for the 7×7 network. Results show that the proposed method achieves the best SCDP compared to other benchmarks although the SCDP of the proposed method gradually decreases and converges to that of JT-MPC for high spectral efficiency usage. The proposed method's superior performance is particularly obvious when the spectral efficiency usage is on the low side, which corresponds to the case with higher network densifi-

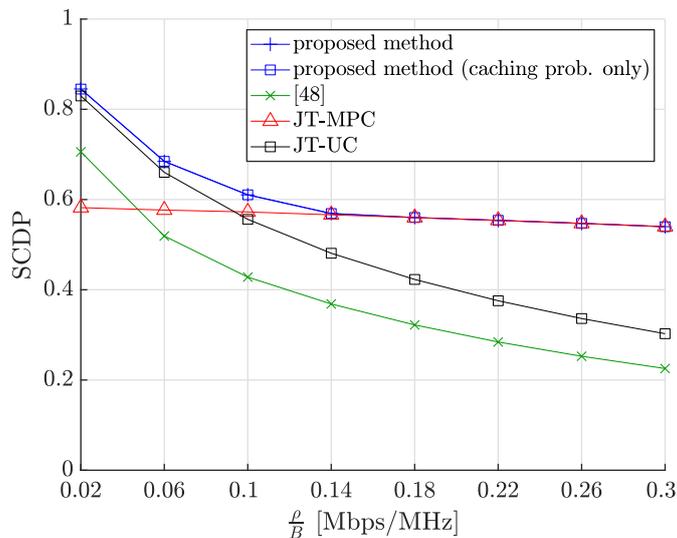


Figure 3.9: The SCDP results.

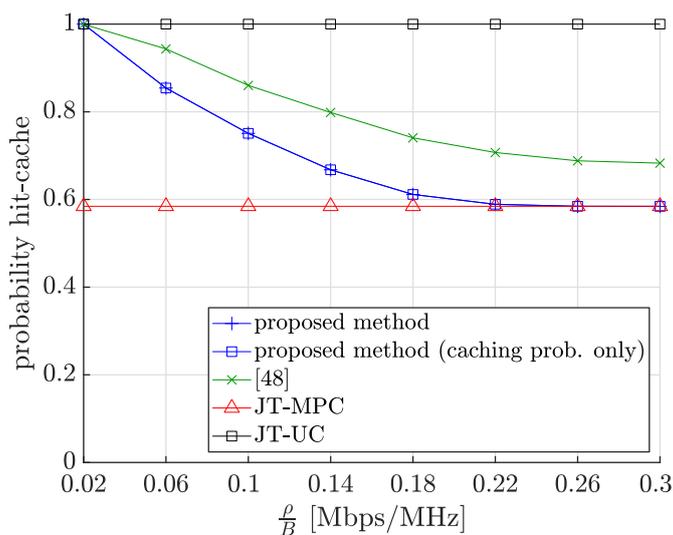


Figure 3.10: The probability hit-cache results.

cation. Also, it is expected that at high spectrum efficiency usage, JT-MPC tends to be optimal so it makes sense to see that the proposed method converges to JT-MPC. The hit-cache probability results in Fig. 3.10 show the different approaches being taken by the various methods. As we can see, for JT-MPC, its hit-cache probability and SCDP are similar, which suggests that in this scheme, whenever there is a hit-cache, it will likely be successfully delivered. On the contrary, for

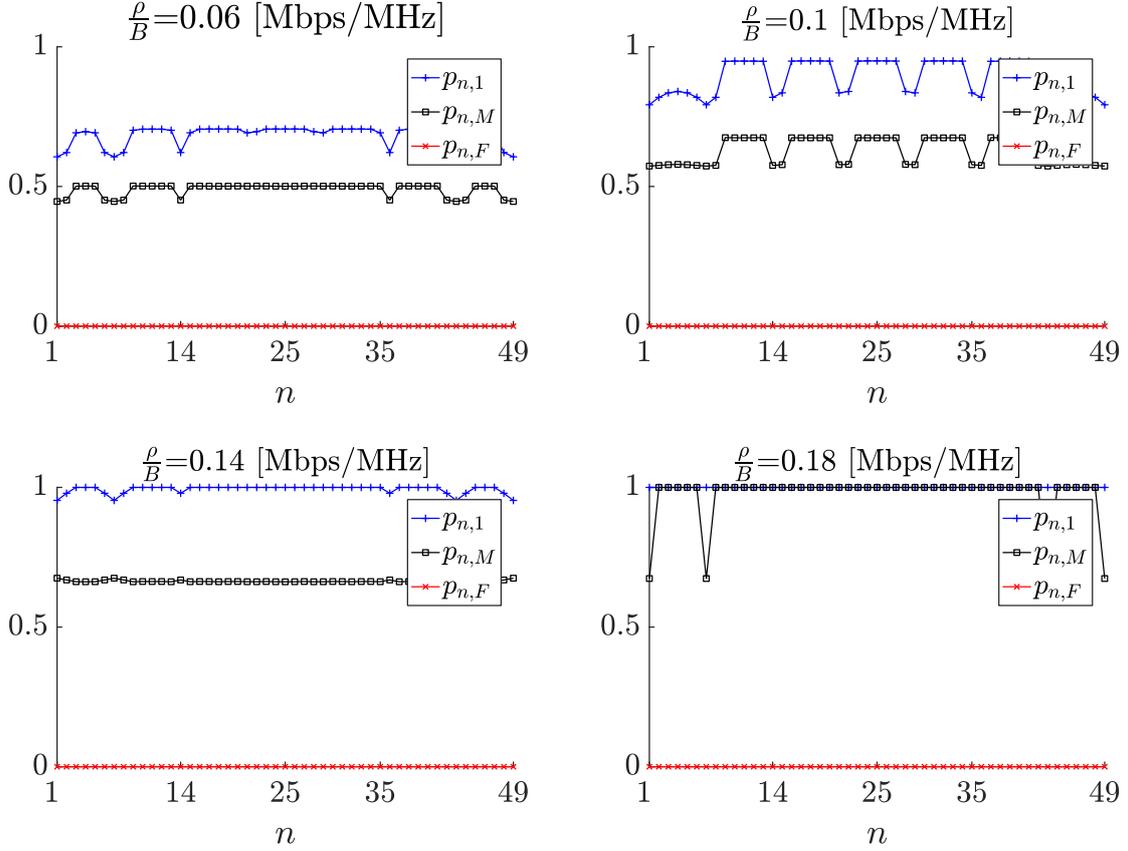


Figure 3.11: The probabilistic content caching policy by the proposed model for various ρ/B . The x-axis shows the index of the UCL while the y-axis corresponds to the content index.

JT-UC, it has a hit-cache probability of one, but not all the contents will be successfully delivered. Therefore, we can also observe that for the proposed method, it is able to increase the hit-cache probability while ensuring that almost all contents are delivered successfully and this is the reason why the proposed method is able to enhance the SCDP. The resulting sets of content caching probabilities from our proposed method are shown in Fig. 3.11 for some representative cases of work-load ρ/B and content indices $f = [1, M, F]$. When user density is high, cooperation gain outweighs diversity gain by storing multiple copies of the same content. A low user density operates to exploit the diversity gain by storing more distinct contents, as reported in Fig. 7. Also, a higher rate requirement ρ will amplify the benefit of cooperation gain and prefer a more biased caching strategy

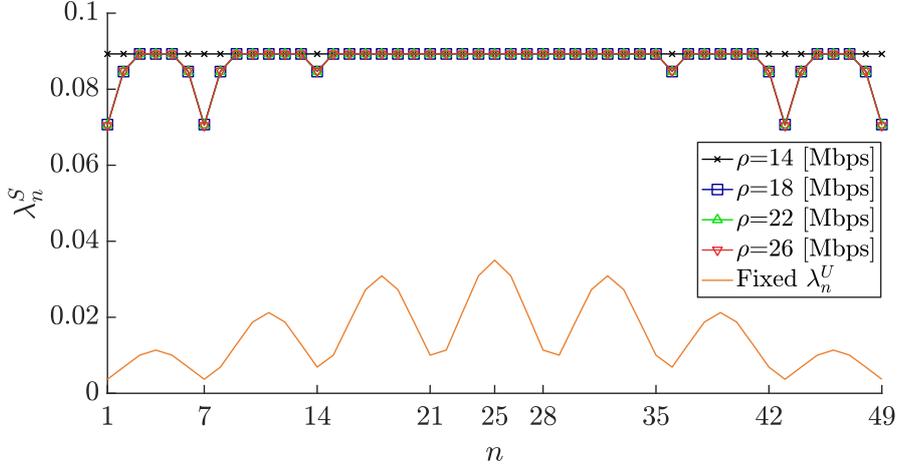


Figure 3.12: The SBS density $\{\lambda_n^S\}$ by the proposed method.

based on content popularity while a smaller ρ will favour a more uniform caching strategy to benefit from content diversity. It can be observed in Fig. 3.12 that the edge node density tends to follow the user density for content delivery. This is particularly clear at higher ρ . The reason is that at higher ρ the caching strategy tends to exploit more the cooperation gain, and decrease the edge node density in areas with low UE density, as shown in Fig. 3.12, to reduce the local number of cooperating nodes.

3.5.2 Network Energy Consumption

Knowing that idling caching nodes can help reducing interference, it is anticipated that the proposed algorithm can not only improve SCDP but also achieve energy saving. It is worth noticing that our objective function (i.e., a lower bound for the target global SCDP) does not explicitly take into account any measure of energy consumption. Caching nodes density and probabilistic content caching model also have a strong effect on the employed bandwidth for content transmission, and the amount of consumed energy by a network is dependent on the bandwidth over which the power is spread. To compare all the considered approaches, we provide the relevant results normalised by that achieved by the proposed method.

Fig. 3.13 shows the total network energy consumption of all the methods normalised by that of the proposed method. As we can see, all the benchmarks except

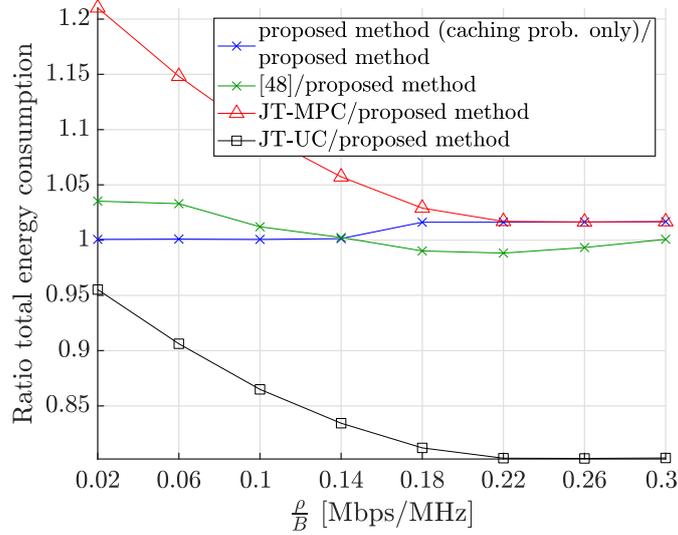


Figure 3.13: Average total network power consumption for various work-load values ρ/B .

JT-UC spend more total power than the proposed method, while it is also important to note that the proposed method has the best SCDP out of all the methods. In addition, although JT-UC spends the least overall power consumption, it has a much worse SCDP than the proposed method, as has been demonstrated before. From the results in this figure, we can compare the energy consumption performance with and without optimising the SBS density. Recall from the results in Fig. 3.9 that optimising the SBS density does not seem to provide any additional benefit for SCDP. We have now identified that the benefit of optimising the SBS density comes in terms of energy consumption.

3.5.3 Optimality for the Proposed Method

The proposed method finds a stationary point for maximising the lower-bound of the SCDP. To understand the optimality of the proposed method, Fig. 3.14 provides the results for the SCDP obtained by the proposed method and that obtained by the function from the optimisation toolbox of MATLAB. Due to the high computational complexity of `GlobalSearch`, we are restricted to consider only a simple 3×3 edge caching network. It can be noticed that at $\frac{\rho}{B} = [0.06, 0.14]$ [Mbps/MHz], the results for the proposed method and `GlobalSearch` depart only

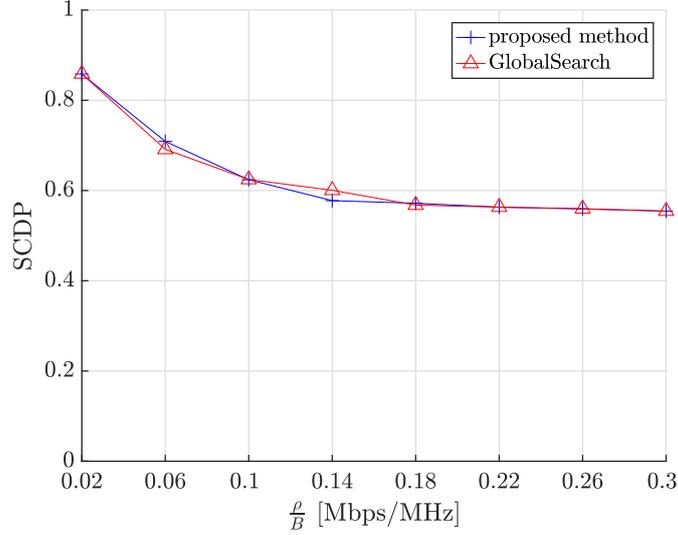


Figure 3.14: Comparison of SCDP results for different methods in a 3×3 network.

very slightly. For the other cases both methods appear to have achieved the same SCDP performance. Based on these results, it is believed that the proposed algorithm is effective to obtain the near-optimal solution.

3.6 Conclusions

In this chapter, we investigated the joint optimisation of the SBS caching node density and the content caching probability for an ultra-dense content caching network where the user density is non-homogeneous. We considered a simple cooperation strategy for delivering the contents from the active SBSs and the only interference control mechanism is to idle SBSs. The optimisation has been performed to maximise the lower bound of the SCDP using the steepest ascent algorithm. Simulation results have illustrated that significant performance improvement in terms of the SCDP can be obtained by the proposed algorithm over conventional approaches, and revealed that the optimised SBS density and content caching probabilities can adapt very well to the non-homogeneous user spatial density. An analytic derivation of the user-load has been derived and shown to be affected by a set of key network parameters. From the numerical results of the sub-optimal sets for the SBS density λ_n^S , it can be observed that the network's choice to idle

some of its components result is beneficial to the SCDP metric (although very narrow improvements have been shown when $\lambda_n^S < \bar{\lambda}^S$). This indicates that over-densification of a network can be detrimental to the performance metric, while obvious benefits are obtained in terms of energy consumptions.

As the network edge node density appears to be operating as a balance of energy consumptions and performance, it is of great importance to deepen the knowledge of the effects of this network parameter.

Chapter 4

Fundamental Network Energy Operating Point

4.1 Introduction and Problem Statement

The massive increase in the number of network components to overcome future demand for network capacity comes at the cost of higher network energy consumption [66, 67]. Moreover, the interference caused by unplanned and massive node displacement in UDN has been indicated as a major bottleneck for performance [68, 69]. Interference saturation is a significant hurdle for content transmission which brings the network to make great resource expenses to finally achieve its most performing state. To obtain network efficiency metrics is becoming more important as the trend of network densification becomes a reality [70, 71]. A theoretic-centered approach which can capture the network EE response to different edge node density becomes essential with the intention to provide high-performance QoS with cost-effective energy expense.

4.1.1 Literature Review

High densities are required to serve massive number of requests, and to strike the optimum balance between performance and efficiency remains an open problem. In [72], *i.e.*, Chapter 3, a limit on the sub-optimal density was proposed for jointly transmitting contents which maximises the overall SCDP, accounting for the re-

source management and interference pattern control. The direct consequence is that a control over network over-densification can also be beneficial to the users' QoS. From [73], the EE of a network is defined as the average network throughput per unit area over the energy consumption per unit area, which is then numerically maximised to determine an optimum operating tier's density. The studies [74–77] adopted the same interpretation of EE. In [74], an analysis was conducted on the obtained EE against the tier density and number of antennas at the BS. Also, [75] confirmed over the importance of the densification of smaller cells over macro cells in terms of EE for a SISO system. However, when aggregated to the measure of network capacity, misleading results are obtained since the adopted measure of energy consumption per unit area is an averaged value which should instead be considered as random variable when de-conditioning the overall EE metric. In [76], caching at BSs was investigated as a tool to improve the network EE and the advantage of caching at smaller cells was shown.

In this chapter, a novel approach for the evaluation of the network EE is envisioned. In [73–77], a ratio of expected values was considered. However, the number of nodes which contribute to energy expenses of the network also contributes to the interfering pattern, therefore affecting the network throughput. As a direct consequence, it is important to define the network EE as the expected value of a ratio of two dependent random quantities.

The contribution of this chapter includes a closed-form expression of the network EE for a given edge node density. By attaining the maximum of the EE metric, a global maximum conversion point (MCP) is found to deliver the highest efficiency to the operators. A comparison with the pure SCDP maximization is investigated and the distance between the most efficient and most performing network states is indicated as an important parameter. This means that the network can arbitrarily idle some of its nodes at the advantage of much lower energy consumptions.

The remainder of this chapter is as follows. Section 4.2 describes the model design. The proposed metric is elaborated and illustrated in Section 4.3 while the associated results are reported in Section 4.4. Finally, conclusions and future work are presented in Section 4.5.

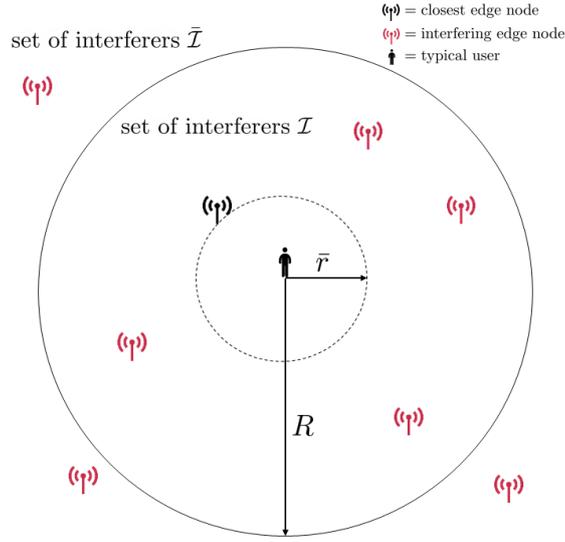


Figure 4.1: Visual representation of the network model, with the interfering set \mathcal{I} and $\bar{\mathcal{I}}$, closest user-node distance \bar{r} and CSA of radius R .

4.2 The Model

A typical user served by the network nodes in the denser tier of a downlink HetNet is considered. These pieces of equipment are responsible of delivering the users' requests directly at the edge of the network. Without loss of generality, no local cache at the nodes is considered and the user-node association relies on a proximity principle, *i.e.*, the user is associated to the closest SBS. The focus of this chapter is indeed beyond the local availability of the physical copy of the content. However, the introduction of a cache-enabled network is illustrated as an extension in this chapter. The interfering region is regulated by the distance of the closest node to the typical user \bar{r} . The sets \mathcal{I} and $\bar{\mathcal{I}}$ are used to denote the interferers within the circular sector of area $\pi(R^2 - \bar{r}^2)$ and from the outer CSA space, respectively. In this chapter, we refer to CSA as the circular region centered at the typical user and of radius R . For the sake of analysis, we consider each transmitter to be able to make use of the full available bandwidth for each content transmission. The activity of an higher tier, *i.e.* MBSs, to serve a request is considered to have null contribution to the network performance metric. The focus of this work is on a single tier only. Typically, when no network nodes are found within the CSA, a

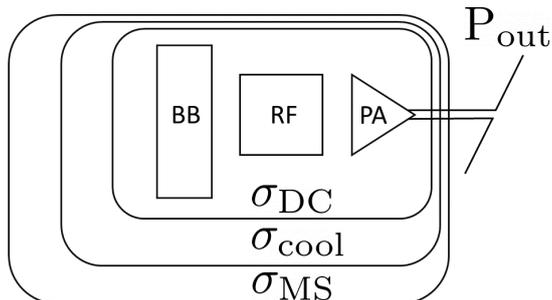


Figure 4.2: Block diagram of BS transceiver.

less dense tier is responsible for content delivery. However, its contribution to the SCDP is not considered at this stage as it requires the energy expenses of an MBS to retrieve a content to be properly quantified, greatly increasing the complexity of the model.

The energy expenses model of a single network node is inspired from [78]. Namely, a sum of three terms which account for the energy groups of a transmitter, *i.e.* power amplifier, radio-frequency transceiver and baseband interface, is accounted. The tier's single node energy expense can be written as

$$e_A = \frac{P_{\text{out}}/\eta_{\text{PA}} + P_{\text{RF}} + P_{\text{BB}}}{(1 - \sigma_{\text{DC}})(1 - \sigma_{\text{cool}})(1 - \sigma_{\text{MS}})} [W], \quad (4.1)$$

where σ_{DC} , σ_{cool} and σ_{MS} indicate the loss factors for DC-DC power supply, main supply and active cooling of the BS transceiver and η_{PA} is the power efficiency. A block diagram of a generic BS transceiver is depicted in Fig. 4.2.

To greatly simplify the analysis of this chapter, it is considered that no content transmission can be performed within 1 [m] of distance from the typical user. Fig. 4.3 gives the cdf of the shortest user-node distance as its analytical derivation, *i.e.*, $\Pr(\bar{r} < r) = 2\pi\lambda^S \int_0^{\bar{r}} r e^{-\lambda^S \pi r^2} dr$, and from its empirical validation obtained by means of a monte-carlo simulation. From the conducted analysis, it has been detected that $\Pr(\bar{r} < 1) = 2\pi\lambda^S \int_0^1 r e^{-\lambda^S \pi r^2} dr \approx 0.03$ with $\lambda^S = 0.01$. Hence, the following investigation is a close approximation of the target measure which allows particularly accessible expressions.

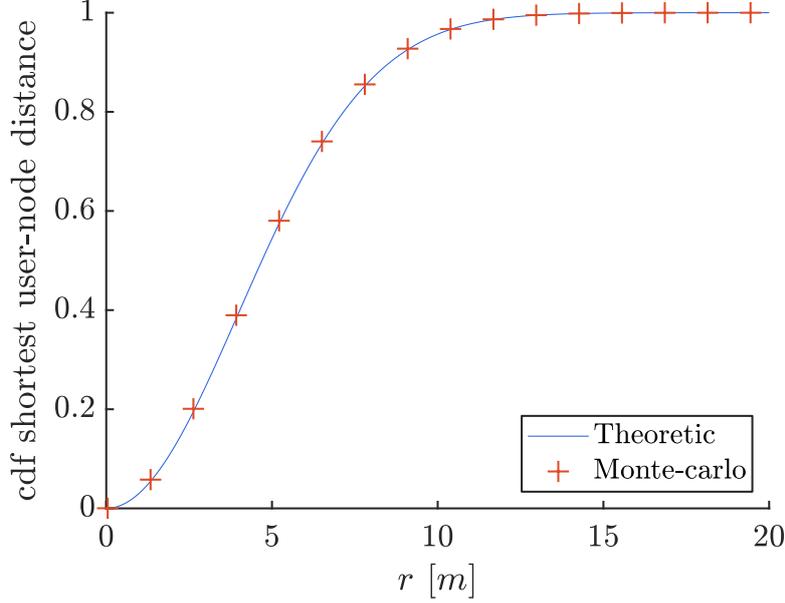


Figure 4.3: Theoretic and empirical measurements of the cdf for the random shortest user-node distance.

4.3 The Proposed Performance Metric

In this section, an analytical expression for the EE is derived. When the maximisation of the SCDP is conducted, the network energy expenses are usually not fully accounted. This often results in minor SCDP increments at the expenses of an enormous amount of consumed energy, especially when the network is interference saturated. The proposed performance metric considers the random ratio between the experienced SCDP and the consumed energy by the nodes within the CSA of the typical user. The ratio of the two terms, dependent on the random realisations of the considered PPPs, writes as

$$\left\{ \frac{\Pr \left(\frac{P^S |h|^2 \bar{r}^{-\alpha}}{\sum_{i=0}^{|\mathcal{I}|} P^S |h_i|^2 r_i^{-\alpha} + \sum_{i=0}^{|\bar{\mathcal{I}}|} P^S |h_i|^2 r_i^{-\alpha} + \sum_{m=0}^{|\mathcal{I}^M|} P^M |h_m|^2 r_m^{-\alpha} + W} > \tilde{\rho} \right)}{\sum_{i=0}^{|\mathcal{I}|+1} [\mathbb{1}_{\mathcal{A}}(i)e_A + \mathbb{1}_{\mathcal{O}}(i)e_O]} \right\}, \quad (4.2)$$

where e_A and e_O are respectively the node energy consumption when active or idle, whose selection depends on an indicator function which discretises among

them according to the event the index i belongs to the set of active \mathcal{A} or idled \mathcal{O} nodes. Moreover, \mathcal{I} and $\bar{\mathcal{I}}$ indicate the interfering sets of nodes within and outside the CSA respectively, with the SINR threshold $\tilde{\rho} = 2^{\frac{\beta}{B}} - 1$ and where the set of edge nodes $|\mathcal{I}| + 1$ includes the designated transmitter within the CSA in charge of delivering the message to the typical user. Unfortunately, the random ratio (4.2) does not permit a closed-form for the node density λ^S and does not allow to explore the properties associated to the MCP. However, in the ultra-dense context and during peaks of traffic time, we can consider all the nodes to be active and contributing to the interfering pattern, hence consuming e_A [W]. Our analytical results are therefore intended to depict a situation of high network activity. With these assumptions, the proposed performance metric can be regarded as a lower-bound of (4.2). We define the EE as

$$\begin{aligned}
 \mathcal{G}(\lambda^S) &= \mathbb{E} \left[\rho \frac{\text{SCDP}}{\sum_{|\mathcal{I}|+1} e_A} \right] \left[\frac{\text{Mbps}}{\text{Watt}} \right] \\
 &= \rho \omega \mathbb{E} \left[\frac{\Pr \left(\frac{P^S |h|^2 \bar{r}^{-\alpha}}{\sum_{i=0}^{|\mathcal{I}|} P^S |h_i|^2 r_i^{-\alpha} + \sum_{i=0}^{|\bar{\mathcal{I}}|} P^S |h_i|^2 r_i^{-\alpha} + \sum_{m=0}^{|\mathcal{I}^M|} P^M |h_m|^2 r_m^{-\alpha} + W} > \tilde{\rho} \right)}{\sum_{i=0}^{|\mathcal{I}|+1} e_A} \right] \left[\frac{\text{Mbps}}{\text{Watt}} \right] \\
 &\leq \rho \omega \mathbb{E} \left[\frac{\Pr \left(\frac{P^S |h|^2 \bar{r}^{-\alpha}}{\sum_{i=0}^{|\mathcal{I}|} P^S |h_i|^2 r_i^{-\alpha} + \sum_{i=0}^{|\bar{\mathcal{I}}|} P^S |h_i|^2 r_i^{-\alpha} + \sum_{m=0}^{|\mathcal{I}^M|} P^M |h_m|^2 r_m^{-\alpha} + W} > \tilde{\rho} \right)}{\sum_{i=0}^{|\mathcal{I}|+1} [\mathbb{1}_{\mathcal{A}}(i) e_A + \mathbb{1}_{\mathcal{O}}(i) e_{\mathcal{O}}]} \right] \left[\frac{\text{Mbps}}{\text{Watt}} \right], \tag{4.3}
 \end{aligned}$$

with the scaling factor $\omega = \left(1 - e^{-\lambda^S \pi R^2}\right)$ which indicates the probability at least one node exists within the CSA and is a direct application of the void probability for ϕ^S . In this work, the focus is on a single network tier and the contribution of an upper MBS tier is considered as interference to the SBS transmission. In particular, the energy expenses from a ubiquitous MBS tier have to be properly addressed to allow a fair investigation over the denser lower tiers. Secondly, the performance metric in (4.3) stands as a simpler handle on the problem. The proposed metric \mathcal{G} can be interpreted as an improved measure of the efficiency of the network. It can easily be seen that $\mathcal{G} \in [0, \rho/e_A]$ [Mbps/Watt] achieves the minimum when

the SCDP reaches zero and attains its maximum when both the SCDP stands at its maximum value and when one single SBS is actively transmitting to the typical user within the CSA. The maximum achievable value by \mathcal{G} leads to the conclusion that the upper limit of this performance metric is determined by the energy consumption of the tier's single node, *i.e.* e_A [W], considering the users' target bit-rate fixed. According to (4.1), the choice of the type of nodes that compose the tier makes the upper-bound to vary. Specifically, lower values of efficiency can be possibly attained when more powerful BS types are considered following the same intuition from [75].

The operating density at which \mathcal{G} is maximized is defined as the MCP of the network, *i.e.*, $\lambda^* \triangleq \operatorname{argmax} \mathcal{G}(\lambda^S)$. The convenience in considering \mathcal{G} over the traditional SCDP metric lies in the advantage of knowing the most efficient tier's operating point. More insights over the most efficient and most performing states of the network's tier will be lately discussed in Section 4.3.2.

The proposed metric in (4.3) can be averaged out from its dependencies, using the steps being discussed in Appendix B.1, and is given by

$$\mathcal{G}(\lambda^S) = \frac{2\pi\rho}{e_A} \left(1 - e^{-\lambda^S \pi R^2}\right) \int_1^R \bar{r} \mathcal{L}_{\mathcal{I}^M} \left(\frac{P^M \tilde{\rho} \bar{r}^\alpha}{P^S}, \bar{r} \right) e^{-\lambda^S (\pi R^2 + \xi_2(\bar{r}))} \frac{e^{\lambda^S \xi_1(\bar{r})} - 1}{\xi_1(\bar{r})} e^{-\frac{\tilde{\rho} W \bar{r}^\alpha}{P^S}} d\bar{r}, \quad (4.4)$$

where $\xi_1(\bar{r}) = 2\pi \int_{\bar{r}}^R \frac{r}{\tilde{\rho} r^{-\alpha} + 1} dr$ and $\xi_2(\bar{r}) = 2\pi \int_R^\infty \left[1 - \frac{1}{1 + \tilde{\rho} r^{-\alpha}}\right] r dr$ and with the Laplace functional for the femto-tier interfering PPP in (4.4) expressed by means of its components to ease future analysis.

Following the Leibniz integral rule¹, the derivative of \mathcal{G} with respect to λ^S stands as

$$\frac{\partial \mathcal{G}}{\partial \lambda^S}(\lambda^S) = \frac{2\pi\rho}{e_A} \int_1^R \bar{r} \frac{e^{-\frac{\tilde{\rho} W \bar{r}^\alpha}{P^S}}}{\xi_1(\bar{r})} \mathcal{L}_{\mathcal{I}^M} \left(\frac{P^M \tilde{\rho} \bar{r}^\alpha}{P^S} \right) e^{-\lambda^S (\pi R^2 + \xi_2(\bar{r}))} \times \left[\begin{aligned} & \left((\pi R^2 + \xi_2(\bar{r})) - e^{\xi_1(\bar{r}) \lambda^S} (\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})) \right) - \\ & e^{-\lambda^S \pi R^2} \left((2\pi R^2 + \xi_2(\bar{r})) - e^{\xi_1(\bar{r}) \lambda^S} (2\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})) \right) \end{aligned} \right] d\bar{r}. \quad (4.5)$$

¹Leibniz rule for differentiation under the integral sign states that, given constant integration bounds, the following special case holds $\frac{\partial}{\partial x} \int_a^b f(x, t) dt = \int_a^b \frac{\partial}{\partial x} f(x, t) dt$ given that both $f(x, t)$ and its partial derivative are continuous in t .

Some terms of the integrand function can be re-arranged as

$$\begin{aligned} \pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r}) &= \pi R^2 + 2\pi \int_R^\infty \left[1 - \frac{1}{1 + \tilde{\rho}\bar{r}^\alpha \hat{r}^{-\alpha}} \right] \hat{r} d\hat{r} - 2\pi \int_{\bar{r}}^R \frac{\hat{r}}{1 + \rho\bar{r}^{-\alpha} \hat{r}^{-\alpha}} d\hat{r} \\ &= \pi \bar{r}^2 + 2\pi \int_{\bar{r}}^\infty \left[1 - \frac{1}{1 + \tilde{\rho}\bar{r}^\alpha \hat{r}^{-\alpha}} \right] \hat{r} d\hat{r} > 0, \end{aligned} \quad (4.6)$$

from which the inequality between the scaling terms $\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})$ and $\pi R^2 + \xi_2(\bar{r})$ stands

$$\pi R^2 + 2\pi \int_R^\infty \left[1 - \frac{1}{1 + \tilde{\rho}\bar{r}^\alpha \hat{r}^{-\alpha}} \right] \hat{r} d\hat{r} \geq \pi \bar{r}^2 + 2\pi \int_{\bar{r}}^\infty \left[1 - \frac{1}{1 + \tilde{\rho}\bar{r}^\alpha \hat{r}^{-\alpha}} \right] \hat{r} d\hat{r}, \quad (4.7)$$

where the equality holds when $\bar{r} = R$. The same conclusions can be drawn for the scaling terms terms in (4.5), with the resulting inequality $2\pi R^2 + \xi_2(\bar{r}) \geq 2\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})$. It is important to find where and how many stationary points characterise \mathcal{G} . The following Theorem 8 stands as an analytical result for the existence of a unique stationary point for \mathcal{G} within $\lambda^S \in (0, \infty)$, thus confirming the solution λ^* , such that $\frac{\partial \mathcal{G}}{\partial \lambda^S}(\lambda^*) = 0$ is the global maximum. The detailed proof of Theorem 8 is given in Appendix B.2.

Theorem 8. *The performance metric $\mathcal{G}(\lambda^S)$ has three distinct stationary points: (i) a zero solution such that $\lambda^S = 0$, (ii) an infinite solution when $\lambda^S \rightarrow \infty$, and third solution (iii), which also corresponds to the network tier's MCP, within the range $\lambda^* \in [\lambda_{\Psi_1}^-, \infty)$, where*

$$\lambda_{\Psi_1}^- = \frac{\ln \left(\frac{\pi R^2 + \xi_2(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})} \right)}{\xi_1(\bar{r})} \Bigg|_{\bar{r}=R^-}.$$

with $R^- = R - \epsilon$ for $\epsilon > 0$ an arbitrary small quantity such that $\lambda_{\Psi_1}^-$ exists.

The analysis conducted in Theorem 8 has some important implications over the results obtained in this chapter. From (B.3), it can be observed that the intersection of the indicated functions $\Psi_1(\lambda^S)$ and $\Psi_2(\lambda^S)$ does not depend on the energy expenses of the network tier's node e_A . This implies that the MCP computed for different network tiers changes as a function of (i) the transmitting

Algorithm 2 Bisection search for MCP

```

1: initialize  $u \leftarrow \frac{\partial \mathcal{G}}{\partial \lambda^S}(0)$ 
2:  $\delta \leftarrow \lambda^+$ 
3:  $\lambda^* \leftarrow \frac{\lambda^- + \lambda^+}{2}$ 
4: while  $\delta > \lambda^+ / 10^5$  do
5:    $u \leftarrow \frac{\partial \mathcal{G}}{\partial \lambda^S}(\lambda^*)$ 
6:   if  $u < 0$  then
7:      $\lambda^* = \lambda^* - \delta$ 
8:   else
9:      $\lambda^* = \lambda^* + \delta$ 
10:  end if
11:   $\delta \rightarrow \delta / 2$ 
12: end while

```

power of the network tier under investigation, (ii) the transmitting power of nodes from different tiers and (iii) the interference produced at different tiers of the HetNet. Having established the uniqueness of $\lambda^* \triangleq \operatorname{argmax} \mathcal{G}(\lambda^S)$, the subsequent focus is on obtaining this value.

4.3.1 The Employed Numerical Solution

Due to the absence of a closed form for \mathcal{G} with respect to \bar{r} , the MCP can only be obtained by means of a numerical search. A bisection search, whose pseudo-code is indicated in Algorithm 2, is employed to find an arbitrary close approximation of the MCP. The proposed bisection search is meant to find a solution to the maximisation of the proposed utility function

$$\begin{aligned}
 (\mathbb{P}_0) : \quad & \lambda^* \triangleq \operatorname{argmax} \quad \mathcal{G}(\lambda^S) \\
 & \text{subject to} \quad 0 \leq \lambda^S,
 \end{aligned} \tag{4.8}$$

and to the SCDP maximisation

$$\begin{aligned}
 (\mathbb{P}_1) : \quad & \bar{\lambda}^* \triangleq \operatorname{argmax} \quad \mathcal{G}^{\text{SCDP}}(\lambda^S) \\
 & \text{subject to} \quad 0 \leq \lambda^S,
 \end{aligned} \tag{4.9}$$

where

$$\begin{aligned} \mathcal{G}^{\text{SCDP}}(\lambda^{\text{S}}) &= \omega \mathbb{E} \left[e^{-\frac{\hat{p}}{PS\bar{r}^{-\alpha}} (\sum_{i=0}^{|\mathcal{I}|} P^{\text{S}} |h_i|^2 r_i^{-\alpha} + \sum_{i=0}^{|\bar{\mathcal{I}}|} P^{\text{S}} |h_i|^2 r_i^{-\alpha} + \sum_{m=0}^{|\mathcal{I}^{\text{M}}|} P^{\text{M}} |h_m|^2 r_m^{-\alpha} + W)} \right] \\ &= 2\pi\omega \int_1^R \bar{r} \lambda^{\text{S}} e^{-\lambda^{\text{S}} (\pi\bar{r}^2 + \xi_3(\bar{r}))} \mathcal{L}_{\mathcal{I}^{\text{M}}} \left(\frac{P^{\text{M}} \tilde{\rho} \bar{r}^{-\alpha}}{PS}, \bar{r} \right) e^{-\frac{\hat{p}W\bar{r}^{-\alpha}}{PS}} d\bar{r} \end{aligned} \quad (4.10)$$

is the pure SCDP analytical form whose derivation can be found in Appendix B.3 and with $\omega = \left(1 - e^{-\lambda^{\text{S}} \pi R^2}\right)$ standing as the void probability in the CSA applied to the SBSs' PPP. The indicated quantities λ^- and λ^+ in Algorithm 2 are directly obtained from $\lambda_{\Psi_1}^-$ and $\lambda_{\Psi_2}^+$ (in Appendix B.2) for (\mathbb{P}_0) and from $\lambda_{\Psi_2^{\text{SCDP}}}^-$ and $\lambda_{\Psi_2^{\text{SCDP}}}^+$ (in Appendix B.4) for (\mathbb{P}_1) . It can be demonstrated that the same outline from Theorem 8 can be applied to $\mathcal{G}^{\text{SCDP}}$ to draw the same conclusions over the uniqueness of its solution $\bar{\lambda}^*$, with the detailed derivation provided in Appendix B.4. From the application of Theorem 8 to $\mathcal{G}^{\text{SCDP}}(\lambda^{\text{S}})$, it can be obtained that $\lambda_{\Psi_2^{\text{SCDP}}}^+ < \bar{\lambda}^* < \infty$, with

$$\lambda_{\Psi_2^{\text{SCDP}}}^+ = \ln \left(\frac{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})}{\pi \bar{r}^2 + \xi_3(\bar{r})} \right) \frac{1}{\pi R^2} \Big|_{\bar{r}=1}.$$

This element is necessary to prove the lower-bound between the sub-optimal solutions $\lambda^* < \bar{\lambda}^*$, which will be demonstrated in the next section.

4.3.2 Network Energy Dualism

Having proved that a unique density solution exists which maximises the EE performance of a network tier, this section will analytically demonstrate that the optimal λ^* stands as a lower-bound for $\bar{\lambda}^*$. The existing lower-bound gives rise to a gap $\Delta = \bar{\lambda}^* - \lambda^*$ whose interpretation will be discussed as well. The bounding is proved in Theorem 9 and its derivation discussed in detail in Appendix B.5

Theorem 9. *The density operating point at which a mobile network reaches its optimal energy-to-SCDP conversion is always a lower-bound of the density at which the maximum SCDP is experienced, i.e. $\lambda^* < \bar{\lambda}^*$.*

The gap between the two measures is referred as $\Delta\lambda = \bar{\lambda}^* - \lambda^*$ and it can

be interpreted as a measure of the density distance between the most performing and most efficient operating points of the network. This leads to the conclusion that by considering a generic density $\lambda^* \leq \lambda^S \leq \bar{\lambda}^*$, the operator can choose between a more efficient or better performing network state by oscillating within the indicated range of values. The cost for an operator when moving from λ^S to $\bar{\lambda}^*$ is in terms of energy expenses, which directly yields from the average number of nodes of the PPP defined by the intensity function λ^S as $\mathbb{E} [|\phi(\lambda^S)|] = \lambda^S \pi R^2$. Hence, the network energy expenses are linearly increasing with λ^S , as the average number of pieces of equipment linearly depends on the displacement density. To an increase of $\Delta\lambda$, more expense is needed to transition from the most efficient to the most performing density operating point of the network. On the other side, from the user's point of view, a lower density of network nodes comes at the cost of a higher latency time.

4.3.3 Extension to Cache-Enabled HetNet

The presented model can easily be extended to a cache-enabled HetNet. To do so, the density λ^S can be substituted by the scaled term $\lambda^S \leftarrow p_f \lambda^S$ with p_f being the probability of caching the f -th content. Theorem 8 is still valid when the MCP is searched for a cache-enabled extended network and the performance metric is expressed as

$$\mathcal{G}(\lambda^S) = \sum_f^F \hat{p}_f \frac{2\pi\rho}{e_A} \omega \int_1^R \bar{r} \mathcal{L}_{TM} \left(\frac{P^M \tilde{\rho} \bar{r}^\alpha}{PS} \right) e^{-p_f \lambda^S (\pi R^2 + \xi_2(\bar{r}))} \frac{e^{p_f \lambda^S \xi_1(\bar{r})} - 1}{\xi_1(\bar{r})} e^{-\frac{\bar{\rho} W \bar{r}^\alpha}{P^S}} d\bar{r},$$

where $\omega = \left(1 - e^{-p_f \lambda^S \pi R^2}\right)$.

4.4 Numerical Results and Analysis

The obtained performance and energy consumptions resulting from (\mathbb{P}_0) and (\mathbb{P}_1) are illustrated and commented in this section, against the target bit-rate ρ . In this way, it is intended to highlight the distance of the investigated network tier operating points as the content transmitting conditions become more demanding.

Variable	Value	Description
W	-174 [dBm]	Noise thermal power
α	3	Path-loss exponent
B	30 [MHz]	Bandwidth
λ^M	10^{-4} [unit/ m^2]	Density MBS tier
P^S	17 [dBm]	SBS transmitting power
P^M	46 [dBm]	MBS transmitting power
P_{RX}	0.3 [W]	Consumed power radio-frequency transceiver
P_{BB}	2.5 [W]	Consumed power baseband interface
η_{PA}	.044	Transceiver power efficiency
σ_{DC}	.9	Loss factor DC-DC power supply
σ_{cool}	0	Loss factor from cooling
σ_{MS}	.11	Loss factor from main supply

Table 4.1: The network and femto-cell energy consumption parameters.

The energy consumption for a femto-cell network tier node, following the power consumption model in (4.1), is considered in Table 4.1 together with the baseline network parameters. The following discussion will focus on the sub-optimal values for λ^* and $\bar{\lambda}^*$, the time-latency for a content chunk being correctly transmitted, the energy consumption over the CSA and the comparison of the proposed upgraded EE measure with the benchmark results obtained from the EE definition employed in [73–77].

4.4.1 Sub-optimal Network Tier Densities Vs. ρ

In Fig. 4.4, the optimal densities are shown for (\mathbb{P}_0) and (\mathbb{P}_1) , together with the associated SCDP values. As expected from the analysis conducted in Section 4.3.2, the values obtained for $\bar{\lambda}^*$ give an upper-bound of those obtained for λ^* . Similarly, the obtained SCDP values from (4.10), computed for the sub-optimal densities λ^* and $\bar{\lambda}^*$, show the gap in QoS perceived at the typical user. The maximum value for $\Delta\lambda$ has been detected at $\rho = 9$ [Mbps], with a progressive decrease as ρ is incremented. From an overall observation on the domain of ρ , it can be concluded that the values of $\bar{\lambda}^*$ are much larger than λ^* . This indicates that enormous energy expenses are behind the QoS increase from $\text{SCDP}(\lambda^*)$ to $\text{SCDP}(\bar{\lambda}^*)$. As more demanding target bit-rate values are required, the network's

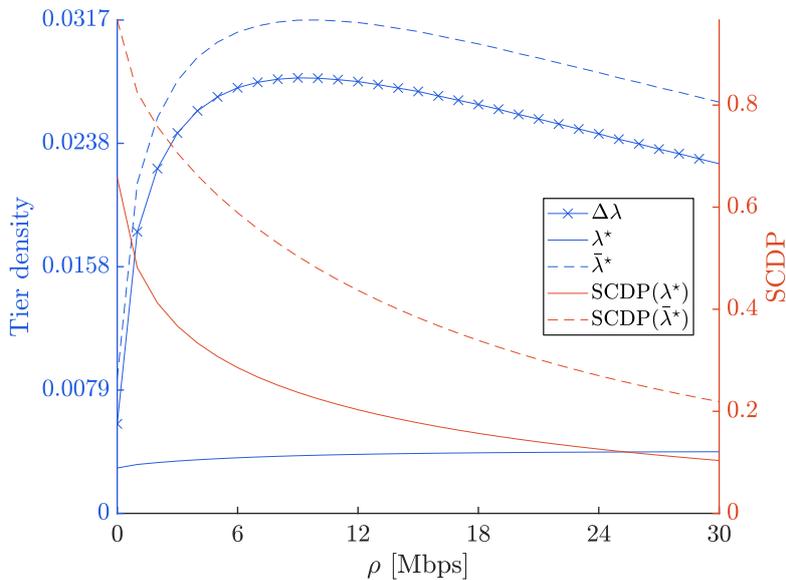


Figure 4.4: Sub-optimal network tiers' densities and associated SCDP values.

choice for $\bar{\lambda}^*$ is to decrease the network tier density in order to regulate the level of interference, when λ^* continues to slightly increase as ρ is incremented. Therefore, it can be noted that $\Delta\lambda$ decreases as more demanding conditions are required for the transmission.

4.4.2 Energy Expenses Measures

The energy expenses within the CSA are discussed in this section. The numerical results are obtained as a direct application of the average size of a PPP, defined by $\mathbb{E}[|\phi(\lambda\pi R^2)|] = \lambda\pi R^2$. As a consequence, it can be noted in Fig. 4.5 that the energy expenses within the CSA follow the values of sub-optimal densities discussed in Section 4.4.1. From this analysis, it can be seen that massive energy costs are associated to even a minor SCDP increase and thus the importance of the knowledge of the most efficient network's tier operating point can be employed to strike the optimal efficiency/performance balance by the operator. This confirms the validity of the proposed metric to obtain sufficient knowledge of the network to allow the operators to make use of the best tools to run their networks.

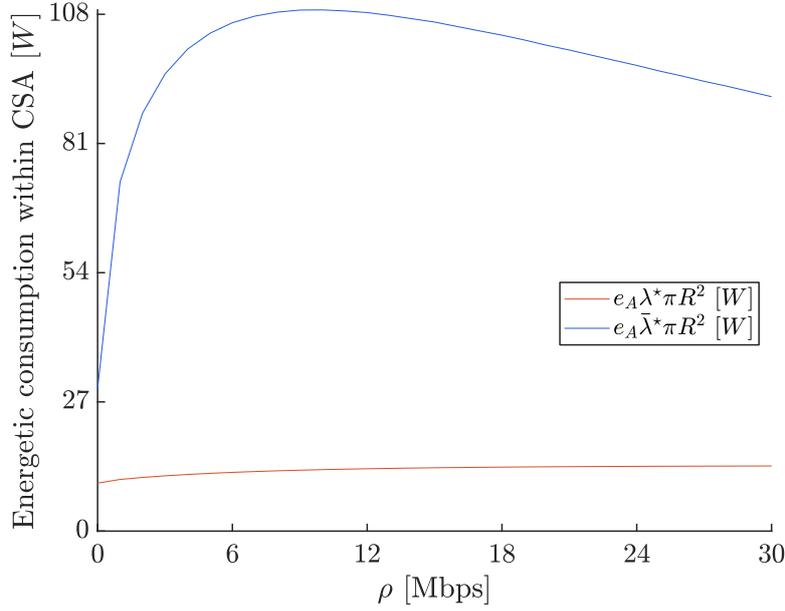


Figure 4.5: Energy expenses for λ^* and $\bar{\lambda}^*$ within the CSA.

4.4.3 Latency

The averaged values of the random content elapsed transmission time are given in Fig. 4.7. The routine from Algorithm 3 is employed to obtain the random value of latency-time, such that the average can be computed over 10^5 instances. A content request is initiated and a uniform distributed random variable $u \in [0, 1]$ is generated to be compared with the corresponding obtained SCDP value from Fig. 4.4. A transmission is considered to be successful if $u > \text{SCDP}$. Differently, the same content is re-transmitted, with the associated latency time increase. Inspired by the indicated 3GPP LTE sub-frame design [59], the latency time is increased by a 10 [msec] factor each time the transmission is unsuccessful. The obtained averaged values highlight higher latency time for the maximization over the EE, as expected. This is a direct consequence of the lower SCDP values. This gap affects the users' QoS and the given values provide a numerical feedback on the detrimental effects which derives from operating the most efficient state of a network's tier.

Algorithm 3 Calculation process for average content latency-time

- 1: initialize $u \leftarrow 0$
 - 2: initialize content latency-time $T \leftarrow 10$ [msec]
 - 3: **while** $u < \text{SCDP}(\lambda)$ **do**
 - 4: generate $u \sim \mathcal{U} \in [0, 1]$
 - 5: **if** $u < \text{SCDP}(\lambda)$ **then**
 - 6: $T \leftarrow T + 10$ [msec]
 - 7: **end if**
 - 8: **end while**
-

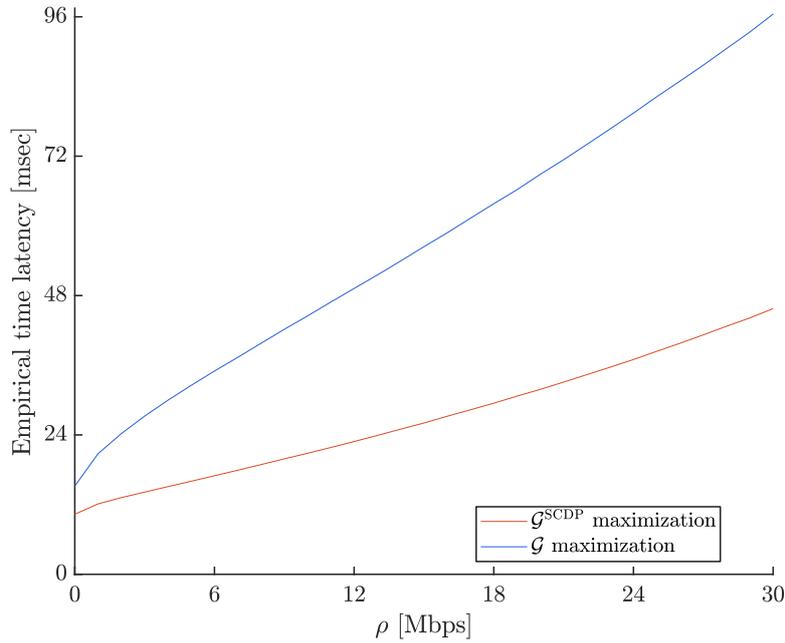


Figure 4.6: Averaged latency time for λ^* and $\bar{\lambda}^*$.

4.4.4 Comparison among EE Approaches

In this section, the numerical evaluation of the EE, *i.e.* $\mathcal{G}(\lambda^S)$, has been reported for the obtained values of λ^* , $\bar{\lambda}^*$ and the results from the maximisation of the network EE, as employed in [73–77], are compared as the baseline. The sub-optimal benchmark density values $\hat{\lambda}^*$ are obtained by applying the bisection search in Algorithm 2 to maximise the network EE defined as the ratio of the average

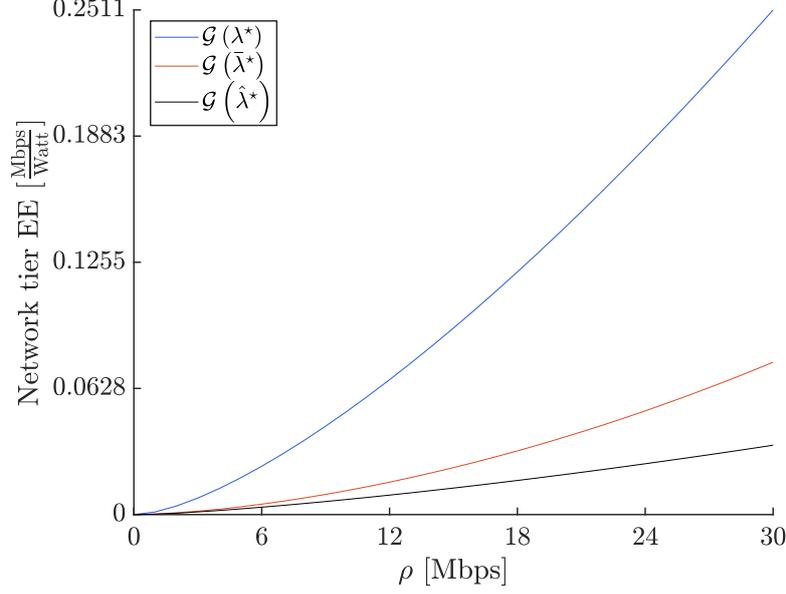


Figure 4.7: Comparison among network tier's EE.

SCDP over the average network consumption as

$$\hat{\lambda}^* \triangleq \operatorname{argmax}_{\lambda^S} \rho \frac{\mathbb{E} [\mathcal{G}^{\text{SCDP}}(\lambda^S)]}{\mathbb{E} [\sum_{|I|} e_A]}, \quad \text{subject to } 0 \leq \lambda^S,$$

where

$$\frac{\mathbb{E} [\mathcal{G}^{\text{SCDP}}(\lambda^S)]}{\mathbb{E} [\sum_{|I|} e_A]} = \frac{2\pi\omega \int_1^R \bar{r} \lambda^S e^{-\lambda^S(\pi\bar{r}^2 + \xi_3(\bar{r}))} \mathcal{L}_{\mathcal{I}^M} \left(\frac{P^M \bar{\rho} \bar{r}^\alpha}{P^S}, \bar{r} \right) e^{-\frac{\bar{\rho} W \bar{r}^\alpha}{P^S}} d\bar{r}}{e_A \lambda^S \pi R^2}.$$

As expected, the values of $\mathcal{G}(\lambda^*)$ stand on higher values than $\mathcal{G}(\bar{\lambda}^*)$. This indicates a remarkable difference in efficiency when considering the proposed metric, which increases as more demanding transmitting conditions are investigated. It can be noted that the benchmark $\mathcal{G}(\hat{\lambda}^*)$ obtains the lowest value of EE, validating the importance of considering the proposed metric.

4.5 Conclusions

A novel definition of network EE is derived in this chapter as the expected value of the random ratio of the SCDP over the CSA energy consumptions. It is demonstrated a closed-form of the proposed metric for the random number of nodes contributing to both energy expenses and interfering pattern is accessible. This in turns has allowed to deepen our understanding of the network tier EE and we found the MCP of a network tier is always upper-bounded by the optimal density which maximizes the SCDP. The examination of the EE is conducted against the results of the pure SCDP metric and a network dualism between performance and efficiency is found and discussed.

Chapter 5

On LOS Contribution to UDN

5.1 Introduction and Problem Statement

Spectrum release of higher frequency bands for 5G means that direct LOS wireless links are exploited for the best network capacity [79]. As network densification is considered as the most promising strategy to meet the 1000-fold increase in traffic demand, LOS wireless links are expected to occur more frequently. A wide set of 5G technologies depends on the network ubiquity of its components. In UDNs, LOS links are more likely to occur, and more power can potentially be released as both desired and interfering power. While a substantial amount of research has been conducted under Rayleigh distributed channel conditions (with no LOS), it becomes more important to investigate approximations which can analytically describe the LOS contribution. In previous chapters, we modeled channel fading as Rayleigh distributed. In this chapter, the aim is to incorporate the LOS contribution into the performance of a cache-enabled UDN.

5.1.1 Literature Review

While it is well understood that LOS channel paths improve performance, it is unfortunately extremely difficult to account for their contributions to the network performance [80]. In [81], the authors considered an exponential series approximation to model the non-central Chi-squared distribution which derives from LOS Rician faded channels. However, the approximation only holds for Rician faded

channel power of the desired signal and Rayleigh faded signals from the interferers, and it fails to account for the interferers which also have LOS links. Furthermore, [82] employed a distance-dependent probabilistic rule to distinguish between LOS and NLOS communication channels, and a more suitable approximation of the interference term was proposed by approximating the resulting non-homogeneous PPPs from the application of the distance-dependent LOS model with an homogeneous counterpart. The limitation of this result is nevertheless that the approximation is complex and unable to be used to gain insight for network optimization. Moreover, JT or any form of cooperative transmission from SBSs which has become an essential feature for UDNs [60] was not considered. Technically speaking, the mathematical challenge to account for Rician LOS channel paths is to handle the Bessel function of the first kind in the channel power distribution [83]. There is a pressing need to analytically quantify network metrics such as SCDP for UDNs with a mixture of LOS and NLOS links that could help optimise network parameters such as content caching probability, SBS spatial density, and etc. [84], which motivates our work.

In this chapter, our objective is to consider a more realistic scenario where both the desirable links and interference links can come from LOS and NLOS paths randomly. We focus on analysing the SCDP performance of the UDN using JT in this scenario. Our contribution is an approximation for the SCDP for the mixed LOS/NLOS scenarios by separating the contributions of a full NLOS network and that from an LOS network. The proposed approximation can be interpreted as a correction factor for previous results which only account for Rayleigh distributed small-scale channel fading. By means of closed-form expressions, the access to some key network parameters is uncovered and the relations among those variables that mostly affect the LOS components are revealed. Our model permits the use of any arbitrary LOS probability function, and the simulation results show the tightness of the proposed approximation over a wide set of LOS probability functions. Hence, the novelty in this chapter can be regarded as (i) a suitable approximated scaling factor to extend the results of full-NLOS networks, (ii) the random nature of this scaling factor is averaged out in closed-form for its main key dependencies and (iii) the gaps with the empirical target SCDP are finally shown and discussed.

The remainder of this chapter is organised as follows. Section 5.2 describes

the network design and content delivery scheme. In Section 5.3 the proposed approximation is elaborated and the network trade-offs are discussed. Numerical validations of the proposed approach are reported in Section 5.4 and Section 5.5 concludes this chapter.

5.2 Network Model and Content Delivery

A downlink two tiers HetNet in which a layer of densely populated SBSs is placed on top of a tier of MBSs serving the UEs is considered. Independent and homogeneous PPPs are employed to model the multi-tiers network nodes, considered over the network space \mathbb{R}^2 . These pieces of equipment are supposed to respond to the requests of the UEs, spatially modeled as an independent homogeneous PPP with homogeneous intensity function over the network space \mathbb{R}^2 . In this chapter, the concept of typical user is considered. Also, the CSA is defined as an Euclidean ball $\mathcal{B}(0, R)$, centered at the origin of the Euclidean space and of radius R . The adopted user-node association scheme works as follows: the content request is firstly passed from the typical user to its CSA; if a hit-cache is experienced, the content can be directly transmitted from the SBSs at the network edge; if the request experiences a missed-cache the closest MBS is in charge of the content delivery. In this case, for simplicity, we consider the MBS to be able to meet all the possible content requests, with no extra costs in terms of latency and power consumption. In case multiple hit-caches are experienced, this means that several SBSs have the content to serve the request and a non-coherent JT scheme over those SBSs is performed to transmit the same required content over the same portion of bandwidth [72, 85]. It is reasonable to impose a practical limit on the maximum number of cooperating SBS nodes for JT, say K_f^{\max} for transmission of content f .

To perform the content transmission of a generic f -th content over the same portion of bandwidth among network's nodes, the available bandwidth for the transmission is split by the user-load $\Xi_f^S = \max\{\xi_1, \dots, \xi_{K_f}\}$, where ξ_k stands as the user-load experienced by the k -th SBS from the set of $K_f \leq K_f^{\max}$ cooperating nodes. Similarly, Ξ_f^M indicates the user-load of the closest MBS. At the typical

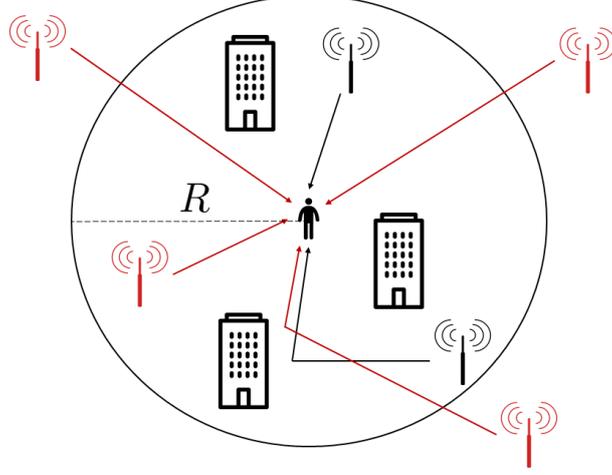


Figure 5.1: Visual representation of the network model, with interfering power indicated in red and desired power from the cooperating set K_f in black.

user, the averaged target SCDP over the whole wordbook is given by

$$\text{SCDP} = \sum_{f=1}^F \hat{p}_f \left(a_f \underbrace{\Pr \left(\gamma_f^S > \bar{\rho}_f^S \right)}_{\Gamma_f^S} + (1 - a_f) \underbrace{\Pr \left(\gamma^M > \bar{\rho}_f^M \right)}_{\Gamma_f^M} \right), \quad (5.1)$$

where $a_f = (1 - e^{-\lambda^S p_f \pi R^2})$ comes as a direct application of the void probability such that there is at least one SBS that has cached the required content within $\mathcal{B}(0, R)$, with Γ_f^S and Γ_f^M respectively standing as the SCDP when a hit-cache or a missed-cache is experienced, and finally with $\bar{\rho}_f^S = 2^{\frac{\alpha}{B}} \Xi_f^S - 1$ and $\bar{\rho}_f^M = 2^{\frac{\alpha}{B}} \Xi_f^M - 1$ which stand as the SINR thresholds. Indicating the NLOS and LOS channel fading respectively as nL and L, and using in and out to refer to the nodes within and outside the Euclidean ball $\mathcal{B}(0, R)$, we can write the SINR from the cooperating

SBS nodes as

$$\gamma_f^S = \frac{D_f^S}{\left(I_f^{S,\text{in},\text{nL}} + I_f^{S,\text{in},\text{L}} + I^{S,\text{out},\text{nL}} + I^{S,\text{out},\text{L}} + I^{\text{M},\text{nL}} + I^{\text{M},\text{L}} + W \right)} = \frac{\left| \sum_{k \in \phi_f^{S,\text{nL}} \cap \mathcal{B}(0,R)} h_k r_k^{-\alpha/2} + \sum_{k \in \phi_f^{S,\text{L}} \cap \mathcal{B}(0,R)} h'_k r_k^{-\alpha/2} \right|^2}{\sum_{i \in \phi_{-f}^{S,\text{nL}} \cap \mathcal{B}(0,R)} |h_i|^2 r_i^{-\alpha} + \sum_{i \in \phi_{-f}^{S,\text{L}} \cap \mathcal{B}(0,R)} |h'_i|^2 r_i^{-\alpha} + \sum_{i \in \phi^{S,\text{nL}} \cap \mathcal{B}(R,\infty)} |h_i|^2 r_i^{-\alpha} + \sum_{i \in \phi^{S,\text{L}} \cap \mathcal{B}(R,\infty)} |h'_i|^2 r_i^{-\alpha} + \sum_{i \in \phi^{\text{M},\text{nL}}} |h_i|^2 r_i^{-\alpha} + \sum_{i \in \phi^{\text{M},\text{L}}} |h'_i|^2 r_i^{-\alpha} + W} \quad (5.2)$$

and the SINR from the closest MBS as

$$\gamma^{\text{M}} = \frac{D^{\text{M}}}{I^{\text{S},\text{nL}} + I^{\text{S},\text{L}} + I^{\text{M},\text{nL}} + I^{\text{M},\text{L}} + W}$$

$$= \begin{cases} \frac{|h_k|^2 r_k^{-\alpha}}{\sum_{i \in \phi^{\text{S},\text{nL}}} |h_i|^2 r_i^{-\alpha} + \sum_{i \in \phi^{\text{S},\text{L}}} |h'_i|^2 r_i^{-\alpha} + \sum_{i \in \phi^{\text{M},\text{nL}} \setminus \mathcal{B}(\bar{r},\infty)} |h_i|^2 r_i^{-\alpha} + \sum_{i \in \phi^{\text{M},\text{L}} \setminus \mathcal{B}(\bar{r},\infty)} |h'_i|^2 r_i^{-\alpha} + W}, & \text{if NLOS} \quad (5.3a) \\ \frac{|h'_k|^2 r_k^{-\alpha}}{\sum_{i \in \phi^{\text{S},\text{nL}}} |h_i|^2 r_i^{-\alpha} + \sum_{i \in \phi^{\text{S},\text{L}}} |h'_i|^2 r_i^{-\alpha} + \sum_{i \in \phi^{\text{M},\text{nL}} \setminus \mathcal{B}(\bar{r},\infty)} |h_i|^2 r_i^{-\alpha} + \sum_{i \in \phi^{\text{M},\text{L}} \setminus \mathcal{B}(\bar{r},\infty)} |h'_i|^2 r_i^{-\alpha} + W}, & \text{if LOS} \quad (5.3b) \end{cases}$$

where the coefficients h and h' indicate, respectively, the Rayleigh and Rician fading channel coefficients. In (5.3a) and (5.3b), \bar{r} indicates the distance of the closest MBS from the typical user and the interfering pattern is determined by those SBSs that have at least one mobile users within and by all the MBSs in the circular sector $\mathcal{B}(\bar{r}, \infty)$. In (5.2), the interfering pattern is determined by the union of the set of SBS edge nodes with at least one users within their CSA, the set of those edge nodes that experience a missed-cache within $\mathcal{B}(0, R)$ and from the MBS tier as if they are all active. Apparently, when a number of cooperating edge nodes exceeds K_f^{max} , the two interfering regions are determined by $\bar{r} \leq R$, with \bar{r} denoting the distance of the farthest cooperating SBS edge node.

The set of nodes $\phi_{-f}^{S,[\text{L},\text{nL}]}$ in (5.2) refers respectively to those SBSs with a LOS and NLOS link where the f -th content has not been cached and that are associated to at least one UEs. For ease of subsequent analysis, we refer to $K_f^{S,\text{L}}$ and $K_f^{S,\text{nL}}$, respectively, as the random number of cooperating nodes whose channel links

experience LOS and NLOS small-scale channel fading, namely, the snapshots of the two PPPs $\phi_f^{S,L} \cap \mathcal{B}(0, R)$ and $\phi_f^{S,nL} \cap \mathcal{B}(0, R)$, after having applied a LOS thinning function to the initial ϕ^S process. Following the indications from Section 2.3.1, the mean and standard deviation of the modeling Gaussian random variables for the channel coefficients are respectively referred as $\mu_h = (k/(k+1))^{1/2}$ and $\sigma_h = (2(k+1))^{-1/2}$.

5.2.1 The LOS Thinning Function

A space dependent thinning transformation is applied to ϕ^S and ϕ^M , in order to model the spatial displacement of LOS communications. As a result, the modeling non-homogeneous PPP for LOS communications $\phi^{S,L}$ has an intensity function $p(r)\lambda^S$, with r being the node-to-user link-distance. On the other hand, the intensity function of the point process for those nodes with NLOS is $(1 - p(r))\lambda^S$.

The independent thinning rule $p(r)$ is strictly correlated to the surrounding environment. We therefore expect this function to dramatically change according to the city's architecture, as indicated by the 3GPP standard [39]. When empirical measures are instead accounted, scaling issues are inevitably associated. For what concerns the following analysis, the actual shape of the thinning function is irrelevant as the proposed approach applies to all possible shape and forms of the thinning function. However, some observations over the hypothetical nature of this function can be taken, to have a tool to arbitrarily control the signal shadowing. The thinning rule employed in our model is

$$p(r) = e^{-\frac{(r-\mu)^2}{2\sigma^2}}, \quad (5.4)$$

with the scaling factor $\sqrt{2\pi\sigma^2}$ being applied to the original Gaussian-like pdf to allow $p(r \rightarrow 0) \rightarrow 1$. The function (5.4) allows to arbitrarily change from open or narrow spaces by means of its shape parameter σ . From (5.4), we consider $\mu = 0$ as we account for the typical user located at the origin of a two dimensional Euclidean space and we vary the factor σ^2 to mimic the obstruction effect on the proposed performance metric. The adopted function can easily be changed with any arbitrary LOS probability rule.

5.3 The Analytical Findings

5.3.1 Essential Tools for Γ_f^S and Γ_f^M Approximations

Here, we present our approximations for Γ_f^S and Γ_f^M . Due to the property of a sum of normal distributed random variables, the desired power term in (5.2) can be obtained as

$$D_f^S = \left| \sum_{k \in K_f^{S,nL}} h_k r_k^{-\alpha/2} + \sum_{k \in K_f^{S,L}} h'_k r_k^{-\alpha/2} \right|^2 = |H + H'|^2 = |\bar{H}|^2,$$

where $H \sim \mathcal{CN}(0, \sum_{k \in K_f^{S,nL}} r_k^{-\alpha})$, and $H' = X' + iY'$ with

$$X' \sim \mathcal{N} \left(\mu_h \sum_{k \in K_f^{S,L}} r_k^{-\alpha/2}, \sigma_h^2 \sum_{k \in K_f^{S,L}} r_k^{-\alpha} \right)$$

and

$$Y' \sim \mathcal{N} \left(0, \sigma_h^2 \sum_{k \in K_f^{S,L}} r_k^{-\alpha} \right).$$

Consequently, $\bar{H} = \bar{X} + i\bar{Y}$ where

$$\bar{X} \sim \mathcal{N} \left(\underbrace{\mu_h \sum_{k \in K_f^{S,L}} r_k^{-\alpha/2}}_{\mu_L}, \sigma_h^2 \sum_{k \in K_f^{S,L}} r_k^{-\alpha} + \frac{1}{2} \sum_{k \in K_f^{S,nL}} r_k^{-\alpha} \right),$$

$$\bar{Y} \sim \mathcal{N} \left(0, \sigma_h^2 \sum_{k \in K_f^{S,L}} r_k^{-\alpha} + \frac{1}{2} \sum_{k \in K_f^{S,nL}} r_k^{-\alpha} \right).$$

We can rewrite the variance of \bar{X} and \bar{Y} as

$$\begin{aligned} \mathbb{V}_{[\bar{X}, \bar{Y}]} &= \frac{1}{2} \left(\left(1 - 1 + \frac{1}{k+1} \right) \sum_{k \in K_f^{\text{S,L}}} r_k^{-\alpha} + \sum_{k \in K_f^{\text{S,nL}}} r_k^{-\alpha} \right) \\ &= \frac{1}{2} \left(\underbrace{\sum_{k \in K_f^{\text{S,L}}} r_k^{-\alpha} + \sum_{k \in K_f^{\text{S,nL}}} r_k^{-\alpha}}_{\sigma_{\text{P}}^2 = \sum_{k \in \phi_f \cap \mathcal{B}(0, R)} r_k^{-\alpha}} - \frac{k}{k+1} \underbrace{\sum_{k \in K_f^{\text{S,L}}} r_k^{-\alpha}}_{\sigma_{\text{L}}^2} \right), \end{aligned} \quad (5.5)$$

where the scaling factor for σ_{L}^2 can be written as μ_h^2 , following the definition of the mean for the Gaussian distributed modelling variable for Rician distributed channel coefficient h' . Therefore the desired power can finally be written in terms of non-complex, zero-mean Gaussian random variables as

$$\begin{aligned} D_f^{\text{S}} &= |\bar{H}|^2 = (\bar{X}^2 + \bar{Y}^2) = \\ &= \left(\underbrace{\sigma_{\text{P}}^2 (X^2 + Y^2)}_{D_f^{\text{S,nL}}} - \underbrace{\mu^2 \sigma_{\text{L}}^2 (X^2 + Y^2) + 2\mu_{\text{L}} \mu \sqrt{\sigma_{\text{P}}^2 - \mu^2 \sigma_{\text{L}}^2} X + \mu_{\text{L}}^2 \mu^2}_{\Delta D_f^{\text{S}} | D_f^{\text{S,nL}}} \right), \end{aligned} \quad (5.6)$$

in which $X \sim Y \sim \mathcal{N}(0, 1/2)$. Note in (5.6) that we have decoupled the term that describes the desired power from a full NLOS network $D_f^{\text{S,nL}}$ from the contribution of the LOS components indicated with $\Delta D_f^{\text{S}} | D_f^{\text{S,nL}}$. It can be seen that the term σ_{P}^2 can be interpreted as the sum of the power attenuation factors of the cooperating SBSs, with no distinction made on the experienced NLOS/LOS small-scale channel fading. In other words, this term coincides with the sum of the contributions of the power attenuation for the desired content as if the network is only composed by NLOS links.

The interfering power from the SBSs in (5.2) is now studied in terms of non-complex normal random variables $X \sim Y \sim \mathcal{N}(0, 1/2)$. The detailed derivation is discussed in Appendix C.1 and and shortly reported in the following for the set of

Variable	Description
μ_h	Mean for LOS small-scale channel fading
σ_h	Standard deviation for LOS small-scale channel fading
μ_L	Sum of amplitude losses LOS components
σ_L^2	Sum of power losses LOS components
σ_P^2	Sum of power losses full NLOS network
σ	Control parameter for LOS thinning function $p(r)$

Table 5.1: Resume of key variables.

interfering PPPs $\phi_{-f}^{S,nL} \cap \mathcal{B}(0, R)$, $\phi_{-f}^{S,L} \cap \mathcal{B}(0, R)$, $\phi^{S,nL} \cap \mathcal{B}(R, \infty)$ and $\phi^{S,L} \cap \mathcal{B}(R, \infty)$

$$\begin{aligned}
 I_f^{S,in,nL} + I_f^{S,in,L} + I^{S,out,nL} + I^{S,out,L} &= \underbrace{\sum_{i \in \phi_{-f}^{S,nL} \cap \mathcal{B}(0,R)} (X^2 + Y^2) r_i^{-\alpha}}_{I_f^{S,in}} + \underbrace{\sum_{i \in \phi^S \cap \mathcal{B}(R,\infty)} (X^2 + Y^2) r_i^{-\alpha}}_{I^{S,out}} \\
 &+ \underbrace{\sum_{i \in \phi_{-f}^{S,L} \cap \mathcal{B}(0,R)} \left(-\frac{k}{k+1} (X^2 + Y^2) + \frac{2\sqrt{k}}{k+1} X + \frac{k}{k+1} \right) r_i^{-\alpha}}_{\Delta I_f^{S,in} | I_f^{S,in}} \\
 &+ \underbrace{\sum_{i \in \phi^{S,L} \cap \mathcal{B}(R,\infty)} \left(-\frac{k}{k+1} (X^2 + Y^2) + \frac{2\sqrt{k}}{k+1} X + \frac{k}{k+1} \right) r_i^{-\alpha}}_{\Delta I^{S,out} | I^{S,out}},
 \end{aligned} \tag{5.7}$$

where the terms $I_f^{S,in}$ and $I^{S,out}$ do not distinguish among NLOS /LOS small-scale channel fading and with $\Delta I_f^{S,in} | I_f^{S,in}$ and $\Delta I^{S,out} | I^{S,out}$, respectively, standing for the conditioned terms that describe the LOS contributions. Similarly, for what concerns the interference from the MBS in (5.2), we can apply the same intuitions in Appendix C.1 for $\phi^{M,nL}$ and $\phi^{M,L}$ to obtain

$$I^{M,nL} + I^{M,L} = \underbrace{\sum_{i \in \phi^M} (X^2 + Y^2) r_i^{-\alpha}}_{I^M} + \underbrace{\sum_{i \in \phi^{M,L}} \left(-\frac{k}{k+1} (X^2 + Y^2) + \frac{2\sqrt{k}}{k+1} X + \frac{k}{k+1} \right) r_i^{-\alpha}}_{\Delta I^M | I^M}. \tag{5.8}$$

The same analysis is now undertaken for what concerns the MBS transmission. The desired power term in γ^M from (5.3b) can be interpreted, following the guidelines in Appendix. C.1 as

$$D^M = |h'|^2 r^{-\alpha} = \underbrace{(X^2 + Y^2) r_i^{-\alpha}}_{D^{M,nL}} + \underbrace{\left(-\frac{k}{k+1}(X^2 + Y^2) + \frac{2\sqrt{k}}{k+1}X + \frac{k}{k+1} \right) r_i^{-\alpha}}_{\Delta D^M}. \quad (5.9)$$

Note that for (5.3a), no LOS contribution is experienced for the term of the desired power and the result would coincide with the $D^{M,nL}$ component in (5.9).

We now apply the separation of the full NLOS term to the interference perceived when the closest MBS is transmitting as previously obtained, which gives, following the result in Appendix C.1 for $\phi^{S,nL}$, $\phi^{S,L}$, $\phi^{M,nL} \setminus \mathcal{B}(\bar{r}, \infty)$ and $\phi^{M,L} \setminus \mathcal{B}(\bar{r}, \infty)$

$$\begin{aligned} I^{S,L} + I^{S,nL} + I^{M,nL} + I^{M,L} &= \underbrace{\sum_{i \in \phi^S} (X^2 + Y^2) r_i^{-\alpha}}_{I^S} + \underbrace{\sum_{i \in \phi^M \setminus \mathcal{B}(\bar{r}, \infty)} (X^2 + Y^2) r_i^{-\alpha}}_{I^M} \\ &+ \underbrace{\sum_{i \in \phi^{S,L}} \left(-\frac{k}{k+1}(X^2 + Y^2) + \frac{2\sqrt{k}}{k+1}X + \frac{k}{k+1} \right) r_i^{-\alpha}}_{\Delta I^S | I^S} \\ &+ \underbrace{\sum_{i \in \phi^{M,L} \setminus \mathcal{B}(\bar{r}, \infty)} \left(-\frac{k}{k+1}(X^2 + Y^2) + \frac{2\sqrt{k}}{k+1}X + \frac{k}{k+1} \right) r_i^{-\alpha}}_{\Delta I^M | I^M}. \end{aligned} \quad (5.10)$$

5.3.2 The Approximations

We have now the tools by which the proposed approximations can be retrieved. Hence, in this section, the main contribution of this work is presented.

The term γ_f^S can now be rewritten by isolating the contribution given by a full NLOS network from the contribution given by LOS communications. We therefore obtain that an approximation for γ_f^S can be attained by considering the conditioned terms to be independent from their respective conditioning factors.

Thus, the SINR γ_f^S is approximated as

$$\gamma_f^S \approx \tilde{\gamma}_f^S = \frac{D_f^{S,nL} + \Delta D_f^S}{I_f^{S,in} + I^{S,out} + \Delta I_f^{S,in} + \Delta I^{S,out} + I^M + \Delta I^M + W}, \quad (5.11)$$

where the terms $\Delta I_f^{S,in}|I_f^{S,in}$, $\Delta I^{S,out}|I^{S,out}$ and $\Delta I^M|I^M$ were initially conditioned on the resulting Gaussian distributed random variables X , Y , and the point processes $\phi_f^{S,L} \cap \mathcal{B}(0, R)$, $\phi_f^{S,L} \cap \mathcal{B}(R, \infty)$ and $\phi^{M,L}$.

We claim that the same approximation is valid for γ^M when the closest MBS is associated to the typical user. Then,

$$\gamma^M \approx \tilde{\gamma}^M = \frac{D^{M,nL} + \Delta D^M}{I^S + \Delta I^S + I^M + \Delta I^M + W}, \quad (5.12)$$

with $\Delta D^M = 0$ in case the wireless link is NLOS as in (5.3a), where the terms $\Delta I^S|I^S$ and $\Delta I^M|I^M$ were initially conditioned on the resulting Gaussian distributed random variables X , Y , and the point processes $\phi^{S,L}$ and $\phi^{M,L}$. This allows us to uncondition the SCDP independently as a full NLOS network, and separately account for the LOS contribution. The approximated SCDP values of $\Gamma_f^S \approx \bar{\Gamma}_f^S$ and $\Gamma_f^M \approx \bar{\Gamma}_f^M$ are finally obtained as

$$\begin{aligned} \bar{\Gamma}_f^S = & \underbrace{\mathbb{E} \left[\exp \left(-\frac{\bar{\rho}_f^S}{\sigma_P^2} (I_f^{S,in} + I^{S,out} + I^M + W) \right) \right]}_{\text{typical NLOS network}} \times \\ & \underbrace{\mathbb{E} \left[\exp \left(-\frac{\bar{\rho}_f^S}{\sigma_P^2} (\Delta I_f^{S,in} + \Delta I^{S,out} + \Delta I^M) + \frac{\Delta D_f^S}{\sigma_P^2} \right) \right]}_{\text{LOS contribution}}, \end{aligned} \quad (5.13)$$

$$\bar{\Gamma}_f^M = \underbrace{\mathbb{E} \left[\exp \left(-\frac{\bar{\rho}_f^M}{\bar{r}^{-\alpha}} (I^S + I^M + W) \right) \right]}_{\text{typical NLOS network}} \underbrace{\mathbb{E} \left[\exp \left(-\frac{\bar{\rho}_f^M}{\bar{r}^{-\alpha}} (\Delta I^S + \Delta I^M) + \frac{\Delta D^M}{\bar{r}^{-\alpha}} \right) \right]}_{\text{LOS contribution}}. \quad (5.14)$$

The complete SCDP approximation at the typical user from (5.1) is now attained

as

$$\text{SCDP} \approx \overline{\text{SCDP}} = \sum_f^F \hat{p}_f \left(a_f \bar{\Gamma}_f^S + (1 - a_f) \bar{\Gamma}_f^M \right). \quad (5.15)$$

The probabilities $\bar{\Gamma}_f^S$ and $\bar{\Gamma}_f^M$ can assume higher values than one, given by the LOS contribution. Therefore, during simulations, we have to respect the support of the exponentially distributed random variable $[0, \infty]$ and fix the probability to one when the argument of the exponential function assumes positive values following $\bar{\Gamma}_f^S = \min(\bar{\Gamma}_f^S, 1)$ and $\bar{\Gamma}_f^M = \min(\bar{\Gamma}_f^M, 1)$.

The generic Laplacian functionals been utilised in the subsequent derivations are shortly reported

$$\begin{aligned} \mathcal{L}^{\Phi, L}(s, a, b, \Phi) &= \mathbb{E} \left[\exp \left(s \sum_{i \in \phi} \left(-\frac{k}{k+1} (X^2 + Y^2) + \frac{2\sqrt{k}}{k+1} X + \frac{k}{k+1} \right) r_i^{-\alpha} \right) \right] \\ &\stackrel{a}{=} \mathbb{E} \left[\prod_{i \in \phi} \exp \left(s \left(-\frac{k}{k+1} (X^2 + Y^2) + \frac{2\sqrt{k}}{k+1} X + \frac{k}{k+1} \right) r_i^{-\alpha} \right) \right] \\ &\stackrel{b}{=} \exp \left(-2\pi\lambda \left(\int_a^b \left(1 - \frac{e^{\frac{ksr^{-\alpha}(sr^{-\alpha}+1)}{ksr^{-\alpha}+k+1}} (k+1)}{ksr^{-\alpha} + k + 1} \right) rp(r) dr \right) \right) \end{aligned} \quad (5.16)$$

and

$$\begin{aligned} \mathcal{L}^{\Phi, nL}(s, a, b, \Phi) &= \mathbb{E} \left[\exp \left(s \sum_{i \in \phi} (X^2 + Y^2) r_i^{-\alpha} \right) \right] \stackrel{a}{=} \mathbb{E} \left[\prod_{i \in \phi} \exp \left(s (X^2 + Y^2) r_i^{-\alpha} \right) \right] \\ &\stackrel{b}{=} \exp \left(-2\pi\lambda \int_a^b \left(1 - \frac{1}{sr^{-\alpha} + 1} \right) r dr \right), \end{aligned} \quad (5.17)$$

where, in both (5.16) and (5.17), (a) is from the independence of the considered PPPs and (b) is the result of the pgfl of the interfering PPPs.

In the following, we show that our solution is able to enhance the understanding of the effects of LOS communications. The expected value over the random variables X and Y (thus the channel small-scale fading) for the LOS contribution

of the desired power of $\bar{\Gamma}_f^S$ can be obtained, after re-arranging its terms, as

$$\mathbb{E}_{X,Y} \left[\exp \left(\frac{\Delta D_f^S}{\sigma_P^2} \right) \right] = \exp \left(\frac{2\mu_L^2 \mu^2}{\mu^2 \sigma_L^2 + \sigma_P^2} \right) \frac{\sigma_P^2}{\sigma_P^2 + \mu^2 \sigma_L^2} \quad (5.18)$$

by averaging out from $X \sim Y \sim \mathcal{N}(0, 1/2)$.

For what concerns the interfering LOS contribution of $\bar{\Gamma}_f^S$, the Laplace transform and pgfl can be employed. We therefore obtain

$$\begin{aligned} \mathbb{E} \left[\exp \left(-\frac{\bar{\rho}_f^S}{\sigma_P^2} \left(\Delta I_f^{S,\text{in}} + \Delta I^{S,\text{out}} + \Delta I^M \right) \right) \right] \\ = \left(\begin{array}{l} \mathcal{L}^{\Delta S,\text{in}} \left(s, 0, R, \phi_{-f}^{S,L} \cap \mathcal{B}(0, R) \right) \\ \times \mathcal{L}^{\Delta S,\text{out}} \left(s, R, \infty, \phi^{S,L} \cap \mathcal{B}(R, \infty) \right) \\ \times \mathcal{L}^{\Delta M} \left(s, 0, \infty, \phi^{M,L} \right) \end{array} \right) \Bigg|_{s=\frac{\bar{\rho}_f^S}{\sigma_P^2}}, \quad (5.19) \end{aligned}$$

where $\mathcal{L}^{\Delta S,\text{in}}$, $\mathcal{L}^{\Delta S,\text{out}}$ and $\mathcal{L}^{\Delta M}$ can be drawn as the Laplacian functional of the LOS interfering contribution in (5.16). The NLOS term in (5.13) can easily be retrieved as

$$\begin{aligned} \mathbb{E} \left[\exp \left(-\frac{\bar{\rho}_f^S}{\sigma_P^2} \left(I_f^{S,\text{in}} + I^{S,\text{out}} + I^M + W \right) \right) \right] \\ = e^{-sW} \left(\begin{array}{l} \mathcal{L}^{S,\text{in}} \left(s, 0, R, \phi_{-f}^S \cap \mathcal{B}(0, R) \right) \\ \times \mathcal{L}^{S,\text{out}} \left(s, R, \infty, \phi^S \cap \mathcal{B}(R, \infty) \right) \\ \times \mathcal{L}^M \left(s, 0, \infty, \phi^M \right) \end{array} \right) \Bigg|_{s=\frac{\bar{\rho}_f^S}{\sigma_P^2}}, \quad (5.20) \end{aligned}$$

where $\mathcal{L}^{S,\text{in}}$, $\mathcal{L}^{S,\text{out}}$ and \mathcal{L}^M are drawn from the Laplacian functional of the NLOS interfering terms in (5.17). By replacing (5.18), (5.19) and (5.20) in (5.13), we have obtained the proposed approximation for $\bar{\Gamma}_f^S$.

Similarly, the de-conditioned derivation from X and Y of the LOS contribution

for the desired power in $\bar{\Gamma}_f^M$ can be computed to yield

$$\mathbb{E}_{X,Y} \left[\exp \left(\frac{\Delta D^M}{\bar{r}^{-\alpha}} \right) \right] = \exp \left(\frac{k\bar{r}^{-\alpha}(\bar{r}^{-\alpha} + 1)}{k\bar{r}^{-\alpha} + k + 1} \right) \frac{k + 1}{k\bar{r}^{-\alpha} + k + 1}. \quad (5.21)$$

In case the closest MBS does not experience an LOS link with the typical user, we would simply consider $\Delta D^M = 0$. For what concerns the LOS interfering contribution, we can apply the Laplace transform of the interference as

$$\mathbb{E} \left[\exp \left(- \frac{\bar{\rho}_f^M}{\bar{r}^{-\alpha}} (\Delta I^S + \Delta I^M) \right) \right] = \quad (5.22)$$

$$\left(\mathcal{L}^{\Delta S} (s, 0, \infty, \phi^{S,L}) \mathcal{L}^{\Delta M} (s, \bar{r}, \infty, \phi^{M,L} \setminus \mathcal{B}(\bar{r}, \infty)) \right) \Big|_{s=\frac{\bar{\rho}_f^M}{\bar{r}^{-\alpha}}}, \quad (5.23)$$

where $\mathcal{L}^{\Delta S}$ and $\mathcal{L}^{\Delta M}$ are obtained from the Laplacian in (5.16) with \bar{r} standing as the distance of the closest serving MBS. Similarly, the interfering term of a typical NLOS network can be written as

$$\mathbb{E} \left[\exp \left(- \frac{\bar{\rho}_f^M}{\bar{r}^{-\alpha}} (I^S + I^M + W) \right) \right] = \quad (5.24)$$

$$e^{-\frac{\bar{\rho}_f^M}{\bar{r}^{-\alpha}} W} \left(\mathcal{L}^S (s, 0, \infty, \phi^S) \mathcal{L}^M (s, \bar{r}, \infty, \phi^M \setminus \mathcal{B}(\bar{r}, \infty)) \right) \Big|_{s=\frac{\bar{\rho}_f^M}{\bar{r}^{-\alpha}}}, \quad (5.25)$$

where \mathcal{L}^S and \mathcal{L}^M are obtained from (5.17).

By replacing (5.21), (5.22) and (5.24) in (5.14), we have derived our approximation $\bar{\Gamma}^M$. Note that in case the approximated terms in (5.11) and (5.12) were not considered, the derived SCDP in (5.13) and (5.14) would write as a non solvable integral which involves a modified Bessel function of the first kind as in [83]. Also, it is known that a closed-form solution for jointly transmitting contents has not been found yet, even for the simple full NLOS network [50,55,72]. Thus, when maximising the SCDP (or minimising the outage probability) over some network parameters, an iterative solution is generally required to achieve a sub-optimal feasible set of decision variables. Based on our proposed approximation, existing solutions in the literature which aim to find sub-optimal network parameters can be updated by scaling the SCDP for a full NLOS network with the proposed LOS contribution as in (5.13) and (5.14). As a final remark, it can be observed that

the derived approximations can easily be integrated to the SCDP derivation in (3.11). More precisely, the LOS contributions derived in (5.18) and (5.19) can be applied as scaling factors to the full NLOS term (3.11), with minor changes to be accounted to the derived approximations in order to be included to the framework in Chapter 3.

5.3.3 Network Trade-off

From the approximated $\bar{\Gamma}_f^S$ and $\bar{\Gamma}_f^M$, some intuitions over the effects of LOS communications can be retrieved. Some preliminary insights can be obtained at this stage.

Conjecture 1. *The limit for $k \rightarrow 0$ for the approximated SCDP in (5.15) results in a full NLOS network.*

When we consider the limit to infinity of k , it can be seen that the mean and variance of the generating Gaussian random variables for the Rician distributed channel fading coefficient tend to $\mu_h \rightarrow 1$ and $\sigma_h^2 \rightarrow 0$. As a consequence, the channel power fading for the LOS link no longer acts like a random variable, due to the collapse of the variance.

Conjecture 2. *The limit for $k \rightarrow \infty$ for the approximated SCDP in (5.15) conditioned on Rayleigh distributed small-scale channel power fading for the interference results in a scaling factor to be applied to the typical NLOS contribution in order to account the LOS for the desired power.*

In particular, referring to **Conjecture 2**, the typical full NLOS SCDP terms in (5.13) and (5.14) have to be scaled, respectively, by

$$\lim_{k \rightarrow \infty} \left[\exp \left(\frac{2\mu_L^2 \mu^2}{\mu^2 \sigma_L^2 + \sigma_P^2} \right) \frac{\sigma_P^2}{\sigma_P^2 + \mu^2 \sigma_L^2} \right] = \exp \left(\frac{2\mu_L^2}{\sigma_L^2 + \sigma_P^2} \right) \frac{\sigma_P^2}{\sigma_P^2 + \sigma_L^2}, \quad (5.26)$$

$$\lim_{k \rightarrow \infty} \left[\exp \left(\frac{k\bar{r}^{-\alpha}(\bar{r}^{-\alpha} + 1)}{k\bar{r}^{-\alpha} + k + 1} \right) \frac{k + 1}{k\bar{r}^{-\alpha} + k + 1} \right] = \frac{\exp(\bar{r}^{-\alpha})}{\bar{r}^{-\alpha} + 1}. \quad (5.27)$$

By means of **Conjecture 1** and **Conjecture 2**, we are able to understand the SCDP results for $k \in [0, \infty]$.

Some intuitions over the effects of LOS wireless links on the SCDP can be attained at this point. The LOS contribution from the cooperating edge nodes in (5.18) can be easily proved to be upper-bounded by the particular case in (5.26), with the same result standing for (5.21) and (5.27). However, the term (5.26) can be considered, especially under a UDN scenario, by the case that all the cooperating nodes experience LOS links. Thus, from (5.26), we can write

$$\exp\left(\frac{2\mu_L^2}{\sigma_L^2 + \sigma_P^2}\right) \frac{\sigma_P^2}{\sigma_P^2 + \sigma_L^2} \leq \frac{1}{2} \exp\left(\left(\frac{\mu_L}{\sigma_L}\right)^2\right) \Bigg|_{\sigma_P^2 \rightarrow \sigma_L^2}. \quad (5.28)$$

It can be seen that the latest upper-bound on the LOS contribution from the set of cooperating nodes is exponentially dependent on the ratio of the received amplitude-to-power attenuation from the cooperating nodes. This quantity is therefore highly sensible on the chosen path-loss model. The result from **Conjecture 2** works as an upper-bound over the received LOS desired power. Therefore, we will employ this quantity in subsequent sections to investigate the gap for a full NLOS network derived in **Conjecture 1**. Similar conclusions can be attained for what concerns the LOS contribution from the MBS, with the particular case for $k \rightarrow \infty$ employed as an upper-bound of the desired signal power.

5.3.4 Extension to Double-Slope Path-Loss Model

In this chapter, a single-slope path-loss model is considered for simplicity. However, an extension to double or multi-slope path-loss model can be attained with minor changes applied to the previous derivations. To provide some guidelines to achieve this result, we consider a double-slope path-loss model. Consider the difference of path-loss coefficients as $\Delta\alpha = \alpha^{\text{nL}} - \alpha^{\text{L}}$, where α^{nL} and α^{L} stand respectively for the NLOS and LOS links. To introduce a double-slope path-loss in our model, it is necessary to change the scaling factor in (5.5) to $(r^{-\Delta\alpha} - r^{-\Delta\alpha} + \frac{1}{k+1})$ and to include this term inside the sum for the cooperating edge nodes with an LOS wireless link. As a result, we can decouple the derivation as a full NLOS and LOS contributions, as previously obtained.

5.4 Numerical Results

In this section, numerical results are illustrated to demonstrate the validity of our SCDP approximation under various network parameters. The results are obtained by employing a UC probabilistic caching strategy and the baseline network values are reported in Table 5.2.

Variable	Value	Description
λ^S	5×10^{-4} [unit/ m^2]	Initial caching node density
λ^M	10^{-6} [unit/ m^2]	Upper caching node density limit
λ^U	10^{-4} [unit/ m^2]	Total number of users
R	25 [m]	Radius of CSA
P^S	26 [dBm]	Transmitting power at the edge node
P^M	43 [dBm]	Transmitting power at the edge node
W	-174 [dBm]	Thermal noise power
F	20	Wordbook size
M	3	Cache size
v	1	Skewness factor content popularity
α	3	Path-loss coefficient
B	40 [MHz]	Bandwidth
ρ	2 [Mbps]	Target bit-rate
K_f^{\max}	3	Maximum number of cooperating edge nodes

Table 5.2: The network parameters.

5.4.1 Validation against λ^S

In Fig. 5.2, results are provided to compare the target SCDP with the proposed approximation in (5.15), against the edge node density λ^S . For comparison, the results for a full Rayleigh network together with **Conjecture 1** have been included to give an estimation of the existing gap between full NLOS and mixed Rayleigh/Rician networks. Results show that the proposed approximation matches with the exact result of the case of a full Rayleigh network when $k = 0$. With a higher probability to encounter LOS links, results begin to see some gaps between

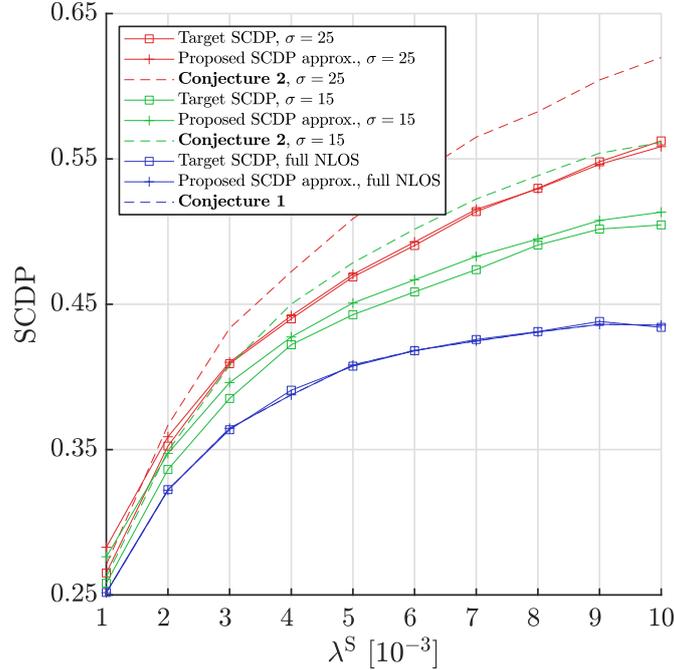


Figure 5.2: Target and approximated SCDP against λ^S .

the SCDP approximation and the exact result for $\sigma > 0$, as λ^S increases. However, the proposed approximation is largely accurate for the entire range of λ^S .

5.4.2 Validation against Rician k -factor

The power ratio between the LOS component and the shadowed multi-paths plays an important role when addressing the SCDP. In Fig. 5.3, the target and approximated SCDP are investigated for different values of the Rician k -factor. An initial increasing trend can be noticed as the Rician k -factor is incremented until the SCDP measure converges. As the k -factor increases the desired power benefits from the contribution from the cooperating nodes until it balances with the destructive contribution of the interferers. From Fig. 5.3, it can be observed that the results for **Conjecture 2** stand on higher values than the associated target and approximated SCDP measures, over the whole considered domain of k -factor values. As discussed in Section 5.3.3, the expression for **Conjecture 2** is an upper limit of the LOS contribution for both the target and approximated SCDP. Further, the obtained values for **Conjecture 2** are not dependent on the increase

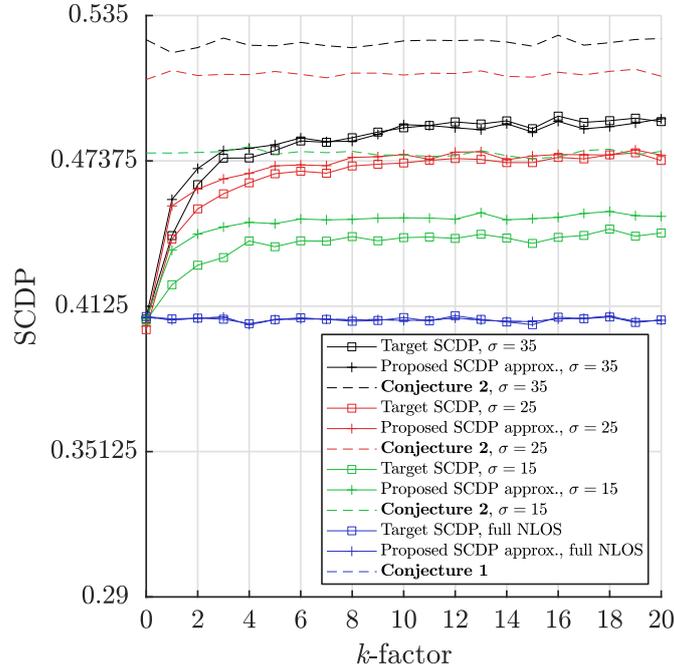


Figure 5.3: Target and approximated SCDP against Rician k -factor.

of the k -factor, because calculated for $k \rightarrow \infty$. It can be seen, that for better SINR conditions, a smaller gap between the target and approximated SCDPs is obtained, revealing higher accuracy of the proposed approximation when better transmitting conditions are experienced.

5.4.3 Effective Bandwidth and User Density

In our network model, user density has a crucial impact on the available frequency bands and possibly our approximation. Our model catches the number of simultaneous content transmission through the terms $\bar{\rho}_f^S$ and $\bar{\rho}_f^M$ in (5.1). Therefore, a study of how the proposed approximation performs according to the available bandwidth for each content transmission is essential. As the transmission to the typical user occurs if the request is previously cached, each edge node has a maximum of M simultaneous content requests that can be simultaneously performed. In Fig. 5.4, we illustrate the SCDP against λ^U . As expected, it can be seen that a decrease of the overall performance is experienced, as a consequence of the increasingly higher number of simultaneous content transmissions. The reduction of

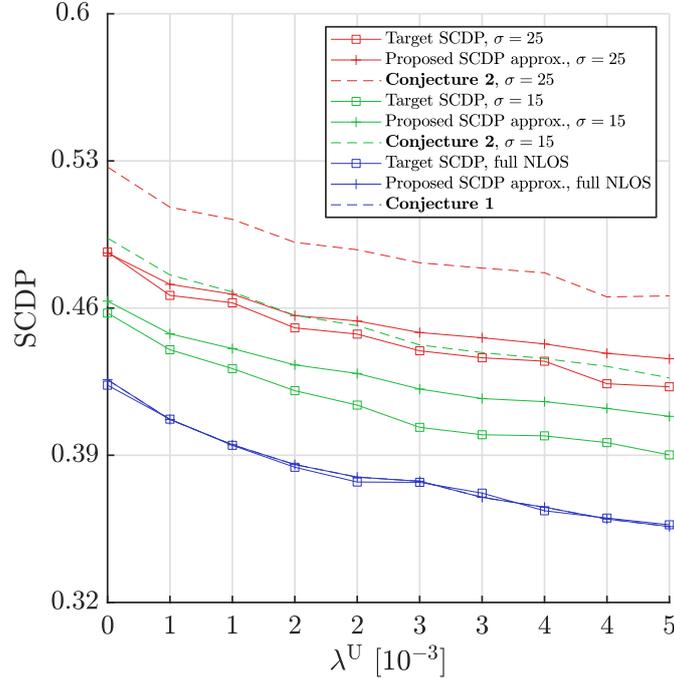


Figure 5.4: Target and approximated SCDP against λ^U .

the available bandwidth for the single content transmission is observed to slightly increase the gap among the proposed approximation and the target SCDP.

5.4.4 Maximum Number of Cooperating Nodes

Cooperation among nodes has been indicated as a promising strategy to achieve the performance standards of future mobile networks. It is important to have our proposed approximation to be as tight as possible as the maximum number of cooperating nodes increases. In general, due to the overhead necessary to synchronise JT, a maximum number of cooperating SBS edge nodes is generally fixed. In the following, we consider an increased radius for the content searching ball $\mathcal{B}(0, R)$ of $R = 35 [m]$ and SBS density of $\lambda^S = 10^{-3}$ such that the average number of cooperating nodes for content f stands as $\mathbb{E} [|\phi_f^S \cap (\mathcal{B}(0, R))|] = p_f \lambda^S \pi R^2 \approx 5.77$. The numerical results are given in Fig. 5.5. We see that our proposed approximation follows the target measure, especially at higher σ . When low values of K^{\max} are considered, it is more likely that some nodes that can potentially take part to the

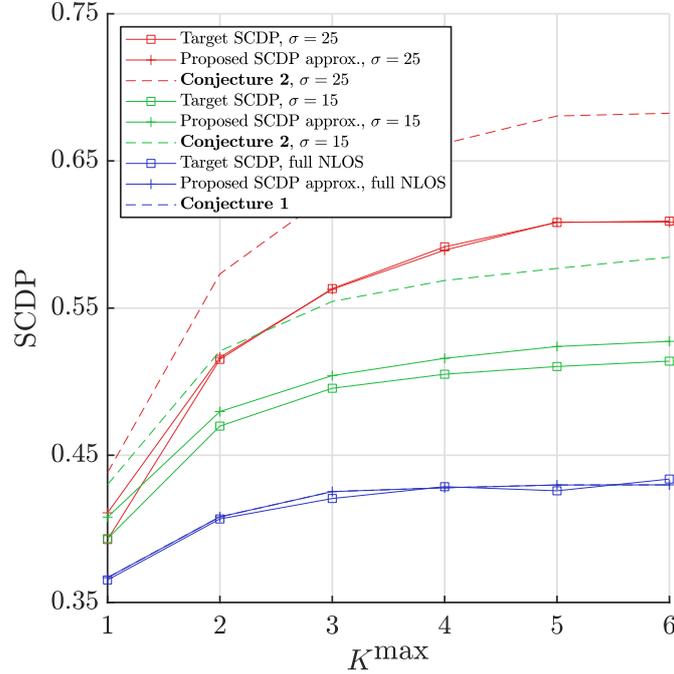


Figure 5.5: Target and approximated SCDP against maximum number of cooperating edge nodes.

content delivery are not included in the cooperating set of edge nodes but, in some cases, are considered as additional interferers. This reason is behind the increase of SCDP against K^{\max} . On top of it, the results highlight the role of σ in terms of SCDP performance and the gap. As K^{\max} increases, the higher probability of experiencing LOS wireless links for higher σ values results in higher SCDP performance and gives a smaller gap with its target SCDP. Therefore, under better SINR conditions an higher increase of SCDP is expected, and a smaller gap is experienced, highlighting the better level of accuracy of the proposed approximation. A saturation of SCDP performance at higher values of K^{\max} indicates the balance between desired and interfering power and is due to the chosen average number of cooperating nodes for content f .

5.4.5 The Proposed Approximation against σ

The intensity of σ gives an indication of the percentage of LOS wireless links over the elements of a snapshot of ϕ^S . Also, as an indicative point of view, the

higher the σ the better the SINR conditions because the closer SBSs experience LOS wireless links. The results show that the gap between the proposed SCDP approximation and its target is smaller as σ increases. Since at higher σ the gap is reduced, the generation of the gap has to be linked to the nature of the wireless link of the cooperating nodes, and in particular, regardless of whether the majority of them experience LOS wireless links with their associated receiver, better levels of accuracy of our proposed approximation are seen. This analysis reveals the trade off between performance and experienced gap in our simulations, showing that for better SCDP performance and thus better SINR values, a better approximation can be achieved.

5.5 Conclusions

In this chapter, an accurate approximation for a two-tier mixed Rayleigh/Rician small-scale channel faded UDN with non-coherent JT has been proposed and validated through numerical results. In particular, a scaling factor to SCDP to account for Rician channel power fading in closed form is derived, which then permits our analysis to SCDP for the mixed fading UDN with LOS and NLOS links with any arbitrary LOS probability function and a dual-slope path-loss model. The results in Section 5.4 revealed that the proposed de-conditioning technique over the LOS contribution is highly suitable to match the target SCDP. However, the employed procedure allows to simplify the tractability of the target SCDP, unlocking some insights over the contribution of LOS wireless links. As a direct consequence, the proposed approximation can be employed to extend existing Rayleigh limited mobile networks to a more general small-scale channel modeling.

Chapter 6

Conclusions

Network densification unlocked unprecedented improvements toward the objective of meeting the expected 1000-fold growth in traffic demand, bridging the gap with future requests of network capacity. However, the unconstrained massive deployment of network nodes can turn out to be destructive in terms of QoS performance, or extremely expensive in terms of energy consumptions. The presented work is an extensive investigation over nodes idling strategies, resource management at the network edge and analytical findings which help exploit future UDNs.

6.1 Summary of the Thesis Outcomes

The contributions of the presented work can be summarised into three main improvements.

In Chapter 3, a sub-optimal resource management policy is obtained which maximises the average users' SCDP. A network binning design has been utilised to allow the derivation of local SCDP measures, which is then combined to obtain a global SCDP evaluation. The overall analysis has been conditioned on the event of the presence of a copy of the request being cached within the CSA. This conditioning factor has forced the analysis on a set of novel interpretations of the network. First, an updated probabilistic content caching design is elaborated and validated by means of empirical analysis. Also, a zero truncated Poisson pmf is utilised to model the distribution of the edge caching nodes based on the condi-

tioning event. Due to the non-existence of a closed-form for CoMP transmitting techniques, Jensen's inequality has been resorted to provide a lower-bound of the target SCDP. The lower-bound is maximised over the content caching probabilities and caching node density. The introduced degree of freedom of network choices, *i.e.*, the possibility to model its active edge caching nodes, allows new understanding on optimal choices for UDNs. It has been observed that when more demanding transmitting conditions are experienced, the network prefers to concentrate its limited edge resources to favour the local area from which the maximum contribution to the SCDP is given. The numerical results from the optimisation of the SCDP have shown that over-densification can be harmful to the network performance metric while being extremely expensive when a huge number of nodes is deployed. The interference saturation at the network's edge results in a bottleneck for the network performance when over-densification is accounted. In some cases, a lower SCDP is acceptable with the intention not to deteriorate neighbouring content transmissions, thus providing an overall benefit to the set of users. The proposed approach has been shown to be able to discretise among distinct content generating densities, providing local sets of solutions which maximise a global SCDP metric.

Having acknowledged the influence of the SCDP maximization over the local network resource management policies, the successive work has extended the role of the edge node density to strike the optimal balance between performance and efficiency. Chapter 4 emphasises the importance of a more accurate definition for the network EE, by considering the expected value of the random ratio between the SCDP and energy consumption within the CSA. In particular, as a closed form of the metric exists with respect to the edge node density, some fundamental results over the MCP have been discussed. Numerical simulations have been compared between the updated proposed definition of network EE against the typically considered interpretation from the literature. The MCP is then proved to stand as a lower-bound of the density which maximises the network SCDP. This permits to define a fundamental distance between the densities, which is then used to describe the energy expenses to move from the most efficient network state to the best performing state.

In Chapter 5, the final main contribution from this work lies in a tight and

scalable approximation for including LOS wireless links under a cooperative transmitting scheme. The approach can be easily extended to a non-cooperative scenario and to any arbitrary LOS thinning function. A suitable approximation is proposed which results from separating the components of the single contributing elements of the SCDP and dropping their dependencies to be considered as independent random variables. This allows closed-form solutions to previously intractable equations, unlocking insights associated to the LOS contribution to a full NLOS mobile wireless network. The technique also allows great scalability, as it can be utilised to extend current literature to a more general scenario.

6.1.1 Overview of Parameters' Choice for Numerical Simulations

In this section, the choices over the adopted network parameters to obtain simulations' results are discussed. The section focuses on a set of quantities which are indicated to have a considerable impact on the obtained numerical values. These decisions are herein analysed, with their effect on the obtained performance commented in detail.

- **Available Bandwidth for Contents Transmission:** The available bandwidth for content transmission B can be observed to change across the presented work. In Chapter 3, a bandwidth of $B = 100$ [MHz] is considered, while the adopted values of available frequency bands stand at $B = 30$ [MHz] and $B = 40$ [MHz] in Chapter 4 and Chapter 5 respectively. A wider frequency spectrum for simulations in Chapter 3 is necessary to simultaneously accommodate a larger number content requests. If lower values of B were considered, the lack of resources would have not allowed to perform multiple concurrent transmissions, representing a serious bottleneck to the network performance. It is important to highlight that the upper-bound for the number of simultaneously transmitted contents is set to correspond to the cache size, *i.e.* $M = 3$, with the bandwidth uniformly partitioned among the requests. This allows each content delivery to be performed over approximately 33 [MHz], which is coherent with the simulation choices in Chapter 4 and Chapter 5. Moreover, given the high density for users considered in Chapter

3, the case of having M requests to be simultaneously accommodated occurs to be most likely situation incurred by the network nodes. As a conclusion, the choice of making use of a wider spectrum for transmission in Chapter 3 ensures the coherence of the outcomes across the work.

- **Wordbook Size:** The wordbook size F can be noticed to vary between Chapter 3 and Chapter 5. The choice of this parameter in Chapter 3 has been greatly affected by the computational complexity involved with the search of the sub-optimal strategy. Differently, the wider workbook size in Chapter 5 is allowed by the low complexity of the objective function. As a result of having more contents, the optimisation algorithm is able to partition the network resources more precisely, thus improving the observed performance. In other words, the optimisation is allowed to refine the search of the optimal value of the objective function, improving the overall results. It is expected that an improved solution can be obtained from Chapter 3 when a wider wordbook is considered.
- **Network Node Density:** The employed SBS density λ^S is also seen to be changed across the thesis work. In Chapter 3, the optimisation is allowed to adjust local node density within a range of values whose upper-bound is controlled by a fixed value, *i.e.* $0 \leq \lambda_n^S \leq \bar{\lambda}^S$. The choice of the intensity of upper-bound has been made to encourage cooperation among nodes and study its detrimental effects on the overall performance. Setting a lower intensity of the upper-bound would not allow to conclude over the benefit of idling a portion of caching helpers. In Chapter 5 the focus lies on the experienced channel power gain, as a combination of LOS and NLOS links and a lower network node density is considered. Conversely, if higher values for λ^S were accounted, an increase in the likelihood of LOS wireless links would be seen. As a consequence of that, the presented results would be skewed towards a setup with a strong dominance of LOS links, which does not allow to conclude over the validity of the presented derivations. It is important to highlight that the work in Chapter 5 can be easily extended to the case of higher node density, as indicated by the analysis conducted in Section 5.3.3.

6.2 Further Developments

There are future directions that deserve further effort.

- **Extension to a multi-objective metric** (Chapter 3): Multi-objective optimisation problems involve more than one objective function to be optimised simultaneously. Typically, a single solution which returns the optimal value for all the considered objective functions is not possible. To find a trade-off between the conflicting objective functions is the aim of this extension of the work. Future research efforts will focus on obtaining the set of Pareto optimal solutions, which would greatly improve the understanding of how different metrics increase at the expenses of concurrent objective functions. This direction of the future work will therefore address the metrics elaborated in this thesis, *i.e.* SCDP and EE, and introduce additional performance indicators such as time latency and channel capacity. When the costs of fetching a generic content are correctly introduced, the inclusion of MBSs would provide more insights on the optimal network policy of usage of its limited network resources.
- **Relaxation assumption of proposed EE** (Chapter 4): The hypothesis of fully active SBS accounted to derive a closed-form to the EE metric can be relaxed to include idled nodes which contribute much less to the overall energy consumptions. The optimal ratio of active and idled edge network tier's nodes can be therefore improved. This however comes at the cost of a more complex objective function.
- **Extension of obtained results with LOS approximation** (Chapter 5): The obtained numerical values from the optimizations of the utility functions in Chapter 3 and Chapter 4 can be updated with the introduction of the proposed approximation for the LOS contribution given in Chapter 5. Intuitively, as iterative solutions have been employed in both chapters, the introduction of the new scaling coefficient to the objective function can easily be applied.

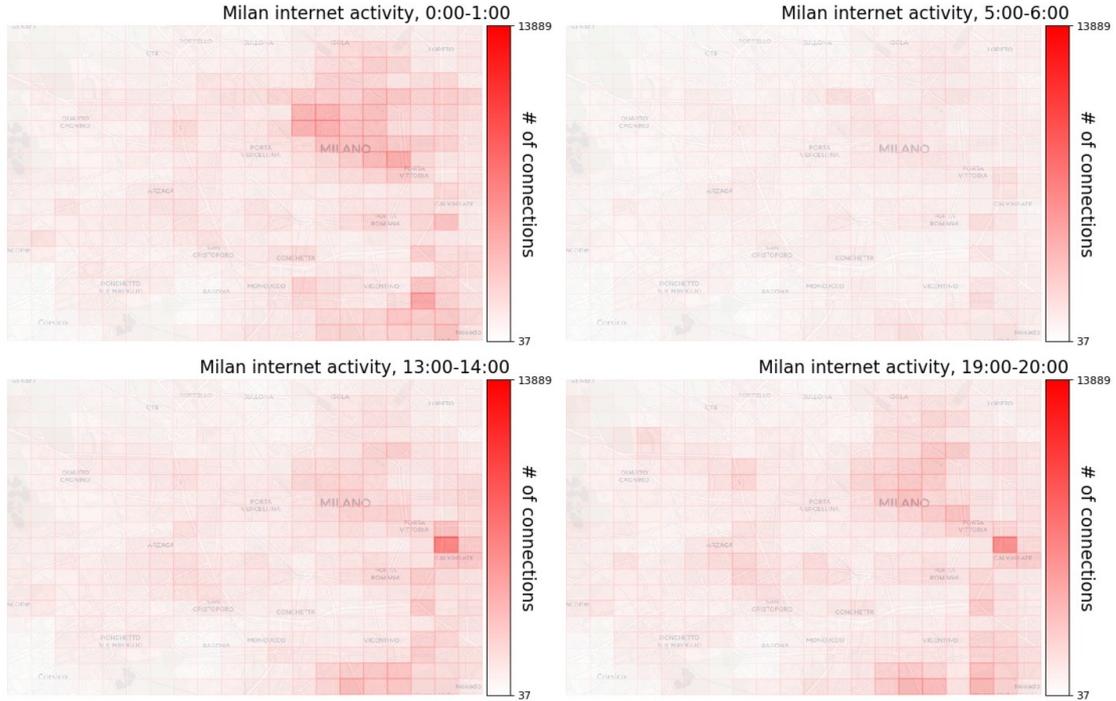


Figure 6.1: Users' activity urban area of Milan.

- Application of reinforcement learning techniques:** The derivation of closed-form expressions of the network metrics are limited to the simplest models. Thus, it is often unfeasible to generate optimal strategies for a wide set of initial network parameters and a relevant effort has been spent by academia to address the problem of time-varying optimal strategies [86–88]. Hence, the introduction of machine learning techniques can provide on-line solutions to chase the time-varying optimal network resource management policy. In this thesis, we considered some network's parameters to be fixed with respect to time. This assumption could be made as the investigations herein discussed focused on finding optimal solutions for a particular state of the network. However, it is true that parameters, *e.g.*, content popularity and content generating density are functions of time over longer time frames. Unsupervised learning has recently attracted great attention from scientific literature [89–92]. As preliminary study of the problem, the data-set [93] has been considered and the number of internet connections performed over the

urban area of Milan are reported in Fig. 6.1 for four representative one-hour ranges on the day 1st November 2013. The number of internet connections in [93] are provided with the precise coordinates in latitude and longitude of the vertices of the binned urban area, whose side length is approximately 235 [m]. The dynamic content generation process for a 20×20 bins network can be seen to be remarkably dynamic from Fig. 6.1. The study of evolving solutions is an interesting development, to investigate optimal strategies over time which can be learned through local and independent observation of the environment, performed by each node of the UDN tier.

Multi-agent learning has attracted great interest due to recent improvements which allowed its application for a variety of different tasks [94–96]. Authors in [97] made use of a multi-agent setup to optimise a centralised metric that estimates the joint action-value function as a non-linear combination of the per-agent values, conditioned on local observations of the environment. The problem in [97] can be adapted to cache-enabled wireless networks. Network nodes can be considered as independent decision-making agents that interact with the environment by means of the establishment of user-node associations and consecutive content transmission. The initial steps for the application of multi-agent learning to cache-enabled networks are described in the following.

A set of K agents is defined as an instance of a PPP ϕ^S . Time is split into time-frames, at the begin of which the agents take one out of three possible choices, after having performed a local and independent observation of the environment. The set of possible agent’s choices is defined by

1. **action 0** (normal activity): the edge node keeps its hardware components active as content requests are expected. When an hit-cache occurs, the edge node delivers the required contents to the UE.
2. **action 1** (update current cache): the set of cached contents can become obsolete over time. The agent, *i.e.* the caching node, refers to the closest MBS to update its local memory with new contents. When selecting this action, the edge node behaves as a UE, consuming the network resources for a future benefit.

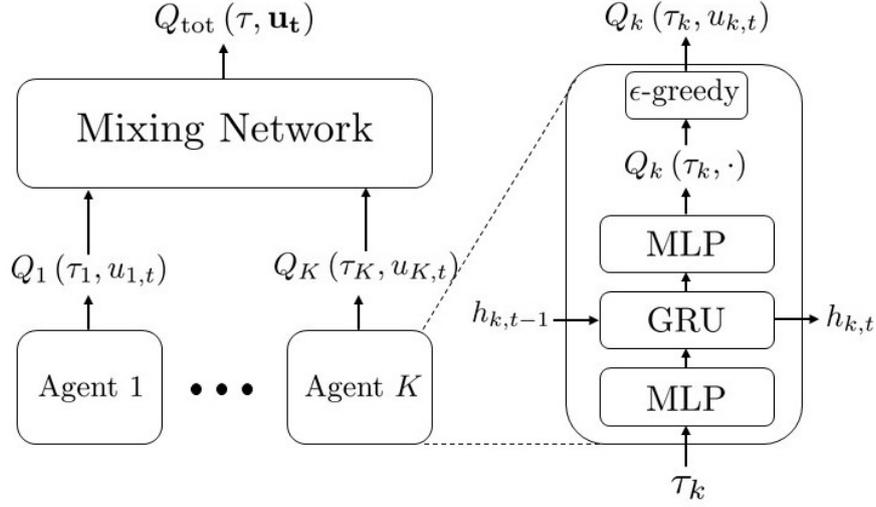


Figure 6.2: The multi-agent reinforcement learning architecture, where the agent network is composed by a combination of multilayer perceptron, gated recurrent networks and a probabilistic ϵ -greedy decision making process.

3. **action 2** (sleep mode): low activity of the surrounding environment can trigger this choice, with the edge node that turns off its hardware components to consume the less possible amount of power.

The k -th SBS's observation of the environment, *i.e.* τ_k , is now described as a collection of historical information gathered within a fixed time-horizon T_o . More rigorously, the vector τ_k is formed by

1. Content requests' hit-cache ratio within the time-horizon T_o ,
2. Content requests' missed-cache ratio within the time-horizon T_o ,
3. Approximation of user density $\hat{\lambda}_k^U$, *i.e.* $\frac{\text{number of received content requests}}{T_o \times \text{SBS activity region}}$,
4. Approximation of active SBS density $\hat{\lambda}_k^S$, *i.e.* $\frac{\text{time spent performing action 0}}{T_o}$,
5. Last chosen action, *i.e.* $u_{k,t-1}$.

The whole set of quantities to describe the current state of the environment can be easily obtained by each agent, without the need of extra links for communication among nodes. An overall knowledge of the network state can be obtained at the MBS level. Namely, when **action 1** is selected, the

collected information from the agent is sent to the associated MBS, from which a regional state of the environment can be determined.

The neural network consists of two main components: (i) the agent architecture, which generates the agent's approximation of its action-value function, *i.e.* Q_k , and (ii) a mixing network, in charge of performing the approximation of the agents' collective actions-value function, *i.e.* Q_{tot} . The design of the integrated network components is presented in Figure 6.2.

The agents learn by means of the back-propagation of the gradients which operate to minimize a loss function. For improved convergence, it is required to make use of a target network, which systematically updates its weights and keeps them constant for an arbitrary number of iterations, and a replay memory, which stores the historical experience of the network. Given θ the matrix which contains the current weights of the neural network and θ^- which stores the weights of the target neural network, a standard option for the loss function is

$$\mathcal{L}(\theta) = \sum_{d=1}^D [(y_d - Q_{\text{tot}}(\tau_d, \mathbf{u}_d; \theta))^2] \quad (6.1)$$

where $y_d = r_d + \gamma \max_{\mathbf{u}'} Q_{\text{tot}}(\tau', \mathbf{u}'; \theta^-)$, with D , γ and r_d which stand as the sampling batch size from the replay memory, the learning rate and the evaluation of the reward respectively, with $\max_{\mathbf{u}'} Q_{\text{tot}}(\tau', \mathbf{u}'; \theta^-)$ that returns the maximum value of the sampled action-value function and finally with τ' , \mathbf{u}' that stand respectively as the collective agents' observations and actions relative to the successive state of the environment. The design of the reward function is a complicated task which greatly impacts the agents' learning.

The overall work presented in this thesis is intended to further develop the understanding of future ultra-dense mobile networks through a series of analytical findings supported by empirical validations. The intention is that this work will be useful for telecommunications engineers to achieve always better results toward the out-of-reach goal of infinite capacity.

Appendix A

A.1 Proof of Theorem.7

The proposed Jensen's lower bound is conditioned on a set of random variables as shown in (3.23). Due to independence between the interference power and the desired signal power, we can work out the expected value in (3.23) by separately deriving the following two independent terms

$$\varphi_{n,f}^I(\boldsymbol{\lambda}^S, p_{n,f}) = \mathbb{E}_{\mathcal{I}_{n,f}} [\mathcal{I}_{n,f} + W], \quad (\text{A.1a})$$

$$\varphi_{n,f}^D(\boldsymbol{\lambda}_n^S, \mathbf{p}_n) = \mathbb{E}_{\substack{\{r_{n,k}\}_{k \in \phi_{n,f}}, \\ K_{n,f}, \Xi_{n,f}}} \left[\frac{2^{\frac{\rho \Xi_{n,f}}{B}} - 1}{\sum_{k=1}^{K_{n,f}} r_{n,k}^{-\alpha}} \right]. \quad (\text{A.1b})$$

For what concerns (A.1a), the derivation of (A.2) is reported to articulate the expected value on the variables we need to average out. Hence,

$$\mathbb{E}_{\mathcal{I}_{n,f}} [\mathcal{I}_{n,f} + W] = \mathbb{E}_{\substack{\{h_{n,k}\}_{\forall k}, \\ \{r_{n,k}\}_{\forall k}, \\ \bar{\phi}_{n,-f}, \tilde{\phi}_i}} \left[\sum_{\bar{k} \in \bar{\phi}_{n,-f}} |h_{n,\bar{k}}|^2 r_{n,\bar{k}}^{-\alpha} + \sum_{i \neq n} \omega_i \sum_{\bar{k} \in \tilde{\phi}_i} |h_{i,\bar{k}}|^2 r_{i,\bar{k}}^{-\alpha} \right] + W. \quad (\text{A.2})$$

For simplicity we will now refer to \bar{n} to indicate the random cardinality of the set of interfering caching nodes that have not cached the f -th content within the n -th CSA whose mean is $\mathbb{E}[\bar{\phi}_{n,-f}] = \bar{\mu}_{n,f} = \lambda_n^S (1 - p_{n,f}) 4d^2$, see (3.4b). Similarly, \tilde{n}_i stands for the set of interferers for the i -th CSA such that $i \neq n$ and $\mathbb{E}[\tilde{\phi}_i] = \tilde{\mu}_i = \lambda_i^S 4d^2$ as per (3.4c).

As a result, we obtain from (A.2) that

$$\mathbb{E}_{\mathcal{I}_{n,f}} [\mathcal{I}_{n,f} + W] = \mathbb{E}_{\bar{n}, h_n, r_n} [\bar{n} |h_n|^2 r_n^{-\alpha}] + \sum_{i \neq n}^N \omega_i \mathbb{E}_{\tilde{n}_i, h_i, r_i} [\tilde{n}_i |h_i|^2 r_i^{-\alpha}] + W, \quad (\text{A.3})$$

where the subscripts \bar{k} and \tilde{k} have been dropped for conciseness as long as the independence of the terms of both sums allows to consider each term of the sums independently.

It is known that the distribution of the squared absolute value of a circular symmetric Gaussian random variable writes as an exponential distribution $\exp(-x)$. Also, it is easy to see that the expected value for the standard exponential random variable $|h|^2$ results to be $\int_0^\infty x \exp(-x) dx = 1$. Thus, we can uncondition (A.3) with respect to the link distances and cardinalities of the two sets as

$$\varphi_{n,f}^{\text{I}}(\boldsymbol{\lambda}^{\text{S}}, p_{n,f}) = \mathbb{E}_{\mathcal{I}_{n,f}} [\mathcal{I}_{n,f} + W] = \bar{\mu}_{n,f} \bar{J}(n) + \sum_{i \neq n}^N \omega_i \tilde{\mu}_i \tilde{J}(i) + W, \quad (\text{A.4})$$

where

$$\begin{cases} \bar{J}(n) = \frac{1}{4d^2} \left(\iint_{\mathcal{D}_n \setminus \mathcal{B}_0} (x^2 + y^2)^{-\alpha/2} dx dy + \pi \right), \\ \tilde{J}(i) = \frac{1}{4d^2} \iint_{\mathcal{D}_i} (x^2 + y^2)^{-\alpha/2} dx dy, \end{cases} \quad (\text{A.5})$$

where \mathcal{B}_0 denotes the circle of unit radius centered at the UCL under investigation.

As such, we have averaged out all the random variables previously highlighted in (A.2) and therefore retrieved the expected value in (A.1a).

From Section 3.3.2, we define the user-load of a set of jointly cooperating caching nodes $K_{n,f}$ at the n -th UCL for content f as

$$\Xi_{n,f} = \max\{\xi_{n,f,1}, \xi_{n,f,2}, \dots, \xi_{n,f,K_{n,f}}\}. \quad (\text{A.6})$$

We can write the pmf of $\Xi_{n,f}$ in terms of the user-load perceived by the single caching node $\xi_{n,f}$ as

$$\Pr(\Xi_{n,f} = m) = F_{\xi_{n,f}}(m)^{K_{n,f}} - F_{\xi_{n,f}}(m-1)^{K_{n,f}},$$

where we have the number of cooperating nodes $K_{n,f}$ whose mean is $\mathbb{E}[\phi_{n,f}] = \mu_{n,f} = \lambda_n^S p_{n,f} 4d^2$ as per (3.4a). Therefore, we can work out (A.1b) as

$$\begin{aligned}
& \mathbb{E}_{\substack{\{r_{n,k}\}_{k \in \phi_{n,f}}, \\ K_{n,f}, \Xi_{n,f}}} \left[\frac{2^{\frac{\rho \Xi_{n,f}}{B}} - 1}{\sum_{k=1}^{K_{n,f}} r_{n,k}^{-\alpha}} \right] \\
&= \mathbb{E}_{\{r_{n,k}\}_{k \in \phi_{n,f}}, K_{n,f}} \left[\sum_{m=1}^M \frac{2^{\frac{\rho m}{B}} - 1}{\sum_{k=1}^{K_{n,f}} r_{n,k}^{-\alpha}} \Pr(\Xi_{n,f} = m) \right] \\
&= \mathbb{E}_{\{r_{n,k}\}_{k \in \phi_{n,f}}, K_{n,f}} \left[\sum_{m=1}^M \frac{2^{\frac{\rho m}{B}} - 1}{\sum_{k=1}^{K_{n,f}} r_{n,k}^{-\alpha}} (F_{\xi_{n,f}}(m)^{K_{n,f}} - F_{\xi_{n,f}}(m-1)^{K_{n,f}}) \right] \\
&= \mathbb{E}_{\{r_{n,k}\}_{k \in \phi_{n,f}}, K_{n,f}} \left[\int_0^\infty e^{-t \sum_{k=1}^{K_{n,f}} r_{n,k}^{-\alpha}} dt \sum_{m=1}^M (2^{\frac{\rho m}{B}} - 1) (F_{\xi_{n,f}}(m)^{K_{n,f}} - F_{\xi_{n,f}}(m-1)^{K_{n,f}}) \right] \\
&= \mathbb{E}_{\{r_{n,k}\}_{k \in \phi_{n,f}}, K_{n,f}} \left[\int_0^\infty \prod_{k=1}^{K_{n,f}} e^{-tr_{n,k}^{-\alpha}} dt \sum_{m=1}^M (2^{\frac{\rho m}{B}} - 1) (F_{\xi_{n,f}}(m)^{K_{n,f}} - F_{\xi_{n,f}}(m-1)^{K_{n,f}}) \right] \\
&\stackrel{a}{=} \mathbb{E}_{K_{n,f}} \left[\sum_{m=1}^M (2^{\frac{\rho m}{B}} - 1) \times \int_0^\infty J_n(t)^{K_{n,f}} (F_{\xi_{n,f}}(m)^{K_{n,f}} - F_{\xi_{n,f}}(m-1)^{K_{n,f}}) dt \right] \\
&\stackrel{b}{=} \sum_{m=1}^M (2^{\frac{\rho m}{B}} - 1) \int_0^\infty \Theta(t, m) dt,
\end{aligned} \tag{A.7}$$

where $\Theta(t, m)$ is defined as

$$\Theta(t, m) = \sum_{K_{n,f}=1}^{\infty} J_n(t)^{K_{n,f}} (F_{\xi_{n,f}}(m)^{K_{n,f}} - F_{\xi_{n,f}}(m-1)^{K_{n,f}}) \times \frac{\mu_{n,f}^{K_{n,f}}}{(e^{\mu_{n,f}} - 1) K_{n,f}!}. \tag{A.8}$$

Regarding (A.7), (a) follows from averaging over the link distance and (b) follows from averaging over the ZTP distribution of the set of the cooperating nodes. It is easy to see that for $K_{n,f} = 0$, we have

$$J_n(t)^0 (F_{\xi_{n,f}}(m)^0 - F_{\xi_{n,f}}(m-1)^0) \frac{\mu_{n,f}^0}{(e^{\mu_{n,f}} - 1) 0!} = 0.$$

Therefore, we can express the sum over $K_{n,f}$ to simply start from 0. This allows

us to write the unconditioning part over $K_{n,f}$ as a series for exponential functions. Thus, we get

$$\begin{aligned}\Theta(t, m) &= \frac{1}{e^{\mu_{n,f}} - 1} \sum_{K_{n,f}=0}^{\infty} \left[\frac{(\mu_{n,f} F_{\xi_{n,f}}(m) J_n(t))^{K_{n,f}}}{K_{n,f}!} - \frac{(\mu_{n,f} F_{\xi_{n,f}}(m-1) J_n(t))^{K_{n,f}}}{K_{n,f}!} \right] \\ &= \frac{1}{e^{\mu_{n,f}} - 1} \left(e^{\mu_{n,f} F_{\xi_{n,f}}(m) J_n(t)} - e^{\mu_{n,f} F_{\xi_{n,f}}(m-1) J_n(t)} \right).\end{aligned}\tag{A.9}$$

By substituting (A.9) into (A.7), $\varphi_{n,f}^D(\lambda_n^S, \mathbf{p}_n)$ is finally defined as

$$\begin{aligned}\varphi_{n,f}^D(\lambda_n^S, \mathbf{p}_n) &= \mathbb{E}_{\{r_{n,k}\}_{k \in \phi_{n,f}}, K_{n,f}, \Xi_{n,f}} \left[\frac{2^{\frac{\rho \Xi_{n,f}}{B}} - 1}{\sum_{k=1}^{K_{n,f}} r_{n,k}^{-\alpha}} \right] = \sum_{m=1}^M \left(e^{\frac{\rho m}{B}} - 1 \right) \int_0^{\infty} \Theta(t, m) dt \\ &= \frac{1}{e^{\mu_{n,f}} - 1} \int_0^{\infty} \sum_{m=1}^M \left(2^{\frac{\rho m}{B}} - 1 \right) \left(e^{\mu_{n,f} F_{\xi_{n,f}}(m) J_n(t)} - e^{\mu_{n,f} F_{\xi_{n,f}}(m-1) J_n(t)} \right) dt.\end{aligned}$$

Appendix B

B.1 Proof of (4.4)

The proposed metric is dependent on a set of network parameters. In the following, the de-conditioning process is analysed to derive an analytical form of the metric. The probability of the SINR ratio to be higher than a threshold can be directly obtained as

$$\Pr \left(\frac{P^S |h|^2 \bar{r}^{-\alpha}}{\sum_{i=0}^{|\mathcal{I}|} P^S |h_i|^2 r_i^{-\alpha} + \sum_{\bar{i}=0}^{|\bar{\mathcal{I}}|} P^S |h_{\bar{i}}|^2 r_{\bar{i}}^{-\alpha} + \sum_{m=0}^{|\mathcal{I}^M|} P^M |h_m|^2 r_m^{-\alpha} + W} > \tilde{\rho} \right)$$

$$= e^{-\frac{\tilde{\rho}}{P^S \bar{r}^{-\alpha}} \left(\sum_{i=0}^{|\mathcal{I}|} P^S |h_i|^2 r_i^{-\alpha} + \sum_{\bar{i}=0}^{|\bar{\mathcal{I}}|} P^S |h_{\bar{i}}|^2 r_{\bar{i}}^{-\alpha} + \sum_{m=0}^{|\mathcal{I}^M|} P^M |h_m|^2 r_m^{-\alpha} + W \right)},$$

given that $|h_i|^2 \sim \exp(1)$. Additionally, as long as the expected value of the channel power gain of the link of the desired information source is independent on the energy expenses of the performance metric, it can be independently performed

as indicated. Consequently, from (4.3) it can be obtained that

$$\begin{aligned}
\mathcal{G} &= \rho \left(1 - e^{-\lambda^S \pi R^2}\right) \mathbb{E} \left[\frac{e^{-\frac{\tilde{\rho}}{PS\bar{r}^{-\alpha}} \left(\sum_{i=0}^{|\mathcal{I}|} PS|h_i|^2 r_i^{-\alpha} + \sum_{i=0}^{|\bar{\mathcal{I}}|} PS|h_i|^2 r_i^{-\alpha} + \sum_{m=0}^{|\mathcal{I}^M|} P^M|h_m|^2 r_m^{-\alpha} + W \right)}}{\sum_{i=0}^{|\mathcal{I}|+1} e_A} \right] \\
&= \rho \left(1 - e^{-\lambda^S \pi R^2}\right) \mathbb{E}_{\bar{r}} \left[\mathbb{E}_{\mathcal{I},r} \left[\frac{1}{(|\mathcal{I}|+1)} \prod_{i=0}^{|\mathcal{I}|} \frac{1}{1 + \tilde{\rho}\bar{r}^\alpha r_i^{-\alpha}} \middle| \bar{r} \right] \frac{e^{-\frac{\tilde{\rho}W\bar{r}^\alpha}{PS}}}{e_A} \times \right. \\
&\quad \left. \underbrace{\mathbb{E}_{\bar{\mathcal{I}},|h|^2,r} \left[e^{-\frac{\tilde{\rho}\bar{r}^\alpha}{PS} \left(\sum_{i=0}^{|\bar{\mathcal{I}}|} PS|h_i|^2 r_i^{-\alpha} \right)} \middle| \bar{r} \right]}_{\mathcal{L}_{\bar{\mathcal{I}}} \left(\frac{\tilde{\rho}\bar{r}^\alpha}{PS} \right)} \underbrace{\mathbb{E}_{\mathcal{I}^M,|h|^2,r} \left[e^{-\frac{\tilde{\rho}\bar{r}^\alpha}{PS} \left(\sum_{m=0}^{|\mathcal{I}^M|} P^M|h_m|^2 r_m^{-\alpha} \right)} \middle| \bar{r} \right]}_{\mathcal{L}_{\mathcal{I}^M} \left(\frac{\tilde{\rho}\bar{r}^\alpha P^M}{PS} \right)} \right], \tag{B.1}
\end{aligned}$$

where the random variables are grouped with the intention to exploit the Laplacian functionals of the outer CSA and upper MBS tier interference, the inner CSA interference is arranged with the energy expenses term successively averaged out, the small-scale channel fading for the inner CSA interferences averaged out as $|h|^2 \sim \exp(1)$ and the interfering contribution from the set of nodes \mathcal{I} is finally described as a product of functions given the independent property for PPP. The obtained three expected values are conditioned on the user-node distance \bar{r} and they are independently averaged out in the following. The expected value over the inner CSA interfering point process is further developed as

$$\begin{aligned}
\mathbb{E}_{\mathcal{I},r} \left[\frac{1}{(|\mathcal{I}|+1)} \prod_{i=0}^{|\mathcal{I}|} \frac{1}{1 + \tilde{\rho}\bar{r}^\alpha r_i^{-\alpha}} \middle| \bar{r} \right] &\stackrel{a}{=} \mathbb{E}_{\mathcal{I}} \left[\frac{1}{(|\mathcal{I}|+1)} \prod_{i=0}^{|\mathcal{I}|} \mathbb{E}_r \left[\frac{1}{1 + \tilde{\rho}\bar{r}^\alpha r_i^{-\alpha}} \right] \middle| \bar{r} \right] \\
&\stackrel{b}{=} \mathbb{E}_{\mathcal{I}} \left[\sum_{i=0}^{\infty} \frac{\left(\frac{\lambda^S \pi (R^2 - \bar{r}^2)}{\pi (R^2 - \bar{r}^2)} 2\pi \int_{\bar{r}}^R \frac{r}{\tilde{\rho}\bar{r}^\alpha r^{-\alpha} + 1} dr \right)^i}{i!(i+1)} e^{-\lambda^S \pi (R^2 - \bar{r}^2)} \middle| \bar{r} \right] \\
&\stackrel{c}{=} \left[\frac{e^{2\pi\lambda^S \int_{\bar{r}}^R \frac{r}{\tilde{\rho}\bar{r}^\alpha r^{-\alpha} + 1} dr} - 1}{2\pi\lambda^S \int_{\bar{r}}^R \frac{r}{\tilde{\rho}\bar{r}^\alpha r^{-\alpha} + 1} dr} e^{-\lambda^S \pi (R^2 - \bar{r}^2)} \middle| \bar{r} \right]
\end{aligned}$$

where (a) comes from the independence of the PPP, (b) is from the de-conditioning over r and by introducing the pmf of the random number of nodes and (c) is given that the sum over the size of the interfering set lies into a closed form given the geometric series $\sum_{i=0}^{\infty} \frac{a^i}{i!(i+1)} = \frac{e^a - 1}{a}$. The Laplace functional $\mathcal{L}_{\bar{\mathcal{I}}}$ is obtained as

$$\begin{aligned} \mathcal{L}_{\bar{\mathcal{I}}}\left(\frac{\tilde{\rho}\bar{r}^{-\alpha}}{PS}\right) &= \mathbb{E}_{\bar{\mathcal{I}},r,|h|^2} \left[e^{-\frac{\tilde{\rho}\bar{r}^{-\alpha}}{PS} \left(\sum_{i=0}^{|\bar{\mathcal{I}}|} PS |h_{\bar{i}}|^2 r_{\bar{i}}^{-\alpha} \right)} \right] \stackrel{a}{=} \mathbb{E}_{\bar{\mathcal{I}}} \left[\prod_{\bar{i}=0}^{\bar{\mathcal{I}}} \mathbb{E}_{r,|h|^2} \left[e^{-\tilde{\rho}\bar{r}^{-\alpha} |h|^2 r^{-\alpha}} \right] \right] \\ &\stackrel{b}{=} \mathbb{E}_{\bar{\mathcal{I}}} \left[\prod_{\bar{i}=0}^{\bar{\mathcal{I}}} \mathbb{E}_r \left[\frac{1}{1 + \tilde{\rho}\bar{r}^{-\alpha} r^{-\alpha}} \right] \right] \\ &\stackrel{c}{=} \exp \left(-2\pi\lambda^S \int_R^\infty \left[1 - \frac{1}{1 + \tilde{\rho}\bar{r}^{-\alpha} r^{-\alpha}} \right] r \, dr \right), \end{aligned}$$

where (a) is derived from the independence property for a PPP, (b) is the result of averaging out the small-scale Rayleigh distributed fading distributed and (c) stands as the pgfl of the interfering process $\phi_{\bar{\mathcal{I}}}$. For what concerns $\mathcal{L}_{\mathcal{I}^M}$, we have

$$\begin{aligned} \mathcal{L}_{\mathcal{I}^M}\left(\frac{\tilde{\rho}\bar{r}^{-\alpha} P^M}{PS}\right) &= \mathbb{E}_{\mathcal{I}^M,r,|h|^2} \left[e^{-\frac{\tilde{\rho}\bar{r}^{-\alpha}}{PS} \left(\sum_{i=0}^{|\mathcal{I}^M|} P^M |h_{\bar{i}}|^2 r_{\bar{i}}^{-\alpha} \right)} \right] \\ &= \exp \left(-2\pi\lambda^M \int_0^\infty \left[1 - \frac{1}{1 + \frac{P^M}{PS} \tilde{\rho}\bar{r}^{-\alpha} r^{-\alpha}} \right] r \, dr \right), \end{aligned}$$

which follows from the same steps of $\mathcal{L}_{\bar{\mathcal{I}}}(s)$. We can now de-condition the metric from the random distance of the associated node. By integrating over the distribution of the link distance from the typical user to its closest network's node, *i.e.*,

$f(\bar{r}) = 2\pi\bar{r}\lambda^S \exp(-\lambda^S\pi\bar{r}^2)$, it can be achieved

$$\begin{aligned}
\mathcal{G}(\lambda^S) &= \rho \frac{2\pi\lambda^S}{e_A} \left(1 - e^{-\lambda^S\pi R^2}\right) \times \\
&\int_1^R \bar{r} e^{-\lambda^S\pi\bar{r}^2} \mathcal{L}_{\bar{I}}\left(\frac{\tilde{\rho}\bar{r}^\alpha}{PS}\right) \mathcal{L}_{\mathcal{I}^M}\left(\frac{\tilde{\rho}\bar{r}^\alpha P^M}{PS}\right) \frac{e^{2\pi\lambda^S \int_{\bar{r}}^R \frac{r}{1+\tilde{\rho}r^\alpha r^{-\alpha}} dr} - 1}{2\pi\lambda^S \int_{\bar{r}}^R \frac{r}{1+\tilde{\rho}r^\alpha r^{-\alpha}} dr} e^{-\lambda^S\pi(R^2-\bar{r}^2)} e^{-\frac{\tilde{\rho}W\bar{r}^\alpha}{PS}} d\bar{r} \\
&= \rho \frac{2\pi}{e_A} \left(1 - e^{-\lambda^S\pi R^2}\right) \times \\
&\int_1^R \bar{r} \mathcal{L}_{\bar{I}}\left(\frac{\tilde{\rho}\bar{r}^\alpha}{PS}\right) \mathcal{L}_{\mathcal{I}^M}\left(\frac{\tilde{\rho}\bar{r}^\alpha P^M}{PS}\right) \frac{e^{2\pi\lambda^S \int_{\bar{r}}^R \frac{r}{1+\tilde{\rho}r^\alpha r^{-\alpha}} dr} - 1}{2\pi \int_{\bar{r}}^R \frac{r}{1+\tilde{\rho}r^\alpha r^{-\alpha}} dr} e^{-\lambda^S\pi R^2} e^{-\frac{\tilde{\rho}W\bar{r}^\alpha}{PS}} d\bar{r},
\end{aligned} \tag{B.2}$$

which finally writes as the intended performance metric in (4.4).

B.2 Proof of Theorem 8

It is straight forward to acknowledge the existence of solution (i) directly from (4.5). The extended form of $\frac{\partial \mathcal{G}}{\partial \lambda^S}(0)$ can be written as an infinite sum of zero valued functions as

$$\frac{\partial \mathcal{G}}{\partial \lambda^S}(0) = \rho \frac{2\pi}{e_A} \int_1^R \bar{r} \frac{e^{-\frac{\tilde{\rho}W\bar{r}^\alpha}{PS}}}{\xi_1(\bar{r})} \mathcal{L}^M\left(\frac{P^M \tilde{\rho}\bar{r}^\alpha}{PS}\right) [0] d\bar{r} = 0.$$

Hence, (i) is proved. For ease of illustration, (4.5) is re-arranged to convert the search of zeros in $\frac{\partial \mathcal{G}}{\partial \lambda^S}(\lambda^S)$ to the examination of the intersecting points of two functions $\Psi_1(\lambda^S)$ and $\Psi_2(\lambda^S)$, defined as

$$\begin{aligned}
\frac{\partial \mathcal{G}}{\partial \lambda^S}(\lambda^S) = 0 &\rightarrow \Psi_1(\lambda^S) = \Psi_2(\lambda^S) \rightarrow \\
&\int_1^R \psi(\bar{r}, \lambda^S) \left[(\pi R^2 + \xi_2(\bar{r})) - e^{\xi_1(\bar{r})\lambda^S} (\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})) \right] d\bar{r} \\
&= \int_1^R \psi(\bar{r}, \lambda^S) e^{-\lambda^S\pi R^2} \left[(2\pi R^2 + \xi_2(\bar{r})) - e^{\xi_1(\bar{r})\lambda^S} (2\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})) \right] d\bar{r},
\end{aligned} \tag{B.3}$$

with

$$\psi(\bar{r}, \lambda^S) = \bar{r} \frac{e^{-\frac{\bar{\rho} W \bar{r}^\alpha}{PS}}}{\xi_1(\bar{r})} \mathcal{L}_{\mathcal{I}^M} \left(\frac{P^M \tilde{\rho} \bar{r}^\alpha}{PS} \right) e^{-\lambda^S (\pi R^2 + \xi_2(\bar{r}))}.$$

Three different density driven behaviours for the functions $\Psi_1(\lambda^S)$ and $\Psi_2(\lambda^S)$ can be detected. Specifically, both integrals assume positive values, as it can be observed from (4.7), for $\lambda^S \in [0, \lambda^-)$ where λ^- is a threshold such that $\lambda^- > 0$. Within a finite range of values $\lambda^S \in [\lambda^-, \lambda^+]$, the resulting integrals $\Psi_1(\lambda^S)$ and $\Psi_2(\lambda^S)$ stand as a combination of either positive and negative contributions. When $\lambda^S \in (\lambda^+, \infty)$, the integrals are infinite sums of negative terms, thus resulting in negative values for $\Psi_1(\lambda^S)$ and $\Psi_2(\lambda^S)$. For what concerns $\Psi_1(\lambda^S)$, these regions are defined by the function

$$\lambda_{\Psi_1}(\bar{r}) = \frac{\ln \left(\frac{\pi R^2 + \xi_2(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})} \right)}{\xi_1(\bar{r})} \quad (\text{B.4})$$

from which the following density thresholds can be obtained as

$$\lambda_{\Psi_1}^- = \min \lambda_{\Psi_1}(\bar{r}) = \left. \frac{\ln \left(\frac{\pi R^2 + \xi_2(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})} \right)}{\xi_1(\bar{r})} \right|_{\bar{r}=R^-},$$

$$\lambda_{\Psi_1}^+ = \max \lambda_{\Psi_1}(\bar{r}) = \left. \frac{\ln \left(\frac{\pi R^2 + \xi_2(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})} \right)}{\xi_1(\bar{r})} \right|_{\bar{r}=1},$$

in such a way that the integration of $\Psi_1(\lambda^S)$ is positive for $\lambda^S < \lambda_{\Psi_1}^-$ and negative for $\lambda^S > \lambda_{\Psi_1}^+$. The same conclusions can be attained for $\Psi_2(\lambda^S)$ by considering the function

$$\lambda_{\Psi_2}(\bar{r}) = \frac{\ln \left(\frac{2\pi R^2 + \xi_2(\bar{r})}{2\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})} \right)}{\xi_1(\bar{r})} \quad (\text{B.5})$$

with the following thresholds

$$\lambda_{\Psi_2}^- = \min \lambda_{\Psi_2}(\bar{r}) = \left. \frac{\ln \left(\frac{2\pi R^2 + \xi_2(\bar{r})}{2\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})} \right)}{\xi_1(\bar{r})} \right|_{\bar{r}=R^-},$$

$$\lambda_{\Psi_2}^+ = \min \lambda_{\Psi_2}(\bar{r}) = \frac{\ln\left(\frac{2\pi R^2 + \xi_2(\bar{r})}{2\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})}\right)}{\xi_1(\bar{r})} \Bigg|_{\bar{r}=1}.$$

The proof of the decreasing monotonic trend of $\lambda_{\Psi_1}(\bar{r})$ and $\lambda_{\Psi_2}(\bar{r})$ is shown in Appendix B.6. It is straight-forward to notice that

$$\lambda_{\Psi_1}(\bar{r}) \geq \lambda_{\Psi_2}(\bar{r}) \rightarrow \frac{\ln\left(\frac{\pi R^2 + \xi_2(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})}\right)}{\xi_1(\bar{r})} \geq \frac{\ln\left(\frac{2\pi R^2 + \xi_2(\bar{r})}{2\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})}\right)}{\xi_1(\bar{r})} \quad \forall \bar{r} \in [1, R),$$

where the equality holds for $\bar{r} = R$ and with the direct consequence that $\lambda_{\Psi_1}^- \geq \lambda_{\Psi_2}^+ > \lambda_{\Psi_2}^-$. This leads to the conclusion that $\Psi_1(\lambda^S)$ represents a strict upper-bound for $\Psi_2(\lambda^S)$ in the range $0 < \lambda^S < \lambda_{\Psi_1}^-$. As a direct result, no stationary point for \mathcal{G} lies within $0 < \lambda^S < \lambda_{\Psi_1}^-$. The following limits hold

$$\begin{aligned} \lim_{\lambda^S \rightarrow \infty} \Psi_1(\lambda^S) &= 0^-, \\ \lim_{\lambda^S \rightarrow \infty} \Psi_2(\lambda^S) &= 0^-. \end{aligned} \tag{B.6}$$

Note that $\Psi_2(\lambda^S)$ tends to zero more rapidly than $\Psi_1(\lambda^S)$, allowing us to conclude that $\Psi_2(\lambda^S)$ stands as an upper-bound for $\Psi_1(\lambda^S)$ for higher λ^S . With that, the infinite solution (ii) can be seen directly from (B.6) as the two functions can be arbitrarily close to each other when λ^S increases. Since the change of upper-bounding function occurs from lower to higher density values, it is valid to say that the two functions have a minimum of one intersection point and that it lies within $\lambda_{\Psi_1}^- < \lambda^S < \infty$. From $\lambda^S = \lambda_{\Psi_1}^-$, we can increase λ^S until a solution λ^\dagger is found such that

$$\Psi_1(\lambda^\dagger) = \Psi_2(\lambda^\dagger).$$

Based on the fact that $\lambda_{\Psi_1}^- \geq \lambda_{\Psi_2}^+ > \lambda_{\Psi_2}^-$ it is legitimate to state that the intersection point λ^\dagger occurs with $\Psi_1(\lambda^\dagger) < 0$ and $\Psi_2(\lambda^\dagger) < 0$. More specifically, given $\lambda^\dagger > \lambda_{\Psi_2}^+$, it is known from the previous analysis that $\Psi_2(\lambda^\dagger + \epsilon) < 0$, for any $\epsilon > 0$. With this in mind, it is necessary and sufficient to prove that $|\Psi_2(\lambda^\dagger + \epsilon)| < |\Psi_1(\lambda^\dagger + \epsilon)|$ for $\epsilon \in [0, \infty)$ such that it can be concluded that $\Psi_1(\lambda^\dagger)$ and $\Psi_2(\lambda^\dagger)$ have one only intersection within $\lambda_{\Psi_1}^- < \lambda^S < \infty$. In other words, proving that $|\Psi_2(\lambda^\dagger + \epsilon)| < |\Psi_1(\lambda^\dagger + \epsilon)|$ for $\epsilon \in [0, \infty)$ would indeed con-

firm $\Psi_2(\lambda^\dagger + \epsilon)$ is an upper-bound of $\Psi_1(\lambda^\dagger + \epsilon)$ from $\lambda^S \in [\lambda^\dagger, \infty)$. Hence,

$$\begin{aligned}\Psi_1(\lambda^\dagger + \epsilon) &= \int_1^R \psi(\bar{r}, \lambda^\dagger + \epsilon) \left[(\pi R^2 + \xi_2(\bar{r})) - e^{\xi_1(\bar{r})(\lambda^\dagger + \epsilon)} (\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})) \right] d\bar{r} \\ &= \int_1^R \psi(\bar{r}, \lambda^\dagger) e^{-\epsilon(\pi R^2 + \xi_2(\bar{r}))} \left[(\pi R^2 + \xi_2(\bar{r})) - e^{\xi_1(\bar{r})(\lambda^\dagger + \epsilon)} (\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})) \right] d\bar{r}\end{aligned}$$

and

$$\begin{aligned}\Psi_2(\lambda^\dagger + \epsilon) &= \int_1^R \psi(\bar{r}, \lambda^\dagger + \epsilon) e^{-(\lambda^\dagger + \epsilon)\pi R^2} \left[(2\pi R^2 + \xi_2(\bar{r})) - e^{\xi_1(\bar{r})(\lambda^\dagger + \epsilon)} (2\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})) \right] d\bar{r} \\ &= e^{-\epsilon\pi R^2} \int_1^R \psi(\bar{r}, \lambda^\dagger) e^{-\lambda^\dagger\pi R^2} e^{-\epsilon(\pi R^2 + \xi_2(\bar{r}))} \left[\begin{array}{c} (\pi R^2 + \xi_2(\bar{r})) - \\ e^{\xi_1(\bar{r})(\lambda^\dagger + \epsilon)} (\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})) - \\ \pi R^2 (e^{(\lambda^\dagger + \epsilon)\xi_1(\bar{r})} - 1) \end{array} \right] d\bar{r} \\ &= e^{-(\epsilon + \lambda^\dagger)\pi R^2} \underbrace{\int_1^R \psi(\bar{r}, \lambda^\dagger) e^{-\epsilon(\pi R^2 + \xi_2(\bar{r}))} \left[\begin{array}{c} (\pi R^2 + \xi_2(\bar{r})) - \\ e^{\xi_1(\bar{r})(\lambda^\dagger + \epsilon)} (\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})) \end{array} \right] d\bar{r}}_{\Psi_1(\lambda^\dagger)} \\ &\quad - e^{-(\epsilon + \lambda^\dagger)\pi R^2} \int_1^R \psi(\bar{r}, \lambda^\dagger) e^{-\epsilon(\pi R^2 + \xi_2(\bar{r}))} \pi R^2 (1 - e^{(\lambda^\dagger + \epsilon)\xi_1(\bar{r})}) d\bar{r},\end{aligned}$$

which leads to $|\Psi_2(\lambda^\dagger + \epsilon)| < |\Psi_1(\lambda^\dagger + \epsilon)|$, for any $\epsilon > 0$, concluding that one unique solution lies within $\lambda^S \in [\lambda_{\Psi_1}^-, \infty)$, *i.e.*, (iii).

B.3 Proof of (4.10)

From the conditioned analytical SCDP, we obtain

$$\begin{aligned}
\mathcal{G}^{\text{SCDP}}(\lambda^S) &= \left(1 - e^{-\lambda^S \pi R^2}\right) \mathbb{E} \left[e^{-\frac{\tilde{\rho}}{P^S \tilde{r}^{-\alpha}} \left(\sum_{i=0}^{|\mathcal{I}|} P^S |h_i|^2 r_i^{-\alpha} + \sum_{i=0}^{|\tilde{\mathcal{I}}|} P^S |h_{\tilde{i}}|^2 r_{\tilde{i}}^{-\alpha} + \sum_{m=0}^{|\mathcal{I}^M|} P^M |h_m|^2 r_m^{-\alpha} + W \right)} \right] \\
&\stackrel{a}{=} \left(1 - e^{-\lambda^S \pi R^2}\right) \times \\
&\mathbb{E}_{\tilde{r}} \left[\underbrace{\mathbb{E}_{\mathcal{I}+\tilde{\mathcal{I}},r,|h|^2} \left[e^{-\tilde{\rho} \sum_{i=0}^{|\mathcal{I}+\tilde{\mathcal{I}}|} |h_i|^2 r_i^{-\alpha}} \middle| \tilde{r} \right]}_{\mathcal{L}_{\mathcal{I}+\tilde{\mathcal{I}}}(\tilde{\rho} \tilde{r}^\alpha)} \underbrace{\mathbb{E}_{\mathcal{I}^M,r,|h|^2} \left[e^{-\frac{P^M \tilde{\rho} \tilde{r}^\alpha}{P^S} \sum_{m=0}^{|\mathcal{I}^M|} |h_m|^2 r_m^{-\alpha}} \middle| \tilde{r} \right]}_{\mathcal{L}_{\mathcal{I}^M}\left(\frac{P^M \tilde{\rho} \tilde{r}^\alpha}{P^S}\right)} e^{-\frac{\tilde{\rho}}{P^S \tilde{r}^{-\alpha}} W} \right], \tag{B.7}
\end{aligned}$$

where in (a) the inner and outer CSA interferers are considered as a single PPP, and the interference contributions are expressed as the Laplace functionals of the SBS and MBS tiers. Regarding $\mathcal{L}_{\mathcal{I}+\tilde{\mathcal{I}}}(s)$, it can be retrieved as

$$\begin{aligned}
\mathcal{L}_{\mathcal{I}+\tilde{\mathcal{I}}}(s) &= \mathbb{E}_{\mathcal{I}+\tilde{\mathcal{I}},r,|h|^2} \left[e^{-s \left(\sum_{i=0}^{|\mathcal{I}+\tilde{\mathcal{I}}|} |h_i|^2 r_i^{-\alpha} \right)} \right] \stackrel{a}{=} \mathbb{E}_{\mathcal{I}+\tilde{\mathcal{I}}} \left[\prod_{i=0}^{|\mathcal{I}+\tilde{\mathcal{I}}|} \mathbb{E}_{r,|h|^2} \left[e^{-s |h|^2 r^{-\alpha}} \right] \right] \\
&\stackrel{b}{=} \mathbb{E}_{\mathcal{I}+\tilde{\mathcal{I}}} \left[\prod_{i=0}^{|\mathcal{I}+\tilde{\mathcal{I}}|} \mathbb{E}_r \left[\frac{1}{1 + s r^{-\alpha}} \right] \right] \\
&\stackrel{c}{=} \exp \left(-2\pi \lambda^S \int_{\tilde{r}}^{\infty} \left[1 - \frac{1}{1 + s r^{-\alpha}} \right] r \, dr \right),
\end{aligned}$$

where (a) is derived from the independence property for a PPP, (b) is the result of averaging out the small-scale Rayleigh distributed fading, whose power gain is a standard exponential random variable $|h|^2 \sim \exp(1)$ and (c) stands as the pgfl of the interfering process $\phi_{\mathcal{I}+\tilde{\mathcal{I}}}$. For what concerns $\mathcal{L}_{\mathcal{I}^M}$, we have

$$\mathcal{L}_{\mathcal{I}^M}(s) = \mathbb{E}_{\mathcal{I}^M,r,|h|^2} \left[e^{-s \left(\sum_{i=0}^{|\mathcal{I}^M|} |h_i|^2 r_i^{-\alpha} \right)} \right] = \exp \left(-2\pi \lambda^M \int_0^{\infty} \left[1 - \frac{1}{1 + s r^{-\alpha}} \right] r \, dr \right),$$

which follows from the same steps in the derivation of $\mathcal{L}_{\mathcal{I}+\bar{\mathcal{I}}}(s)$. By integrating over the distribution of the link distance from the typical user to its closest network's node, *i.e.*, $f(\bar{r}) = 2\pi\bar{r}\lambda^S \exp(-\lambda^S\pi\bar{r}^2)$, we have

$$\begin{aligned} \mathcal{G}^{\text{SCDP}}(\lambda^S) &= \left(1 - e^{-\lambda^S\pi R^2}\right) \int_1^R 2\pi\bar{r}\lambda^S e^{-\lambda^S\pi\bar{r}^2} \mathcal{L}_{\mathcal{I}+\bar{\mathcal{I}}}(s) \Big|_{s=\tilde{\rho}\bar{r}^\alpha} \mathcal{L}_{\mathcal{I}^{\text{M}}}(s) \Big|_{s=\frac{P^{\text{M}}\tilde{\rho}\bar{r}^\alpha}{P^{\text{S}}}} e^{-\frac{\tilde{\rho}W\bar{r}^\alpha}{P^{\text{S}}}} d\bar{r} \\ &= 2\pi\lambda^S \left(1 - e^{-\lambda^S\pi R^2}\right) \int_1^R \bar{r} e^{-\lambda^S(\pi\bar{r}^2+\xi_3(\bar{r}))} \mathcal{L}_{\mathcal{I}^{\text{M}}}(s) \Big|_{s=\frac{P^{\text{M}}\tilde{\rho}\bar{r}^\alpha}{P^{\text{S}}}} e^{-\frac{\tilde{\rho}W\bar{r}^\alpha}{P^{\text{S}}}} d\bar{r}, \end{aligned}$$

with $\xi_3(\bar{r}) = 2\pi \int_{\bar{r}}^\infty \left[1 - \frac{1}{1+\tilde{\rho}\bar{r}^\alpha r^{-\alpha}}\right] r dr$.

B.4 Theorem 8 applied to $\mathcal{G}^{\text{SCDP}}$

The derivative of the metric $\mathcal{G}^{\text{SCDP}}(\lambda^S)$ is simply obtained as

$$\begin{aligned} \frac{\partial \mathcal{G}^{\text{SCDP}}}{\partial \lambda^S}(\lambda^S) &= 2\pi \int_1^R \bar{r} e^{-\frac{\tilde{\rho}W\bar{r}^\alpha}{P^{\text{S}}}} \mathcal{L}_{\mathcal{I}^{\text{M}}}\left(\frac{P^{\text{M}}\tilde{\rho}\bar{r}^\alpha}{P^{\text{S}}}\right) \left(e^{-\lambda^S(\pi\bar{r}^2+\xi_3(\bar{r}))} - e^{-\lambda^S(\pi R^2+\pi\bar{r}^2+\xi_3(\bar{r}))} \right) d\bar{r} - \\ &2\pi\lambda^S \int_1^R \bar{r} e^{-\frac{\tilde{\rho}W\bar{r}^\alpha}{P^{\text{S}}}} \mathcal{L}_{\mathcal{I}^{\text{M}}}\left(\frac{P^{\text{M}}\tilde{\rho}\bar{r}^\alpha}{P^{\text{S}}}\right) \left(\begin{array}{c} (\pi\bar{r}^2 + \xi_3(\bar{r})) e^{-\lambda^S(\pi\bar{r}^2+\xi_3(\bar{r}))} - \\ (\pi R^2 + \pi\bar{r}^2 + \xi_3(\bar{r})) e^{-\lambda^S(\pi R^2+\pi\bar{r}^2+\xi_3(\bar{r}))} \end{array} \right) d\bar{r}, \end{aligned} \quad (\text{B.8})$$

where $\xi_3(\bar{r}) = 2\pi \int_{\bar{r}}^\infty \left[1 - \frac{1}{1+\tilde{\rho}\bar{r}^\alpha r^{-\alpha}}\right] r dr$. It is easy to notice from (B.8) that the first solution (*i*) does not hold, *i.e.*, $\frac{\partial \mathcal{G}^{\text{SCDP}}}{\partial \lambda^S}(0) > 0$, concluding the stationary points can be found for $\lambda^S > 0$. The problem of finding the zeros of the derivative is translated to the search of the intersections by defining Ψ_1^{SCDP} and Ψ_2^{SCDP} as

$$\begin{aligned} \frac{\partial \mathcal{G}^{\text{SCDP}}}{\partial \lambda^S}(\lambda^S) = 0 &\rightarrow \Psi_1^{\text{SCDP}}(\lambda^S) = \Psi_2^{\text{SCDP}}(\lambda^S) \rightarrow \\ &\int_1^R \bar{r} e^{-\frac{\tilde{\rho}W\bar{r}^\alpha}{P^{\text{S}}}} \mathcal{L}_{\mathcal{I}^{\text{M}}}\left(\frac{P^{\text{M}}\tilde{\rho}\bar{r}^\alpha}{P^{\text{S}}}\right) \left(e^{-\lambda^S(\pi\bar{r}^2+\xi_3(\bar{r}))} - e^{-\lambda^S(\pi R^2+\pi\bar{r}^2+\xi_3(\bar{r}))} \right) d\bar{r} = \\ &\lambda^S \int_1^R \bar{r} e^{-\frac{\tilde{\rho}W\bar{r}^\alpha}{P^{\text{S}}}} \mathcal{L}_{\mathcal{I}^{\text{M}}}\left(\frac{P^{\text{M}}\tilde{\rho}\bar{r}^\alpha}{P^{\text{S}}}\right) \left(\begin{array}{c} (\pi\bar{r}^2 + \xi_3(\bar{r})) e^{-\lambda^S(\pi\bar{r}^2+\xi_3(\bar{r}))} - \\ (\pi R^2 + \pi\bar{r}^2 + \xi_3(\bar{r})) e^{-\lambda^S(\pi R^2+\pi\bar{r}^2+\xi_3(\bar{r}))} \end{array} \right) d\bar{r}, \end{aligned} \quad (\text{B.9})$$

where it can be noted $\Psi_1^{\text{SCDP}}(\lambda^{\text{S}}) \geq 0$, $\forall \lambda^{\text{S}} \in [0, \infty)$. The following function detects the density driven behaviour of $\Psi_2^{\text{SCDP}}(\lambda^{\text{S}})$

$$\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r}) = \ln \left(\frac{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})}{\pi \bar{r}^2 + \xi_3(\bar{r})} \right) \frac{1}{\pi R^2}, \quad (\text{B.10})$$

from which the following thresholds can be obtained

$$\lambda_{\Psi_2^{\text{SCDP}}}^- = \min \lambda_{\Psi_2^{\text{SCDP}}}(\bar{r}) = \ln \left(\frac{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})}{\pi \bar{r}^2 + \xi_3(\bar{r})} \right) \frac{1}{\pi R^2} \Bigg|_{\bar{r}=R},$$

$$\lambda_{\Psi_2^{\text{SCDP}}}^+ = \max \lambda_{\Psi_2^{\text{SCDP}}}(\bar{r}) = \ln \left(\frac{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})}{\pi \bar{r}^2 + \xi_3(\bar{r})} \right) \frac{1}{\pi R^2} \Bigg|_{\bar{r}=1},$$

in such a way $\Psi_2^{\text{SCDP}}(\lambda^{\text{S}}) < 0$ for $\lambda^{\text{S}} \in [0, \lambda_{\Psi_2^{\text{SCDP}}}^-)$ and $\Psi_2^{\text{SCDP}}(\lambda^{\text{S}}) > 0$ for $\lambda^{\text{S}} \in (\lambda_{\Psi_2^{\text{SCDP}}}^+, \infty)$. The monotonic decreasing trend of $\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r})$ with respect to \bar{r} is demonstrated in Appendix B.7. Given the expressions for $\lambda_{\Psi_2^{\text{SCDP}}}^-$ and $\lambda_{\Psi_2^{\text{SCDP}}}^+$, it can be concluded that the intersection in (B.9) lies within $\lambda_{\Psi_2^{\text{SCDP}}}^+ \leq \bar{\lambda}^* < \infty$. From $\lambda^{\text{S}} = \lambda_{\Psi_2^{\text{SCDP}}}^+$ we can increase λ^{S} until a solution λ^\dagger is found

$$\Psi_1^{\text{SCDP}}(\lambda^\dagger) = \Psi_2^{\text{SCDP}}(\lambda^\dagger).$$

It is legitimate to state that the intersection point λ^\dagger occurs with $\Psi_1^{\text{SCDP}}(\lambda^\dagger + \epsilon) > 0$ and $\Psi_2^{\text{SCDP}}(\lambda^\dagger + \epsilon) > 0$ for $\epsilon \in [0, \infty)$, since $\lambda^\dagger > \lambda_{\Psi_2^{\text{SCDP}}}^+$. The following limits hold

$$\lim_{\lambda^{\text{S}} \rightarrow \infty} \Psi_1^{\text{SCDP}}(\lambda^{\text{S}}) = 0^+,$$

$$\lim_{\lambda^{\text{S}} \rightarrow \infty} \Psi_2^{\text{SCDP}}(\lambda^{\text{S}}) = 0^+, \quad (\text{B.11})$$

with $\Psi_1^{\text{SCDP}}(\lambda^{\text{S}})$ attaining the zero more rapidly than $\Psi_2^{\text{SCDP}}(\lambda^{\text{S}})$. From (B.11), the solution (ii) can be easily observed as the two functions get arbitrarily close to each other as λ^{S} is increase. From the previous analysis we know that $\Psi_1^{\text{SCDP}}(\lambda^{\text{S}})$ is an upper-bound for $\Psi_2^{\text{SCDP}}(\lambda^{\text{S}})$ when $0 < \lambda^{\text{S}} < \lambda_{\Psi_2^{\text{SCDP}}}^+$. With this in mind, it is necessary and sufficient to prove that $|\Psi_1^{\text{SCDP}}(\lambda^\dagger + \epsilon)| < |\Psi_2^{\text{SCDP}}(\lambda^\dagger + \epsilon)|$ for $\epsilon \in [0, \infty)$ such that it can be concluded $\Psi_1^{\text{SCDP}}(\lambda^\dagger)$ and $\Psi_2^{\text{SCDP}}(\lambda^\dagger)$ have one only

intersection within $\lambda_{\Psi_2^{\text{SCDP}}}^+ < \lambda^S < \infty$. Hence,

$$\begin{aligned}\Psi_1^{\text{SCDP}}(\lambda^\dagger + \epsilon) &= e^{-\epsilon} \int_1^R \bar{r} e^{-\frac{\bar{\rho} W \bar{r}^\alpha}{PS}} \mathcal{L}_{\mathcal{IM}} \left(\frac{P^M \tilde{\rho} \bar{r}^\alpha}{PS} \right) \left(e^{-\lambda^\dagger(\pi \bar{r}^2 + \xi_3(\bar{r}))} - e^{-\lambda^\dagger(\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r}))} \right) d\bar{r} \\ &= e^{-\epsilon} \Psi_1^{\text{SCDP}}(\lambda^\dagger),\end{aligned}$$

and

$$\begin{aligned}\Psi_2^{\text{SCDP}}(\lambda^\dagger + \epsilon) &= (\lambda^\dagger + \epsilon) \int_1^R \bar{r} e^{-\frac{\bar{\rho} W \bar{r}^\alpha}{PS}} \mathcal{L}_{\mathcal{IM}} \left(\frac{P^M \tilde{\rho} \bar{r}^\alpha}{PS} \right) \left(\begin{array}{c} (\pi \bar{r}^2 + \xi_3(\bar{r})) e^{-(\lambda^\dagger + \epsilon)(\pi \bar{r}^2 + \xi_3(\bar{r}))} - \\ (\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})) e^{-(\lambda^\dagger + \epsilon)(\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r}))} \end{array} \right) d\bar{r} \\ &= \lambda^\dagger e^{-\epsilon} \int_1^R \bar{r} e^{-\frac{\bar{\rho} W \bar{r}^\alpha}{PS}} \mathcal{L}_{\mathcal{IM}} \left(\frac{P^M \tilde{\rho} \bar{r}^\alpha}{PS} \right) \left(\begin{array}{c} (\pi \bar{r}^2 + \xi_3(\bar{r})) e^{-\lambda^\dagger(\pi \bar{r}^2 + \xi_3(\bar{r}))} - \\ (\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})) e^{-\lambda^\dagger(\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r}))} \end{array} \right) d\bar{r} \\ &\quad + \epsilon e^{-\epsilon} \int_1^R \bar{r} e^{-\frac{\bar{\rho} W \bar{r}^\alpha}{PS}} \mathcal{L}_{\mathcal{IM}} \left(\frac{P^M \tilde{\rho} \bar{r}^\alpha}{PS} \right) \left(\begin{array}{c} (\pi \bar{r}^2 + \xi_3(\bar{r})) e^{-\lambda^\dagger(\pi \bar{r}^2 + \xi_3(\bar{r}))} - \\ (\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})) e^{-\lambda^\dagger(\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r}))} \end{array} \right) d\bar{r} \\ &= e^{-\epsilon} \Psi_2^{\text{SCDP}}(\lambda^\dagger) \left(1 + \frac{\epsilon}{\lambda^\dagger} \right).\end{aligned}$$

Knowing that, by assumption, equality holds $\Psi_1^{\text{SCDP}}(\lambda^\dagger) = \Psi_2^{\text{SCDP}}(\lambda^\dagger)$, it can be obtained

$$\begin{aligned}|e^{-\epsilon} \Psi_1^{\text{SCDP}}(\lambda^\dagger)| &< \left| e^{-\epsilon} \Psi_2^{\text{SCDP}}(\lambda^\dagger) \left(1 + \frac{\epsilon}{\lambda^\dagger} \right) \right| \\ 1 &< \left| \left(1 + \frac{\epsilon}{\lambda^\dagger} \right) \right|,\end{aligned}$$

given that $\frac{\epsilon}{\lambda^\dagger} > 0$. It is demonstrated that one unique solution lies within $\lambda_{\Psi_2^{\text{SCDP}}}^+ < \bar{\lambda}^* < \infty$, *i.e.*, (iii).

B.5 Proof of Theorem 9

Using Theorem 8 for $\mathcal{G}^{\text{SCDP}}(\lambda^S)$, we have $\bar{\lambda}^* > \lambda_{\Psi_2^{\text{SCDP}}}^+ > \lambda_{\Psi_2^{\text{SCDP}}}^-$. It can be demonstrated that $\lambda_{\Psi_2^{\text{SCDP}}} \geq \lambda_{\Psi_2}$, (where $\lambda_{\Psi_2^{\text{SCDP}}}$ is from (B.10) and λ_{Ψ_2} is derived

from (B.5)) so

$$\begin{aligned} \ln \left(\frac{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})}{\pi \bar{r}^2 + \xi_3(\bar{r})} \right) \frac{1}{\pi R^2} &\geq \ln \left(\frac{\pi R^2 + \xi_2(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})} \right) \frac{1}{\xi_1(\bar{r})} \\ \left(\frac{\pi \bar{r}^2 + \xi_3(\bar{r})}{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})} \right)^{-\xi_1(\bar{r})} &\geq \left(\frac{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})}{\pi R^2 + \xi_2(\bar{r})} \right)^{-\pi R^2}, \end{aligned}$$

which follows from knowing $\xi_1(\bar{r}) > 0 \forall \bar{r} \in [1, R)$ and by applying the exponential function on both sides. As the exponent of the left hand-side can be upper-bounded as $\xi_1(\bar{r}) \leq \pi(R^2 - \bar{r}^2)$, it is necessary to prove that

$$\begin{aligned} \left(\frac{\pi \bar{r}^2 + \xi_3(\bar{r})}{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})} \right)^{-\pi(R^2 - \bar{r}^2)} &\geq \left(\frac{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})}{\pi R^2 + \xi_2(\bar{r})} \right)^{-\pi R^2} \\ \left(\frac{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})}{\pi \bar{r}^2 + \xi_3(\bar{r})} \right)^{-\pi \bar{r}^2} &\geq \left(\frac{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})}{\pi R^2 + \xi_2(\bar{r})} \right)^{-\pi R^2}, \end{aligned}$$

where some terms been simplified given that $\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r}) = \pi \bar{r}^2 + \xi_3(\bar{r})$. Moreover, knowing that $\pi R^2 + \xi_2(\bar{r}) > \pi \bar{r}^2 + \xi_3(\bar{r}) \forall \bar{r} \in [1, R)$, the inequality holds and thus the upper-bound $\lambda_{\Psi_2^{\text{SCDP}}} \geq \lambda_{\Psi_2}$ is strictly valid $\forall \bar{r} \in [1, R)$ with the equality standing at $\bar{r} = R$.

The direct consequence of this upper-bound is that, given $\bar{\lambda}^* > \lambda_{\Psi_2^{\text{SCDP}}}^+ > \lambda_{\Psi_2}^+$, both $\Psi_1(\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r})) < 0$ and $\Psi_2(\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r})) < 0$, previously defined in (B.3). Consequently, it is necessary to prove the condition $|\Psi_2(\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r}))| < |\Psi_1(\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r}))|$ to demonstrate that the intersection $\Psi_1^{\text{SCDP}}(\bar{\lambda}^*) = \Psi_2^{\text{SCDP}}(\bar{\lambda}^*)$ occurs for $\bar{\lambda}^* > \lambda^*$. The indicated condition is easily satisfied if $\Psi_2(\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r}))$ stands as an upper-bound of $\Psi_1(\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r}))$. To begin with, we write

$$\begin{aligned} \Psi_1(\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r})) &= \int_1^R \psi(\bar{r}) \left(\frac{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})}{\pi \bar{r}^2 + \xi_3(\bar{r})} \right)^{\frac{\pi R^2 + \xi_2(\bar{r})}{\pi R^2}} \times \\ &\left[(\pi R^2 + \xi_2(\bar{r})) - \left(\frac{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})}{\pi \bar{r}^2 + \xi_3(\bar{r})} \right)^{-\frac{\xi_1(\bar{r})}{\pi R^2}} (\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})) \right] d\bar{r}, \end{aligned} \quad (\text{B.12})$$

$$\Psi_2 \left(\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r}) \right) = \int_1^R \psi(\bar{r}) \left(\frac{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})}{\pi \bar{r}^2 + \xi_3(\bar{r})} \right)^{\frac{\pi R^2 + \xi_2(\bar{r})}{\pi R^2}} \left(\frac{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})}{\pi \bar{r}^2 + \xi_3(\bar{r})} \right)^{-1} \times$$

$$\left[(2\pi R^2 + \xi_2(\bar{r})) - \left(\frac{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})}{\pi \bar{r}^2 + \xi_3(\bar{r})} \right)^{-\frac{\xi_1(\bar{r})}{\pi R^2}} (2\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})) \right] d\bar{r}, \quad (\text{B.13})$$

with

$$\psi(\bar{r}) = \bar{r} \frac{e^{-\frac{\partial W_{\bar{r}}^\alpha}{PS}}}{\xi_1(\bar{r})} \mathcal{L}_{\mathcal{I}^M} \left(\frac{P^M \tilde{\rho} \bar{r}^\alpha}{PS} \right).$$

From (B.12) and (B.13), it is necessary and sufficient for the integrand function of $\Psi_2 \left(\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r}) \right)$ to be a lower-bound of the integrand function of $\Psi_1 \left(\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r}) \right)$ to validate the inequality on the absolute values of the two function. This occurs for

$$\left[(2\pi R^2 + \xi_2(\bar{r})) - \left(\frac{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})}{\pi \bar{r}^2 + \xi_3(\bar{r})} \right)^{-\frac{\xi_1(\bar{r})}{\pi R^2}} (2\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})) \right] \left(\frac{\pi \bar{r}^2 + \xi_3(\bar{r})}{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})} \right)$$

$$> \left[(\pi R^2 + \xi_2(\bar{r})) - \left(\frac{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})}{\pi \bar{r}^2 + \xi_3(\bar{r})} \right)^{-\frac{\xi_1(\bar{r})}{\pi R^2}} (\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})) \right],$$

which after some straight-forward steps, it is equivalent to the following inequality

$$\pi R^2 \left[1 - \left(\frac{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})}{\pi \bar{r}^2 + \xi_3(\bar{r})} \right)^{-\frac{\xi_1(\bar{r})}{\pi R^2}} \right] \left(\frac{\pi \bar{r}^2 + \xi_3(\bar{r})}{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})} \right) > 0. \quad (\text{B.14})$$

It is easy to see (B.14) always holds as the product of three non-negative components.

B.6 Proof of Monotonic Decreasing Trend for $\lambda_{\Psi_1}(\bar{r})$ and $\lambda_{\Psi_2}(\bar{r})$

Consider the derivative performed over generic functions

$$\frac{\partial}{\partial \bar{r}} \left(\ln \left(\frac{f(\bar{r})}{g(\bar{r})} \right) \frac{1}{h(\bar{r})} \right) = \frac{\frac{h(\bar{r}) \frac{\partial}{\partial \bar{r}} f(\bar{r})}{f(\bar{r})} - \ln \left(\frac{f(\bar{r})}{g(\bar{r})} \right) \frac{\partial}{\partial \bar{r}} h(\bar{r}) - \frac{h(\bar{r}) \frac{\partial}{\partial \bar{r}} g(\bar{r})}{g(\bar{r})}}{h(\bar{r})^2}. \quad (\text{B.15})$$

It is straight-forward to observe that (B.15) is always negative whether the following inequality holds

$$h(\bar{r}) \left(\frac{\frac{\partial}{\partial \bar{r}} f(\bar{r})}{f(\bar{r})} - \frac{\frac{\partial}{\partial \bar{r}} g(\bar{r})}{g(\bar{r})} \right) < \ln \left(\frac{f(\bar{r})}{g(\bar{r})} \right) \frac{\partial}{\partial \bar{r}} h(\bar{r}). \quad (\text{B.16})$$

The left side of the expression (B.16) is expressed with the terms of $\lambda_{\Psi_1}(\bar{r})$ and the following result is obtained

$$\xi_1(\bar{r}) \left(\frac{\frac{\partial}{\partial \bar{r}} \xi_2(\bar{r})}{\pi R^2 + \xi_2(\bar{r})} - \frac{\frac{\partial}{\partial \bar{r}} \xi_2(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})} + \frac{\frac{\partial}{\partial \bar{r}} \xi_1(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})} \right),$$

where it is easy to notice

$$\xi_1(\bar{r}) \left(\frac{\frac{\partial}{\partial \bar{r}} \xi_2(\bar{r})}{\pi R^2 + \xi_2(\bar{r})} - \frac{\frac{\partial}{\partial \bar{r}} \xi_2(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})} \right) < 0,$$

given $\xi_1(\bar{r}) > 0$ and $\frac{\partial}{\partial \bar{r}} \xi_2(\bar{r}) > 0$. With this in mind, the focus is now on the right side of the expression (B.16) to stand as a positive quantity, with the residual term from (B.6) included as

$$\frac{\partial}{\partial \bar{r}} \xi_1(\bar{r}) \left(\ln \left(\frac{\pi R^2 + \xi_2(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})} \right) - \frac{\xi_1(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})} \right) > 0, \quad (\text{B.17})$$

such that, given $\frac{\partial}{\partial \bar{r}} \xi_1(\bar{r}) < 0$, the inequality holds when

$$\ln \left(\frac{\pi R^2 + \xi_2(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})} \right) - \frac{\xi_1(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})} < 0.$$

The logarithmic function can be upper-bounded as $\ln(x) \leq \frac{x-1}{\sqrt{x}}$. Therefore, (B.17) can be considered to be satisfied whether the following is true

$$\frac{\frac{\xi_1(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})}}{\sqrt{\frac{\pi R^2 + \xi_2(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})}}} - \frac{\xi_1(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})} < 0$$

which is easy to notice it is true since

$$\sqrt{\frac{\pi R^2 + \xi_2(\bar{r})}{\pi R^2 + \xi_2(\bar{r}) - \xi_1(\bar{r})}} > 1.$$

To fully complete this proof it is necessary to demonstrate the assumptions $\frac{\partial}{\partial \bar{r}} \xi_1(\bar{r}) < 0$ and $\frac{\partial}{\partial \bar{r}} \xi_2(\bar{r}) > 0$. To demonstrate the decreasing slope of $\xi_1(\bar{r})$, Leibniz integral rule is resorted and the following result is achieved

$$\frac{\partial}{\partial \bar{r}} \xi_1(\bar{r}) = -\frac{\bar{r}}{1 + \tilde{\rho}} - \int_{\bar{r}}^R \frac{\alpha \tilde{\rho} r^{\alpha+1} \bar{r}^{\alpha-1}}{(1 + \tilde{\rho} r^\alpha r^{-\alpha})^2} dr < 0 \quad \forall \bar{r},$$

which can be easily seen to always stand negative value. The same technique can not be applied to compute the derivative of $\xi_2(\bar{r})$ as the integral function has an infinite upper bound. Given $\bar{r}_1 < \bar{r}_2$, it can be written

$$\begin{aligned} \xi_2(\bar{r}_1) - \xi_2(\bar{r}_2) &= 2\pi \int_R^\infty \left[1 - \frac{1}{\tilde{\rho} \bar{r}_1^\alpha r^{-\alpha} + 1} \right] r dr - 2\pi \int_R^\infty \left[1 - \frac{1}{\tilde{\rho} \bar{r}_2^\alpha r^{-\alpha} + 1} \right] r dr \\ &= 2\pi \int_R^\infty \left[1 - \frac{1}{\tilde{\rho} \bar{r}_1^\alpha r^{-\alpha} + 1} \right] r - \left[1 - \frac{1}{\tilde{\rho} \bar{r}_2^\alpha r^{-\alpha} + 1} \right] r dr \\ &= 2\pi \int_R^\infty \frac{\tilde{\rho} r^{-\alpha+1} (\bar{r}_1^\alpha - \bar{r}_2^\alpha)}{(\tilde{\rho} \bar{r}_2^\alpha r^{-\alpha} + 1) (\tilde{\rho} \bar{r}_1^\alpha r^{-\alpha} + 1)} dr < 0, \quad \forall \bar{r}_1 < \bar{r}_2, \end{aligned}$$

which confirms $\xi_2(\bar{r})$ is a strictly monotonic increasing function for \bar{r} . The same study of function can be easily extended to $\lambda_{\Psi_2}(\bar{r})$, with the same final conclusions.

B.7 Proof of Monotonic Decreasing Trend for

$$\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r})$$

The derivative of $\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r})$ is

$$\frac{\partial}{\partial \bar{r}} \lambda_{\Psi_2^{\text{SCDP}}}(\bar{r}) = \frac{\frac{\partial}{\partial \bar{r}}(\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r}))}{\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})} - \frac{\frac{\partial}{\partial \bar{r}}(\pi \bar{r}^2 + \xi_3(\bar{r}))}{\pi \bar{r}^2 + \xi_3(\bar{r})}, \quad (\text{B.18})$$

such that $\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r})$ is monotonic decreasing when the numerator in (B.18) is negative. Therefore, the monotonic decreasing slope of $\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r})$ depends on the condition

$$\frac{(\pi \bar{r}^2 + \xi_3(\bar{r})) \frac{\partial}{\partial \bar{r}}(\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})) - (\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})) \frac{\partial}{\partial \bar{r}}(\pi \bar{r}^2 + \xi_3(\bar{r}))}{(\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r}))(\pi \bar{r}^2 + \xi_3(\bar{r}))} < 0.$$

Given that $(\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})) > 0$ and $(\pi \bar{r}^2 + \xi_3(\bar{r})) > 0$, the focus is now on the sign of the numerator such that

$$\begin{aligned} & (\pi \bar{r}^2 + \xi_3(\bar{r})) \frac{\partial}{\partial \bar{r}}(\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})) - (\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})) \frac{\partial}{\partial \bar{r}}(\pi \bar{r}^2 + \xi_3(\bar{r})) < 0 \\ & (\pi \bar{r}^2 + \xi_3(\bar{r})) \left(2\pi \bar{r} + \frac{\partial}{\partial \bar{r}}(\xi_3(\bar{r})) \right) - (\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})) \left(2\pi \bar{r} + \frac{\partial}{\partial \bar{r}}(\xi_3(\bar{r})) \right) < 0 \\ & (\pi \bar{r}^2 + \xi_3(\bar{r})) - (\pi R^2 + \pi \bar{r}^2 + \xi_3(\bar{r})) = -\pi R^2 < 0, \end{aligned}$$

from which it can be concluded that $\lambda_{\Psi_2^{\text{SCDP}}}(\bar{r})$ is strictly monotonic decreasing with a constant rate.

Appendix C

C.1 Decoupling Interfering Terms

Given two PPPs defined over the same portion of space which refer to the received power from LOS and NLOS terms, ϕ^L and ϕ^{nL} respectively, the resulting interfering power can be developed as the summation of a term which refers to the whole received power as if no differentiation is conducted on the kind of wireless link and a second term which takes into account the contribution from the LOS term. In particular, the LOS contribution is dependent to the power of full NLOS network. That is,

$$\begin{aligned}
 I^{nL} + I^L &= \sum_{i \in \phi^{nL}} |h_i|^2 r_i^{-\alpha} + \sum_{i \in \phi^L} |h'_i|^2 r_i^{-\alpha} \\
 &= \sum_{i \in \phi^{nL}} (X^2 + Y^2) r_i^{-\alpha} + \sum_{i \in \phi^L} ((\sqrt{2}\sigma_h X + \mu_h)^2 + (\sqrt{2}\sigma_h Y)^2) r_i^{-\alpha} \\
 &= \sum_{i \in \phi^{nL}} (X^2 + Y^2) r_i^{-\alpha} + \sum_{i \in \phi^L} (2\sigma_h^2 X^2 + 2\sigma_h^2 Y^2 + 2\sqrt{2}\mu_h\sigma_h X + \mu_h^2) r_i^{-\alpha} \\
 &= \sum_{i \in \phi^{nL}} (X^2 + Y^2) r_i^{-\alpha} + \sum_{i \in \phi^L} ((1 - 1 + 2\sigma_h^2)(X^2 + Y^2) + 2\sqrt{2}\mu_h\sigma_h X + \mu_h^2) r_i^{-\alpha} \\
 &\stackrel{a}{=} \underbrace{\sum_{i \in \phi} (X^2 + Y^2) r_i^{-\alpha}}_I + \underbrace{\sum_{i \in \phi^L} \left(-\frac{k}{k+1}(X^2 + Y^2) + \frac{2\sqrt{k}}{k+1}X + \frac{k}{k+1} \right) r_i^{-\alpha}}_{\Delta I^L | I},
 \end{aligned}$$

where (a) comes from substituting the values of σ_h and μ_h with the k -factor and from re-arranging the terms of the two sums. The obtained derivation stands as

the separated contributions from a full NLOS network I and the contribution given by its LOS components $\Delta I^L|I$.

References

- [1] Index Cisco Visual Networking. Global mobile data traffic forecast update, 2015–2020. *White Paper*, 3, Feb. 2016.
- [2] Akhil Gupta and Rakesh Kumar Jha. A survey of 5G network: Architecture and emerging technologies. *IEEE Access*, 3:1206–1232, July 2015.
- [3] Nisha Panwar, Shantanu Sharma, and Awadhesh Kumar Singh. A survey on 5G: The next generation of mobile communication. *Physical Communication*, 18:64–84, 2016.
- [4] Godfrey Anuga Akpakwu, Bruno J Silva, Gerhard P Hancke, and Adnan M Abu-Mahfouz. A survey on 5G networks for the internet of things: Communication technologies and challenges. *IEEE Access*, 6:3619–3647, 2017.
- [5] António Morgado, Kazi Mohammed Saidul Huq, Shahid Mumtaz, and Jonathan Rodriguez. A survey of 5G technologies: regulatory, standardization and industrial perspectives. *Digital Communications and Networks*, 4(2):87–97, 2018.
- [6] GSMA Intelligence. *The 5G Guide a Reference for Operators*, April 2019.
- [7] Mahmoud Kamel, Walaa Hamouda, and Amr Youssef. Ultra-dense networks: A survey. *IEEE Communications Surveys & Tutorials*, 18(4):2522–2545, 2016.
- [8] Mugen Peng, Yong Li, Zhongyuan Zhao, and Chonggang Wang. System architecture and key technologies for 5G heterogeneous cloud radio access networks. *IEEE Network*, 29(2):6–14, Mar. 2015.

-
- [9] *Service Requirements for the 5G System*. document 3GPP TS 22.261, Jan. 2018.
- [10] Muhammad Ali Naeem, Shahrudin Awang Nor, Suhaidi Hassan, and Byung-Seo Kim. Performances of probabilistic caching strategies in content centric networking. *IEEE Access*, 6:58807–58825, 2018.
- [11] Ibrahim Abdullahi, Suki Arif, and Suhaidi Hassan. Survey on caching approaches in information centric networking. *Journal of Network and Computer Applications*, 56:48–59, 2015.
- [12] Guoqiang Zhang, Yang Li, and Tao Lin. Caching in information centric networking: A survey. *Computer Networks*, 57(16):3128–3141, 2013.
- [13] Xiaofei Wang, Min Chen, Tarik Taleb, Adlen Ksentini, and Victor CM Leung. Cache in the air: Exploiting content caching and delivery techniques for 5G systems. *IEEE Communications Magazine*, 52(2):131–139, 2014.
- [14] Kuikui Li, Zhiyong Chen, and Meixia Tao. Probabilistic caching in multi-tier heterogeneous networks: Optimization and tradeoff. In *2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–6. IEEE, 2017.
- [15] Ejder Baştuğ, Mehdi Bennis, Marios Kountouris, and Mérouane Debbah. Cache-enabled small cell networks: Modeling and tradeoffs. *EURASIP Journal on Wireless Communications and Networking*, 2015(1):1–11, 2015.
- [16] Chenchen Yang, Yao Yao, Zhiyong Chen, and Bin Xia. Analysis on cache-enabled wireless heterogeneous networks. *IEEE Transactions on Wireless Communications*, 15(1):131–145, 2015.
- [17] Mona Jaber, Muhammad Ali Imran, Rahim Tafazolli, and Anvar Tukmanov. 5G backhaul challenges and emerging research directions: A survey. *IEEE Access*, 4:1743–1766, 2016.
- [18] Md Maruf Ahamed and Saleh Faruque. *Broadband Communications Networks: Recent Advances and Lessons from Practice*. IntechOpen, 2018.

-
- [19] Mona Jaber. *Realistic 5G backhaul*. PhD thesis, University of Surrey, 2017.
- [20] QUALCOMM Incorporated. *Rising to meet the 1000x mobile data challenge*, 2012.
- [21] Naga Bhushan, Junyi Li, Durga Malladi, Rob Gilmore, Dean Brenner, Aleksandar Damnjanovic, Ravi Teja Sukhavasi, Chirag Patel, and Stefan Geirhofer. Network densification: the dominant theme for wireless evolution into 5G. *IEEE Communications Magazine*, 52(2):82–89, 2014.
- [22] David López-Pérez, Ming Ding, Holger Claussen, and Amir H Jafari. Towards 1 gbps/ue in cellular systems: Understanding ultra-dense small cell deployments. *IEEE Communications Surveys & Tutorials*, 17(4):2078–2101, 2015.
- [23] Sumudu Samarakoon, Mehdi Bennis, Walid Saad, Mérouane Debbah, and Matti Latva-Aho. Ultra dense small cell networks: Turning density into energy efficiency. *IEEE Journal on Selected Areas in Communications*, 34(5):1267–1280, 2016.
- [24] Edwin Mugume and Daniel KC So. Deployment optimization of small cell networks with sleep mode. *IEEE Transactions on Vehicular Technology*, 68(10):10174–10186, 2019.
- [25] Chuan Ma, Ming Ding, David López-Pérez, Zihuai Lin, Jun Li, and Guoqiang Mao. Performance analysis of the idle mode capability in a dense heterogeneous cellular network. *IEEE Transactions on Communications*, 66(9):3959–3973, 2018.
- [26] Chuan Ma, Ming Ding, He Chen, Zihuai Lin, Guoqiang Mao, and David López-Pérez. On the performance of multi-tier heterogeneous cellular networks with idle mode capability. In *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2018.
- [27] Jie Xu, Ling Qiu, and Chengwen Yu. Improving network energy efficiency through cooperative idling in the multi-cell systems. *EURASIP Journal on Wireless Communications and Networking*, 2011(1):165, 2011.

- [28] Jeffrey G Andrews, Radha Krishna Ganti, Martin Haenggi, Nihar Jindal, and Steven Weber. A primer on spatial modeling and analysis in wireless networks. *IEEE Communications Magazine*, 48(11):156–163, 2010.
- [29] Robert W Heath, Marios Kountouris, and Tianyang Bai. Modeling heterogeneous network interference using poisson point processes. *IEEE Transactions on Signal Processing*, 61(16):4114–4126, 2013.
- [30] Wei Lu and Marco Di Renzo. Stochastic geometry modeling of cellular networks: Analysis, simulation and experimental validation. In *Proceedings of the 18th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 179–188, 2015.
- [31] Norman Campbell. The study of discontinuous phenomena. In *Proceedings of the Cambridge Philosophical Society*, volume 15, pages 117–136, 1909.
- [32] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE mobile computing and communications review*, 5(1):3–55, 2001.
- [33] Claude Elwood Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949.
- [34] Carlos R Cunha, Azer Bestavros, and Mark E Crovella. Characteristics of WWW client-based traces. Technical report, Boston University Computer Science Department, 1995.
- [35] Ghaleb Abdulla, Edward A Fox, Marc Abrams, and Stephen Williams. WWW proxy traffic characterization with application to caching. Technical report, Department of Computer Science, Virginia Polytechnic Institute & State University, 1997.
- [36] Venkata N Padmanabhan and Lili Qiu. The content and access dynamics of a busy web site: Findings and implications. In *Proceedings of the conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, pages 111–123, 2000.

-
- [37] Dimitrios N Serpanos, George Karakostas, and Wayne Hendrix Wolf. Effective caching of web objects using zipf's law. In *2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast Changing World of Multimedia (Cat. No. 00TH8532)*, volume 2, pages 727–730. IEEE, 2000.
- [38] Theodore S Rappaport et al. *Wireless communications: principles and practice*, volume 2. New Jersey: Prentice Hall PTR, 1996.
- [39] 3GPP. Study on channel model for frequencies from 0.5 to 100 ghz. Technical report, 3GPP TR 38.901, version 14.0.0, Release 14, 2017.
- [40] Steven M Kay. *Fundamentals of statistical signal processing*. Prentice Hall PTR, 1993.
- [41] Bartłomiej Blaszczyszyn and Anastasios Giovanidis. Optimal geographic caching in cellular networks. In *IEEE International Conference on Communications (ICC)*, pages 3358–3363. IEEE, 2015.
- [42] Mugen Peng, Dong Liang, Yao Wei, Jian Li, and Hsiao-Hwa Chen. Self-configuration and self-optimization in lte-advanced heterogeneous networks. *IEEE Communications Magazine*, 51(5):36–45, 2013.
- [43] Aamod Khandekar, Naga Bhushan, Ji Tingfang, and Vieri Vanghi. Lte-advanced: Heterogeneous networks. In *2010 European wireless conference (EW)*, pages 978–982. IEEE, 2010.
- [44] Mugen Peng, Chonggang Wang, Jian Li, Hongyu Xiang, and Vincent Lau. Recent advances in underlay heterogeneous networks: Interference control, resource allocation, and self-organization. *IEEE Communications Surveys & Tutorials*, 17(2):700–729, 2015.
- [45] Han-Shin Jo, Young Jin Sang, Ping Xia, and Jeffrey G Andrews. Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis. *IEEE Transactions on Wireless Communications*, 11(10):3484–3495, 2012.

-
- [46] Mugen Peng, Yuan Li, Jiamo Jiang, Jian Li, and Chonggang Wang. Heterogeneous cloud radio access networks: A new perspective for enhancing spectral and energy efficiencies. *IEEE Wireless Communications*, 21(6):126–135, 2014.
- [47] Seong Ho Chae and Wan Choi. Caching placement in stochastic wireless caching helper networks: Channel selection diversity via caching. *IEEE Transactions on Wireless Communications*, 15(10):6626–6637, 2016.
- [48] Ying Cui, Dongdong Jiang, and Yueping Wu. Analysis and optimization of caching and multicasting in large-scale cache-enabled wireless networks. *IEEE Transactions on Wireless Communications*, 15(7):5101–5112, 2016.
- [49] Zheng Chen, Jemin Lee, Tony QS Quek, and Marios Kountouris. Cooperative caching and transmission design in cluster-centric small cell networks. *IEEE Transactions on Wireless Communications*, 16(5):3401–3415, 2017.
- [50] Wanli Wen, Ying Cui, Fu-Chun Zheng, Shi Jin, and Yanxiang Jiang. Random caching based cooperative transmission in heterogeneous wireless networks. *IEEE Transactions on Communications*, 66(7):2809–2825, 2018.
- [51] Seong Ho Chae, Tony QS Quek, and Wan Choi. Content placement for wireless cooperative caching helpers: A tradeoff between cooperative gain and content diversity gain. *IEEE Transactions on Wireless Communications*, 16(10):6795–6807, 2017.
- [52] Mark Newton and John Thompson. Classification and generation of non-uniform user distributions for cellular multi-hop networks. In *2006 IEEE International Conference on Communications*, volume 10, pages 4549–4553. IEEE, 2006.
- [53] Sami Baroudi and Yousef R Shayan. Analytical evaluation of outage probability based on signal to interference ratio for gaussian-distributed users. In *2015 IEEE Symposium on Computers and Communication (ISCC)*, pages 841–844. IEEE, 2015.
- [54] Sami Baroudi and Yousef R Shayan. Percentage of gaussianly distributed users with adequate quality of service in a circular cell. In *2014 IEEE 27th Canadian*

-
- Conference on Electrical and Computer Engineering (CCECE)*, pages 1–4. IEEE, 2014.
- [55] Ramy Amer, Hesham ElSawy, Jacek Kibilda, M Majid Butt, and Nicola Marchetti. Cooperative transmission and probabilistic caching for clustered d2d networks. In *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2019.
- [56] Ejder Baştuğ, Mehdi Bennis, Marios Kountouris, and Mérouane Debbah. Cache-enabled small cell networks: Modeling and tradeoffs. *EURASIP Journal on Wireless Communications and Networking*, 2015(1):1–11, 2015.
- [57] Junliang Ye, Xiaohu Ge, Guoqiang Mao, and Yi Zhong. 5G ultradense networks with nonuniform distributed users. *IEEE Transactions on Vehicular Technology*, 67(3):2660–2670, 2017.
- [58] Antonis Gotsis, Stelios Stefanatos, and Angeliki Alexiou. Ultradense networks: The new wireless frontier for enabling 5g access. *IEEE Vehicular Technology Magazine*, 11(2):71–78, 2016.
- [59] 3GPP. Technical specification group radio access network; evolved universal terrestrial radio access (e-utra); further advancements for e-utra physical layer aspects. Technical report, 3GPP TS 36.211, version 12.9.0, Release 124, 2016.
- [60] Gaurav Nigam, Paolo Minero, and Martin Haenggi. Coordinated multipoint joint transmission in heterogeneous networks. *IEEE Transactions on Communications*, 62(11):4134–4146, 2014.
- [61] Gaurav Nigam, Paolo Minero, and Martin Haenggi. Coordinated multipoint in heterogeneous networks: A stochastic geometry approach. In *IEEE Globecom Workshops (GC Wkshps)*, pages 145–150. IEEE, 2013.
- [62] P. Lin, K. S. Khan, Qingyang Song, and A. Jamalipour. Caching in heterogeneous ultradense 5G networks: A comprehensive cooperation approach. *IEEE Vehicular Technology Magazine*, 14:22–32, 2019.

-
- [63] GJ Foschini, K Karakayali, and RA Valenzuela. Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency. *IEEE Proceedings-Communications*, 153(4):548–555, 2006.
- [64] Larry Armijo. Minimization of functions having lipschitz continuous first partial derivatives. *Pacific Journal of mathematics*, 16(1):1–3, 1966.
- [65] Stephen Boyd and Lieven Vandenbergh. *Convex optimization*. Cambridge university press, 2004.
- [66] Yuzhou Li, Yu Zhang, Kai Luo, Tao Jiang, Zan Li, and Wei Peng. Ultra-dense hetnets meet big data: Green frameworks, techniques, and approaches. *IEEE Communications Magazine*, 56(6):56–63, 2018.
- [67] Yuan Wu, Li Ping Qian, Jianchao Zheng, Haibo Zhou, and Xuemin Sherman Shen. Green-oriented traffic offloading through dual connectivity in future heterogeneous small cell networks. *IEEE Communications Magazine*, 56(5):140–147, 2018.
- [68] Jaesung Park and Heejung Byun. Autonomous transmission power decision strategy for energy efficient operation of a dense small cell network. *Wireless Communications and Mobile Computing*, 2018, 2018.
- [69] Yanzan Sun, Han Xu, Shunqing Zhang, Yating Wu, Tao Wang, Yong Fang, and Shugong Xu. Joint optimization of interference coordination parameters and base-station density for energy-efficient heterogeneous networks. *Sensors*, 19(9):2154, 2019.
- [70] Xiaofei Wang, Athanasios V Vasilakos, Min Chen, Yunhao Liu, and Ted Taekyoung Kwon. A survey of green mobile networks: Opportunities and challenges. *Mobile Networks and Applications*, 17(1):4–20, 2012.
- [71] Sofie Lambert, Prasanth Ananth, Peter Vetter, Ka-Lun Lee, Jie Li, Xin Yin, Hungkei Chow, Jean-Patrick Gelas, Laurent Lefèvre, Dominique Chiaroni, et al. Road to energy-efficient optical access: greentouch final results. *Journal of Optical Communications and Networking*, 8(11):878–892, 2016.

-
- [72] Emanuele Gruppi, Kai-Kit Wong, Mohammud Z Bocus, and Woon Hau Chin. Ultra dense edge caching networks with arbitrary user spatial density. *IEEE Transactions on Wireless Communications*, 2020.
- [73] Tony QS Quek, Wang Chi Cheung, and Marios Kountouris. Energy efficiency analysis of two-tier heterogeneous networks. In *17th European Wireless 2011-Sustainable Wireless Technologies*, pages 1–5. VDE, 2011.
- [74] Dong Liu and Chenyang Yang. Energy efficiency of downlink networks with caching at base stations. *IEEE Journal on Selected Areas in Communications*, 34(4):907–922, 2016.
- [75] Qi Ren, Jiancun Fan, Xinmin Luo, Zhikun Xu, and Yami Chen. Analysis of spectral and energy efficiency in ultra-dense network. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, pages 2812–2817. IEEE, 2015.
- [76] Chang Li, Jun Zhang, and Khaled B Letaief. Throughput and energy efficiency analysis of small cell networks with multi-antenna base stations. *IEEE Transactions on Wireless Communications*, 13(5):2505–2517, 2014.
- [77] Tao Chen, Haesik Kim, and Yang Yang. Energy efficiency metrics for green wireless communications. In *2010 International Conference on Wireless Communications & Signal Processing (WCSP)*, pages 1–6. IEEE, 2010.
- [78] Gunther Auer, Vito Giannini, Claude Desset, Istvan Godor, Per Skillermark, Magnus Olsson, Muhammad Ali Imran, Dario Sabella, Manuel J Gonzalez, Oliver Blume, et al. How much energy is needed to run a wireless network? *IEEE Wireless Communications*, 18(5):40–49, 2011.
- [79] Yi Wang, Jian Li, Lei Huang, Yao Jing, Andreas Georgakopoulos, and Panagiotis Demestichas. 5G mobile: Spectrum broadening to higher-frequency bands to support high data rates. *IEEE Vehicular Technology Magazine*, 9(3):39–46, 2014.
- [80] Yassine Hmamouche, Mustapha Benjillali, and Samir Saoudi. A stochastic geometry based approach to tractable 5G rnp with a new h -los model. In

-
- 2019 *IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–8. IEEE, 2019.
- [81] Xiaobin Yang and Abraham O Fapojuwo. Coverage probability analysis of heterogeneous cellular networks in rician/rayleigh fading environments. *IEEE Communications Letters*, 19(7):1197–1200, 2015.
- [82] Kumin Cho, Junman Lee, and Chung G Kang. Low-complexity coverage analysis of downlink cellular network for combined los and nlos propagation. *IEEE Communications Letters*, 23(1):160–163, 2018.
- [83] Youjia Chen, Ming Ding, and David López-Pérez. Performance of ultra-dense networks with a generalized multipath fading. *IEEE Wireless Communications Letters*, 8(5):1419–1422, 2019.
- [84] Ming Ding, Peng Wang, David López-Pérez, Guoqiang Mao, and Zihuai Lin. Performance impact of los and nlos transmissions in dense cellular networks. *IEEE Transactions on Wireless Communications*, 15(3):2365–2380, 2015.
- [85] Ralph Tanbourgi, Sarabjot Singh, Jeffrey G Andrews, and Friedrich K Jondral. A tractable model for noncoherent joint-transmission base station cooperation. *IEEE Transactions on Wireless Communications*, 13(9):4959–4973, 2014.
- [86] Michael J Neely. Energy optimal control for time-varying wireless networks. *IEEE transactions on Information Theory*, 52(7):2915–2934, 2006.
- [87] Michael J Neely, Eytan Modiano, and Charles E Rohrs. Dynamic power allocation and routing for time varying wireless networks. In *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428)*, volume 1, pages 745–755. IEEE, 2003.
- [88] Yuhua Xu, Jinlong Wang, Qihui Wu, Jianchao Zheng, Liang Shen, and Alan Anpalagan. Dynamic spectrum access in time-varying environment: Distributed learning beyond expectation optimization. *IEEE Transactions on Communications*, 65(12):5305–5318, 2017.

-
- [89] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.
- [90] Alireza Sadeghi, Fatemeh Sheikholeslami, and Georgios B Giannakis. Optimal and scalable caching for 5G using reinforcement learning of space-time popularities. *IEEE Journal of Selected Topics in Signal Processing*, 12(1):180–190, 2017.
- [91] Chongjie Zhang and Victor Lesser. Coordinated multi-agent reinforcement learning in networked distributed pomdps. In *Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.
- [92] Samuel O Somuyiwa, András György, and Deniz Gündüz. A reinforcement-learning approach to proactive caching in wireless networks. *IEEE Journal on Selected Areas in Communications*, 36(6):1331–1344, 2018.
- [93] Telecom Italia. Telecommunications - SMS, Call, Internet - MI, 2015.
- [94] Fuqiang Yao and Luliang Jia. A collaborative multi-agent reinforcement learning anti-jamming algorithm in wireless networks. *IEEE Wireless Communications Letters*, 8(4):1024–1027, 2019.
- [95] Navid Naderializadeh, Jaroslaw J Sydir, Meryem Simsek, and Hosein Nikopour. Resource management in wireless networks via multi-agent deep reinforcement learning. *IEEE Transactions on Wireless Communications*, 20(6):3507–3523, 2021.
- [96] Amal Feriani and Ekram Hossain. Single and multi-agent deep reinforcement learning for ai-enabled wireless networks: A tutorial. *IEEE Communications Surveys & Tutorials*, 2021.
- [97] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning. In *International Conference on Machine Learning*, pages 4295–4304. PMLR, 2018.