# Evaluation and optimization of the quality of service perceived by mobile users for new services in cellular networks

Miodrag Jovanovic

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2015-ENST-0052

EDITE - ED 130

# Doctorat ParisTech

# T H È S E

**pour obtenir le grade de docteur délivré par**

## TELECOM ParisTech

### Spécialité "Informatique et Réseau"

*présentée et soutenue publiquement par*

**Miodrag JOVANOVIC**

le 11 Septembre 2015

## Titre

# Evaluation et optimisation de la qualité de service perçue par les utilisateurs mobiles pour les nouveaux services dans les réseaux cellulaires

Directeur de thèse: **Bartek BLASZCZYSZYN**
Co-encadrement de la thèse: **Mohamed KARRAY**

**Jury**

| | |
|---|---|
| **M. Laurent DECREUSEFOND**, Professeur, Télécom ParisTech | Président du jury |
| **M. Sem BORST**, Professeur, Technische Universiteit Eindhoven et Bell Labs | Rapporteur |
| **M. Martin HAENGGI**, Professeur, University of Notre Dame | Rapporteur |
| **M. Mérouane DEBBAH**, Professeur, Supélec et Huawei | Examinateur |
| **M. James ROBERTS**, Professeur, IRT-SystemX | Examinateur |
| **M. Bartek BLASZCZYSZYN**, Professeur, Ecole normale supérieure et INRIA | Directeur de thèse |

**TELECOM ParisTech**
école de l'Institut Mines-Télécom - membre de ParisTech
46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

THÈSE

*To my family and friends*

**RESUME: L'objectif de cette thèse est de développer des outils et des méthodes pour l'évaluation de la qualité de service (Quality of Service - QoS) perçue par les utilisateurs, en fonction de la demande de trafic, dans les réseaux cellulaire sans fil moderne. Ce problème complexe, directement liée au dimensionnement du réseau, implique la modélisation des processus dynamiques à plusieurs échelles de temps, qui en raison de leurs nature aléatoire se prêtent à la formalisation probabiliste.**

**Tout d'abord, sur la base de la théorie de l'information, nous capturons les performances d'un seul lien entre une station de base et un utilisateur dans un réseau cellulaire avec des canaux orthogonaux et la technologie MIMO. Nous prouvons et utilisons certaines bornes inférieures de la capacité ergodique en vue de la théorie de l'information d'un tel lien, qui prend aussi en compte la variabilité du canal rapide causée par la propagation des trajets multiples. Ces bornes donnent une base solide pour l'évaluation plus profonde de la qualité de service perçue par les utilisateurs.**

**Ensuite, on considère plusieurs utilisateurs (éventuellement mobiles), arrivant dans le réseau et demandant un service. Nous considérons des services (élastiques) à débit variable dans lesquels les transmissions de certaines quantités de données sont réalisées d'une manière "best-effort", ou services à débit constant, dans lesquels une certaine vitesse de transmission doit être maintenue pendant les périodes demandées. Sur la base de la théorie des files d'attente, on capture cette demande du trafic et processus de service à l'aide des modèles appropriés (multi-classes) de partage du processeur (processor-sharing PS) ou modèle de perte. Dans cette thèse, nous adaptons les modèles PS existants et développons un nouveau modèle de perte pour le trafic streaming de transmission sans fil, où les bornes théoriques (au regard de la théorie de l'information) mentionnées ci-dessus de la capacité des liens simples décrivent les taux de service instantanés des utilisateurs. Les modèles multi-classes sont utilisés pour capturer l'hétérogénéité spatiale des canaux utilisateur. Ceux-ci dépendent de l'emplacement géographique de l'utilisateur et du phénomène de "shadowing" de propagation.**

**Enfin, au-dessus des processus de file d'attente théoriques, il faut tenir compte d'un réseau multicellulaire, dont les stations de base ne sont pas nécessairement régulièrement placées, et dont la géométrie est en outre perturbée par la phénomène de shadowing. Nous abordons cet aspect aléatoire en utilisant des modèles de géométrie stochastique, notamment processus de Poisson ponctuels et le formalisme de Palm appliqué à la cellule typique du réseau. En appliquant l'approche triple mentionnée ci-dessus, censée à représenter tous les mécanismes cruciaux et les paramètres de l'ingénierie des réseaux cellulaires (tels que LTE - Long Term Evolution), nous établissons des relations macroscopiques entre la demande de trafic et les métriques de la qualité de service perçue par les utilisateurs pour certains services à débit binaire élastiques et constants. Ces relations sont obtenues principalement d'une manière semi-analytique, c'est-à-dire qu'elles concernent des simulations statiques d'un processus ponctuel de Poisson (modélisation des emplacements des stations de base). Ceci afin d'évaluer ses caractéristiques qui ne se prêtent pas aux expressions analytiques.**

**Plus précisément, en ce qui concerne le trafic de données (le service de débit binaire élastique), nous capturons l'interférence inter-cellule, rendant les modèles des files d'attente PS de cellules individuelles dépendantes, via un système d'équations de charge des cellules. Ces équations permettent de déterminer le débit moyen par utilisateur, le nombre moyen d'utilisateurs et la charge moyenne de la cellule dans un grand réseau, en fonction de la demande du trafic. La distribution spatiale de ces métriques de QoS dans le réseau est également étudiée. Nous validons notre approche en comparant les résultats obtenus avec ceux mesurés à partir de traces du réseau réel. Nous observons une concordance remarquable entre les prédictions du modèle et les données statistiques recueillies dans plusieurs scénarios de déploiement.**

**En ce qui concerne les services de débit binaire constants, nous proposons un nouveau modèle stochastique pour évaluer la fréquence et le nombre d'interruptions lors de streaming en temps réel en fonction des conditions radio utilisateur. Nous l'utilisons pour étudier les métriques de la qualité de service en fonction des conditions radio utilisateur dans les réseaux LTE.**

**Tous les résultats établis ici contribuent au développement de méthodes de dimensionnement de réseau et sont actuellement utilisés dans les outils internes d'Orange pour les calculs de capacité du réseau.**

**MOTS-CLEFS: QoS; LTE; débit; charge; théorie des files d'attente; géométrie stochastique; mesures; 3GPP**

**ABSTRACT: The goal of this thesis is to develop tools and methods for the evaluation of the QoS (Quality of Service) perceived by users, as a function of the traffic demand, in modern wireless cellular networks. This complex problem, directly related to network dimensioning, involves modeling dynamic processes at several time-scales, which due to their randomness are amenable to probabilistic formalization.**

**Firstly, on the ground of information theory, we capture the performance of a single link between a base station and a user in the context of a cellular network with orthogonal channels and MIMO technology. We prove and use some lower bounds of the information-theoretic ergodic capacity of such a link, which account also for the fast channel variability caused by multi-path propagation. These bounds give robust basis for further user QoS evaluation.**

**Next, one considers several (possibly mobile) users, arriving in the network and requesting some service from it. We consider variable (elastic) bit-rate services, in which transmissions of some amounts of data are realized in a best-effort manner, or constant bit-rate services, in which a certain transmission rate needs to be maintained during requested times. On the ground of queuing theory, one captures this traffic demand and service process using appropriate (multi-class) processor sharing (PS) or loss models. In this thesis, we adapt existing PS models and develop a new loss model for wireless streaming traffic, in which the aforementioned information-theoretic capacities of single links describe the instantaneous user service rates. The multi-class models are used to capture the spatial heterogeneity of user channels, which depends on the user geographic locations and propagation shadowing phenomenon.**

**Finally, on top of the queueing-theoretic processes, one needs to consider a multi-cellular network, whose base stations are not necessarily regularly placed, and whose geometry is further perturbed by the shadowing phenomenon. We address this randomness aspect by using some models from stochastic geometry, notably Poisson point processes and Palm formalism applied to the typical cell of the network.**

**Applying the above three-fold approach, supposed to represent all crucial mechanisms and engineering parameters of cellular networks (such as LTE), we establish some macroscopic relations between the traffic demand and the user QoS metrics for some elastic and constant bit-rate services. These relations are mostly obtained in a semi-analytic way, i.e., they only involve static simulations of a Poisson point process (modeling the locations of base stations) in order to evaluate its characteristics which are not amenable to analytic expressions.**

**More precisely, regarding the data traffic (the elastic bit-rate service), we capture the inter-cell interference, making the PS queue models of individual cells dependent, via some system of cell-load equations. These equations allow one to determine the mean user throughput, the mean number of users and the mean cell load in a large network, as a function of the traffic demand. The spatial distribution of these QoS metrics in the network is also studied. We validate our approach by comparing the obtained results with those measured from live-network traces. We observe a remarkably good agreement between the model predictions and the statistical data collected in several deployment scenarios.**

**Regarding constant bit-rate services, we propose a new stochastic model to evaluate the frequency and the number of interruptions during real-time streaming calls in function of user radio conditions. We use it to study the quality of service metrics in function of user radio conditions in LTE networks.**

**All established results contribute to the development of network dimensioning methods and are currently used in Orange internal tools for network capacity calculations.**

**KEY-WORDS: QoS; LTE; throughput; load; queueing theory; stochastic geometry; measures; 3GPP**

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Thesis motivation

There is a need for simple, yet realistic methods for the evaluation of the *quality of service* (QoS) in *wireless networks* capturing both the spatial distribution of the elements of the network and the temporal dynamics of users and having a limited number of parameters. This can be obtained by decomposing the problem into three layers corresponding to different time-scales, which are addressed on the ground of information theory, queuing theory and stochastic geometry. Firstly, information theory studies the performance of a single radio link accounting particularly for the signal variations due to multi-path fading. Once the link performance is characterized, resources (power, bandwidth etc.) are allocated to the users while accounting for their mutual interference. This can be modelled by an appropriate service policy on the ground of queuing theory which accounts next for the users' arrivals, mobility and departures and allows appropriate time averages. Finally, stochastic geometry is used to model network, i.e. base stations spatial pattern and shadowing.

Individual elements of the above puzzle (i.e. information theory, queuing theory and stochastic geometry models) are often studied and optimized separately. The main specificity of the methodology proposed in this thesis is a global approach that combines these three elements. In doing so, it is necessary to separate carefully the time scales of different elements of the network dynamics.

We apply an information-theoretic modeling of a link layer between a user and a base station. In this context we show that the worst additive noise is the white Gaussian one and establish a lower bound for the link capacity. Further the modeling approach consists in representing the configuration of users (positions, call durations or volumes, allocated resources) as a random object (point pattern with associated random variables), which evolves in time. The quality of service perceived by the users may then be expressed as a function of the stationary state of this process and thus will depend only on its distribution parameters. This approach often allows one for an explicit evaluation of the key QoS characteristics and for efficient optimization of the network cost and capacity. Some examples from other research works which prove the pertinence of this approach, can be found in [49], [78], [24], [68], [42], [10], [54]. We use homogeneous spatial Poisson point processes to model base station positions and apply results from stochastic geometry to evaluate QoS.

The performance of wireless cellular networks is often evaluated in terms of parameters such as the spectral efficiency [7] (in particular within 3GPP [3]) or the outage probability [44]. However from the point of view of an operator, it is even more important to calculate the QoS

perceived by the users; and in particular to relate this QoS to the key network parameters such as the traffic demand, the cell radius, the transmitted power, etc. This relation is crucial for the network dimensioning; i.e., evaluating the minimal number of base stations (more generally, the required network setup) assuring some QoS (for some given traffic demand). This permits in particular to minimize the network cost. The probabilistic approach described above often allows an explicit evaluation of the key characteristics such as users QoS and for efficient optimization of the network dimensioning.

Classically, the services are classified into two main categories:

- Variable bit-rate (VBR); e.g., mail, ftp. Users aim to transmit some given volume of data at a bit-rate which may be decided by the network
- Constant bit-rate (CBR); e.g., voice, video conferencing. Users require some given (constant) bit-rate for some duration. In this case the requested bit-rates may sometimes exceed the available capacity, a situation usually called congestion. CBR services do not tolerate temporary interruptions of their transmissions. Consequently, if congestion occurs, the network blocks (i.e., refuses the access to) new calls and/or drops (i.e., interrupts definitely) some other calls during their transmissions.

When we account for calls' arrivals, mobility and departures, the QoS perceived by the users (in the long run of the network) is different for each of the above traffic classes. For VBR connections, the QoS may be defined in terms of the mean throughput or delay per user [25], [59]. For CBR calls, the main QoS indicators are the blocking and dropping probabilities [10], [42], [59]. The research done in the last few years permitted to build efficient methods to calculate these QoS indicators (see for example [6]). These methods are based respectively on processor sharing for VBR and on multi-class Erlang models for CBR services.

However, new multimedia services are gaining interest in wireless cellular networks, especially streaming services [45]. Streaming connections require some given bit-rate for some duration [68], [75]. Thus congestion may occur (when the bit-rates requested by the users in the network exceed the available capacity). All streaming calls are admitted, but, as a counterpart, they tolerate temporary interruptions of their transmission. We distinguish two sub-classes:

- Streaming-RT (Real-Time): e.g. mobile TV, RTP streaming. When congestion occurs, the corresponding portions of some calls are definitely lost, but the call is not dropped.
- Streaming-NRT (Non-Real-Time): e.g., streaming-video (youtube, dailymotion, on demand video) on the web. When congestion occurs, the corresponding portions of calls are delayed.

For the streaming users, the QoS is related to the frequency of the interruption of their calls and the durations of these interruptions. These performance measures depend strongly on the mobility of users, as mobility increases the variability of the radio conditions.

## 1.2 Thesis contribution and structure

This thesis comprises three technical chapters. In Chapter 2 we focus on the first element of the analytic approach: single link quality between a user and a base station. We examine radio links in cellular networks such as LTE (Long Term Evolution) and HSDPA (High-Speed Downlink Packet Access) and take into account MIMO (Multiple Input Multiple Output) and OFDM (Orthogonal Frequency Division Multiplex) technologies. The principal result is a lower

theoretical bound for a single link quality which is very close to the exact link quality and which is tractable analytically; i.e. calculated in a simple manner.

Once expressions for link quality are developed, we take into account user dynamics using queuing theory. Finally, we apply the results from stochastic geometry to model the spatial configuration of network resources and users. In Chapter 3 and Chapter 4 we apply the above mentioned methodology to evaluate QoS for the different types of services. Namely, in Chapter 3 we develop a method for the user throughput estimation in large cellular networks, regular or irregular, for VBR traffic. In addition, we could estimate the mean number of users and the mean cell load in the network. All the above means account for the disparity of different base stations and traffic randomness over some period of time (one hour, for example). Further, we are able to estimate the spatial CDF (Cumulative Distribution Function) of mean user throughput, mean number of users and cell load over all base stations (averaged in time).

The core of the mathematical modeling in Chapter 3 was to capture the dependence between the traffic demand and the interference in cellular networks with orthogonal channels (in time and/or frequency). We did this using the aforementioned probabilistic tools in order to get analytically tractable and simple relations, but in a manner that reflects the physical behaviour of the system. It turns out that the dependence between the traffic demand and the interference is well captured via a *fixed-point problem*. Solving this problem, we get all elements to evaluate QoS as function of traffic demand.

In Chapter 4 the evaluation of the QoS for real-time streaming is presented. The number and duration of interruptions are calculated as function of traffic demand and radio conditions, i.e. SINR (Signal to Interference and Noise Ratio). Hexagonal cellular network with orthogonal channels is considered. The stochastic analysis is based on Poisson processes representation of the traffic and Palm formalism related to the typical call. The results can be used to estimate the QoS for this type of traffic, but also for network dimensioning. In fact, based on the spatial distribution of radio conditions we can deduce the QoS at all positions in a cell (area served by one base station) for a given traffic demand. On the other hand the spatial distribution of SINR depends on cell radius. So, determining a constraint on QoS, one can deduce what is the necessary cell radius to satisfy this constraint on QoS. This manipulation can be done for any value of traffic demand, which is the cellular network dimensioning.

Chapter 2 is based on the following publications [61] and [63]. Chapter 3 is constructed from the following articles [62], [15], [57] and [16] and the Chapter 4 is an adapted version of [17].

The results of the thesis are used in Orange tools, such as the operational tool Utrandim, for the dimensioning of wireless cellular networks. They are also used to study the spectral and energy efficiencies and the required emitted power in these networks.

# Chapter 2

# Link quality

## 2.1 Introduction

In this Chapter we are interested in the link quality analysis. A link is a communication channel between two or more communicating devices. We need it to develop a global analytic approach to the performance evaluation of wireless cellular networks and especially LTE networks.

The link performance in an OFDM (used in LTE) cellular network with MIMO antennas may be studied using two methodologies. Information theory considers the ultimate performance of best possible coding schemes and looks for mathematical formulae to describe this performance. Real systems, such as 3GPP (*3rd Generation Partnership Project*) [3], deploy suboptimal coding schemes which are usually evaluated by simulation.

A key link characteristic of OFDM cellular networks is the *peak bit-rate* at each location defined as the maximal bit-rate a user can get at the considered location from his serving base station. The objective of the present Chapter is to establish some closed form information theoretic bounds for the peak bit-rate in OFDM cellular networks with MIMO and compare them to real system performance predicted by simulation and estimated from field measurements.

We describe a simple model of a MIMO cellular network which permits to obtain an analytical expression of users' bit-rates, which are feasible from the information theory point of view. This expression accounts for the variety of MIMO configurations (numbers of transmitting and receiving antennas) and radio conditions (SINR). This expression is compared to practical LTE performance evaluated by 3GPP simulations for different cases including the so-called calibration case [3]. The comparison shows that the analytical expression may be adjusted to the practical performance by a multiplicative coefficient, which depends on the MIMO configuration but not on the SINR. Additionally, we show the progress margin for potential evolution of the technology.

The capacity of the MIMO channel without interference is known. Accounting for the extra-cell interference, Proposition 2 gives a lower bound for the downlink capacity in a multi-cell OFDM network with MIMO antennas. This bound relies on the observation made in Proposition 3 that the worst additive noise for the capacity of the MIMO flat-fading complex-valued channel is the white Gaussian one. In order to make the established lower bound more explicit, we give an asymptotic approximation based on random matrix theory and derive also a further lower bound from Jensen's inequality.

Finally we build bounds for the peak bit-rate of the MMSE (Minimum Mean Square Error) scheme currently implemented in operational networks as well as its improvement MMSE-SIC (Successive Interference Cancellation).

## 2.2 Related work

Telatar [91] gives the information theoretic capacity of a MIMO channel with fading and AWGN (Additive White Gaussian Noise). Different MIMO configurations are compared for this channel by Foschini and Gans [46]. Blum et al. [22] study the capacity of a MIMO cellular network with flat Rayleigh fading. Clark et al. [32] show that in an OFDM system with a sufficiently large number of sub-carriers, the capacity with respect to Rayleigh fading is approximately normally distributed. Tulino and Verdu [94] apply random matrix theory to analyze this capacity. Random matrix theory is useful to study CDMA as for example in [93], [95] and [43].

The 3GPP [3] evaluates the performance of LTE systems by simulation. Goldsmith and Chua [53] observed that real coding schemes performance may be described by a modification of the famous $\log_2(1+\mathrm{SNR})$ Shannon's formula. Mogensen et al. [73] have observed that the LTE capacity in the AWGN context is well approximated by this formula with a multiplicative coefficient. These ideas will be extended in the present Chapter to MIMO cellular networks with fading.

Explicit expressions for the capacity of AWGN channels are well known. Interference in wireless cellular networks is not necessarily Gaussian nor white (the term *white* means that the samples are independent and identically distributed). The explicit expression for capacity in such context is not known. To circumvent this difficulty, a possible idea is to look for a lower bound and check whether it is tight enough to meet a desired precision. Using a result of Shannon [84, Theorem 18], it may be shown that, in a SISO channel, the worst additive noise process with given power is AWGN. This result is extended to a network with relays by Shomorony and Avestimehr [85]. Diggavi and Cover [40] study the worst noise process for an additive channel under covariance constraints and characterize the so-called saddle-point input and noise distributions for the mutual information [40, Theorem II.1]. Girnyk et al. [51] calculate the asymptotic sum-rate of uplink MIMO cellular network.

Note that the fact that the worst additive noise is Gaussian may be derived from [40, Theorem II.1]; but the input and the noise vectors are real-valued there whereas we shall consider complex-valued random vectors. The whiteness of the worst noise process proved in Proposition 3 does not follow immediately from the aforementioned result either.

## 2.3 OFDM Cellular network with MIMO

### 2.3.1 Network model

We consider a wireless network composed of several base stations (BS). The power transmitted by each BS is limited to some given maximal value. The network operates the *Orthogonal Frequency-Division Multiple Access* (OFDMA) linked to OFDM, which we describe now. The frequency spectrum allocated to the considered network is divided into a given number of sub-carriers, which are made available to all base stations. Each BS allocates disjoint subsets of the sub-carriers to its users. Each user is served by a single BS and receives only other-BS *interference*; that is the sum of powers emitted by other BS on the sub-carriers allocated to him by his serving BS. We consider *multiple input and multiple output* (MIMO) antennas. More precisely, BS are equipped with $t_A$ transmitting antennas whereas users have $r_A$ receiving antennas and each BS uses all its transmitting antennas to serve a given user.

We assume that the bandwidth of each sub-carrier is smaller than the *coherence frequency* of the channel, so we can consider that the *fading* in each sub-carrier is *flat* [23]. That is, the output of the channel at a given time depends on the input only at the same instant of time. Indeed, the use of a *cyclic prefix* in OFDM permits to transform the frequency selective fading

channel into a set of parallel *flat fading* channels [92, §3.4.4]. We don't make any assumption on the *correlation* of the fading processes corresponding to different subcarriers for a given user and a given BS. However, the fading processes for different users or base stations are assumed independent.

Time is divided into time-slots of length smaller than the *coherence time* of the channel, so that, for a given sub-carrier, the fading remains *constant during each time-slot* and the fading process in different time-slots may be assumed *ergodic*. Such model for fading generalizes the so-called *quasi-static* model where the fading process at different time-slots is assumed to be independent and identically distributed. We shall always assume that the receiver knows the fading.

The codeword duration equals the time-slot, which is assumed sufficiently large so that the *capacity* (*peak bit-rate*) within each time-slot may be defined in the *asymptotic* sense of the information theory.

Users perform *single user detection*; thus the interference from other BS is added to AWGN. The statistical properties of the interference are not known a priori since they depend on the coding of other BS.

Consider a user served by a BS indexed by $u$. For a given sub-carrier and time-slot, we consider the following discrete-time model of the OFDM channel with MIMO [23, Equation (7)]

$$Y_n = L_u^{-1/2} H_u \Psi_n + \Theta_n + \mathcal{I}_n, \quad n \in \mathbb{N} \tag{2.1}$$

where $n$ is a discrete-time index, $Y_n \in \mathbb{C}^{r_A}$ is the channel output, $\Psi_n \in \mathbb{C}^{t_A}$ is the channel input signal, $H_u$ is a complex matrix of dimension $r_A \times t_A$ representing the fading with the serving BS $u$, $\mathcal{I}_n \in \mathbb{C}^{r_A}$ is the interference, $\Theta_1, \Theta_2, \ldots$ are i.i.d. random noises with values in $\mathbb{C}^{r_A}$ such that each $\Theta_n$ is circularly-symmetric Gaussian with covariance matrix $E[\Theta_n \Theta_n^*] = \mathcal{N} I_{r_A}$ where $\Theta_n^*$ designates the transpose complex conjugate of $\Theta_n$, $\mathcal{N}$ is a given positive constant and $I_{r_A}$ is the identity matrix of dimension $r_A$, and $L_u$ is the *propagation loss* due to *distance* and *shadowing* between the user and BS $u$. The propagation loss $L_u$ is the ratio between the emitted and received powers, hence the factor $L_u^{-1/2}$ in Equation (2.1).The interference equals to

$$\mathcal{I}_n = \sum_{v \neq u} L_v^{-1/2} H_v \Psi_{v,n}$$

where the sum is over the interfering BS $v \neq u$, $\Psi_{v,n}$ is the transmitted signal by the interfering BS $v$, $H_v$ represents fading for BS $v$, and $L_v$ is the propagation loss due to distance and shadowing between the user and BS $v$.

We make the following probabilistic assumptions:

(H1) All the channel input signals are centred; i.e. $E[\Psi_n] = 0$.

(H2) The signals transmitted by different antennas including multiple antennas of the same BS are independent. Let $P$ be the power transmitted by each BS in a given sub-carrier aggregated over all the $t_A$ transmitters. Assume that this power is equally partitioned between the $t_A$ transmitting antennas; each one emitting a power $P/t_A$. This assumption is justified by the last statement in Proposition 3.

(H3) The fading matrices $H_v$ are constant for all channel uses $n \in \mathbb{N}$ within a given time-slot and sub-carrier. For a given BS $v$, the fading matrix $H_v$ is resampled across different time-slots, and we assume that it follows a stationary and ergodic sequence of random matrices. Moreover, these processes are independent across different BS $v$.

(H4) Users are motionless at the considered information theoretic time-scale; that is $L_v$ are constant for all BS $v$ and time-slots.

### 2.3.2 Link capacity given fading

In this section, we focus on a given time-slot and sub-carrier; that is, the expectation $E[\cdot]$ is taken *with respect to the distribution of the transmitted signals and noise*. By Assumptions (H1-H2), the covariance matrix of the transmitted signals are

$$E[\Psi_n \Psi_n^*] = E\left[\Psi_{v,n}\Psi_{v,n}^*\right] = \frac{P}{t_A} I_{t_A} \tag{2.2}$$

and the covariance matrix of the interference equals

$$\begin{aligned} E[\mathcal{I}_n \mathcal{I}_n^*] &= E\left[\sum_{v\neq u} L_v^{-1} H_v \Psi_{v,n}\Psi_{v,n}^* H_v^*\right] \\ &= \sum_{v\neq u} L_v^{-1} H_v E\left[\Psi_{v,n}\Psi_{v,n}^*\right] H_v^* = \frac{P}{t_A}\sum_{v\neq u}\frac{H_v H_v^*}{L_v} \end{aligned}$$

Noise and interference are assumed independent, thus the covariance matrix of $\Sigma_n = \Theta_n + \mathcal{I}_n$ is

$$\Delta = E[\Sigma_n \Sigma_n^*] = \mathcal{N} I_{r_A} + \frac{P}{t_A}\sum_{v\neq u}\frac{H_v H_v^*}{L_v} \tag{2.3}$$

The capacity of a channel may be interpreted as a maximal average bitrate sustainable in long communication time. The capacity $C$ is defined in Section 2.6.2.

**Proposition 1** *The capacity $C$ of the OFDM channel with MIMO (2.1) with power constraint (2.2) is lower bounded by*

$$C \geq \log_2 \det\left(I_{r_A} + \frac{P}{t_A}\frac{H_u H_u^*}{L_u}\Delta^{-1}\right) \tag{2.4}$$

*where the noise plus interference covariance matrix $\Delta$ is given by (2.3).*

**Proof.** The mathematical background and proof are defended in Section 2.6.2. ■

We call the right-hand side of the above equation[1] *feasible* bit-rate for the considered user. Since our assumptions (H1)-(H5) are the same for all users, we get similar expressions for the feasible bit-rates of the other users and this *collection of bit-rates* of the different users is *feasible*.

**Remark 1** Continuous-time. *Consider a continuous-time model of the channel (2.1). Let $w$ be the bandwidth of the considered sub-carrier. The results in the discrete-time extend to the continuous-time case, but the capacity bounds, such as the right-hand side of (2.4), should be multiplied by the bandwidth $w$ of the considered sub-carrier.*

### 2.3.3 Ergodic capacity

Consider now a given sub-carrier and multiple time-slots. Recall that we assumed that the fading matrices are ergodic across different time-slots. Then, by the ergodic theorem, the capacity averaged over a large number of time-slots approaches the *ergodic capacity* $E[C]$ where the expectation is taken *with respect to the fading distribution*.

[1] which is consistent with [22, Equation (2)]

**Corollary 1** *The ergodic capacity of the OFDM channel with MIMO (2.1) is lower bounded by*

$$E\left[C\right] \geq E\left[\log_2 \det\left(I_{r_A} + \frac{P}{t_A}\frac{H_u H_u^*}{L_u}\Delta^{-1}\right)\right] \tag{2.5}$$

*where* $\Delta$ *is given by (2.3).*

**Proof.** The result follows by taking the expectation of Equation (2.4) with respect to the fading. ■

### Asymptotic bound

The right-hand side of the Equation (2.5) may be approximated using the following asymptotic result when the number of transmitting and receiving antennas goes to infinity. As will be shown in Appendix 2.1.3, the approximation remains reasonable even for a moderate number of antennas.

**Lemma 1** *[69, Appendix] Assume that the fading matrix of each base station has i.i.d. centred components with variance* 1*. (Recall that we have already assumed that* $H_v$ *are independent across* $v$*.) Assume that* $t_A, r_A \to \infty$ *such that* $\frac{t_A}{r_A} \to Q$*, then*

$$\frac{1}{r_A}\log\det\left(I_{r_A} + \frac{P}{t_A}\frac{H_u H_u^*}{L_u}\left(\mathcal{N} I_{r_A} + \frac{P}{t_A}\sum_{v\neq u}\frac{H_v H_v^*}{L_v}\right)^{-1}\right) \tag{2.6}$$

*converges almost surely to*

$$\begin{aligned} &Q\sum_{v\neq u}\log\left(\frac{L_v + \frac{P}{\mathcal{N}}\frac{\eta_1}{Q}}{L_v + \frac{P}{\mathcal{N}}\frac{\eta_2}{Q}}\right) + Q\log\left(1 + \frac{P}{\mathcal{N}L_u}\frac{\eta_1}{Q}\right) \\ &+ \log\left(\frac{\eta_2}{\eta_1}\right) + \eta_1 - \eta_2 \end{aligned} \tag{2.7}$$

*where* $\eta_1$ *and* $\eta_2$ *are respectively solutions of*

$$\eta_1 + \sum_v \frac{P\eta_1}{\frac{P}{Q}\eta_1 + \mathcal{N}L_v} = 1$$

$$\eta_2 + \sum_{v\neq u} \frac{P\eta_2}{\frac{P}{Q}\eta_2 + \mathcal{N}L_v} = 1$$

The above expressions involve solutions of two non linear equations, which require the knowledge of the received powers from all interfering base stations. In what follows we will establish another lower bound for the capacity, whose evaluation is much simpler, as simple as the evaluation of the capacity of the AWGN channel, and requires only the knowledge of the interference power aggregated over all the interfering base stations. We shall compare the two bounds numerically in Section 2.3.3 below.

**Jensen's bound**

The following proposition gives a lower bound for the ergodic capacity under the assumption that the covariance of the fading matrix $H_v$ equals identity; that is

$$E\left[H_v H_v^*\right] = I_{r_A}, \quad \text{for all BS } v$$

which means in particular that the fadings of two different transmitting antennas are uncorrelated.

**Proposition 2** *Assume that $E\left[H_v H_v^*\right] = I_{r_A}$, for all base station $v$, then the ergodic capacity $E[C]$ of the channel (2.1) is lower bounded by*

$$E[C] \geq E\left[\log_2 \det\left(I_{r_A} + \mathrm{SINR} H_u H_u^*\right)\right] \tag{2.8}$$

*where*

$$\mathrm{SINR} = \frac{\left(P/t_A\right)/L_u}{\mathcal{N} + \left(P/t_A\right)\sum_{v \neq u} 1/L_v} \tag{2.9}$$

The SINR in the above equation can be seen as the *Signal to Interference and Noise Ratio* per transmitting antenna.

**Proof.** Let $E\left[\cdot|H_u\right]$ designate the expectation conditionally to $H_u$. By the properties of the conditional expectation we have

$$E[C] = E\left[E\left[C|H_u\right]\right]$$

Equation (2.4) implies that

$$E\left[C|H_u\right] \geq E\left[\log_2 \det\left(I_{r_A} + \frac{P}{t_A}\frac{H_u H_u^*}{L_u}\Delta^{-1}\right)\middle| H_u\right]$$

where $\Delta$ is given by (2.3).

Using Jensen's inequality and convexity of the function $A \mapsto \log_2 \det\left(I_{r_A} + \frac{P}{t_A} H H^* A^{-1}\right)$ on the set of positive definite matrices of $\mathbb{C}^{r_A \times r_A}$ (cf. [40, Lemma II.3]), we deduce that

$$E\left[C|H_u\right] \geq E\left[\log_2 \det\left(I_{r_A} + \mathrm{SINR} H_u H_u^*\right)|\, H_u\right]$$

where the SINR is given by (2.9). Thus

$$\begin{aligned} E[C] &= E\left[E\left[C|H_u\right]\right] \\ &\geq E\left[E\left[\log_2\left(1 + \mathrm{SINR} H_u H_u^*\right)|H_u\right]\right] \\ &= E\left[\log_2\left(1 + \mathrm{SINR} H_u H_u^*\right)\right] \end{aligned}$$

■

**Remark 2** *Note that the right-hand side of (2.8) represents the capacity of a MIMO channel with AWGN channel and i.i.d. circularly symmetric Gaussian fading given in Telatar [91, Theorem 1]. Thus, it may be calculated using the analytic formula given in [91, Theorem 2] or approximated with the help of the asymptotic result of Lemma 4 stated in the Appendix as follows*

$$\begin{aligned} E[C] &\geq E\left[\log_2 \det\left(I_{r_A} + \mathrm{SINR} H_u H_u^*\right)\right] \\ &\simeq \frac{r_A}{\log(2)} \mathcal{C}\left(t_A \times \mathrm{SINR}, \frac{t_A}{r_A}\right) \end{aligned} \tag{2.10}$$

*where $\mathcal{C}$ is given by (2.30).*

**Remark 3** Time/frequency diversity. *Averaging over a large number of time-slots corresponds to exploiting the so-called* time-diversity*, which is suitable for the analysis of the performance of a* variable bit-rate *traffic as observed in [30, §I]. Consider now a given time-slot and large number $n$ of sub-carriers. Assume that the fadings for different sub-carriers are i.i.d., then again, by the law of large numbers, the capacity of a large number $n$ of sub-carriers approaches the ergodic capacity. Thus the ergodic capacity is also appropriate for a* constant bit-rate *traffic provided the number of sub-carriers allocated to each user is large enough. If the number of sub-carriers allocated to each user is not sufficiently large, then, as observed in [30, §I], a relevant performance indicator is the* outage probability*, defined as the probability that the capacity in a given time-slot is smaller than the desired bit-rate $r$. Evaluation of this latter characteristic is not in the scope of this thesis.*

### Comparison of the lower bounds

We aim now to compare numerically the bounds (2.7) and (2.8). In this regard, we consider a hexagonal cell surrounded by 6 neighboring base stations. The distance between two base stations is 0.5km and the distance propagation law, i.e. path-loss is $l(r) = (Kr)^{\beta}$ where $K = 7764$, $\beta = 3.52$ which are the typical values in urban areas. We consider that a noise power equals $\mathcal{N} = -93$dBm, standard deviation of shadowing of 8dB and a transmission power of the base station $P = 58.5$dBm. We consider 2 receiving antennas and a number of transmitting antennas $t_A \in \{1, 2, 8\}$. Figure 2.1 gives the capacity lower bounds (2.7) and (2.8) called, respectively, asymptotic and Jensen bound, as function of the distance between the user and the central base station. This figure shows that the two bounds are close to each other.

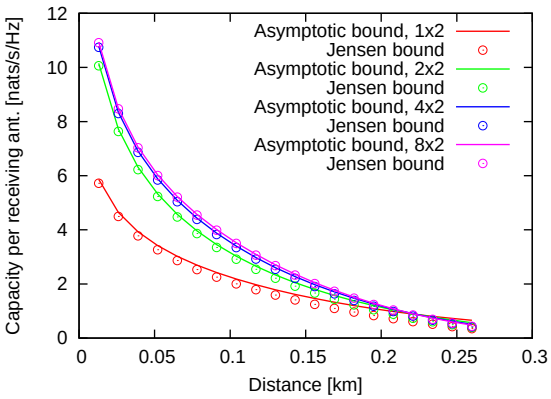


Figure 2.1: Jensen (2.8) and asymptotic (2.7) lower bounds for the capacity (peak bit-rate) of the MIMO flat fading channel with additive noise in the downlink of an OFDMA cellular network. Capacity as function of the distance between the user and his serving base station.

## 2.4 MMSE

The linear MMSE (Minimum Mean Square Error) decoder of the channel (2.1) means that the receiver estimates the transmitted signal $\Psi_n$ using a linear transformation $\Psi_n$ of the received signal $Y_n$ minimizing the error $E\left[\left\|\Psi_n - \hat{\Psi}_n\right\|^2\right]$. MMSE-SIC (Successive Interference Cancella-

tion) consists in decoding successively the $t_A$ transmitting antennas while suppressing recursively the previously decoded signals. The objective of the present section is to establish lower bounds for the capacity of MMSE and MMSE-SIC for a cellular network where interference is neither white nor Gaussian.

### 2.4.1 MMSE capacity given fading

We consider a single time-slot and sub-carrier in this section; thus the fading is assumed given. It follows from the general theory of linear estimation [28, §3.3] that

$$\hat{\Psi_n} = \Gamma_{\Psi_n Y_n} \Gamma_{Y_n}^{-1} Y_n, \quad n \in \mathbb{N} \tag{2.11}$$

where $\Gamma_{Y_n} = E\left[Y_n Y_n^*\right] = \Gamma_\Sigma + \frac{P}{t_A L_u} HH^*$ is the covariance matrix of $Y_n$ and $\Gamma_{\Psi_n Y_n} = E\left[\Psi_n Y_n^*\right] = \frac{P}{t_A L_u^{1/2}} H^*$ is the covariance matrix of $\Psi_n$ and $Y_n$. Equation (2.11) is in fact a system of $t_A$ equations corresponding to the estimation of the signals emitted by the different transmitting antennas of the serving BS. More specifically, denoting by $\Psi_n(k)$ the signal emitted by the $k$-th antenna and $\hat{\Psi}_n(k)$ the corresponding estimation, Equation (2.11) decomposes into

$$\hat{\Psi}_n(k) = \Gamma_{\Psi_n(k) Y_n} \Gamma_{Y_n}^{-1} Y_n, \quad n \in \mathbb{N}$$

where $\Gamma_{\Psi_n(k) Y_n} = E\left[\Psi_n(k) Y_n^*\right]$. The above equation may be written in the form

$$\hat{\Psi}_n(k) = \alpha^* \Psi_n(k) + z_n$$

where $\alpha \in \mathbb{C}$ and $z_n$ is a random variable with values in $\mathbb{C}$. The above expression may be seen as the input-output relation of an additive channel corresponding to the $k$-th transmitting antenna. It is shown in [92, Equation (8.67)] that the corresponding signal to noise power ratio equals

$$\mathrm{SNR}_k = \frac{E\left[\left|\alpha^* \Psi_n(k)\right|^2\right]}{E\left[\left|z_n\right|^2\right]} = \frac{P}{t_A L_u} h_k^* \Gamma_k^{-1} h_k, \quad k = 1, \ldots, t_A \tag{2.12}$$

where $h_k$ is the $k$-th column of the fading matrix $H_u$ and

$$\Gamma_k = \Delta + \sum_{i=1, i \neq k}^{t_A} \frac{P}{t_A L_u} h_i h_i^*, \quad k = 1, \ldots, t_A$$

where $\Delta$ is given by (2.3). It follows from Corollary 2 that the capacity of the $k$-th transmitting antenna (when considering interference from other antennas as well as from other BS) is lower bounded by

$$C_k \geq \log_2\left(1 + \mathrm{SNR}_k\right), \quad k = 1, \ldots, t_A$$

Thus the capacity of the channel is lower bounded by

$$C_{\mathrm{MMSE}} = \sum_{k=1}^{t_A} C_k \geq \sum_{k=1}^{t_A} \log_2\left(1 + \mathrm{SNR}_k\right) \tag{2.13}$$

### 2.4.2 MMSE ergodic capacity

Consider now a given sub-carrier and multiple time-slots. The capacity of the channel is then the expectation of the above capacity (2.13) with respect to fading; thus

$$\begin{aligned} E\left[C_{\mathrm{MMSE}}\right] &\geq \sum_{k=1}^{t_A} E\left[\log_2\left(1+\mathrm{SNR}_k\right)\right] \\ &= \sum_{k=1}^{t_A} E\left[E\left[\log_2\left(1+\mathrm{SNR}_k\right)\middle| H_u\right]\right] \end{aligned}$$

Using [40, Lemma II.3] and Jensen's inequality, it follows that

$$E\left[C_{\mathrm{MMSE}}\right] \geq \sum_{k=1}^{t_A} E\left[\log_2\left(1+\frac{P}{t_A L_u} h_k^* \bar{\Gamma}_k^{-1} h_k\right)\right] \tag{2.14}$$

where

$$\bar{\Gamma}_k = E\left[\Gamma_k | H_u\right] = \left(\mathcal{N} + \frac{P}{t_A}\sum_{v\neq u}\frac{1}{L_v}\right) I_{r_A} + \sum_{i=1, i\neq k}^{t_A} \frac{P}{t_A L_u} h_i h_i^*$$

The right-hand side of (2.14) may be evaluated numerically using Monte Carlo method based on samples of the fading matrix $H_u$.

### 2.4.3 MMSE-SIC

As we said previously, MMSE-SIC consists in decoding successively the $t_A$ transmitting antennas, but before decoding the signal from a given antenna we suppress the previously decoded signals. Thus the channel for the $k$-th transmitting antenna is an additive channel with SNR given by (2.12) where the matrix $\Gamma_k$ is now given by

$$\Gamma_k = \Delta + \sum_{i=k+1}^{t_A} \frac{P}{t_A L_u} h_i h_i^*, \quad k = 1, \ldots, t_A$$

The lower bound (2.13) of capacity given the fading remains valid with the above modification of SNR. The lower bound (2.14) of the ergodic capacity holds also true with

$$\bar{\Gamma}_k = \left(\mathcal{N} + \frac{P}{t_A}\sum_{v\neq u}\frac{1}{L_v}\right) I_{r_A} + \sum_{i=k+1}^{t_A} \frac{P}{t_A L_u} h_i h_i^*$$

The proof is based on Jensen inequality and follows the same lines as for MMSE. It follows from [92, Equation (8.71)] that

$$\begin{aligned} &\sum_{k=1}^{t_A} E\left[\log_2\left(1+\frac{P}{t_A L_u} h_k^* \bar{\Gamma}_k^{-1} h_k\right)\right] \\ &= E\left[\log_2 \det\left(I_{r_A} + \mathrm{SINR} H_u H_u^*\right)\right] \end{aligned}$$

where SINR is given by (2.9). Note that the right-hand sides of the above equation and Equation (2.8) are equal; that is we retrieve the same capacity lower bound as for the original channel (2.1).

## 2.5 Numerical results for the link capacity

The objective of the present section is to compare the theoretical expressions established in the previous section to real field measurements and to some simulation compliant with the 3GPP recommendation [3].

### 2.5.1 Link layer model calibration

We consider firstly a user served by a base station through an additive white Gaussian noise (AWGN) SISO channel neglecting fading and interference for the moment. The user gets ideally (i.e. in the asymptotic sense of information theory) a bit-rate given by the famous Shannon's formula $w \log_2(1 + \mathrm{SNR})$ where $w$ is the bandwidth allocated to the considered user and SNR is the signal to noise power ratio. In order to get rid of the dependence of the bit-rate on the bandwidth, we define the *spectral efficiency* as the ratio of the bit-rate to the bandwidth which equals $\log_2(1 + \mathrm{SNR})$ in the AWGN context.

Mogensen et al. [73], [53] and the 3GPP [4, §A.2] have observed that the LTE system spectral efficiency in this AWGN context is well approximated by

$$C \simeq c \log_2(1 + q\mathrm{SNR}) \tag{2.15}$$

for some constant $c < 1$ and $q$ accounting on the one hand for the gap between the practical coding schemes and the optimal ones and on the other hand for the loss of capacity due to signalling. This observation shall be confirmed and the typical value of $c$ and $q$ for LTE will be given.

First, we will *calibrate* these parameters $c$ and $q$ for real coding schemes considering the simplest AWGN SISO channel, and then use them in the analysis of the MIMO channel with fading and interference.

Note that the relative difference $1 - c$ for $q = 1$ between the Shannon's limit and the practical LTE system may be seen as a progress margin for potential evolution of the technology in the AWGN context.

According to [2, §6.8], [35, p.155] LTE signalling consumes about 30% of the available capacity. On the other hand, different *M-QAM modulations* with $M \in \{4, 16, 64\}$ are used with link adaptation and a target block error rate $10^{-2}$. Moreover, *CRC* and *turbo coding* are implemented. The 3GPP [4, §A.2] shows that the bit-rate of LTE is about 25% smaller than the Shannon capacity. In order to account for these losses, we assume

$$c = (1 - 0.3) \times (1 - 0.25) \simeq 0.5, \quad \text{and } q = 1 \tag{2.16}$$

in Equation (2.15).

Figure 2.2 shows that the *analytic approximation* (2.15) of the SISO capacity with the values of $c$ and $q$ proposed in (2.16) fits well the results of Orange's link simulation tool in AWGN. Thus we will retain $c = 0.5$ and $q = 1$ to weigh, respectively, the capacity and the SINR in the subsequent analysis of the MIMO channel with fading and interference.

More specifically, using the modified AWGN formula (2.15), the capacity lower bound (2.10) becomes

$$E[C] \geq c \frac{r_A}{\log(2)} \mathcal{C}\left(t_A \times \mathrm{SINR}, \frac{t_A}{r_A}\right) \tag{2.17}$$

where the parameter $c$ is given by (2.16) and the function $\mathcal{C}$ is given by Equation (2.30).

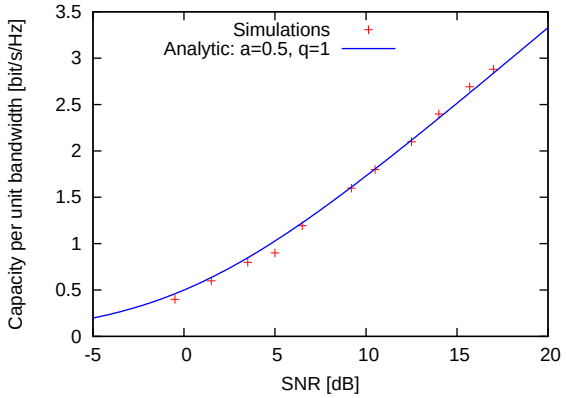


Figure 2.2: Performance of SISO without fading evaluated using analytic approximation (2.15) and link simulation.

### 2.5.2 Comparison to simulation

We compare now our theoretical bound to 3GPP simulation [3]. The simulation is carried out under the following assumptions. There are 2 transmitting and 2 receiving antennas. Each base station always transmits at its maximal power. The receiver is MMSE-SIC (interference cancellation), the channel is 3GPP Spatial Channel Model (SCM) and users' speed is 3km/h. Several realizations of the user positions, shadowing losses and fading channels are generated. For each user location and each shadowing realization, the capacity is averaged over about 1000 fadings samples. Moreover, the value of the SINR including only the distance and the shadowing effects (and not fading) is also given. The simulation accounts for correlations between individual MIMO sub-antennas.

Figure 2.3 shows the simulation results compared to the analytic bound for MMSE-SIC with correlated antennas. Observe that the analytic curves agree with simulation results; but there is more variability in simulation due to the averaging over fading which does not yet converge to the ergodic capacity.

### 2.5.3 Comparison to measurements

We take measurements from the city of Marseille. These measurements are collected by dedicated users in the downlink of Orange's experimental LTE network composed of 75 cells each having 2 transmitting antennas. The mobiles used for measurements have also 2 receiving antennas. Carrier frequency is 2.6GHz, bandwidth is 20MHz.

Figure 2.4 shows these measurements compared to the analytic bounds (2.14) and (2.8) for MMSE and MMSE-SIC respectively.

The curve 'MMSE Correl' shows the results of the MMSE scheme with correlations between individual MIMO antennas; more precisely we assume a correlation factor of 0.3 for transmitting antennas and 0.9 for receiving antennas as proposed by 3GPP [5, §B.2.3]. We observe that this assumption fits the real performance of the current network. The curve 'MMSE-SIC Correl' predicts the performance of the MMSE-SIC scheme still with correlated antennas. If technology allows for decorrelated antennas, then the performance can reach the values predicted by "MMSE Uncorr' and 'MMSE-SIC Uncorr' depending on the used scheme. Recall that MMSE-SIC with decorrelated antennas gives the full MIMO capacity and that the corresponding analytic

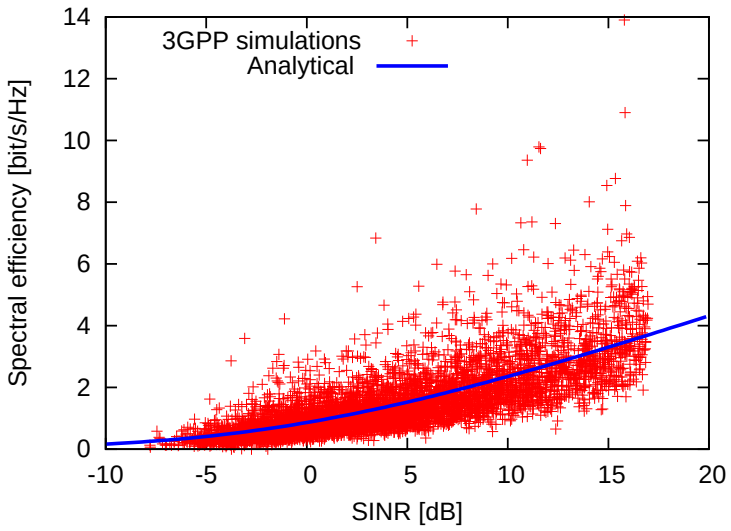


Figure 2.3: Analytic relation of the peak bit-rate to SINR compared to 3GPP simulation

bound (2.8) is well approximated by the asymptotic expression (2.17) as shown in Section 2.3.3.

### 2.5.4 Approximate link quality estimation via simulations

The goal is to use the Orange's internal 3GPP simulators mentioned earlier and develop a quick and simple estimation of link performance for different configurations of MIMO. Indeed the signalling loss depends on the number of transmitting and receiving antennas and consequently the weighting constants $c$ and $q$. It is about $40/168 = 24\%$, $48/168 = 29\%$ and $52/168 = 31\%$ respectively for SIMO $1 \times 2$, MIMO $2 \times 2$ and MIMO $4 \times 2$ (see [2, §6.8], [35, p.155]). Here, all the considered cases have a MRC (Maximum Ratio Combining) receiver, except the MIMO $4 \times 2$ case which has a MMSE receiver, which is different compared to Section 2.5.2, where MMSE-SIC is used. At the base station side, the transmitting antennas are pairwise cross-polar. In the case MIMO $4 \times 2$ the two cross-polar pairs of transmitting antennas are separated by 10 times the wavelength.

In order to simplify the notation, we denote by $\hat{S}$ the analytical (lower bound for the) spectral efficiency given in the right-hand side of (2.8) weighted by the parameter $c = 0.5$ obtained in the previous section; that is

$$\hat{S}\left(\mathrm{SINR}, t_A, r_A\right) = cE\left[\log_2 \det\left(I_{r_A} + H_u H_u^* \mathrm{SINR}\right)\right] \tag{2.18}$$

where SINR is the signal to interference and noise ratio (per transmitting antenna) given by Equation (2.9).

In order to get the practical LTE performance, we make the same kind of comparison as in Section 2.5.2. We consider the output of Orange's simulator compliant with the 3GPP recommendation [3] (see this reference for the details of the simulations) in the so-called *calibration* case. It corresponds to MIMO $1 \times 2$ with *round robin* (RR) scheduler. We consider also other MIMO configurations and *proportional fair* (PF) scheduler, keeping all the other parameters unchanged. In particular, each base station always transmits at its maximal power (*full buffer*).

The spectral efficiency as function of the SINR is compared to the theoretical relation (2.18) or equivalently (2.17). More specifically, we make a linear regression between the spectral efficiency

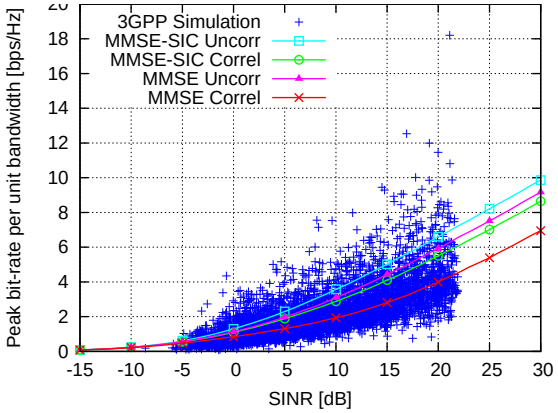


Figure 2.4: Analytic relations of the peak bit-rate to SINR compared to measurements; see Section 2.5.3 for the details.

| MIMO | Scheduler | $b$ | residual stand. dev. | $b'$ |
|---|---|---|---|---|
| $1 \times 2$ | RR | 0.83 | 0.45 | 0.98 |
| $1 \times 2$ | PF | 1.02 | 0.65 | 1.19 |
| $2 \times 2$ | PF | 0.67 | 0.74 | 1.08 |
| $4 \times 2$ | PF | 0.49 | 0.76 | 0.90 |

Table 2.1: Results of the linear fittings.

obtained from simulations and the theoretical efficiency given by Equation (2.18); that is we search for some $b$ such that

$$\hat{s} \simeq b \times \hat{S}\left(\mathrm{SINR}, t_A, r_A\right) \tag{2.19}$$

Table 2.1 gives the results of the linear fitting (2.19); i.e. the values of $b$ and the corresponding residual standard deviation for different MIMO configurations (the first row corresponds to the calibration case [3, Table A.2.2-1]). Moreover, the 95%-confidence interval is about $b \pm 0.01$ for all the studied cases.

Figure 2.5 shows the spectral efficiency as function of the SINR from simulations and from the analytical expression (right-hand side of (2.19)) for the calibration case. Observe again that the analytical expression reproduces well the general tendency of the empirical data obtained from simulations, similar as in Section 2.5.2.

**Remark 4** *In order to simplify the calculations we have also tested a linear regression between the spectral efficiency $\hat{s}$ obtained from simulations and the AWGN expression (2.15). Observe from Equation (2.9) that when noise is dominant against interference, then*

$$\mathrm{SINR} = \frac{\left(P/t_A\right)/L_u}{\mathcal{N}} = \frac{P/L_u}{\mathcal{N}} \times \frac{1}{t_A}$$

*Thus, in this particular case, the term $\frac{P/L_u}{\mathcal{N}}$ in the right-hand side of (2.15) equals* $\mathrm{SINR} \times t_A$. *Then, in the general case, it is natural to look for a fitting in the form*

$$\hat{s} \simeq b' \times c \log_2\left(1 + \mathrm{SINR} \times t_A\right)$$

*The resulting values of $b'$ are indicated in Table 2.1 with residual standard deviations close to those indicated in the fourth column of that table.*

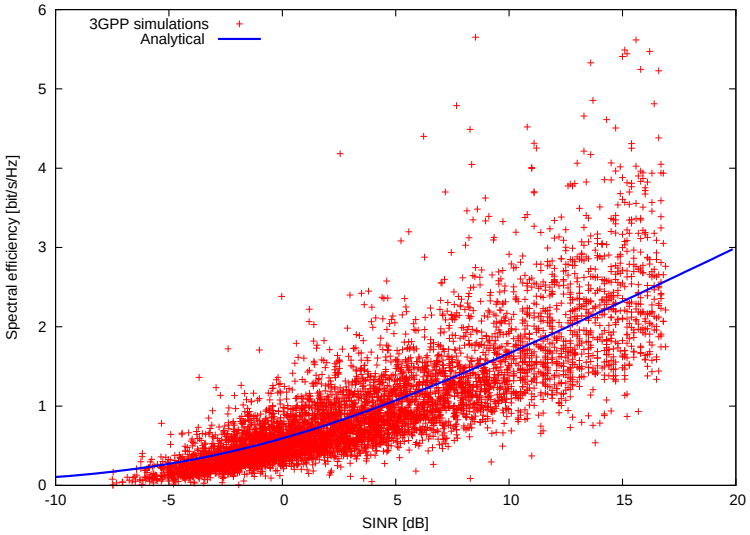


Figure 2.5: Simulations versus the analytical expression (right-hand side of (2.19)) for the calibration case

## 2.6 Link quality observed by a typical user

The objective of the present section is to estimate the spatial distribution of the link performance parameters (SINR, spectral efficiency etc.) observed by a typical user whose location will be randomly chosen in the network. We generate the link performance parameters distribution based on the aforementioned study and in the context of the 3GPP simulation scenario [3]. So, we want to apply the above analysis to produce for example the spatial distribution of the SINR, and in such a manner provide the link quality ingredient as a corner stone in our further Quality of Service (QoS) examination.

### 2.6.1 SINR

For the analytical approach we use a similar geometric pattern of the network (hexagonal) and the same propagation-loss modeling regarding the distance and shadowing effects (fading has been already taken into account on the link level in the previous section) as the 3GPP calibration case [1, Table A.2.1.1-3] and [3, Table A.2.2-1].

More specifically, the frequency carrier is 2GHz. The path-loss model is $l(r) = 128.1 + 37.6 \times \log_{10}(r)$ [in dB]. A supplementary penetration loss of 20dB is added. The shadowing is modeled as a centered log-normal random variable of standard deviation 8dB. The following 2D horizontal antenna pattern is used

$$A(\varphi) = -\min\left(12\left(\frac{\varphi}{\zeta}\right)^2, A_{\mathrm{m}}\right), \quad \zeta = 70^\circ, A_{\mathrm{m}} = 20\mathrm{dB} \tag{2.20}$$

The system bandwidth is $W = 10$MHz, the noise power equals $\mathcal{N} = -95$dBm ($-174$dBm/Hz, noise figure=9dB) and the transmission power of the base station is $P = 60$dBm (46dBm plus $G = 14$dBi of antenna gain). The network is composed of 36 hexagons ($6 \times 6$). Each hexagon comprises three sectors which gives a total of 108 sectors. The distance between the centers of

two neighboring hexagons is 500m. We generate 3600 random user locations uniformly in the network; that is 100 user locations per hexagon on average.

The 3GPP simulations published in [3] are made on a planar network with random locations of the users. In the present study, two network models are considered: either planar or toroidal (to avoid the border effects).

Each mobile is served by the base station with the smallest *propagation-loss* (including distance, shadowing and antenna pattern). In order to facilitate the comparison of our results to those of 3GPP, we define the coupling-gain as the antenna gain $G$ minus propagation-loss $L$ with the serving base station. The cumulative distribution function (CDF)[2] of the coupling gain obtained by 3GPP simulations [3, Figures A.2.2-1 (left)] and by our models are given in Figure 2.6. This figure shows that the results of our planar network are close to those of 3GPP simulations, whereas those of the toroidal network give larger coupling gain. This is due to the fact that in a planar network edge users get smaller coupling gain than in the toroidal one.

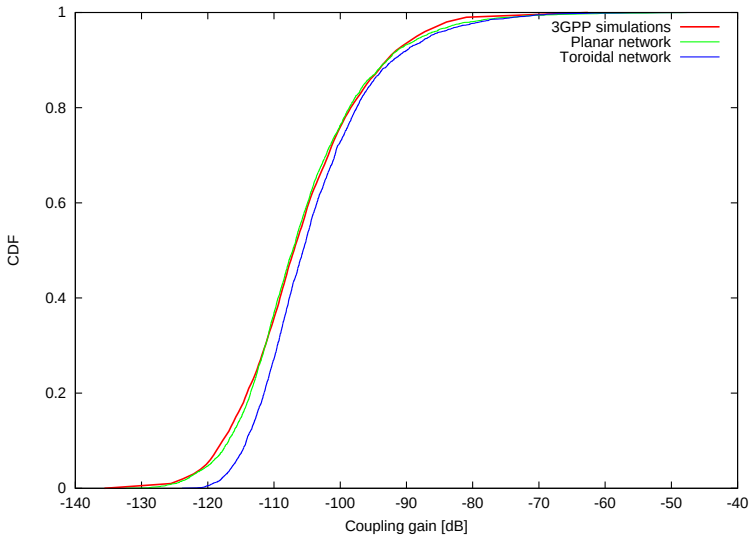


Figure 2.6: CDF of the coupling gain (antenna gain minus propagation loss)

The SINR for each mobile is calculated by Equation (2.9), where $u$ is the index of the serving base station. Figure 2.7 shows the CDF of the SINR coming from 3GPP simulations [3, Figure A.2.2-1 (right)] compared to that resulting from our models. Again our planar model gives closer results to the 3GPP simulations than the toroidal one. Nevertheless, the difference between the SINRs of the toroidal and the planar networks is smaller than 0.5dB.

Figure 2.7 shows that the SINR does not exceed 17dB. Indeed, each mobile served by a given base station (sector) is at least interfered by the two other sectors on the same site. The power received from each of these sectors is at least $10^{-2}$ times that received from the serving BS (this is related to $A_{\mathrm{m}} = 20$dB in Equation (2.20)). The interference to signal ratio is consequently larger than $2 \times 10^{-2}$ i.e. $-17$dB which explains the observed upper limit of the SINR.

**Remark 5** *Observe that the SINR defined by Equation (2.9) is different from the SINR calcu-*

[2] over all the user locations in the network

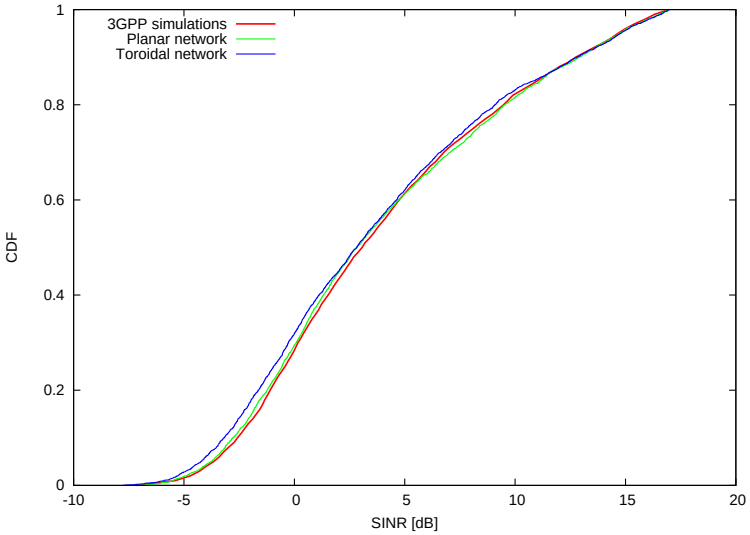


Figure 2.7: CDF of SINR

*lated by 3GPP simulations which equals*

$$\mathrm{SINR}_{3\mathrm{GPP}} = \frac{P/L_u}{\mathcal{N} + P\sum_{v\neq u} 1/L_v}$$

*However, if noise is negligible compared to interference, then the two SINRs are identical. This is the case in the considered urban scenario (small cell radius), so we do not have to distinguish between these two SINRs.*

### 2.6.2 Spectral efficiency

For each mobile we calculate the spectral efficiency corresponding to its SINR by relation (2.19). In order to facilitate the comparison of our results to those of 3GPP, we define the normalized user throughput as the spectral efficiency divided by 10 (this is historically related to the fact there are 10 users per cell in 3GPP simulations). The CDFs of the normalized user throughput obtained by 3GPP simulations [3, Figure A.2.2-3 (left)] and by our model are plotted in Figure 2.8. The 3GPP distribution is more spread than that of our models; this is related to the fact that the 3GPP spectral efficiency represents some variability around the analytic one as shown in Figure 2.5. Moreover, we observe that the results of the planar and toroidal models for the network are close to each other. Thus, the toroidal model is considered for the remaining part of the Section.

Table 2.2 gives the *arithmetic mean* of the spectral efficiencies at the different locations (called cell spectral efficiency) for both, the 3GPP simulations and analytic approach. The results of two methods agree for all the considered MIMO and scheduler configurations.

Note that the results of the simulations given in Table 2.2 are produced by the simulator of Orange which is one of the contributors to 3GPP. The values indicated in [3, Table A.2.2-2] are in fact averaged over the different 3GPP contributors including Orange. In particular, for the calibration case (MIMO $1 \times 2$ with RR scheduler) Orange's result is 1.01 whereas 3GPP average is 1.1. The variability of the results amoung the contributors is partially due to the randomness induced by the shadowing.

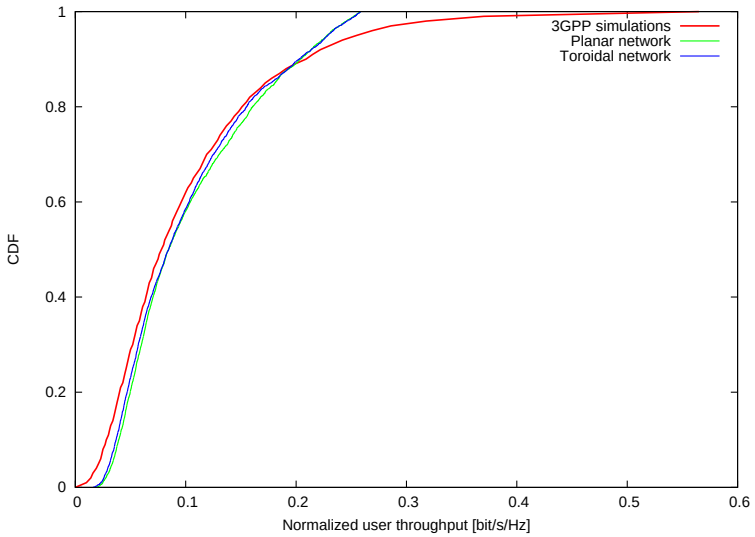


Figure 2.8: CDF of normalized user throughput

| | | Arithmetic mean | | Harmonic mean | |
|---|---|---|---|---|---|
| MIMO | Scheduler | Simus | Analytic | Simus | Analytic |
| $1 \times 2$ | RR | 1.01 | 1.00 | 0.50 | 0.69 |
| $1 \times 2$ | PF | 1.32 | 1.23 | 0.80 | 0.85 |
| $2 \times 2$ | PF | 1.43 | 1.41 | 0.84 | 1.00 |
| $4 \times 2$ | PF | 1.54 | 1.54 | 0.95 | 1.18 |

Table 2.2: Cell spectral efficiency: Comparison of the 3GPP simulations and the analytic results.

Table 2.2 shows the harmonic means of the spectral efficiency obtained from 3GPP simulations and from the analytical expression. The difference may be explained as follows. Recall that the harmonic mean is sensitive to the minimal value of the considered data; for example if one of these data is null then the harmonic mean vanishes. Moreover, Figure 2.5 shows that the 3GPP spectral efficiency represents some variability around (and in particular comprise smaller values than) the analytic curve. This explains why the harmonic means obtained from simulations in Table 2.2 are lower than the analytic ones.

# Appendix

## 2.A Theoretical results: MIMO flat-fading channel with additive noise

We shall establish in what follows a useful lower bound of the capacity of a general additive noise channel, where the noise is not necessarily Gaussian nor white. Our motivation is that interference in wireless networks does not have necessarily these properties.

### 2.1.1 Model

Consider a discrete-time model of a *multiple input and multiple output* (MIMO) channel with $t_A$ transmitting and $r_A$ receiving antennas such that, at each time $n = 1, 2, \ldots$, the channel output $Y_n \in \mathbb{C}^{r_A}$ is related to the channel input $\Psi_n \in \mathbb{C}^{t_A}$ by

$$Y_n = H\Psi_n + \Sigma_n, \quad n = 1, 2, \ldots \tag{2.21}$$

where $H$ is a complex matrix of dimension $r_A \times t_A$ modelling the *fading*, and $\Sigma_1, \Sigma_2, \ldots \in \mathbb{C}^{r_A}$ is the noise process. We assume that the fading matrix $H$ is deterministic. The channel input is subject to a power constraint of the form

$$\frac{1}{n}\sum_{k=1}^{n} \Psi_k^* \Psi_k \leq P, \quad n = 1, 2, \ldots$$

where $P$ is a given positive constant and $\Psi_k^*$ designates the *transpose complex conjugate* of $\Psi_k$. Note that the above constraint concerns the total power aggregated over all the $t_A$ transmitters and averaged over $n$ channel uses. The channel (2.21) is called *MIMO additive noise channel with deterministic fading*.

### 2.1.2 Capacity lower bound

We are interested in the *capacity* of the channel (2.21) when the noise samples $\Sigma_1, \Sigma_2, \ldots$ are not necessarily Gaussian nor independent. We shall in fact establish an explicit lower bound for this capacity.

We begin by some definitions and notation. The identity matrix of dimension $r_A \times r_A$ is denoted by $I_{r_A}$. The *covariance matrix* of a centred random vector $\Psi \in \mathbb{C}^{t_A}$ is denoted by

$$\Gamma_\Psi = E\left[\Psi\Psi^*\right]$$

The covariance matrix of two centred random vectors $\Psi \in \mathbb{C}^{t_A}$ and $Y \in \mathbb{C}^{r_A}$ is denoted by

$$\Gamma_{\Psi Y} = E\left[\Psi Y^*\right]$$

A random vector $\Psi \in \mathbb{C}^n$ is called *circularly-symmetric* if $e^{i\phi}\Psi$ has the same distribution as $\Psi$ for all $\phi \in \mathbb{R}$ which implies that $\Psi$ is centred.

From now on all the considered random vectors are assumed to have *well defined entropies* [56, § 1.3]. For example, if the random vector $\Psi \in \mathbb{C}^n$ has a density $p_\Psi$ with respect to the Lebesgue measure on $\mathbb{C}^n$, then its entropy is defined by $h(\Psi) = -\int_{\mathbb{C}^n} p_\Psi(x) \log p_\Psi(x) dx$ provided the Lebesgue integral is well defined. We denote by $I(\Psi; Y)$ the *mutual information* between two random vectors $\Psi$ and $Y$ which is related to the entropy by [56, Theorem 1.6.2]

$$I(\Psi; Y) = h(\Psi) - h(\Psi|Y) \tag{2.22}$$

where $h(\Psi|Y)$ is the entropy of $\Psi$ conditionally to $Y$.

We give now two preliminary lemmas.

**Lemma 2** *Let $\Psi_1, \Psi_2, \ldots, \Psi_n$ be random vectors in $\mathbb{C}^{t_A}$ and $Y_1, Y_2, \ldots, Y_n$ be random vectors in $\mathbb{C}^{r_A}$. Denote $\Psi^{(n)} = (\Psi_1, \Psi_2, \ldots, \Psi_n)$ and $Y^{(n)} = (Y_1, Y_2, \ldots, Y_n)$. If $\Psi_1, \Psi_2, \ldots, \Psi_n$ are independent, then*

$$I\left(\Psi^{(n)}; Y^{(n)}\right) \geq \sum_{k=1}^{n} I(\Psi_k; Y_k)$$

**Proof.** The mutual information may be expressed in terms of the entropy as follows [56, Theorem 1.6.2]

$$I\left(\Psi^{(n)}; Y^{(n)}\right) = h\left(\Psi^{(n)}\right) - h\left(\Psi^{(n)}|Y^{(n)}\right)$$

Since $\Psi_1, \Psi_2, \ldots, \Psi_n$ are independent, the entropy $h\left(\Psi^{(n)}\right)$ may be decomposed as the sum of the individual entropies [56, Theorem 1.3.2 (h.6)] $h\left(\Psi^{(n)}\right) = \sum_{k=1}^{n} h(\Psi_k)$. On the other hand, the conditional entropy $h\left(\Psi^{(n)}|Y^{(n)}\right)$ may be bounded as follows

$$\begin{aligned} h\left(\Psi^{(n)}|Y^{(n)}\right) &= \sum_{k=1}^{n} h\left(\Psi_k|Y^{(n)}, \Psi_1, \ldots, \Psi_{k-1}\right) \\ &\leq \sum_{k=1}^{n} h\left(\Psi_k|Y^{(n)}\right) \leq \sum_{k=1}^{n} h(\Psi_k|Y_k) \end{aligned}$$

where for the first equality we use [56, Theorem 1.3.2 (h.7)] and for the two above inequalities we use [56, Theorem 1.3.2 (h.7)] and [56, Theorem 1.3.2 (h.5)] respectively. Combining the above three equations, we get the desired result. ■

The following lemma may be seen as an extension of [56, Theorem 1.8.6] or [40, Lemma II.2] to the complex case. Our proof is inspired by [71] and [64].

**Lemma 3** *Consider three random vectors $\Psi \in \mathbb{C}^{t_A}$, $Y, \tilde{Y} \in \mathbb{C}^{r_A}$. Assume that the random vector $\left(\Psi, \tilde{Y}\right)$ is circularly-symmetric Gaussian with the same covariance matrix as $(\Psi, Y)$ and that $\Gamma_Y$ is invertible. Then*

$$I(\Psi; Y) \geq I(\Psi; \tilde{Y})$$

**Proof.** For any deterministic matrix $A \in \mathbb{C}^{t_A \times r_A}$,

$$h(\Psi|Y) = h(\Psi - AY|Y) \leq h(\Psi - AY) \tag{2.23}$$

where for the above inequality we use [56, Theorem 1.3.2]. In particular, taking $A = \Gamma_{\Psi Y}\Gamma_Y^{-1}$ in which case $AY$ is the best quadratic approximation of $\Psi$ by a linear function of $Y$, and letting $U = \Psi - AY$, we get

$$h(\Psi - AY) = h(U) \leq \log\left[\det\left(\pi e \Gamma_U\right)\right] \tag{2.24}$$

Combining the above two inequalities we get $h(\Psi|Y) \leq \log[\det(\pi e \Gamma_U)]$. Then Equation (2.22) implies

$$I(\Psi;Y) \geq h(\Psi) - \log\det(\pi e \Gamma_U) \tag{2.25}$$

Apply now the above arguments with $\tilde{Y}$ in the role of $Y$. Observe that $\tilde{U} = \Psi - A\tilde{Y}$ is circularly-symmetric Gaussian, thus equality holds in (2.24). Moreover, $\tilde{U}$ is independent from $\tilde{Y}$ since $\Gamma_{\tilde{U}\tilde{Y}} = E\left[\tilde{U}\tilde{Y}^*\right] = \Gamma_{\Psi Y} - A\Gamma_Y = 0$, and decorrelation implies independence for circularly-symmetric Gaussian random vectors. Thus equality holds also in (2.23) which shows that

$$I\left(\Psi;\tilde{Y}\right) = h(\Psi) - \log\det(\pi e \Gamma_{\tilde{U}})$$

which combined with the observation that $\Gamma_U = \Gamma_\Psi - \Gamma_{\Psi Y}\Gamma_Y^{-1}\Gamma_{Y\Psi} = \Gamma_{\tilde{U}}$ and (2.22) finishes the proof of the desired inequality. ■

We show now that the above lemmas permit to deduce a lower bound for the capacity of the channel (2.21). The considered channel has memory; i.e. different channels uses are not independent because of the noise samples might be correlated, thus its information capacity $C$ is defined as follows

$$C = \liminf_{n\to\infty} \frac{1}{n} C^{(n)}$$

where

$$C^{(n)} = \sup_{\Psi^{(n)}} \left\{ I\left(\Psi^{(n)};Y^{(n)}\right); \frac{1}{n}\sum_{k=1}^{n} E\left[\Psi_k^*\Psi_k\right] \leq P \right\}$$

where $\Psi^{(n)} = (\Psi_1, \ldots, \Psi_n)$ is a random object with values in $(\mathbb{C}^{t_A})^n$; and $Y^{(n)}$ is the output of the channel associated to the input $\Psi^{(n)}$.

**Proposition 3** *Assume that the covariance matrix $E[\Sigma_k\Sigma_k^*]$ of the noise $\Sigma_k$ is finite for all $k \in \mathbb{N}$ and denote*

$$\Delta_n = \frac{1}{n}\sum_{k=1}^{n} E[\Sigma_k\Sigma_k^*]$$

*Then the information capacity of the channel (2.21), given the fading matrix $H$, is lower bounded by*

$$C \geq \liminf_{n\to\infty} \left[\log_2\det\left(I_{r_A} + \frac{P}{t_A} H H^* \Delta_n^{-1}\right)\right] \tag{2.26}$$

*The above inequality remains true under the additional constraint that the signals emitted by the transmitting antennas are independent and have equal powers.*

**Proof.** Consider independent inputs $\Psi_1, \Psi_2, \ldots$, then by Lemma 2

$$I\left(\Psi^{(n)};Y^{(n)}\right) \geq \sum_{k=1}^{n} I(\Psi_k;Y_k)$$

Assume now that each $\Psi_k$ is circularly-symmetric Gaussian, independent of $\Sigma_k$, and with covariance matrix $E[\Psi_k\Psi_k^*] = \frac{P}{t_A} I_{t_A}$. Note that

$$E[\Psi_k Y_k^*] = E[\Psi_k\Psi_k^* H^*] + E[\Psi_k\Sigma_k^*] = \frac{P}{t_A} H^*$$

and

$$E[Y_k Y_k^*] = E[H\Psi_k\Psi_k^* H^*] + E[\Sigma_k\Sigma_k^*] = \frac{P}{t_A} H H^* + E[\Sigma_k\Sigma_k^*]$$

Consider $\Sigma_k$ circularly-symmetric Gaussian independent from $\Psi_k$ and with the same covariance matrix as $\Sigma_k$. Denoting $\tilde{Y}_k = H\Psi_k + \Sigma_k$, then $\left(\Psi_k, \tilde{Y}_k\right)$ is circularly-symmetric Gaussian with the same covariance matrix as $(\Psi_k, Y_k)$. We deduce from Lemma 3 that

$$\begin{aligned} I(\Psi_k; Y_k) &\geq I(\Psi_k; \tilde{Y}_k) \\ &= h(\tilde{Y}_k) - h(\tilde{Y}_k|\Psi_k) \\ &= h(\tilde{Y}_k) - h(\Sigma_k) \\ &= \log_2 \left[\det\left(\pi e \left(\frac{P}{t_A} HH^* + E[\Sigma_k \Sigma_k^*]\right)\right)\right] \\ &\quad - \log_2 \left[\det\left(\pi e E[\Sigma_k \Sigma_k^*]\right)\right] \\ &= \log_2 \det\left(I_{r_A} + \frac{P}{t_A} HH^* E[\Sigma_k \Sigma_k^*]^{-1}\right) \end{aligned}$$

Using Jensen's inequality and convexity of the function $A \mapsto \log_2 \det\left(I_{r_A} + \frac{P}{t_A} HH^* A^{-1}\right)$ on the set of positive definite matrices of $\mathbb{C}^{r_A \times r_A}$, cf. [40, Lemma II.3], we obtain

$$\frac{1}{n}\sum_{k=1}^{n} I\left(\Psi_k; Y_k\right) \geq \log_2 \det\left(I_{r_A} + \frac{P}{t_A} HH^* \Delta_n^{-1}\right)$$

which concludes the proof of (2.26).

The last statement in the Proposition follows from the fact that the above inequality is proved for inputs $\Psi_1, \Psi_2, \ldots$ such that each $\Psi_k$ is circularly-symmetric Gaussian with covariance matrix $E[\Psi_k \Psi_k^*] = \frac{P}{t_A} I_{t_A}$, which implies that, for each $k \in \mathbb{N}^*$, the components of the vector $\Psi_k$ are independent from each other. ■

We make an observation and give a corollary.

**Remark 6** *Assume that $\Sigma_1, \Sigma_2, \ldots$ have the same covariance matrix $E[\Sigma_k \Sigma_k^*] = \Delta$, then, the right-hand side of (2.26) equals*

$$\log_2 \det\left(I_{r_A} + \frac{P}{t_A} HH^* \Delta^{-1}\right)$$

*Note that the above formula gives also the capacity of a MIMO channel with additive circularly-symmetric Gaussian noise process with independent samples and equi-partition of power between the transmitting antennas.*

The following Corollary of Proposition 3 states that, for a *single input and single output* (SISO) channel $t_A = r_A = 1$, the worst additive noise process distribution (not necessarily white nor Gaussian) for capacity with given second moment, is the *additive white Gaussian noise* (AWGN). This result may be seen as an extension of Gallager's result [47, Theorem 7.4.3] for memoryless channels to the channels with memory. It may also be deduced from Shannon's result [84, Theorem 18]–proved there by the *entropy power inequality*– and from the fact that the *entropy power* is not larger than the average power.

**Corollary 2** *Consider a SISO channel whose input and output, at time $n$, represented by $\Psi_n \in \mathbb{C}$, $Y_n \in \mathbb{C}$ respectively, are related by*

$$Y_n = \Psi_n + \Sigma_n, \quad n = 1, 2, \ldots$$

*where the noise process* $\Sigma_1, \Sigma_2, \ldots \in \mathbb{C}$ *is assumed stationary and satisfies* $E\left[|\Sigma_n|^2\right] = \mathcal{N}$. *Assume that the channel has a power constraint in the form* $\frac{1}{n}\sum_{k=1}^{n}|\Psi_k|^2 \leq P$. *Then the information capacity* $C$ *of the channel is lower bounded by*

$$C \geq \log_2\left(1 + \frac{P}{\mathcal{N}}\right) \tag{2.27}$$

### 2.1.3 Asymptotic analysis

Consider a particular case where the noise samples $\Sigma_1, \Sigma_2, \ldots$ are i.i.d. each being circularly-symmetric Gaussian with covariance matrix $E[\Sigma_n \Sigma_n^*] = \mathcal{N} I_{r_A}$ where $\mathcal{N}$ is a given positive constant. In this case, the right-hand side of (2.26) equals

$$\log_2 \det\left(I_{r_A} + \frac{P}{t_A \mathcal{N}} HH^*\right) = \log_2 \det\left(I_{r_A} + \frac{P}{t_A} HH^*\right) \tag{2.28}$$

where $P = \frac{P}{\mathcal{N}}$ is the signal to noise power ratio (SNR). For given $t_A$ and $r_A$ the capacity (2.28) depends on $H$.

Frequently, one is interested in the *ergodic capacity*, that is the expectation of the capacity with respect to the fading matrix $H$ assumed random with a given distribution. Assume for example that $H$ has i.i.d. components each being circularly-symmetric Gaussian with variance 1. In this case, the expectation $E\left[\log_2 \det\left(I_{r_A} + \frac{P}{t_A} HH^*\right)\right]$ may be calculated with the help of the analytical result given by Telatar [91, Theorem 2].

Alternatively, the capacity (2.28) may be approximated with the help of the following asymptotic result saying that when the number of transmitting and receiving antennas go to infinity, the capacity per receiving antenna converges to a deterministic limit.

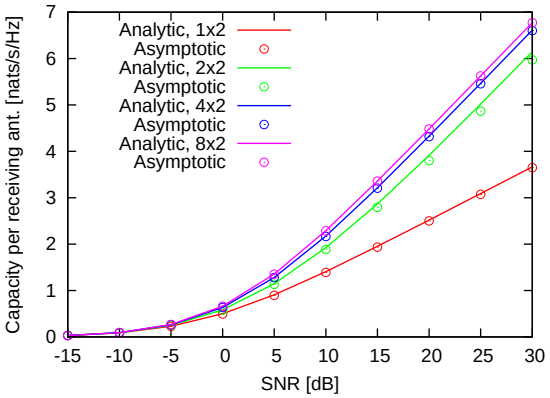


Figure 2.9: Comparison of the analytic capacity and the asymptotic formula

**Lemma 4** *[95, Equations (9), (38)], [79, Appendix] Assume that the fading matrix* $H \in \mathbb{C}^{r_A \times t_A}$ *has i.i.d. components, centred and with variance* 1. *Assume that* $t_A, r_A \to \infty$ *such that* $\frac{t_A}{r_A} \to Q \in \mathbb{R}_+$, *then*

$$\frac{1}{r_A} \log \det\left(I_{r_A} + \frac{P}{t_A} HH^*\right) \to \mathcal{C}(P, Q) \tag{2.29}$$

*almost surely, where*

$$\mathcal{C}(P,Q) := Q\log\left(1+\frac{P}{Q}-\frac{1}{4}\mathcal{F}\left(\frac{P}{Q},Q\right)\right) + \log\left(1+P-\frac{1}{4}\mathcal{F}\left(\frac{P}{Q},Q\right)\right)-\frac{Q}{4P}\mathcal{F}\left(\frac{P}{Q},Q\right) \tag{2.30}$$

*where*

$$\mathcal{F}(\xi,Q) = \left(\sqrt{\xi\left(1+\sqrt{Q}\right)^2+1}-\sqrt{\xi\left(1-\sqrt{Q}\right)^2+1}\right)^2$$

The above asymptotic result gives a good approximation of the expectation $E\left[\frac{1}{r_A}\log\det\left(I_{r_A}+\frac{P}{t_A}HH^*\right)\right]$ even for a small number of antennas as already observed in [69, Table 1] for SNR = 10dB and confirmed for SNR $\in[-15\text{dB}, 30\text{dB}]$ in Figure 2.9 where we consider 2 receiving antennas and a number of transmitting antennas $t_A \in \{1,2,4,8\}$.

# Chapter 3

# User throughput versus traffic demand — global network performance via a fixed point problem

## 3.1 Introduction

The traffic demand in wireless cellular networks is increasing rapidly and is expected to double every year. To respond efficiently to this demand the new generation of mobile cellular systems called LTE is being developed as a successor of the currently deployed 2G and 3G systems.

The deployment of such networks, frequently based on coverage conditions, should now be revised to account for this traffic increase. In particular, a densification is sometimes required. But how many sites are required to satisfy a given traffic demand with a specified quality of service target? This is a *dimensioning problem*. Another important question is to establish a relationship between the QoS perceived by the users, e.g. user throughput and the traffic demand. This would enable for example network operators to know how close their networks are to some "stability" limits. We focus on variable bit-rate (VBR) traffic; that is users requiring some volume of date to transmit at a bit-rate which may be decided by the network. In this case, the traffic demand may be expressed in bit/s/km$^2$ and the quality of service in terms of the mean throughput (in bit/s) offered to users in the long run of arrivals and departures.

The objective of this chapter is to develop an approach based on queueing theory and stochastic geometry, as well as on the previous results of Chapter 2, to tackle the dimensioning and QoS prediction problems in an efficient way.

In the present chapter we continue building the analytical approach and consider two network scenarios: symmetric (regarding the spatial pattern of base stations' positions) or regular and non-symmetric or irregular:

- We account for the dynamics of call arrivals and departures and calculate within this context the QoS perceived by users. This represents a step forward compared to the classical coverage (or static capacity) point of view.

- We continue the idea proposed in [82] of studying the dependence of load on traffic demand

by developing an analytical approach based on queueing theory.

- We show and mathematically capture the dependence between loads of base stations in the network via a *fixed-point problem.*
- We use the obtained results to show how to dimension a symmetric network.
- Exploring the results of Chapter 2 and a queueing approach analysis we apply the results of stochastic geometry and examine the user QoS in non-symmetric network e.g. Poisson, which is more realistic scenario.
- We show how the mean user throughput depends on the traffic demand in large irregular networks and after give the spatial CDF of the following QoS parameters: cell load, mean users number and mean user throughput characterizing each base station and averaged over one particular hour.
- We do an analytic and numerical study for irregular networks, the same one as mentioned in the previous point, considering heterogeneous, multi-tier networks
- We validate our whole approach by comparing our result to these obtained via 3GPP simulations [3] and real-field measurements from Orange network and present some advantages of the developed methods compared to a pure simulation approach.

Most of the time in this chapter we will be interested in the mean user throughput as a key QoS metric. Mean user throughput is a key quality-of-service metric in cellular data networks. It describes the average "speed" of data transfer during a typical data connection. It is usually defined as the ratio of the average number of bits sent (or received) per data request to the average duration of the data transfer. Since coexisting connections in a given network cell share some given cell transmission capacity, mean user throughput depends inherently on the requested data traffic. It also depends on the network architecture (positioning of the base stations) and in fact may significantly vary across different network cells. Moreover, extra-cell interference makes performance of different cells interdependent. Predicting the mean user throughput as function of the mean traffic demand locally (for each cell) and globally in the network (which involves appropriate spatial averaging in conjunction with the temporal one, already present in the classical definition of the throughput) is a key engineering task in cellular communications. We will use this metric in a two-fold manner, as an indicator of the network performance and as a dimensioning constraint.

Little's law allows to calculate the mean user throughput as the ratio of the mean traffic demand (number of bits requested per unit of time) to the mean number of users in the steady state of the network. This argument can be used to express mean user throughput locally in any region of the network. Using this argument along with some others presented in the thesis, we show in Section 3.4 how to perform a network dimensioning (planning) assuming all cells "are the same", i.e. considering a symmetric spatial pattern of base stations. If a network is symmetric, a hexagonal network for example, then all cells are of equal are if we do not consider the shadowing. On the other hand, shadowing is the same random process for all cells in the network. So, any cell is a statistical representative of the network. Consequently, considering only one cell we can capture the network performance in terms of the mean user throughput and study the dimensioning of the network. The obtained results are compared to 3GPP simulations in Section 3.7.2.

Regarding non-symmetric networks, treated in Section 3.5, we introduce the notion of a "typical cell" as a statistical representative of the network. The problem arising here consists in

adequate averaging including the spatial one and not only the temporal one. In this case, base stations form some irregular spatial point pattern.

Spatial averages of point patterns, modeling in our case the geographic locations of base stations, can be studied using the formalism of Palm distributions naturally related to the ergodic results for point processes. Within this setting one considers a typical base station with its typical cell (zone of service) whose probabilistic characteristics correspond to the aforementioned spatial averages of the characteristics for all base stations in the network. Adopting this formalism, we define the *mean user throughput in the infinite ergodic network* as the limit of the ratio of the mean volume of the data request to the mean service duration in a large, increasing to the whole plane, network window. As the main result, we prove that such defined (macroscopic) throughput characteristic is equal to the ratio of the mean traffic demand to the mean number of users in the typical cell of the network. Both these means account for double averaging: over time and network geometry.

In Section 3.7.3 we compare the results regarding the QoS in irregular networks to real-field measurements. More precisely, statistics usually collected in operational networks allow to estimate the mean traffic demand and the mean number of users for each cell and hour of the day. Even if they carry important information about the local network performance, they exhibit important variability over time (24 hours) and network cells; cf. Figure 3.8. This can be explained by the fact that mean user throughput in a particular cell does not depend only on the traffic in this cell, but also on the neighbouring cells. Moreover, the geometry of different cells in a real network may significantly differ. For these two reasons, the family of local (established for each cell) throughput-versus-traffic laws usually exhibits a lot of variability both in real data and in network simulations, and hence does not explain well the macroscopic (network- level) relation between the mean traffic demand and mean user throughput. Finding such a macroscopic relation in irregular networks is an important task for network dimensioning. It is clear that an appropriate spatial averaging analogous to the one considered in the analytic model is necessary to discover such a macroscopic law.

A key element of the analysis of the cellular network is the spatial distribution of the signal-to-interference-and-noise ratio (SINR). We show how this distribution enters into the macroscopic characterization of the throughput. When considering SINR we are able to account for the fact that the base stations which are idling, i.e., have no users to serve, do not contribute to the interference. This makes the performance of different cells interdependent and we take it into account via a system of cell-load *fixed point equations* in Section 3.5.3.

Finally, we show how to amend the model letting it account for the shadowing in the path loss. The latter is known to impact the geometry of the network, in the sense that the serving base station is not necessarily the closest one. It also alters the distribution of the SINR.

## 3.2 Related work

### 3.2.1 Related work regarding the dimensioning problem

The dimensioning of cellular networks is often treated using a *coverage* or *static capacity* approach. Basically one aims to assure that the bit-rate (or the SINR) of a permanent user exceeds some target value with a high probability. To do so, in [52] the (CDF) of the so-called effective SINR ('averaged' over the different OFDM subcarriers) is calculated with the help of a Gaussian approximation. Then this CDF is used to assure the coverage condition. A similar approach is adopted in [66] where other approximations for the CDF of the SINR are proposed. In these works arrivals and departures of users are not considered. This *dynamic context* is taken into account in the present chapter.

The dimensioning problem in this dynamic context may, in principle, be solved using a simulation approach such as that proposed in [3]. There are some other simulation tools (not necessarily compliant with 3GPP) such as LTE-Sim [77] developed by TelematicsLab, LTE simulator developed by University of Vien [72, 86] and LENA tool [12, 13] developed by CTTC.

However the simulation approach requires a huge amount of time and it is useless in the context of dimensioning. Indeed, calculating the users quality of service for a particular network configuration by 3GPP simulations takes up to a few weeks of calculation, and thus the dimensioning problem, which requires tens of such calculations, would take about a year! Analytic alternatives to pure simulations have already been proposed for VBR calls. They are essentially based on queueing theory.

The *dynamics* of user arrivals and departures are taken into account in [25], [54], [58] assuming the *symmetric network pattern and that base stations are always transmitting at their maximal power and* . In this context, the *peak bit-rate* at a given location is defined as the bit-rate obtained by a user assumed alone in the cell. The quality of service perceived by the users in the long run of their arrivals and departures is then calculated using multi-class processor sharing models [33], [26, Proposition 3.1]. The effect of mobility on the users' QoS is studied in [24, §4], [60].

In reality, the base stations emit only when they have at least one user to serve, and thus interference depends on the traffic of other base stations. In order to account for this dependence, the authors of [82] describe a fixed-point problem and propose to solve it iteratively.

### 3.2.2 Related work regarding QoS evaluation

The evaluation of user QoS metrics in cellular networks is a hard problem, but crucial for network operators and equipment manufacturers. It also motivates a lot of engineering and research studies and as dimensioning can be done using similar types of simulations as mentioned above.

A possible analytical approach to this problem is based on the information theoretic characterization of the individual link performance; cf e.g. [53, 73], in conjunction with a queueing theoretic modeling and analysis of the user traffic; cf. e.g. [24–26, 54, 62, 82]. All these works consider some particular aspects of the network and none of them considers a large, irregular multi-cell network. Such a scenario is studied in our approach by using stochastic-geometric tools combined with the two aforementioned theories. As a result, we propose a global, macroscopic approach to the evaluation of the user QoS metrics in cellular networks, which we compare and validate with respect to real network measurements. Stochastic geometry has already been shown to give analytically tractable models of cellular networks, see e.g. [8].

The disparity of cell load and QoS parameters has already been observed in the literature. For example [97] shows temporal and spatial cell load fluctuations in cellular commercial networks. These results are obtained from data collected by the mobile operators. In [76], traffic and cell load disparity are shown graphically. Data are derived from nationwide 3G cellular network and the results are presented from network and user point of view. In [48] the authors analyzed QoS (throughput etc.) perceived by the users using data collected from mobile operators and experiments. QoS parameters such as throughput, latency etc. are also analyzed based on field-measurements in [89]. Cell load and QoS parameters disparity are assumed in many studies treating load balancing. Load balancing consists in the redistribution of load between cells in such a way that all cells are equally loaded. Namely, articles as [74], [41], [96] and [103] present different algorithms for spectrum and energy efficient load balancing. The performance of heterogeneous networks gained a lot of research interest recently, for example see [31] and [50], since their deployment is already commercial and will probably continue to grow. In [70] the authors give an algorithm for network planning implying cell load disparity such that to compensate spatially

non-uniform traffic demand, but they do not give the CDF of cell load and other QoS parameters.

The authors of [62] and [87] describe the dependence between the traffic demand and the interference in wireless cellular networks and show that there is a fixed-point problem in the expressions giving the cell load.

## 3.3 Processor-sharing queue model for one cell scenario

We consider a cell comprising a finite set $\{1, 2, \ldots, J\}$ of possible *locations*. We denote by $R_j$ the *peak bit-rate* at each location $j \in \{1, 2, \ldots, J\}$ of the cell; that is the bit-rate allocated by the base station to the user at this location assuming that: (1) the user is alone in the considered cell; and (2) the other base stations transmit at their maximal powers (this assumption will be revisited later). The peak bit-rate can be e.g. the outcome of the link analysis in Chapter 2.

We describe now the *allocation of the resources* to the different users present in the cell at a given time. Let $x_j$ be the number of users at location $j$ and $x = (x_1, x_2, \ldots, x_J)$ be the vector counting the number of users at each location called *configuration of the users*. Assume that the base station allocates to each user at location $j$ a specific portion of time $\varphi_j$ depending on its location, then such user gets the bit-rate

$$r_j = \varphi_j R_j \tag{3.1}$$

Writing that the sum of the time portions may not exceed 1; i.e. $\sum_{j=1}^{J} x_j \varphi_j \leq 1$, we get the following constraint on the bit-rates which may be allocated by the base station to the different users in its cell

$$\sum_{j=1}^{J} x_j \frac{r_j}{R_j} \leq 1 \tag{3.2}$$

We shall assume that each user gets an equal portion of time $\varphi_j = 1/N$ where $N$ is the total number of users in the cell; then we deduce from Equation (3.1)

$$r_j = \frac{R_j}{N}, \quad j \in \{1, 2, \ldots, J\} \tag{3.3}$$

**Remark 7** *The constraint (3.2) (and the particular allocation (3.3)) may also be obtained by* multiplexing *the users in frequency or codes (or any mixture of time, frequency and code multiplexing). The only condition is that the users are served in a strictly orthogonal way. Moreover, the bit-rates $r_j$ should be understood as an average over a sufficiently long run of the multiplexing.*

We now introduce the *dynamics* of user (call) arrivals and departures. The inter-arrival times at location $j$ are assumed to be exponentially distributed random variables with parameter $\lambda_j$ (average inter-arrival duration equals $1/\lambda_j$). The users arriving to location $j$ require to transmit some volumes of data (in bits) which are i.i.d. random variables of mean $1/\mu_j$, not necessarily exponentially distributed. We assume independence between the inter-arrivals of users and the required volumes. We call $\rho_j := \lambda_j/\mu_j$ the *traffic demand* at location $j$ and $\rho = \sum_{j=1}^{J} \rho_j$ the total traffic demand in the cell. We denote by $\lambda = \sum_{j=1}^{J} \lambda_j$ the total arrival rate.

### 3.3.1 No mobility case

We assume in the present section that the users *do not move* during their calls. The following proposition gives the performance in the long run of the calls arrivals and departures. Denote the set of locations by $\mathbb{D} := \{1, 2, \ldots, J\}$. In order to position our problem in the queueing

theory context, we may view a cell as a *queue* and a location as a *class*. In doing so, a cell may be considered as a *multi-class processor sharing* queue. We define a critical traffic demand as follows:

$$\rho_{\mathrm{c}} := \frac{\rho}{\sum_{j=1}^{J} \rho_j R_j^{-1}} \tag{3.4}$$

**Proposition 4** *The cell is stable if and only if the traffic demand does not exceed the critical traffic demand; that is*

$$\frac{\rho}{\rho_{\mathrm{c}}} = \sum_{j=1}^{J} \rho_j R_j^{-1} < 1 \tag{3.5}$$

*In case of stability, the steady state distribution of the configuration of the users is*

$$\pi\left(x\right) = \left(1 - \frac{\rho}{\rho_{\mathrm{c}}}\right) x_{\mathbb{D}}! \prod_{j \in \mathbb{D}} \frac{\left(\rho_j / R_j\right)^{x_j}}{x_j!}, \quad x \in \mathbb{N}^{\mathbb{D}} \tag{3.6}$$

*where* $x = \left(x_j\right)_{j \in \mathbb{D}}$ *is a vector counting the numbers of users in each location and* $x_{\mathbb{D}} := \sum_{j \in \mathbb{D}} x_j$. *Moreover, the mean number of users, the delay and the throughput per user* at a given location $j \in \{1, 2, \ldots, J\}$ *are respectively given by*

$$\bar{N}_j = \frac{\rho_j}{\left(1 - \frac{\rho}{\rho_{\mathrm{c}}}\right) R_j}, \quad \bar{T}_j = \frac{1}{\left(1 - \frac{\rho}{\rho_{\mathrm{c}}}\right) R_j \mu_j}, \quad \bar{r}_j = \left(1 - \frac{\rho}{\rho_{\mathrm{c}}}\right) R_j \tag{3.7}$$

*and the mean number of users, the delay and the throughput per user* in the cell *in the steady state are respectively given by*

$$\bar{N} = \frac{\rho}{\rho_{\mathrm{c}} - \rho}, \quad \bar{T} = \frac{\rho}{\left(\rho_{\mathrm{c}} - \rho\right) \lambda}, \quad \bar{r} = \rho_{\mathrm{c}} - \rho \tag{3.8}$$

**Remark 8** *It might look cumbersome to express the stability condition (3.5) in terms of* $\frac{\rho}{\rho_{\mathrm{c}}}$. *This is particularly convenient when the traffic demand* $\rho_j$ *at location* $j$ *is parameterized in the following manner:*

$$\rho_j = \rho d_j, \quad j \in \{1, 2, \ldots, J\} \tag{3.9}$$

*where* $d_j$ *is a geographical distribution of traffic and* $\rho$ *is a parameter expressing the total traffic demand. In this case*

$$\rho_{\mathrm{c}} := \left[\sum_{j=1}^{J} d_j R_j^{-1}\right]^{-1} \tag{3.10}$$

*and the condition 3.5 can be written as:*

$$\rho < \rho_{\mathrm{c}} \tag{3.11}$$

*In the remaining part we will always assume traffic demand in the form of (3.9).*

**Proof of Proposition 4.** See the appendix 3.A for a detailed proof in the Markovian case; i.e., when the transmitted volumes are assumed exponentially distributed. In the more general case (when the transmitted volumes are arbitrary distributed) the proof is more involved. For the stability condition (3.11) and the expression (3.6) of the steady state distribution see [33], [26, Proposition 3.1].

The mean number of users, the delay and the throughput expressions may be obtained from [25] or by specializing [58, Example 10] to the current discrete context with no mobility.

We give here an outline of the proof of Equations (3.7) and (3.8). The mean number of users (either in a given location or in the cell) is obtained from the expression (3.6) of the steady state distribution. The delays are then deduced from Little's formula [11]

$$\bar{T}_j = \frac{\bar{N}_j}{\lambda_j}, \quad \bar{T} = \frac{\bar{N}}{\lambda}$$

The throughput per user at a given location $j$ is simply the average volume $1/\mu_j$ divided by the delay $\bar{T}_j$. It remains to show the expression of the throughput per user in the cell; i.e. $\bar{r} = \rho_c - \rho$. To do so, observe that the throughput of the whole cell at the steady state is equal to the traffic demand $\rho$; since at equilibrium the volumes of data incoming to and leaving the cell in the long run should be equal. The throughput per user in the cell is *defined* as the ratio of the cell throughput $\rho$ by the average number of users; that is $\bar{r} = \frac{\rho}{\bar{N}} = \rho_c - \rho$. ■

Note that (3.6) may be written as follows

$$\pi(x) = \left[(1-\rho')\,\rho'^{x_{\mathbb{D}}}\right] \left[ x_{\mathbb{D}}! \prod_{j=1}^{J} \frac{\left(\rho'_j/\rho'\right)^{x_j}}{x_j!} \right], \quad x \in \mathbb{N}^J$$

where $\rho'_j = \rho_j/R_j$ and $\rho' = \rho/\rho_c$. If follows that the distribution of the total number of users in the cell $X_{\mathbb{D}} := \sum_{j=1}^{J} X_j$ is the geometric distribution on $\mathbb{N}$ with parameter $1-\rho' = 1 - \frac{\rho}{\rho_c}$; that is $\Pr(X_{\mathbb{D}} = n) = (1-\rho')\,\rho'^n, n \in \mathbb{N}$. In particular the probability that the cell is not empty equals $\rho' = \frac{\rho}{\rho_c}$ (called *load* of the cell).

Moreover the above expression shows that, given the total number of users $n$, the distribution of the number of users among the different locations is *multinomial* of size $(n, J)$ and parameters $(\rho'_1/\rho', \ldots, \rho'_J/\rho')$; this is equivalent to say that the users are assigned to classes independently of each other, with the probability $\rho'_j/\rho'$ of a given user to be assigned to class $j$.

**Corollary 3** *With the notations of Proposition 4, if $\rho < \rho_c$ then*

$$\bar{r} = \frac{\rho}{\sum_{j=1}^{J} \rho_j \bar{r}_j^{-1}}$$

*and*

$$\bar{T} = \frac{1}{\lambda} \sum_{j=1}^{J} \lambda_j \bar{T}_j$$

*where $\lambda = \sum_{j=1}^{J} \lambda_j$ is the total arrival rate to the cell.*

**Proof.** Straightforward calculations from (3.7) and (3.8). ■

The above corollary shows that the throughput per user in the cell is the *harmonic* mean of the throughputs at the different locations weighted by the traffic demands; whereas the delay per user in the cell is the *arithmetic* mean of the delays at the different locations pondered by the arrival rates. So we should be careful when calculating the average of the quality of service over a cell.

### Mobile categories

A user located at a given geographic location undergoes some *radio conditions*; i.e., some specific propagation losses (due to distance, shadowing and indoor) with the different base stations in the network. Given these radio conditions, the user gets some bit-rate. The relation between the

radio conditions and the bit-rate may be specific to each mobile *category*. We shall use the term *class* to designate not only the geographic location but also the specific mobile's category.

Let $\mathcal{K}$ be the number of mobile categories and $I$ be the number of geographic locations, then a class $j$ is a pair $(i, \kappa)$ where $i \in \{1, 2, \ldots, I\}$ is a geographic location and $\kappa \in \{1, 2, \ldots, \mathcal{K}\}$ is the mobile category. The number of classes is now $J = I \times \mathcal{K}$. With this extended notion of class, the results of Proposition 4 obviously apply; we get in particular the expression for the throughput and delay for each class $j = (i, \kappa)$. The following proposition gives the expressions for the throughput and delay per mobile's category but averaged over the geographic locations.

**Proposition 5** *Assume the stability condition (3.11). For a given mobile's category $\kappa \in \{1, 2, \ldots, \mathcal{K}\}$, the throughput per user in the cell in the steady state is*

$$\bar{r}_\kappa = \frac{\sum_{i=1}^{I} \rho_{i,\kappa}}{\sum_{i=1}^{I} \rho_{i,\kappa} \bar{r}_{i,\kappa}^{-1}}$$

*that is the* harmonic *mean of the throughputs at the different geographic locations weighted by the corresponding traffic demands. The delay per user in the cell in the steady state is*

$$\bar{T}_\kappa = \frac{\sum_{i=1}^{I} \lambda_{i,\kappa} \bar{T}_{i,\kappa}}{\sum_{i=1}^{I} \lambda_{i,\kappa}}$$

*that is the* arithmetic *mean of the delays at the different geographic locations pondered by the corresponding arrival rates.*

**Proof.** The result is obtained by specializing [58, Example 10] to the current discrete context with no mobility. ■

### Connection between traffic demand, QoS, capacity and cell radius

Fixing a target value $\bar{r}$ for the throughput per user in the cell, we deduce from (3.8)

$$\rho_c - \rho = \bar{r}$$

Since $\rho$ and $\rho_c$ are functions of the cell radius, the above equation might be solved with respect to the cell radius. This approach will be further developed in Section 3.4 and illustrated by numerical examples in Section 3.7.2.

We assume, without loss of generality, that the peak bit-rates are sorted in decreasing order; that is $R_1 > R_2 > \ldots > R_J$. Fixing a target value $\bar{r}_J$ of the throughput per user in the cell border, we deduce from (3.7)

$$\frac{\rho}{\rho_c} = 1 - \frac{\bar{r}_J}{R_J}$$

which may be taken as the dimensioning constraint.

Given some $q \in [0, 1]$, let $j_q$ be the $q$-quantile of the traffic distribution $(d_1, d_2, \ldots, d_J)$; i.e. such that

$$\sum_{j=1}^{j_q - 1} d_j < q \leq \sum_{j=1}^{j_q} d_j$$

Then fixing a target value $\bar{r}_{j_q}$ for the throughput per user at location $j_q$, we deduce from (3.7)

$$\frac{\rho}{\rho_c} = 1 - \frac{\bar{r}_{j_q}}{R_{j_q}}$$

which may be taken also as the dimensioning constraint.

**Remark 9** *Note that $j_q$ is not the q-quantile of the proportion of users in the steady state $\left(\bar{N}_1/\bar{N}, \bar{N}_2/\bar{N}, \ldots, \bar{N}_J/\bar{N}\right)$ since from (3.7)*

$$\frac{\bar{N}_j}{\bar{N}} = \frac{\rho_{\mathrm{c}}}{\rho}\frac{\rho_j}{R_j} = d_j\frac{\rho_{\mathrm{c}}}{R_j}$$

*Thus we should not say that a proportion q* of the users *in the steady state would have a throughput larger than $\bar{r}_{j_q}$; but we should say that a proportion q* of the cell area *(weighted by the traffic demand) would have a throughput larger than $\bar{r}_{j_q}$.*

### 3.3.2 Infinite mobility

Now we will study the effect of mobility on performance from the queueing theory point of view. The case when the average user's speed is finite and nonnull is intractable analytically. But it may be bounded by the two extreme cases of no mobility and infinite mobility since mobility improves performance as proved in [24, §4.2.2]. This motivates our study of the infinite mobility case where each user is assumed to move along all the possible locations and thus experiences all the radio conditions during his call (whereas in the no mobility case, the user undergoes a given radio condition).

We assume in the present section that the mean volume of data does not depend on the location; that is $\mu_j \equiv \mu$. We assume also that each user moves according to some ergodic Markov process with invariant distribution $(\varrho_1, \varrho_2, \ldots, \varrho_J)$. Moreover we assume that each user moves so fast that he receives a peak bit-rate averaged over his mobility; that is $\sum_{j=1}^{J} \varrho_j R_j$. Then the bit-rate allocation (3.3) is now replaced by

$$r_j \equiv r := \frac{\sum_{j=1}^{J} \varrho_j R_j}{N}, \quad j \in \{1, 2, \ldots, J\} \tag{3.12}$$

**Proposition 6** *In case of infinite mobility, the cell is stable when*

$$\rho < \rho_{\mathrm{c}}$$

*where*

$$\rho_{\mathrm{c}} := \sum_{j=1}^{J} \varrho_j R_j \tag{3.13}$$

*In case of stability, the mean number of users, the delay and the throughput per user in the cell at the steady state are respectively given by*

$$\bar{N} = \frac{\rho}{\rho_{\mathrm{c}} - \rho}, \quad \bar{T} = \frac{\rho}{(\rho_{\mathrm{c}} - \rho)\,\lambda}, \quad \bar{r} = \rho_{\mathrm{c}} - \rho$$

**Proof.** See [60, Proposition 2]. ■

Note that $\rho_{\mathrm{c}}$ is the arithmetic mean of the peak bit-rates weighted by the mobility distribution $(\varrho_1, \varrho_2, \ldots, \varrho_J)$. Assume that the traffic demand $(\rho_1, \rho_2, \ldots, \rho_J)$ is proportional to the mobility distribution, then, since the arithmetic mean is larger than the harmonic mean, we deduce that the critical traffic with mobility is larger than the critical traffic in the no mobility case which is consistent with [24, §4.2.2].

**Mobile categories**

If there are different mobile categories, then it is natural to assume that mobility holds between the geographic locations $\{1, 2, \ldots, I\}$ but not between categories $\{1, 2, \ldots, \mathcal{K}\}$. We assume that the mean volume of data does not depend on the geographic location but may depend on the mobile's category; that is

$$\mu_{i,\kappa} \equiv \mu_\kappa, \quad i \in \{1, 2, \ldots, I\}, \kappa \in \{1, 2, \ldots, \mathcal{K}\}$$

Again we assume that each user of category $\kappa \in \{1, 2, \ldots, \mathcal{K}\}$ moves so fast that he receives a peak bit-rate averaged over his mobility; that is

$$R_\kappa := \sum_{l=1}^{I} \varrho_{l,\kappa} R_{l,\kappa}, \quad \kappa \in \{1, 2, \ldots, \mathcal{K}\} \tag{3.14}$$

Thus the bit-rate allocation is now

$$r_{i,\kappa} \equiv r_\kappa := \frac{\sum_{l=1}^{I} \varrho_{l,\kappa} R_{l,\kappa}}{N}, \quad i \in \{1, 2, \ldots, I\}, \kappa \in \{1, 2, \ldots, \mathcal{K}\}$$

**Proposition 7** *The cell is stable when*

$$\rho < \rho_{\mathrm{c}} := \frac{\rho}{\sum_{\kappa=1}^{\mathcal{K}} \rho_\kappa R_\kappa^{-1}}$$

*where $R_\kappa$ are given by (3.14) and*

$$\rho_\kappa := \sum_{l=1}^{I} \rho_{l,\kappa}, \quad \kappa \in \{1, 2, \ldots, \mathcal{K}\}$$

*is the cell traffic for category $\kappa$. In case of stability, the mean number of users, the delay and the throughput per user of category $\kappa \in \{1, 2, \ldots, \mathcal{K}\}$ are respectively given by*

$$\bar{N}_\kappa = \frac{\rho_\kappa}{\left(1 - \frac{\rho}{\rho_{\mathrm{c}}}\right) R_\kappa}, \quad \bar{T}_\kappa = \frac{1}{\left(1 - \frac{\rho}{\rho_{\mathrm{c}}}\right) R_\kappa \mu_\kappa}, \quad \bar{r}_\kappa = \left(1 - \frac{\rho}{\rho_{\mathrm{c}}}\right) R_\kappa$$

*and the mean number of users, the delay and the throughput per user in the cell in the steady state are respectively given by*

$$\bar{N} = \frac{\rho}{\rho_{\mathrm{c}} - \rho}, \quad \bar{T} = \frac{\rho}{(\rho_{\mathrm{c}} - \rho)\lambda}, \quad \bar{r} = \rho_{\mathrm{c}} - \rho$$

**Proof.** Observe that the present context is similar to that of Proposition 4 with the categories here in the role of the locations there and where the peak bit-rates are given now by (3.14). The desired results then follow from Proposition 4. ■

## 3.4 Multi-cell scenario: symmetric networks

An LTE cellular network is composed of base stations covering some geographic zone. Each base station transmits at some power limited to some maximal value $P$ and assigns a *specific* portion $w$ of the total system bandwidth $W$ to each user. Here, we consider that all cells have the same form.

Since fading is already averaged out at the link level, the remaining *propagation loss* comprises only the distance and shadowing. Consider a given user and let $L_v$ be his propagation loss with base station $v$. We assume that each user is served by the base station (denoted by index $u$) with the smallest loss; that is $L_u = \inf\{L_v\}$. Assume moreover that each base station transmits a constant power spectral density.

We assume in the present section that each base station transmits at its maximal power $P$ Then the received signal power equals

$$\hat{P} = \frac{w}{W}\frac{P}{L_u}$$

and the interference equal

$$I = \frac{w}{W}\sum_{v\neq u}\frac{P}{L_v}$$

Let $\mathcal{N}$ be the noise power in the system bandwidth, then the SINR per transmitting antenna[1] equals

$$\text{SINR} = \frac{\frac{\hat{P}}{t_A}}{\frac{w}{W}\mathcal{N} + \frac{\mathcal{I}}{t_A}} = \frac{1}{\frac{\mathcal{N}L_u}{t_A P} + f} \tag{3.15}$$

where

$$f := \sum_{v\neq u}\frac{L_u}{L_v} \tag{3.16}$$

is called the *interference factor*. The SINR calculated by Equation (3.15) should be plugged in Equation (2.19) to get the corresponding bit-rate.

### 3.4.1 Cell load

In this Section we will explain the dependence between cells in a given LTE network, and propose a fixed-point equation which can capture this dependance. We assumed in the previous section that the interfering base stations transmit always at their maximal power $P$. In fact a base station does not transmit when there are no users to serve. The power transmitted by base station $v$ is then $1\{X_v(t) \neq 0\}P$ where $X_v(t)$ is the number of users served by base station $v$ at time $t$.

Thus the interference at time $t$ equals

$$\mathcal{I}(t) = \frac{w}{W}\sum_{v\neq u} 1\{X_v(t) \neq 0\}\frac{P}{L_v}$$

Explicit analysis of the multi-cell model with assumption is not possible. Even the stability condition of the network in this case is not yet known. Nevertheless, the full activity assumption made in the previous section gives a useful lower bound for the peak bit-rates and thus a lower bound of the critical traffic demand. We shall make now another model assumption.

[1] See [94, Equation (3.169)].

Invoking the law of large numbers, we may approximate the interference as follows

$$\begin{aligned}\mathcal{I}(t) &\simeq \frac{w}{W}\sum_{v\neq u} E\left[1\left\{X_v(t)\neq 0\right\}\right]\frac{P}{L_v}\\ &= \frac{w}{W}\sum_{v\neq u} \Pr\left(X_v(t)\neq 0\right)\frac{P}{L_v}\\ &= \frac{w}{W}\sum_{v\neq u} \frac{\rho}{\rho_{\mathrm{c}}}\frac{P}{L_v}\end{aligned}$$

where for the third equality we use the observation following Proposition 4. Then the SINR equals

$$\mathrm{SINR} = \frac{1}{\frac{\mathcal{N}L_0}{t_A P} + \frac{\rho}{\rho_{\mathrm{c}}} f}$$

and the corresponding peak bit-rate equals

$$R = W\psi(\mathrm{SINR}) = W\psi\left(\frac{1}{\frac{\mathcal{N}L_0}{t_A P} + \frac{\rho}{\rho_{\mathrm{c}}} f}\right) \tag{3.17}$$

where the function $\psi$ is given for example by (2.17).

Equations (3.4) and (3.13) show that the critical traffic $\rho_{\mathrm{c}}$ is a function of the peak bit-rates which are themselves functions of the critical traffic as shown in the above equation. Thus $\rho_{\mathrm{c}}$ may be obtained by solving a fixed-point problem. For example, in the case of infinite mobility Equation (3.13) implies

$$\rho_{\mathrm{c}} = E[R] = E\left[W\psi\left(\frac{1}{\frac{\mathcal{N}L_0}{t_A P} + \frac{\rho}{\rho_{\mathrm{c}}} f}\right)\right] \tag{3.18}$$

where the expectation is with respect to a user distributed according to the mobility invariant distribution $\varrho$. Once the above fixed-point problem is solved, the ratio

$$\theta := \frac{\rho}{\rho_{\mathrm{c}}} \tag{3.19}$$

is called the *load* of the system. The equation given by (3.18) we call *fixed-point equation*.

**Definition 1** *The following different load situations are considered in conjunction with the* queueing approach*:*

- Adapted (Weighted) interference (load)*: A base station transmits only when it has at least one user to serve.*
- Full interference (load)*: Base stations are always transmitting at their maximal power.*
- Null interference (load)*: Interference is assumed completely cancelled. This corresponds to a cell in isolation.*

**Remark 10** *Note that the load depends on the traffic demand, so we can not consider these two parameters as independent inputs when evaluating the users QoS.*

**Remark 11** *The above queueing analysis is carried out for a typical cell of a network composed of multiple cells assumed statistically equivalent. Indeed, the interference between the different cells is taken into account through the interference factor (3.16) and the solution of the fixed-point problem (3.18).*

## 3.5 Multi-cell irregular networks scenario

In contrast to Section 3.4, here we consider irregular cellular networks, which corresponds better to the commercially deployed networks, especially in urban areas. Irregular means here that a spatial pattern of base stations' locations is non-symmetric and that different base stations radiate different powers (in the first part of the analysis we will consider constant radiation power).

### 3.5.1 Network model

We consider locations $\{v_1, v_2, \ldots\}$ of base stations (BS) on the plane $\mathbb{R}^2$ as a realisation of a point process, which we denote by $\Phi$. [2] We assume that $\Phi$ is stationary and ergodic with positive, finite intensity (mean number of BS per unit of area) $\gamma$. [3]

In order to simplify the presentation, we shall make first the following two assumptions, which will be relaxed in Sections 3.5.3 and 3.5.3, respectively.

1. *There is no shadowing.* The (time-averaged over fading) propagation loss depends only on the distance $r$ between the transmitter and the receiver through a path-loss function $l(r)$, which we assume increasing. The SINR expression remains valid except that instead of propagation loss, we consider only the path-loss.

2. *Full load (interference).* Each base station is always transmitting at some fixed power $P$, common for all stations.

We will also assume throughout the whole Section that *each user is served by the BS which he or she receives with the strongest signal power.* The consequence of the assumption 1 above is that each BS $u \in \Phi$ serves users in a geographic zone $V(u) = \{y \in \mathbb{R}^2 : |y-u| \leq \min_{v\in\Phi, v\neq u} |y-v|\}$ which is called *Voronoi cell* of $u$ in $\Phi$.

For single link performance, we use formula (2.17) from Section 2.5. We assume the same service policy and traffic demand as in Section 3.3. We assume spatially uniform traffic demand, which means that different base stations have different traffic demands to serve. The traffic demand in a given cell equals

$$\rho(v) = \rho |V(v)|, \quad v \in \Phi. \tag{3.20}$$

### 3.5.2 Generalization of processor sharing model

In this section, we generalize the results already presented in Section 3.3. Namely, there we consider the case when the cell is stable, that is (3.11) is always valid. Here, we consider also the possibility that there are unstable cells present in a network. Another difference is that evidently here different base stations serve zones (cells) of different size even without considering the shadowing. For a fixed configuration of BS $\Phi$, the service of users arriving to the cell $V(v)$ of a given BS $v \in \Phi$ can be modeled by an appropriate (spatial) multi-class processor sharing queue, with classes corresponding to different peak bit-rates characterized by user locations $y \in V(v)$. Note also that a consequence of our model assumptions (in particular the full interference assumption 2, inter-cell channel independence and space-time Poisson arrivals) the service processes of different queues are independent.

[2] According to the formalism of the theory of point processes (cf e.g. [37]), a point process is a random measure $\Phi = \sum_j \delta_{X_j}$, where $\delta_x$ denotes the Dirac measure at $x$.

[3] Stationarity means that the distribution of the process is translation invariant, while ergodicity allows to interpret some mathematical expectations as spatial averages of some network characteristics.

We consider now the steady-state number of users served in each cell $V(v)$. [4] The following expressions follow from the queueing-theoretic analysis of the processor sharing systems of each BS $v \in \Phi$, explained in Section 3.3.1, also, cf [25,62] for the details.

- The service process of BS $v \in \Phi$ is stable if and only if its traffic demand does not exceed the critical value that is the harmonic mean of the peak bit-rate over the cell:

$$\rho_{\mathrm{c}}(v) := \frac{|V(v)|}{\int_{V(v)} 1/R(\mathrm{SINR}(y,\Phi))\, dy}\,. \tag{3.21}$$

  Note that $\rho_c(v)$ depends on $v$ and on $\Phi$. The same observation is valid for the subsequent cell characteristics. This definition comes from (3.10).

- The mean user throughput in the given cell, defined as the ratio of the mean volume of the data request $1/\mu$ to the average service time of users *in this cell*, can be expressed as follows

$$r(v) = \max(\rho_{\mathrm{c}}(v) - \rho(v), 0)\,. \tag{3.22}$$

- The mean number of users in steady state of the given cell equals to

$$N(v) = \frac{\rho(v)}{r(v)}\,. \tag{3.23}$$

  Note that $N(v) = \infty$ if $\rho(v) \geq \rho_c(v)$.

- The probability that the given BS is not idling in steady state (has at least one user to serve) equals

$$p(v) = \min(\theta(v), 1)\ , \tag{3.24}$$

  where $\theta(v)$, which we call *cell load*, is defined as

$$\theta(v) := \frac{\rho(v)}{\rho_{\mathrm{c}}(v)}\,. \tag{3.25}$$

Note that the cell is stable if and only if $\theta(v) < 1$ and

$$\theta(v) = \rho \int_{V(v)} 1/R(\mathrm{SINR}(y,\Phi))\, dy\,. \tag{3.26}$$

Moreover,

$$N(v) = \frac{\theta(v)}{1-\theta(v)}\,, \tag{3.27}$$

$$r(v) = \rho(v)(1/\theta(v) - 1) \tag{3.28}$$

provided $\theta(v) < 1$. The function $R$ is the same as in (3.17).

The above expressions allow to express all other characteristics in terms of the traffic demand per cell $\rho(v)$ and the cell load $\theta(v)$.

**Remark 12** *All the above characteristics are* local *network characteristics in the sense that they characterize the service at each BS $v$ and vary over $v \in \Phi$. Real data analysis and simulations for Poisson network models exhibit a lot of variability among these characteristics. In particular, plotting the mean user throughput $r(v)$ as function of the mean traffic demand $\rho(v)$ for different $v \in \Phi$ does not reveal any apparent systematic relation between these two local characteristics; cf. Figure 3.8.*

[4] Note that the (mean) QoS characteristics of users in this state correspond to time-averages of user characteristics.

### 3.5.3 Global network characteristics

In this section we propose some global characteristics of the network allowing to characterize its macroscopic performance. We are particularly interested in finding such a relation between the (per area) traffic demand $\rho$ and the (global) mean user throughput in the network, with this latter characteristic yet to be properly defined.

#### Typical cell of the network

A first, natural idea in this regard is to consider spatial averages of the local characteristics in an increasing network window $A$, say a ball centered at the origin and the radius increasing to infinity. Assuming ergodicity of the point process $\Phi$ of the BS, these averages can be expressed and calculated as Palm-expectations of the respective characteristics of the so called "typical cell" $V(0)$. For example

$$\lim_{|A|\to\infty} 1/\Phi(A) \sum_{v\in A} \rho(v) = \mathbf{E}^0[\rho(0)] = \rho \mathbf{E}^0[|V(0)|]\,. \tag{3.29}$$

The typical cell $V(0)$ is the cell of the BS located at the origin $v = 0$ and being part of the network $\Phi$ distributed according to the *Palm distribution* $\Pr^0$ associated to the original stationary distribution Pr of $\Phi$. In the case of Poisson process, the relation between the Palm and stationary distribution is particularly simple and (according to Slivnyak's theorem) consists just in adding the point $v = 0$ to the stationary pattern $\Phi$.

The convergence analogue to (3.29) holds for each of the previously considered local characteristics $\mathbf{E}^0[\rho_c(0)]$, $\mathbf{E}^0[r(0)]$, $\mathbf{E}^0[N(0)]$, $\mathbf{E}^0[p(0)]$ and $\mathbf{E}^0[\theta(0)]$. The convergence is Pr almost sure and follows from the ergodic theorem for point processes (see [9, Theorem 4.2.1], [37, Theorem 13.4.III]). However, as we will explain in what follows, *not all* of these *mean-typical cell* characteristics have natural interpretations as macroscopic network characteristics.

First, note that the existence of some (even arbitrarily small) fraction of BS $v$ which are not stable (with $\rho(v) \geq \rho_c(v)$, hence $N(v) = \infty$) makes $\mathbf{E}^0[N(0)] = \infty$.

**Remark 13** *For a well dimensioned network one does not expect unstable cells. For a perfectly hexagonal network model* $\Phi$ all *cells are stable or unstable depending on the value of the per-area traffic demand* $\rho$*. An artifact of an infinite, homogeneous, Poisson model* $\Phi$ *is that for arbitrarily small* $\rho$ *there exists a non-zero fraction of BS* $v \in \Phi$*, which are non-stable. This fraction is very small for reasonable* $\rho$*, allowing to use Poisson model to study QoS metrics which, unlike* $\mathbf{E}^0[N(0)]$*, are not "sensitive" to this artifact.*

We will also show in the next section that it is *not* natural to interpret $\mathbf{E}^0[r(0)]$ (which is *not* sensitive to the existence of a small fraction of unstable cells) as the mean user throughput in the network; see Remark 16. Before we give an alternative definition of this latter QoS, let us state the following result, which will be useful in what follows.

**Proposition 8** *We have*

$$\mathbf{E}^0[\rho(0)] = \frac{\rho}{\gamma}\,, \tag{3.30}$$

$$\mathbf{E}^0[\theta(0)] = \frac{\rho}{\gamma}\mathbf{E}[1/R\left(\mathrm{SINR}\left(0,\Phi\right)\right)]\,. \tag{3.31}$$

**Proof.** The first equation is quite intuitive: the average cell area is equal to the inverse of the average number of BS per unit of area. Formally, both equations follow from the inverse formula of Palm calculus [9, Theorem 4.2.1]. In particular, for (3.31) one uses representation (3.26) in conjunction with the inverse formula. ■

**Remark 14** *Note that the expectation in the right-hand-side of (3.31) is taken with respect to the* stationary *distribution of the BS process $\Phi$. It corresponds to the spatial average of the inverse of the peak bit-rate calculated throughout the network. The (only) random variable in this expression is the SINR experienced by the* typical user. *This distribution is usually known in operational networks (estimated from user measurements). It can also be well approximated using a Poisson network model for which its distribution function admits an explicit expression [19].*

### Mean user throughput in the network

Faithful to the usual definition of the mean user throughput as the ratio of the mean volume of the data request to the mean service duration (which we retained at the local, cell level) we aim to define now the mean user throughput in the (whole) network as the ratio of these two quantities taken for increasing network window $A$. However in order to "filter out" the impact of cells which are not stable and avoid undesired degeneration of this characteristic (e.g. for Poisson process; cf. Remark 13) let us consider the union of all stable cells

$$\mathcal{S} := \bigcup_{v\in\Phi:\rho(v)<\rho_c(v)} V(v)\,.$$

Note that the stationarity of $\Phi$ implies the same for the random set $\mathcal{S}$. We denote by $\pi_{\mathcal{S}} = \mathbf{E}[\mathbf{1}(0 \in \mathcal{S})]$ the volume fraction of $\mathcal{S}$ and call it the *stable fraction of the network*. It is equal to the average fraction of the plane covered by the stable cells; cf. [9, Definition 3.4 and the subsequent Remark]. Denote also

$$N^0 := \mathbf{E}^0[N(0)\mathbf{1}(N(0) < \infty)]\,.$$

We are ready now to define the *mean user throughput in the network* $r^0$ as the ratio of the average number of bits per data request to the average duration of the data transfer in the stable part of the network

$$r^0 := \lim_{|A|\to\infty} \frac{1/\mu}{\text{(temporal-)mean service time in } A\cap\mathcal{S}}\,. \tag{3.32}$$

Here is the key result of the typical cell approach. Its proof is given at the end of this section.

**Theorem 1** *For an ergodic network $\Phi$ we have*

$$r^0 = \frac{\rho\,\pi_{\mathcal{S}}}{\gamma N^0}\,. \tag{3.33}$$

**Remark 15** *Equation (3.33) provides a macroscopic relation between the traffic demand and the mean user throughput in the network, which we are primarily looking for. It will be validated by comparison to real data measurements. The quantities $N^0$ and $\pi_{\mathcal{S}}$ do not have explicit analytic expressions analogous to (3.31). Nevertheless they can be estimated from simulations of a given network model $\Phi$. Note that these are static simulations of the network model. No simulation of the traffic demand process is necessary, which greatly simplifies the task. For small and moderate values of the traffic demand (observed in real networks) one obtains $\pi_{\mathcal{S}} \simeq 1$. Moreover, in Section 3.5.4 we will propose some more explicit approximation of $N^0$.*

**Remark 16** *Assume that there are no unstable cells in the network. This is the case e.g. for lattice (say hexagonal) network models with traffic demand $\rho < \rho_c(v) = \rho_c$, where the value of the critical traffic is the same for all cells. Then $\pi_{\mathcal{S}} = 1$, $N^0 = \mathbf{E}^0[N(0)]$ and the relation (3.33) takes form*

$$r^0 = \frac{\rho}{\gamma\mathbf{E}^0[N(0)]} = \frac{\mathbf{E}^0[\rho(0)]}{\mathbf{E}^0[N(0)]}\,. \tag{3.34}$$

*Thus, in general* $r^0 \neq \mathbf{E}^0[r(0)] = \mathbf{E}^0[\rho(0)/N(0)]$. *We want to emphasize that this is not merely a theoretical detail resulting from our (and common) definition of the mean throughput (3.32). The expression* $\mathbf{E}^0[r(0)] = \mathbf{E}^0[\rho(0)/N(0)]$, *which in principle can be considered as another global QoS metric, is in practice difficult to estimate. Indeed, when estimating* $\mathbf{E}^0[r(0)]$ *as the average of the ratio "traffic demand to the number of users" from real data measurements, one needs to give a special treatment to observations which correspond to cells during their idling hours (i.e., with no user, and such observations are not rare in operational networks). Neither skipping nor literal acceptance of these observations captures the right dependence of the mean user throughput on the traffic demand.*

**Proof of Theorem 1.** By Little's law [11] the temporal mean service time $T^{\mathcal{W}}$ of users in any region of the network $\mathcal{W}$, say the union of stable cells with BS in some region $A$, $\mathcal{W} = \bigcup_{v\in A\cap\mathcal{S}} V(v)$, is related to the mean number $N^{\mathcal{W}}$ of the users served in this region $\mathcal{W}$ in the steady state by the equation $N^{\mathcal{W}} = \lambda|\mathcal{W}|T^{\mathcal{W}}$. Consequently, the mean user throughput in this region $\mathcal{W}$ can be expressed as $1/(\mu T^{\mathcal{W}}) = \rho|\mathcal{W}|/N^{\mathcal{W}}$. Using

$$\begin{aligned}\frac{|\mathcal{W}|}{N^{\mathcal{W}}} &= \frac{|\mathcal{W}|}{\sum_{v\in A\cap\mathcal{S}} N(v)}\\ &= \frac{\sum_{v\in A}|V(v)|\mathbf{1}(N(v)<\infty)}{|A|}\frac{|A|}{\sum_{v\in A\cap\mathcal{S}} N(v)}\end{aligned}$$

and again the ergodic theorem for point yprocess $\Phi$, we obtain that the limit in (3.32) is Pr-almost surely equal to $\rho\mathbf{E}^0[|V(0)|\mathbf{1}(N(0)<\infty)]/\mathbf{E}^0[N(0)\mathbf{1}(N(0)<\infty)]$. By the aforementioned inverse formula of Palm calculus we conclude $\mathbf{E}^0[|V(0)|\mathbf{1}(N(0)<\infty)] = \mathbf{E}[\mathbf{1}(0\in\mathcal{S})]/\gamma$. ■

### Cell-load equations

We have to revoke now the full interference assumption 2 made in Section 3.5.1. An amendment is necessary in this matter for the model to be able to predict the real network data; cf numerical examples in Section 3.7. Recall that the consequence of this assumption is that in the expression (3.15) of the SINR all the interfering BS are always transmitting at a given power $P$. In real networks BS transmit only when they serve at least one user. [5] Taking this fact into account in an exact way requires introducing in the denominator of (3.15) the indicators that a given station $v \in \Phi$ at a given time is not idling. This, in consequence, would lead to the probabilistic dependence of the service process at different cells and result in a non-tractable model. In particular, we are not aware of any result regarding the stability of such a family of dependent queues. For this reason, we take into account whether $v$ is idling or not in a simpler way, multiplying its powers $P$ by the *probability* $p(v)$ that it is not idle in the steady state. In other words we modify the expression of the SINR as follows (for base station $u$ as a serving base station)

$$\mathrm{SINR}\,(y,\Phi) := \frac{P/l\,(|y-u|)}{\mathcal{N} + P\sum\limits_{v\in\Phi\setminus\{u\}} p\,(v)\,/l\,(|y-v|)}, \tag{3.35}$$

for $y \in V(v)$, $v \in \Phi$ where $p(v)$ are cell non-idling probabilities given by (3.24). We will see in Section 3.7 that this model, called *weighted interference (load) model*, fits better to real field measurements than the full interference model. The above modification of the model preserves the independence of the processor-sharing queues at different cells given the realization $\Phi$ of the network (thus allowing for the explicit analysis of Section 3.5.2). However the cell loads $\theta(v)$ are

[5] Analysis of more sophisticated power control schemes is beyond the scope of this thesis.

no longer functions of the traffic demand and the SINR experienced in the respective cells, but are related to each other by the following equations that replace (3.26)

$$\theta(v) = \rho \int_{V(v)} \frac{1}{R\left(\frac{P/l(|y-u|)}{\mathcal{N}+P\sum\limits_{v\in\Phi\setminus\{u\}}\min(\theta(v),1)/l(|y-v|)}\right)}\, dy\,. \tag{3.36}$$

We call this system of equations in the unknown cell loads $\{\theta(v)\}_{v\in\Phi}$ the *cell-load equations.*

**Remark 17 (Spatial stability)** *The weighted interference model introduces more "spatial" dependence between the processor sharing queues of different cells, while preserving their "temporal" (conditionally, given $\Phi$) independence. A natural question regarding the existence and uniqueness of the solution of the fixed point problem (3.36) arises. Note that the mapping in the right-hand-side of (3.36) is increasing in all $\theta(v)$, $v \in \Phi$ provided function $R$ is increasing. Using this property it is easy to see that successive iterations of this mapping started off $\theta(v) \equiv 0$ on one hand side and off $\theta(v)$ as in (3.26) (full interference model) on the other side, converge to a minimal and maximal solution of (3.36), respectively. An interesting theoretical question regards the uniqueness of the solution of (3.36), in particular for a random, say Poisson, point process $\Phi$. Answering this question, which we call "spatial stability" of the model, is unfortunately beyond the scope of this thesis.*[6] *The simulation study of the typical cell model, presented in Section 3.7 (where we use Matlab to find a solution of (3.36) for any given finite pattern of base stations $\Phi$) is less stable for larger values of the traffic demand $\rho$.*

In the mean cell approach (cf Section 3.5.4) we take into account the weighted interference model by the following (single) equation in the mean-cell load $\bar{\theta}$

$$\bar{\theta} = \frac{\rho}{\gamma}\mathbf{E}\left[1/R\left(\frac{P/l(|v^*|)}{\mathcal{N}+P\sum_{v\in\Phi\setminus\{v^*\}}\bar{\theta}/l(|y-v|)}\right)\right] \tag{3.37}$$

where $v^*$ is the location of the BS whose cell covers the origin. We solve the above equation with $\bar{\theta}$ as unknown. We will show in the numerical section that the solution of this equation gives a good estimate of the empirical average of the loads $\{\theta(v)\}_{v\in\Phi}$ obtained by solving the system of cell-load equations (3.36) for the simulated model.

**Remark 18 (Pilot channel)** *The cells which are not idle might still emit some power (e.g. in the pilot channel). This can be taken into account by replacing $p(v) = \min(\theta(v), 1)$ in (3.36) by $p(v)(1-\epsilon)+\epsilon$, where $\epsilon$ is the fraction of the power emitted all the time. Similar modification concerns $\bar{\theta}$ in the right-hand-side of (3.37).*

### Shadowing

Until now we were assuming that the propagation loss is only induced by the distance between the transmitter and the receiver. In this section we will briefly explain how the effect of shadowing can be taken into account.

Assume that the shadowing between a given station $v \in \Phi$ and all locations $y \in \mathbb{R}^2$ is modeled by some random field $S_v(y-v)$. That is, we assume the propagation loss between $v$ and $y$ $L_v(y) = \frac{l(|y-v|)}{S_v(y-v)}$. We assume that, given $\Phi$, the random fields $S_v(\cdot)$ are independent

[6] Existence and uniqueness of the solution of a very similar problem (with finite number of stations and a discrete traffic demand) is proved in [87].

across $v \in \Phi$ and identically distributed. In general, we do not need to assume any particular distribution for $S_v(\cdot)$ (neither independence nor the same distribution of $S_v(y)$ across $y$).

The assumption that each user is served by the BS received with the smallest path-loss results in the following modification of the geographic service zone of $u$, which we keep calling "cell"

$$V(u) = \{y \in \mathbb{R}^2 : L_u(y) \leq \min_{v \in \Phi, v \neq u} L_v(y)\}. \tag{3.38}$$

For mathematical consistency we shall assume that, almost surely, the origin belongs to a unique cell (i.e., is not located on any cell boundary).

The SINR at location $y$ can be expressed by (3.15) or (3.35) with $l(|y-v|)$ replaced by $L_v(y)$ for $v \in \Phi$, depending on whether we consider the full interference or the weighted interference model, for $y \in V(v)$, with $V(v)$ defined by (3.38). The same modification regards the cell-load equations (3.36) and (3.37).

All the previous results involving the typical cell remain valid for this modification of the model. In particular, the results of Proposition 8 can be extended to the model with shadowing (where the cell associated to each base station is not necessarily the Voronoi cell) provided the origin 0 belongs to a unique cell almost surely.

Note that the mean cell surface $\mathbf{E}^0[|V(0)|] = 1/\gamma$, and hence the mean traffic demand per cell $\mathbf{E}^0[\rho(0)] = \rho/\gamma$, do not depend at all on the shadowing. The values of other characteristics of the typical and the mean cell will change depending on the distribution of the random shadowing field $S_v(y)$ (both the marginal distributions and the correlation across $y$). An interesting remark in this regard is as follows.

In the full interference model, the mean load of the typical cell and the load of the mean cell (which are by the definition equal $\mathbf{E}^0[\theta(0)] = \bar{\theta}$) depend only on the stationary marginal distribution of $\mathrm{SINR}(0, \Phi)$, cf. (3.40). Hence, it does not depend on the (spatial) correlation of $S_v(y)$ across $y$. Moreover, this distribution is known in the case of the Poisson network and identically distributed marginal shadowing $S_v(y) \sim S$. As explained in [19], in this case $\mathrm{SINR}(0, \Phi)$ in the model with shadowing has the same distribution as in the model without shadowing and the density of stations equal to $\gamma \times \mathbf{E}[S^{2/\beta}]$. In particular, a specific distribution of $S$ and the correlation of $S_v(y)$ across $y$ play no role. This equivalence of the two models (with and without shadowing) is more general, as explained in [19] and [21], and applies also to the mean-cell-load equation (3.37).

**Remark 19** *The impact of the shadowing on the* mean cell model, both in full and weighted interference scenario *in the above Poisson model with $S_v(y) \sim S$ can be summarized as follows. It modifies only the cell load $\bar{\theta}$ and not the traffic demand $\bar{\rho}$. Moreover, multiplying $\gamma$ by $\mathbf{E}[S^{2/\beta}]$ and dividing $\rho$ by the same moment one obtains an equivalent (in terms of all considered characteristics) mean cell without shadowing.*

### 3.5.4 Mean cell

It is tempting to look for a synthetic model which would allow to relate main parameters and QoS metrics of a large irregular cellular network in a simple, yet not simplistic way. The typical cell approach described up to now offers such possibility. In this section we will go a little bit further and propose an even simpler model mimicking that of Section 3.4.1. It consists in considering a *virtual* cell, to which we will assign the parameters and QoS metrics inspired by the analysis of the typical cell. In contrast to the typical cell, our virtual cell is not random and this is why we call it the *mean cell*. Specifically, we define it as a (virtual) cell having the same traffic demand $\bar{\rho}$ and load $\bar{\theta}$ as the typical cell. Note that these two characteristics admit explicit expressions;

cf. Proposition 8 and Remark 14.

$$\bar{\rho} := \mathbf{E}^0 \left[\rho\left(0\right)\right] = \frac{\rho}{\gamma}\,, \tag{3.39}$$

$$\bar{\theta} := \mathbf{E}^0 \left[\theta\left(0\right)\right] = \frac{\rho}{\gamma}\mathbf{E}[1/R\left(\mathrm{SINR}\left(0,\Phi\right)\right)]\,. \tag{3.40}$$

For the remaining characteristics, we assume that they are related to the above two via the relations presented in Section 3.5.2. Specifically, following (3.20) we define the area of the mean cell by $\bar{V} = \bar{\rho}/\rho$ and in analogy to (3.25) we define the critical load of the mean cell as

$$\bar{\rho}_{\mathrm{c}} := \frac{\bar{\rho}}{\bar{\theta}}\,. \tag{3.41}$$

We say that the mean cell is stable if $\bar{\rho} < \bar{\rho}_c$. Inspired by (3.22) we define the user's throughput in the mean cell by

$$\bar{r} := \max\left(\bar{\rho}_{\mathrm{c}} - \bar{\rho}, 0\right)$$

and, as in (3.23), the mean number of users in the mean cell is defined as

$$\bar{N} := \frac{\bar{\rho}}{\bar{r}}\,.$$

We observe the following immediate relations.

**Corollary 4** *The mean cell is stable if and only if $\bar{\theta} < 1$. In this case*

$$\bar{N} = \frac{\bar{\theta}}{1-\bar{\theta}}\,, \tag{3.42}$$

$$\bar{r} = \bar{\rho}(1/\bar{\theta} - 1)\,, \tag{3.43}$$

*which are analogous to (3.27) and (3.28), respectively.*

**Remark 20** *The equation (3.43) provides an alternative macroscopic relation between the traffic demand and the mean user throughput in the network. It is purely analytic; no simulations are required provided one knows the distribution of the SINR of the typical user in (3.40). It will be validated by comparison to real data measurements. We consider it as an approximation of (3.33). It consists in assuming $\bar{N} \simeq N^0/\pi_{\mathcal{S}}$. This latter hypothesis will be also separately validated numerically.*

**Remark 21** *Note that the key characteristic of the mean cell is its load $\bar{\theta}$. In analogy to the load factor of the (classical) M/G/1 processor sharing queue, it characterizes the stability condition, mean number of users and the mean user throughput.*

## 3.6 Heterogeneous networks

The objective of the present section is to extend our model to heterogenous wireless cellular networks comprising different categories of base stations transmitting at distinct powers. This model permits to calculate by static simulation (Monte carlo estimation of some functionals of Poisson point process) the quality of service perceived by the users served by each category of base stations. Analytical approximations are also proposed.

Specifically, we shall build a model for cellular networks comprising macro and micro BS emitting different powers. The user is served by the BS offering the strongest received power

among all the BS in the network. In this context, we calculate the quality of service (QoS) offered to the users by each category of BS (macro and micro). Extending the results from Section 3.5 we develop here also the corresponding *typical cell* and *mean cell* approach globally and for each base station category. Further, we analyse the QoS parameters: mean user throughput, mean number of users and cell load per base station category.

### 3.6.1 Model description

We consider a cellular network comprising different categories of BS characterized by different transmitting powers. In this section we restrict ourselves to a Poisson model. Let $O$ be the finite set of possible categories of base stations. The locations of BS of category $o \in O$ are modeled by a homogeneous Poisson point process of intensity parameter $\gamma_o \in \mathbb{R}_+$ and denoted by $\Phi_o$. The transmitting power of BS of category $o$ is denoted by $P_o \in \mathbb{R}_+$. Let $\Phi$ be the superposition of $\{\Phi_o\}_{o\in O}$. Then $\Phi = \{v_n\}_{n\in\mathbb{N}}$ is a homogeneous Poisson point process of intensity parameter $\gamma = \sum_{o\in O} \gamma_o$.

Inversely, starting from the point process $\Phi$ we may retrieve the processes with the same distribution as $\{\Phi_o\}_{o\in O}$ in the following manner. Let $Z_n$ be i.i.d. marks of $\Phi$ such that

$$\mathbb{P}(Z_n = o) = \frac{\gamma_o}{\gamma}, \quad o \in O, n \in \mathbb{N}^*$$

Then

$$\Phi_o = \sum_{n\in\mathbb{N}^*} \delta_{v_n} \mathbf{1}\{Z_n = o\}, \quad o \in O$$

Much as in Section 3.5 the propagation loss due to distance is a power function $l(x) = (K|x|)^\beta$ where $K > 0$ and $\beta > 2$ are given constants. The shadowing between a given station $v_n \in \Phi$ and all locations $y \in \mathbb{R}^2$ is modeled by some stochastic process $\mathbf{S}_n(y - v_n)$ taking values in $\mathbb{R}_+$. The shadowing stochastic processes $\mathbf{S}_n(\cdot)$ are i.i.d. marks of $\Phi$. We assume that $\mathbf{S}_1(y)$ are identically distributed across $y$. (We do not make any assumption concerning the dependence of $\mathbf{S}_1(y)$ across $y$.) Thus the received power at location $y \in \mathbb{R}^2$ from base station $v_n \in \Phi$ equals

$$\frac{P_{Z_n}\mathbf{S}_n(y - v_n)}{l(|y - v_n|)}$$

The shadowing fields $\mathbf{S}_n(\cdot)$ and the categories of BS $Z_n$ are assumed independent.

In order to simplify the notation, let $\{Z(x)\}_{x\in\mathbb{R}^2}$ be a stochastic process such that $Z(v_n) = Z_n$. In a similar way, let $\{\mathbf{S}_x(\cdot)\}_{x\in\mathbb{R}^2}$ be a stochastic process such that $\mathbf{S}_{v_n}(\cdot) = \mathbf{S}_n(\cdot)$.

Let $L_{v_n}(y)$ be the inverse of the received power at location $y$ from BS $v_n$; that is

$$L_{v_n}(y) = \frac{l(|y - v_n|)}{P_{Z(v_n)}\mathbf{S}_{v_n}(y - v_n)}, \quad n \in \mathbb{N}^* \tag{3.44}$$

Each user is served by the BS offering the strongest received power among all the BS in the network. Then the cell served by BS $u \in \Phi$ is

$$V(u) = \left\{y \in \mathbb{R}^2 : L_u(y) \le L_v(y), \forall v \in \Phi\right\} \tag{3.45}$$

The *signal to interference and noise ratio* (SINR) at location $y$ in the downlink (BS to user) equals

$$\mathrm{SINR}(y, \Phi) = \frac{\frac{1}{L_u(y)}}{\mathcal{N} + \sum_{v\in\Phi\setminus\{u\}} \frac{1}{L_v(y)}}, \quad y \in V(u), u \in \Phi \tag{3.46}$$

where $\mathcal{N}$ is the noise power. Note that the above expression of SINR relies on the full interference model from Definition 1 which will be improved in Section 3.6.4.

We assume that the peak bit-rate at location $y$, defined as the bit-rate of a user located at $y$ when served alone by its BS, is some function $R(\mathrm{SINR})$ given for example in (3.17). Local cell characteristics and traffic dynamics are assumed to be the same as in Section 3.5.2.

### 3.6.2 Typical and mean cell in multi-tier network

We aim to study the distribution of the received powers for a user located at the origin. The inverse of the power received from BS $v_n \in \Phi$ may be deduced from (3.44)

$$L_n := L_{v_n}(0) = \frac{l(|v_n|)}{S_n P_{Z_n}}$$

where $S_n := \mathbf{S}_n(0 - v_n)$ are the shadowing random variables. Let $\hat{\Phi} = \{L_n\}_{n\in\mathbb{N}^*}$ which is a point process on $\mathbb{R}_+$ and $S = S_1$.

**Lemma 5** *Assume that* $\mathbb{E}\left[S^{2/\beta}\right] < \infty$. *Then* $\hat{\Phi} = \{L_n\}_{n\in\mathbb{N}^*}$ *is a Poisson point process with intensity measure*

$$\Gamma(0, m] = am^{2/\beta}$$

*where*

$$a := \frac{\pi\mathbb{E}\left[S^{2/\beta}\right]}{K^2} \sum_{o\in O} \gamma_o P_o^{2/\beta}$$

**Proof.** It follows from our assumptions on the shadowing that the random variables $S_n = \mathbf{S}_n(-v_n)$ are i.i.d. marks of $\Phi$. Introducing $\tilde{S}_n := S_n P_{Z_n}$, we get

$$L_n = \frac{l(|v_n|)}{\tilde{S}_n}$$

Since $\tilde{S}_n := S_n P_{Z_n}$ are i.i.d. marks of $\Phi$ (which may be viewed as *modified shadowing*), then $\hat{\Phi}$ may be obtained by a transformation of $\Phi$ through the kernel

$$p(x, A) = \mathbb{P}\left(\frac{l(x)}{\tilde{S}_1} \in A\right)$$

By the displacement theorem $\hat{\Phi}$ is a Poisson process with intensity measure

$$\begin{aligned}
\Gamma[0, m) &= \int_{\mathbb{R}^2} p(x, [0, m))\gamma dx \\
&= \gamma \int_{\mathbb{R}^2} \mathbb{P}\left(\frac{l(x)}{\tilde{S}_1} \in [0, m)\right) dx \\
&= \gamma \int_{\mathbb{R}^2\times\mathbb{R}} \mathbf{1}\left\{\frac{l(x)}{s} \in [0, m)\right\} dx\mathbb{P}_{\tilde{S}_1}(ds) \\
&= \gamma \int_{\mathbb{R}^2\times\mathbb{R}} \mathbf{1}\left\{|x| < (sm)^{\frac{1}{\beta}} \frac{1}{K}\right\} dx\mathbb{P}_{\tilde{S}_1}(ds) \\
&= \gamma \int_{\mathbb{R}} \frac{\pi (sm)^{2/\beta}}{K^2} \mathbb{P}_{\tilde{S}_1}(ds) = \frac{\gamma\pi}{K^2}\mathbb{E}\left[\tilde{S}_1^{2/\beta}\right] m^{2/\beta}
\end{aligned}$$

Note that

$$\mathbb{E}\left[\tilde{S}_1^{2/\beta}\right] = \mathbb{E}\left[S^{2/\beta}\right]\mathbb{E}\left[P_{Z_n}^{2/\beta}\right] = \mathbb{E}\left[S^{2/\beta}\right]\sum_{o\in O}\gamma_o P_o^{2/\beta}$$

which completes the proof. ■

It follows from the above lemma that the typical user (located at the origin) receives powers from the different BS *as if* the network was a *homogeneous* one, i.e. with a single category of BS transmitting at power

$$P = \left(\sum_{o\in O}\frac{\gamma_o}{\gamma}P_o^{2/\beta}\right)^{\beta/2} \tag{3.47}$$

Let $\hat{\Phi}_o$ be the process of the inverses of the received powers from BS of category $o \in O$. They are independent Poisson point processes with respective intensity measures

$$\Gamma_o(0,m] = a_o m^{2/\beta}, \quad o \in O$$

where

$$a_o = \frac{\pi\mathbb{E}\left[S^{2/\beta}\right]}{K^2}\gamma_o P_o^{2/\beta}, \quad o \in O$$

We may view the whole process $\hat{\Phi}$ as a superposition of the point processes $\left\{\hat{\Phi}_o\right\}_{o\in O}$. Inversely, these latter processes may be obtained from $\hat{\Phi} = \{L_n\}_{n\in\mathbb{N}^*}$ by generating i.i.d. marks $\left\{\hat{Z}_n\right\}_{n\in\mathbb{N}^*}$ such that

$$\mathbb{P}(\hat{Z}_n = o) = \frac{a_o}{a}, \quad o \in O, n \in \mathbb{N}^* \tag{3.48}$$

in which case

$$\hat{\Phi}_o = \sum_{n\in\mathbb{N}^*}\delta_{L_n}\mathbf{1}\left\{\hat{Z}_n = o\right\}, \quad o \in O$$

Let $Z^*$ be the category of the BS serving the user located at the origin and $L^* = \min_{n\in\mathbb{N}^*} L_n$.

**Lemma 6** *The probability that the user at the origin is served by a BS of category $o \in O$ equals*

$$\mathbb{P}(Z^* = o) = \frac{a_o}{a} \tag{3.49}$$

*Moreover, $\hat{\Phi}$ and $Z^*$ are independent and in particular the random variables $L^*$ and $Z^*$ are independent.*

**Proof.** We may assume without loss of generality that the points $\{L_n\}_{n\in\mathbb{N}^*}$ of $\hat{\Phi}$ are sorted in the increasing order. Then

$$\mathbb{P}(Z^* = o) = \mathbb{P}\left(\hat{Z}_1 = o\right) = \frac{a_o}{a}$$

On the other hand, since $Z^* = \hat{Z}_1$, it follows from (3.48) that $Z^*$ is independent from $\hat{\Phi}$. ■

The above lemma shows that the process of powers received from all the BS is independent from the category of the serving base station. In particular, the interference and the signal to interference and noise ratio (SINR) are independent from the category of the serving base station.

**Remark 22** *It follows from Lemma 6 that $\mathbb{P}(L^* > m|Z^* = o) = \mathbb{P}(L^* > m)$; i.e. the distribution of the received power from each category of BS is the same.*

### 3.6.3 Full interference model

The average of the characteristics of the cells described in Section 3.6.1, i.e. Section 3.5.2 over the whole network gives a first global indication of the network performance. We may also average these characteristics per category of BS leading to the notion of *typical cell* for each category of BS. The objective of the present section is to establish the relations between these averages. The Section title relates to the load (interference) scenarios in 1.

#### Typical cell approach

**Traffic and load** We shall average the characteristics of the cells over an increasing sequence of discs, denoted by $A$, of radii going to infinity, exactly in the same way as it is done in Section 3.5.3; for example the global average traffic demand per cell equals

$$\bar{\rho} := \lim_{|A|\to\infty} \frac{1}{\Phi(A)} \sum_{v\in\Phi\cap A} \rho(v)$$

whereas the average traffic demand per cell of category $o$ is

$$\bar{\rho}_o := \lim_{|A|\to\infty} \frac{1}{\Phi_o(A)} \sum_{v\in\Phi_o\cap A} \rho(v) \tag{3.50}$$

We define similarly the global average load $\bar{\theta}$ and the average load $\bar{\theta}_o$ for cells of category $o$.

The following lemma gives the explicit expressions of the average traffic demands per cell both globally and for each category of BS. Palm theory will be useful in the proof of this lemma as well as for upcoming results. The following statements are also inspired by the considerations in Section 3.5.3. In particular we shall make use of the Palm probability $\mathbb{P}^0$ associated to the point process $\Phi$ of base station locations. Let $\mathbb{E}^0[\cdot]$ be the expectation with respect to $\mathbb{P}^0$.

**Lemma 7** *We have*

$$\begin{aligned} \bar{\rho} &= \frac{\rho}{\gamma} \\ \bar{\rho}_o &= \frac{\rho a_o}{\gamma_o a}, \quad o \in O \end{aligned} \tag{3.51}$$

**Proof.** Since the point process $\Phi$ is ergodic, it follows from the ergodic theorem for point processes that [36, Proposition 12.2.VI]

$$\bar{\rho} = \mathbb{E}^0[\rho(0)] = \rho\mathbb{E}^0[|V(0)|] = \frac{\rho}{\gamma}$$

where the second equality is due to (3.20), and the last equality follows for the inverse formula of Palm calculus [9, Theorem 4.2.1] (which may be extended to the case where the cell associated to each BS is not necessarily the Voronoi cell; the only requirement is that the user located at 0 belongs to a unique cell almost surely). Similarly,

$$\begin{aligned} \bar{\rho}_o &= \lim_{|A|\to\infty} \frac{\Phi(A)}{\Phi_o(A)} \frac{1}{\Phi(A)} \sum_{v\in\Phi\cap A} \rho(v)\, 1\{v\in\Phi_o\} \\ &= \frac{\gamma}{\gamma_o}\mathbb{E}^0[\rho(0)\, 1\{0\in\Phi_o\}] \\ &= \mathbb{E}^0[\rho(0)\,|0\in\Phi_o] = \rho\mathbb{E}^0[|V(0)|\,|0\in\Phi_o] \end{aligned} \tag{3.52}$$

Moreover,

$$\begin{aligned} E^0\left[|V(0)|\,|0\in\Phi_o\right] &= \frac{\mathbb{E}^0\left[|V(0)|\times 1\{0\in\Phi_o\}\right]}{\mathbb{P}^0\left(0\in\Phi_o\right)} \\ &= \frac{\frac{1}{\gamma}\mathbb{P}\left(v^*\in\Phi_o\right)}{\mathbb{P}^0\left(0\in\Phi_o\right)} \\ &= \frac{\frac{1}{\gamma}\frac{a_o}{a}}{\frac{\gamma_o}{\gamma}} = \frac{a_o}{\gamma_o a} \end{aligned} \tag{3.53}$$

where the second equality follows from the inverse formula of Palm calculus, $v^*$ designates the BS serving the origin and the third equality follows from Lemma 6. ■

**Remark 23** *We may interpret Equation (3.53) as follows. Consider a sufficiently large area $A$ and denote its area by $|A|$. By the ergodic theorem, there are $\gamma_o|A|$ base stations of category $o$ within $A$. On the other hand, the probability that a user is served by a BS of category $o$ is $\frac{a_o}{a}$ and consequently a portion $\frac{a_o}{a}$ of the area of $A$ is covered by $\gamma_o|A|$ base stations. So, the mean cell area of a base station of category $o$ is*

$$\frac{\frac{a_o}{a}|A|}{\gamma_o|A|} = \frac{a_o}{\gamma_o a}$$

The following lemma shows that the global average cell load is related to the stationary distribution of the SINR for a user located at the origin. It gives also the expression of the average load for cells of each category.

**Proposition 9** *The average loads per cell equal*

$$\bar{\theta} = \frac{\rho}{\gamma}E\left[R^{-1}\left(\mathrm{SINR}\left(0,\Phi\right)\right)\right] \tag{3.54}$$

$$\bar{\theta}_o = \bar{\theta}\frac{\gamma a_o}{\gamma_o a} = \bar{\theta}\frac{P_o^{2/\beta}}{P^{2/\beta}}, \quad o\in O \tag{3.55}$$

*where $P$ is given by (3.47).*

**Proof.** Along the same lines as the proof of (3.52), we have

$$\bar{\theta}_o = \mathbb{E}^0\left[\theta(0)\,|0\in\Phi_o\right]$$

On the other hand, let

$$g\left(y,\Phi(\omega)\right) = \rho R^{-1}\left(\mathrm{SINR}\left(y,\Phi(\omega)\right)\right), \quad y\in\mathbb{R}^2,\omega\in\Omega$$

let $v_y^*$ be the BS serving the user at location $y$, and let $v^*$ be the BS serving the origin, then

$$\begin{aligned} \bar{\theta}_o &= \mathbb{E}^0\left[\theta(0)\,|0\in\Phi_o\right] \\ &= \mathbb{E}^0\left[\theta(0)\,1\{0\in\Phi_o\}\right]/\mathbb{P}^0\left(0\in\Phi_o\right) \\ &= \mathbb{E}^0\left[\int_{V(0)} g\left(y,\Phi\right)1\left\{v_y^*\in\Phi_o\right\}dy\right]/\mathbb{P}^0\left(0\in\Phi_o\right) \\ &= \frac{1}{\gamma}\mathbb{E}\left[g\left(0,\Phi\right)1\{v^*\in\Phi_o\}\right]/\mathbb{P}^0\left(0\in\Phi_o\right) \\ &= \frac{1}{\gamma}\mathbb{E}\left[g\left(0,\Phi\right)\right]\frac{\mathbb{P}\left(v^*\in\Phi_o\right)}{\mathbb{P}^0\left(0\in\Phi_o\right)} = \frac{1}{\gamma}\mathbb{E}\left[g\left(0,\Phi\right)\right]\frac{\gamma a_o}{\gamma_o a} \end{aligned}$$

where the fourth equality follows from the inverse formula of Palm calculus [9, Theorem 4.2.1], the fifth equality follows from Lemma 6 and the last equality follows from (3.49). It follows that

$$\begin{aligned}\bar{\theta} &= \mathbb{E}^0\left[\theta\left(0\right)\right] \\ &= \sum_{o\in O}\mathbb{E}^0\left[\theta\left(0\right)|0\in\Phi_o\right]\mathbb{P}^0\left(0\in\Phi_o\right) \\ &= \sum_{o\in O}\frac{1}{\gamma}\mathbb{E}\left[g\left(0,\Phi\right)\right]\mathbb{P}\left(v^*\in\Phi_o\right) = \frac{1}{\gamma}\mathbb{E}\left[g\left(0,\Phi\right)\right]\end{aligned}$$

■

**Remark 24** Heterogeneous versus homogeneous network. *Surprisingly, the global average cell load in the heterogeneous network is the same as that in the corresponding homogeneous network (where all the BS transmit the same power (3.47)). This observation holds for the current full interference model.*

**Users number and throughput** Note that if for BS $v$, $\theta\left(v\right)\geq 1$ then the corresponding number of users $N\left(v\right)$ is infinite; in which case we say that BS $v$ is instable. The empirical average of the number of users over the cells in the network would then be infinite. In order to avoid this degeneration, we define the global average number of users as

$$\begin{aligned}\bar{N} &:= \lim_{|A|\to\infty}\frac{1}{\Phi\left(A\right)}\sum_{v\in\Phi\cap A}N\left(v\right)1\left\{\theta\left(v\right)<1\right\} \\ &= \mathbb{E}^0\left[N\left(0\right)1\left\{\theta\left(0\right)<1\right\}\right]\end{aligned}$$

where the second equality follows from ergodicity. Similarly, the average number of users for BS of category $o\in O$

$$\begin{aligned}\bar{N}_o &:= \lim_{|A|\to\infty}\frac{1}{\Phi_o\left(A\right)}\sum_{v\in\Phi_o\cap A}N\left(v\right)1\left\{\theta\left(v\right)<1\right\} \\ &= \mathbb{E}^0\left[N\left(0\right)1\left\{\theta\left(0\right)<1\right\}|0\in\Phi_o\right]\end{aligned} \tag{3.56}$$

Let $\mathcal{S}_o$ be the union of stable cells of category $o\in O$; that is

$$\mathcal{S}_o = \bigcup_{v\in\Phi_o:\theta(v)<1}V\left(v\right)$$

and $\mathcal{S}=\bigcup_{o\in O}\mathcal{S}_o$. The user's average throughput is defined as the mean transmitted volume per call; i.e. $1/\mu$, divided by the mean call duration; that is

$$\bar{r} := \lim_{|A|\to\infty}\frac{1/\mu}{\text{mean call duration in } A\cap\mathcal{S}}$$

and for category $o\in O$,

$$\bar{r}_o := \lim_{|A|\to\infty}\frac{1/\mu}{\text{mean call duration in } A\cap\mathcal{S}_o}$$

**Proposition 10** *We have*

$$\begin{aligned} \bar{r} &= \frac{\bar{\rho}}{\bar{N}}\pi \\ \bar{r}_o &= \frac{\bar{\rho}_o}{\bar{N}_o}\pi_o, \quad o \in O \end{aligned} \tag{3.57}$$

*where*

$$\begin{aligned} \pi &= \mathbb{P}\left(\theta\left(v^*\right) < 1\right) \\ \pi_o &= \mathbb{P}\left(\theta\left(v^*\right) < 1 | v^* \in \Phi_o\right), \quad o \in O \end{aligned}$$

*where $v^*$ is the BS serving the origin.*

**Proof.** Let $\mathcal{W}_o = \bigcup_{v \in A \cap \mathcal{S}_o} V\left(v\right)$. Consider call arrivals and departures to $\mathcal{W}_o$ over a sufficiently large time interval. Let $T^{\mathcal{W}_o}$ be the mean call duration in $\mathcal{W}_o$, and let $N^{\mathcal{W}_o}$ be the mean number of users in $\mathcal{W}_o$. By Little's law

$$N^{\mathcal{W}_o} = \lambda \left|\mathcal{W}_o\right| T^{\mathcal{W}_o}$$

Thus the user's throughput in $\mathcal{W}_o$ equals

$$\begin{aligned} \frac{1/\mu}{T^{\mathcal{W}_o}} &= \frac{\rho \left|\mathcal{W}_o\right|}{N^{\mathcal{W}_o}} \\ &= \rho \frac{\sum_{v \in A \cap \mathcal{S}_o} \left|V\left(v\right)\right|}{\sum_{v \in A \cap \mathcal{S}_o} N\left(v\right)} \\ &= \rho \frac{\sum_{v \in A \cap \Phi} \left|V\left(v\right)\right| 1\left\{\theta\left(v\right) < 1, v \in \Phi_o\right\}}{\sum_{v \in A \cap \Phi} N\left(v\right) 1\left\{\theta\left(v\right) < 1, v \in \Phi_o\right\}} \end{aligned}$$

When $|A| \to \infty$, it follows from the ergodic theorem that

$$\bar{r}_o = \rho \frac{\mathbb{E}^0\left[\left|V\left(0\right)\right| 1\left\{\theta\left(0\right) < 1, 0 \in \Phi_o\right\}\right]}{\mathbb{E}^0\left[N\left(0\right) 1\left\{\theta\left(0\right) < 1, 0 \in \Phi_o\right\}\right]}$$

On the other hand, it follows from the inverse formula of Palm calculus that

$$\mathbb{E}^0\left[\left|V\left(0\right)\right| 1\left\{\theta\left(0\right) < 1, 0 \in \Phi_o\right\}\right] = \frac{1}{\gamma}\mathbb{P}\left(\theta\left(v^*\right) < 1, v^* \in \Phi_o\right)$$

Then

$$\begin{aligned} \bar{r}_o &= \frac{\rho}{\gamma} \frac{\mathbb{P}\left(\theta\left(v^*\right) < 1, v^* \in \Phi_o\right)}{\mathbb{P}^0\left(0 \in \Phi_o\right) \bar{N}_o} \\ &= \frac{\rho}{\gamma} \frac{\mathbb{P}\left(v^* \in \Phi_o\right) \mathbb{P}\left(\theta\left(v^*\right) < 1 | v^* \in \Phi_o\right)}{\mathbb{P}^0\left(0 \in \Phi_o\right) \bar{N}_o} \\ &= \frac{\bar{\rho}_o}{\bar{N}_o}\mathbb{P}\left(\theta\left(v^*\right) < 1 | v^* \in \Phi_o\right) = \frac{\bar{\rho}_o}{\bar{N}_o}\pi_o \end{aligned}$$

The expression for $\bar{r}$ may be proved in the same lines as above. ■

Under Palm probability, the cell of the base station located at 0 is usually called *typical cell*. Note that the typical cell has not a concrete existence, but it is rather a useful mathematical tool to analyze or predict the behavior of the empirical averages over many cells. We define both a

global typical cell and also a typical cell per category of BS. The global typical cell has a traffic demand $\bar{\rho}$, a load $\bar{\theta}$, a users number $\bar{N}$ and a user's throughput $\bar{r}$. The typical cell of category $o \in O$ has a traffic demand $\bar{\rho}_o$, a load $\bar{\theta}_o$, a users number $\bar{N}_o$ and a user's throughput $\bar{r}_o$.

Unfortunately, neither the number of users nor the user's throughput in the typical cell have explicit analytic expression. We shall introduce in the following section an approximation called *mean cell* whose characteristics admit explicit expressions and approximate those of the typical cell both globally and for each category of BS.

**Mean cell approximation**

We define a *mean cell* as a virtual cell having traffic demand $\tilde{\rho} := \bar{\rho}$ and load $\tilde{\theta} := \bar{\theta}$. Similarly, we introduce a mean cell for each category $o \in O$ as a virtual cell having traffic demand $\tilde{\rho}_o := \bar{\rho}_o$ and load $\tilde{\theta}_o := \bar{\theta}_o$. The mean cells are assumed to behave like a concrete cell from the queuing theory point of view as described in Sections 3.6.1 and 3.5.2. Specifically, we define the critical load of each mean cell in analogy to (3.25) as

$$\tilde{\rho}_{\mathrm{c}} := \frac{\tilde{\rho}}{\tilde{\theta}}, \quad \tilde{\rho}_{\mathrm{c}_o} := \frac{\tilde{\rho}_o}{\tilde{\theta}_o}, \quad o \in O$$

Inspired by (3.22), the user's throughput in the mean cell is defined as

$$\tilde{r} := \max(\tilde{\rho}_{\mathrm{c}} - \tilde{\rho}, 0), \quad \tilde{r}_o := \max(\tilde{\rho}_{\mathrm{c}_o} - \tilde{\rho}_o, 0), \quad o \in O$$

The users number in the mean cell is defined following (3.23) as

$$\tilde{N} = \frac{\tilde{\rho}}{\tilde{r}}, \quad \tilde{N}_o = \frac{\tilde{\rho}_o}{\tilde{r}_o}, \quad o \in O$$

Combining Equations (3.51) and (3.55), we see that $\tilde{\rho}_{\mathrm{c}_o} = \tilde{\rho}_{\mathrm{c}}$; that is the critical traffic of the mean cell is the same for all the categories of the BS. Moreover, observe that all the characteristics of the mean cell are straightforwardly deduced from its traffic and load which in view of (3.54) depends on the distribution of $\mathrm{S}I\mathrm{NR}(0, \Phi)$ whose analytic expression is given in [19].

We shall evaluate the mean cell approximation (both globally and per category) by comparison to the characteristics of the typical cell obtained both from simulation and from real field measurements.

### 3.6.4 Weighted interference model

**Typical cell approach**

The expression (3.46) of SINR relies on the assumption that the interfering BS are always transmitting at their maximal power. In real networks, the BS transmits only when it serves at least one user. Since the probability of such event is given by (3.24), we modify the expression of SINR as follows, for $y \in V(u), u \in \Phi$,

$$\mathrm{SINR}(y, \Phi) = \frac{\frac{1}{L_u(y)}}{\mathcal{N} + \sum_{v \in \Phi \setminus \{u\}} \frac{\min(\theta(v),1)}{L_v(y)}} \tag{3.58}$$

It follows from Equation (3.25) that

$$\theta(u) = \rho \int_{V(u)} R^{-1} \left( \frac{\frac{1}{L_u(y)}}{\mathcal{N} + \sum_{v \in \Phi \setminus \{u\}} \frac{\min(\theta(v),1)}{L_v(y)}} \right) dy \tag{3.59}$$

which is a system of equations with the loads $\{\theta(v)\}_{v\in\Phi}$ as unknowns. Given a realization of the network, one gets the loads of the different BS by solving numerically the above system, similarly as in 3.5.3. The traffic demand (or equivalently surface) of each cell $\{\rho(v)\}_{v\in\Phi}$ may also be estimated numerically. The other characteristics of each cell are then deduced from the load and traffic demands using the relations in Section 3.6.1 and 3.5.2; specifically, the critical traffic is deduced from (3.25), the user's throughput from (3.22) and the users number from (3.23).

The characteristics of the typical cell for each category of BS may then be computed by taking the empirical averages over the cells of each category as explained in Section 3.6.3. Specifically, the traffic is calculated by (3.50) with a similar formula for the load, the number of users is calculated by (3.56) and the user's throughput is given by (3.57) since Proposition 10 holds true in the present context. Moreover, the results of Proposition 9 remain true; in particular the load of the typical cell of a given category $\bar{\theta}_o$ is related to the global typical cell load $\bar{\theta}$ by Equation (3.55).

We will see in the numerical section that such model, called (load-)*weighted interference model*, fits better to real field measurements than the full interference model. In particular, the fact that the average load of each category of BS increases with its power through a simple law will be validated by real field measurements in the numerical section.

**Mean cell approximation**

In the mean cell approximation, the expression (3.58) of the S$I$NR is modified by replacing the load of each BS by the mean load of the corresponding category; that is

$$\widetilde{\mathrm{SINR}}(y,\Phi)=\frac{\frac{1}{L_u(y)}}{\mathcal{N}+\sum_{o\in O}\tilde{\theta}_o\sum_{v\in\Phi_o\setminus\{u\}}\frac{1}{L_v(y)}}$$

Moreover, by analogy to (3.54) and (3.55) we assume that

$$\begin{aligned}\tilde{\theta}&=\frac{\rho}{\gamma}\mathbb{E}\left[R^{-1}\left(\widetilde{\mathrm{SINR}}(0,\Phi)\right)\right]\\ \tilde{\theta}_o&=\tilde{\theta}\frac{P_o^{2/\beta}}{P^{2/\beta}}\end{aligned}\tag{3.60}$$

Combining the three above equations we deduce that $\tilde{\theta}$ is solution of the following fixed-point equation

$$\tilde{\theta}=\frac{\rho}{\gamma}\mathbb{E}\left[R^{-1}\left(\frac{\frac{1}{L_u(0)}}{\mathcal{N}+\tilde{\theta}\sum_{o\in O}\alpha_o\sum_{v\in\Phi_o\setminus\{u\}}\frac{1}{L_v(0)}}\right)\right]\tag{3.61}$$

where

$$\alpha_o:=\frac{P_o^{2/\beta}}{P^{2/\beta}},\quad o\in O\tag{3.62}$$

Solving the above equation we get the load of the mean cell; then we deduce the load of the mean cell of each category by applying (3.60). The characteristics of the mean cell both globally and for each category are then deduced from its traffic and load by the relations presented in Section 3.6.3.

## 3.7 Numerical results: network dimensioning and QoS estimation

In this Section the following results are presented:

- Firstly, the validation is done for a dynamic context of users arivals and departures from the network by comparison to 3GPP simulations results.
- Then, the results for dynamic context are applied to dimension hexagonal cellular LTE network based on Section 3.4.
- Finally, the previous two tasks are extended to demonstrate the results concerning irregular networks. QoS estimation as function of traffic demand is done for irregular cellular networks and results are compared to real-field measurements.

Throughout this Section we assume that user channels are *intra-cell orthogonal and inter-cell independent*: if BS $u$ serves $n$ users located at $y_1, y_2, \ldots, y_n \in V(u)$ then the bit-rate of the user located at $y_j$ equals to $1/n$ th of its peak bit-rate $\frac{1}{n} R(\mathrm{SINR}(y_j, \Phi))$, $j \in \{1, 2, \ldots, n\}$. [7]The pattern of BS $\Phi$ does not evolve in time.

### 3.7.1 Validation in a dynamic context

The aim is to compare the results of the queueing approach described in Section 3.4.1 to those of 3GPP simulations in a *dynamic context*; i.e. calls arrive and depart from the network and each base station transmits only when it has at least one user to serve. This context is called *FTP traffic model* in [3, §A.2.1.3.1].

As simulation results, we consider the results of tools which are compliant with 3GPP approach [3] (an Orange simulator developed in C++ being one of them). The average (as well confidence intervals at 20% and 80%) of the results of the different contributors to 3GPP will be plotted and compared to our analytical approach (implemented in Matlab).

We begin by describing the subset of the parameters in [1, Table A.2.1.1-3], [3, Table A.2.2-1] which are used in our analytical calculations. The frequency carrier equals 2GHz; the path-loss model is $l(r) = 128.1 + 37.6 \times \log_{10}(r)$ [in dB] (where $r$ is in km); the penetration loss equals 20dB and the shadowing is centered and log-normally distributed with standard deviation 8dB. The antenna pattern in the horizontal plane is $A(\varphi) = -\min\left(12\left(\varphi/\varphi_{3\mathrm{dB}}\right)^2, A_{\mathrm{m}}\right)$ where $\varphi_{3\mathrm{dB}} = 70^\circ, A_{\mathrm{m}} = 20\mathrm{dB}$. The system bandwidth equals $W = 10\mathrm{MHz}$; the noise power is $\mathcal{N} = -95\mathrm{dBm}$ and the base station transmission power equals $P = 60\mathrm{dBm}$ (including antenna gain).

Figure 3.1 gives the load as function of traffic demand per cell resulting from 3GPP simulations and from the queueing approach. For 3GPP simulations, the load is calculated as the fraction of time where a base station has at least one user to serve. The average of the simulation results of the different 3GPP contributors as well the confidence intervals at 20% and 80% are plotted. For the queueing approach, the load is calculated by Equation (3.19) where the critical traffic $\rho_{\mathrm{c}}$ is the solution of the fixed-point problem (3.18). We observe in Figure 3.1 that the two loads calculated by these two methods are close, except when the queueing load is close to 1. Indeed in this case, the system is at its limit of stability and therefore the time averages converge very slowly to their ergodic limits [80, p.114]. This explains why the 3GPP simulations are too time consuming at high loads (up to 3 weeks of calculation) and also the gap between the simulation and queueing loads in Figure 3.1. To get the curves in Figure 3.1, the computing time for the 3GPP simulations is several weeks whereas it is about 1 minute for the queueing approach.

Figure 3.2 gives the mean user throughput as function of traffic demand for the different load situations described in Definition 1 and for 3GPP simulations. In this latter case, the mean user throughput is obtained by averaging the users throughput over the whole simulation time.

[7]This can be achieved using various multiple access schemes, e.g. time (related to HSDPA) or frequency (related to LTE) division.

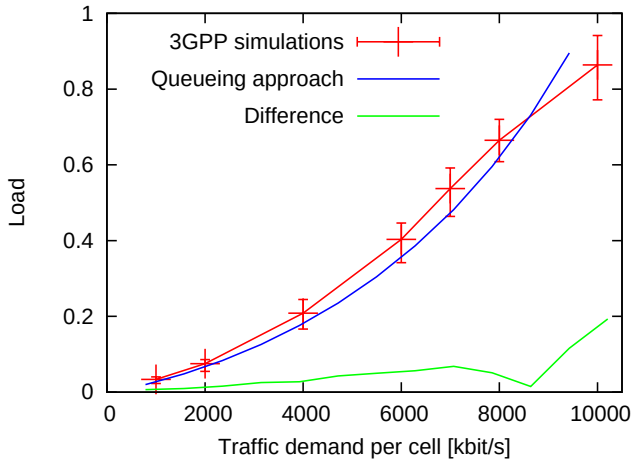


Figure 3.1: Load versus traffic demand per cell

One can observe that the curve obtained from 3GPP simulations and the one derived from the queueing approach with adapted load are close, except for the highest value of traffic demand (which corresponds to a load close to 1). This gap is due to the slow convergence rate of the 3GPP simulations discussed above. On the other hand, as expected, the curves for the full and adapted load converge for the highest value of the traffic demand since this corresponds to the limit of stability of the network (when the user throughput vanishes). Going backward with the values of traffic demand, the difference between these two curves increases up to factor 5. Finally, the null and adapted load curves have the same value for the smallest traffic demand and diverge as the traffic demand increases.

Figure 3.3 gives the 95% quantile of user throughput as function of the traffic demand for 3GPP simulations and the queueing approach. Observe that the quantiles of the 3GPP simulations are smaller than those of the queueing approach with adapted load; nevertheless the two curves have the same tendency. This is related to the fact that peak bit-rates of 3GPP simulations are more dispersed than the analytical ones as shown in Figure 2.5. On the other hand, the null and full load curves agree with the adapted load one for the smallest and the highest values of traffic demand, respectively.

In this Section we want to use all previously developed results and apply them to various numerical examples of cellular network dimensioning and QoS performance evaluation. We will use the results from Section 3.4 to calculate the cell radius necessary to satisfy the traffic demand and some QoS constraint with a given network setup, which is network dimensioning.

Using the model described in Section 3.5 we will estimate the QoS parameters and network performance and compare the results to the real-field measurements.

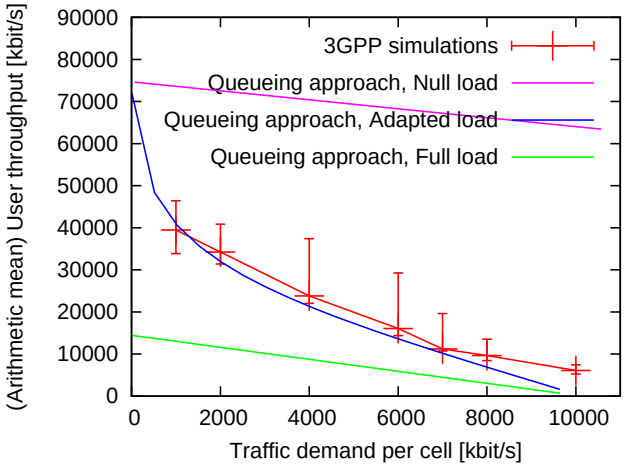


Figure 3.2: Mean user throughput versus traffic demand per cell

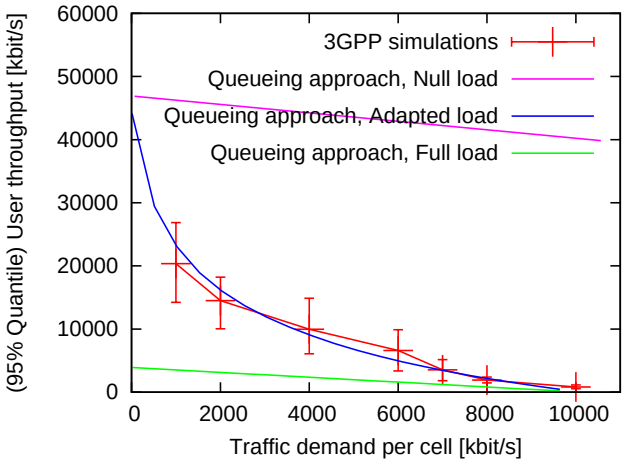


Figure 3.3: 95% quantile of user throughput versus traffic demand

### 3.7.2 Hexagonal LTE network dimensioning

We aim now to illustrate the processor-sharing queueing approach for cellular network dimensioning purposes applied to a multi-cell scenario in symmetric hexagonal networks. In this case all cells are "the same", that is to say, if we take randomly any cell then it represents the typical cell, so we can use the approach described in Section 3.4 for network dimensioning. Figure 3.4 shows mean user throughput as function of cell radius for traffic demand densities equal to 0.1 and 10Mbit/s/km$^2$ and different load situations (see Definition 1). The mean user throughput $\bar{r}$ is calculated by (3.8) where the critical traffic $\rho_c$ is the solution of the fixed-point problem (3.18).

As expected, for each value of the traffic demand, the curves are in decreasing order for respectively the null, adapted and full load situations. Moreover, observe that the null and adapted load curves are close to each other for the small value of traffic demand since in this case the interference is too small. Contrarily, for higher value of traffic demand, interference is

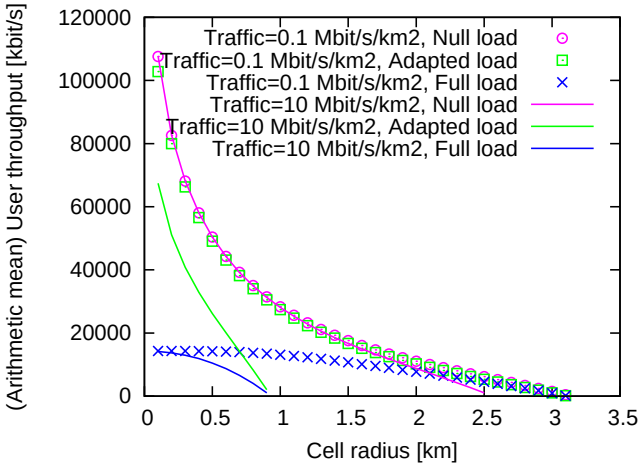


Figure 3.4: Mean user throughput as function of the cell radius for different load situations

significant so that adapted load curve deviates from the null load one. Moreover, as observed previously, the adapted and full load curves converge for the limit of stability of the network (when user throughput vanishes). The computing time to get Figure 3.4 is some minutes, whereas it would require several weeks for 3GPP simulations.

Figure 3.5 shows mean user throughput as function of cell radius for different traffic demand densities for adapted load situation. As expected each curve is decreasing and ultimately vanishes for some critical value of cell radius corresponding to the stability limit of the network. Additionally, when the traffic increases, the curves decrease and the critical cell radius decreases rapidly. On the other hand, the curves for traffic demands of 10 and 100kbit/s/km$^2$ are close to that of null traffic up to the cell radius of 2km which shows that noise is preponderant against interference.

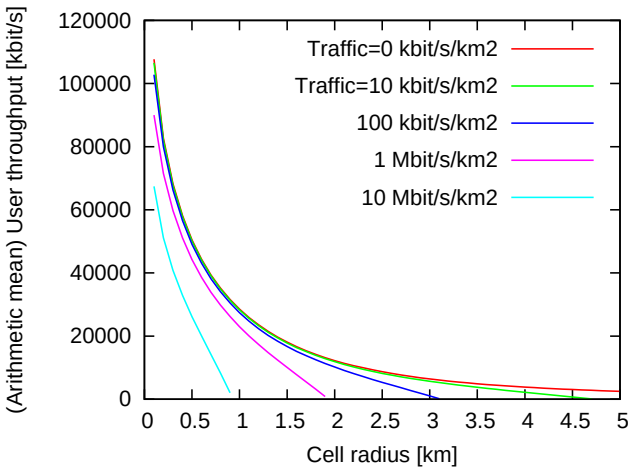


Figure 3.5: Mean user throughput as function of the cell radius for different traffic demand densities (adapted load)

We give now the numerical solution of the dimensioning problem in terms of the cell radius which is more appealing than the number of base stations per unit surface (note, however, that the latter is inversely proportional to the square of the former). Figure 3.6 shows cell radius versus traffic demand density for two target arithmetic means of the user throughput equal to 1 and 10Mbit/s for the three load situations described in Definition 1. For the smaller user throughput, the three curves are close to each other whereas for the larger throughput they differ significantly from each other. The adapted load curve lies between the null and full load ones; and meets each of them for low and high traffic, respectively. This is due to the fact that when traffic increases, the network evolves from a noise-limited to an interference-limited regime.

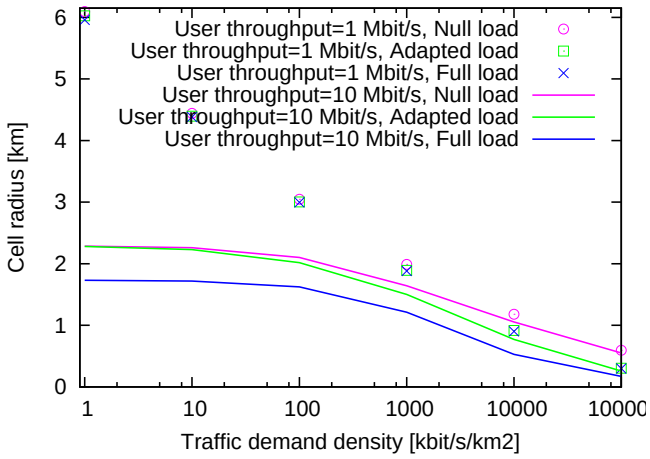


Figure 3.6: Cell radius versus traffic demand density for mean user throughput $10^4$kbit/s

Figure 3.7 shows cell radius versus traffic demand density for different target values of the arithmetic mean of the user throughput. As expected, the cell radius is decreasing with the traffic demand and with the user throughput. Note that, for the three largest user throughputs, the curve comprises a stationary part corresponding to a coverage constraint and a decreasing part corresponding to a capacity constraint.

Figures 3.6 and 3.7 are obtained in few minutes by the analytical approach, whereas they would require several months for 3GPP simulations.

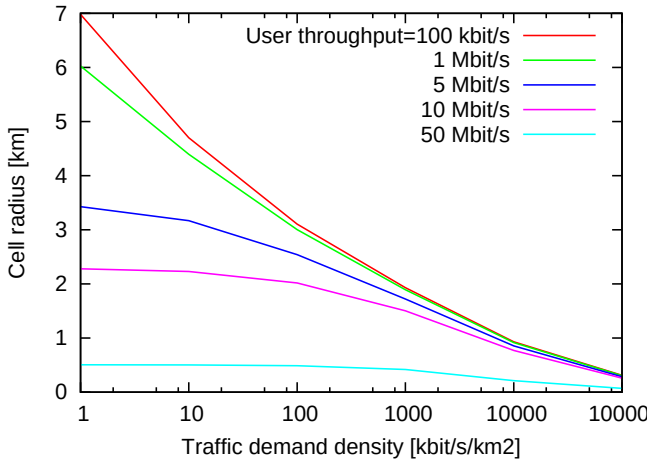


Figure 3.7: Cell radius versus traffic demand density for different mean user throughputs

### 3.7.3 Mean performance estimation of irregular network using Poisson process

Here, we consider the irregular cellular network with orthogonal users' channels as described in Section 3.5 and produce the examples showing the network performance as function of the traffic demand. Also, the figures showing the mean user throughput as function of the traffic demand can be useful not only in the sense of their own explicit importance (gives the information about the QoS in a network), but also for network dimensioning and capacity.

To illustrate the motivation of this work, we present first in Figure 3.8 non-averaged data obtained from the measurements performed in an operational network in some zone of some big city in Europe. More precisely, a *dense urban network zone* consisting of 382 base stations was selected in a big European city, whose locations loosely satisfy the homogeneous spatial Poisson assumption. Ripley's $L$-function, cf [88, page 50], plotted on Figure 3.16, was used to verify this latter assumption. The density of base stations in this dense urban zone is about 4.62 base stations per km$^2$. Later, we will consider also an *urban zone* of a different European city, where the spatial homogeneous Poissonianity of the base station locations can also be retained; cf. Figure 3.16, with roughly four times smaller density of base stations, more precisely 1.15 stations per km$^2$. In both cases the network operates a HSDPA system with MMSE coding.

Different points in this figure correspond to the measurements of the traffic demand and the estimation of the user throughput made by different cells during different hours of the day. No apparent relation between these two quantities can be observed in this way.

In order to uderstand and predict the performance of the network for which we have presented the above data, we will now specify correspondingly our general model and study it using the proposed approach. The obtained results will be compared to the appropriately averaged real field measurements.

Consider the following numerical setup. Assume a Poisson process of BS with intensity $\gamma = 4.62\text{km}^{-2}$ (which corresponds to an average distance between two neighbouring BS of 0.5km). We assume the path-loss function $l(r) = (Kr)^\beta$, with $K = 7117\text{km}^{-1}$, and the path loss exponent $\beta = 3.8$. The propagation model comprises the log-normal shadowing with the logarithmic standard deviation 10dB; cf [20], and the mean spatial correlation distance 0.05km.

The transmission power is $P = 58\text{dBm}$, with a fraction $\epsilon = 10\%$ used in the pilot channel.

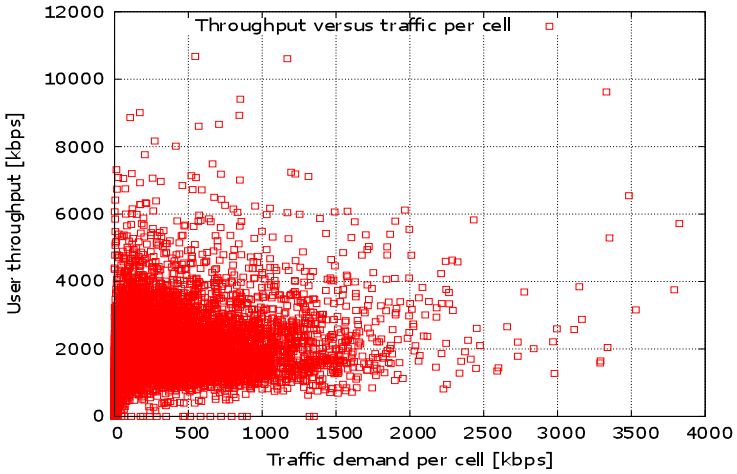


Figure 3.8: Local user throughput versus local traffic demand for some zone (selected to satisfy a spatial homogeneity of the base stations) of an operational cellular network deployed in a big city in Europe. 9288 different points correspond to the measurements made by different sectors of different base stations during 24 different hours of some given day.

The antenna pattern is described in [3, Table A.2.1.1-2]. The noise power is $N = -96$dBm.

We assume the peak bit-rate to be equal to 30% of the ergodic capacity of the AWGN channel with the frequency bandwidth $W = 5$MHz and the Rayleigh fading with mean power $E[|H|^2] = 1$.

Estimations of the typical cell are performed by the simulation of 30 realizations of the Poisson model within a finite observation window, which is taken to be the disc of radius 2.63km. We first average over all BS in this window and then over the model realizations. The empirical standard deviation from the obtained averages will be presented via error-bars.

We shall study now our model using the typical and mean cell approach, assuming first the full interference model and then the weighted one.

### Full interference

We consider first the full interference model (i.e. all BS emit the signal all the time, regardless of whether or not they serve users). Figure 3.9 shows the mean cell load of the typical cell $\mathbf{E}^0[\theta(0)]$ and the stable fraction of the network $\pi_{\mathcal{S}}$ obtained from simulations, as well as the analytically calculated load of the mean cell $\bar{\theta}$, versus mean traffic demand per cell $\rho/\gamma$. We confirm that the typical cell and the mean cell models have the same load. Note that for a traffic demand up to 500kbps per cell we do not observe unstable cells in our simulation window ($\pi_{\mathcal{S}} = 1$).

Figure 3.10 shows the mean number of users per cell in the stable part of the network $N^0/\pi_{\mathcal{S}}$ (obtained from simulations) and the analytically calculated number of users in the mean cell $\bar{N}$ versus mean traffic demand per cell. We have two remarks. For the traffic demand smaller than 500kbps per cell (for which all the simulated cells are stable; $\pi_{\mathcal{S}} = 1$, cf. Figure 3.9), both models predict the same mean number of users per cell. Beyond this value of the traffic demand per cell the estimators of the number of users in the typical cell become inaccurate due to the very rapidly increasing fraction of the unstable region. (Error bars on all figures represent the standard deviation in the averaging over 30 realizations of the Poisson network).

Finally, Figure 3.11 presents the dependence of the mean user throughput in the network

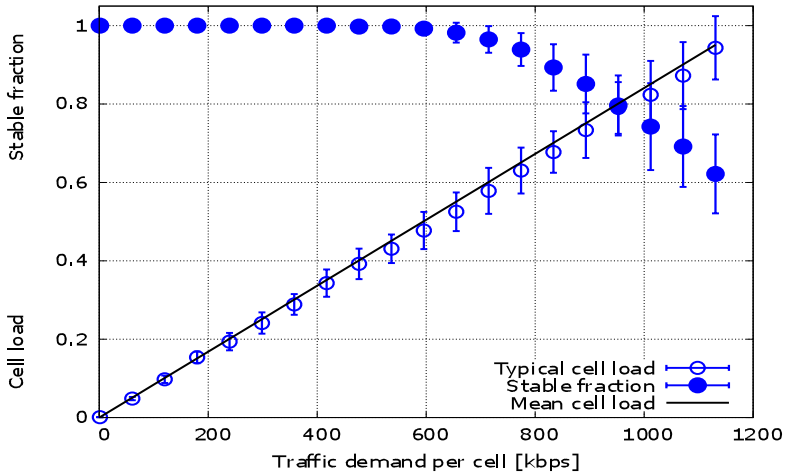


Figure 3.9: Cell load and the stable fraction of the network versus traffic demand per cell in the full interference model.

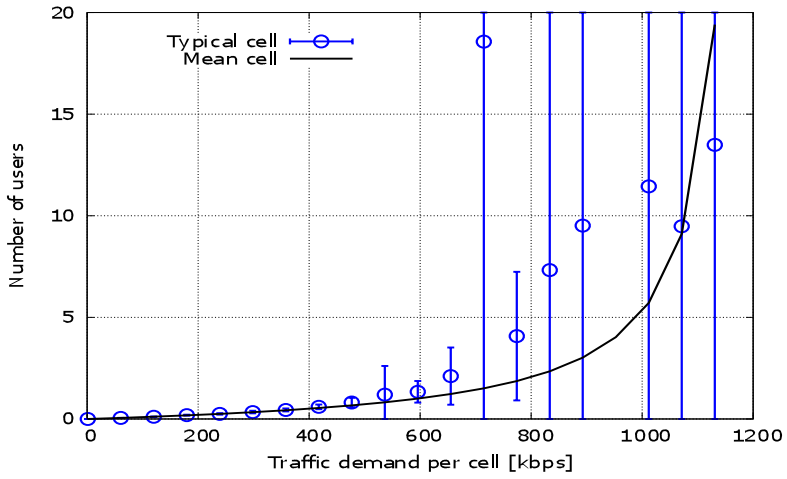


Figure 3.10: Number of users per cell versus traffic demand per cell in the full interference model.

on the mean traffic demand per cell obtained using the two approaches: $r^0$ and for the typical cell and $\bar{r}$ for the mean cell. Again, both models predict the same performance up to roughly 500kbps.

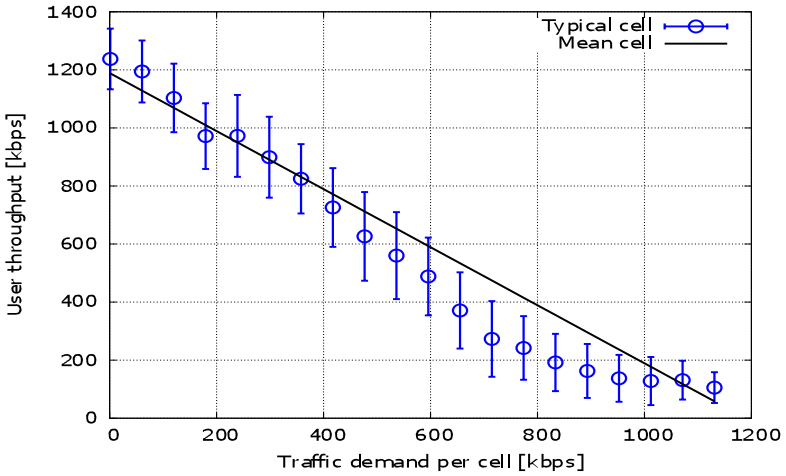


Figure 3.11: Mean user throughput in the network versus traffic demand per cell in the full interference model.

### Weighted interference

We consider now the load-weighted interference model taking into account idling cells. We see in Figures 3.12, 3.13 and 3.14 that the consequence of this (more realistic) assumption is that the cell loads are smaller, a larger fraction $\pi_{\mathcal{S}}$ of the network remains stable, and the two approaches (by the typical cell and by the mean cell) predict similar values of the QoS metrics up to a larger value of the traffic demand per cell, roughly 700kbps. Note that it is in this region that the real network operates for which we present the measurements, and that its performance coincides with the performance metrics calculated using the typical and mean cell approach. More precisely, the field measurements in Figures 3.12, 3.13 and 3.14 correspond to the same day and network zone considered in Figure 3.8.

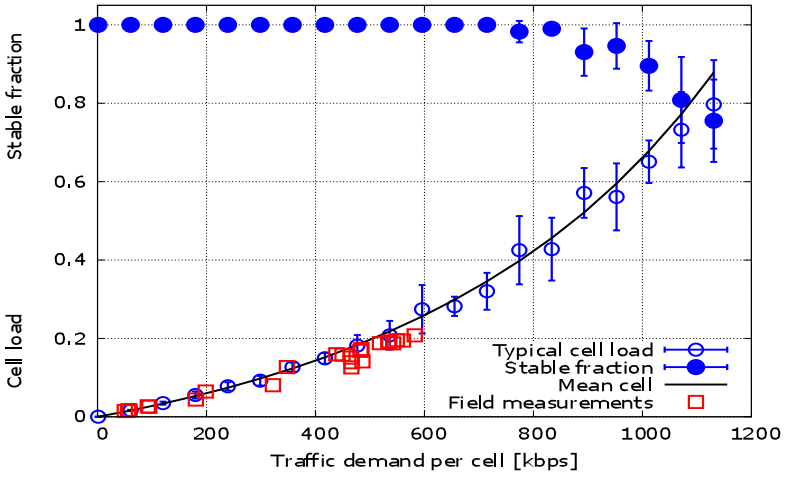


Figure 3.12: Load and the stable fraction of the network versus traffic demand in the weighted interference model. Also, load estimated from real field measurements.

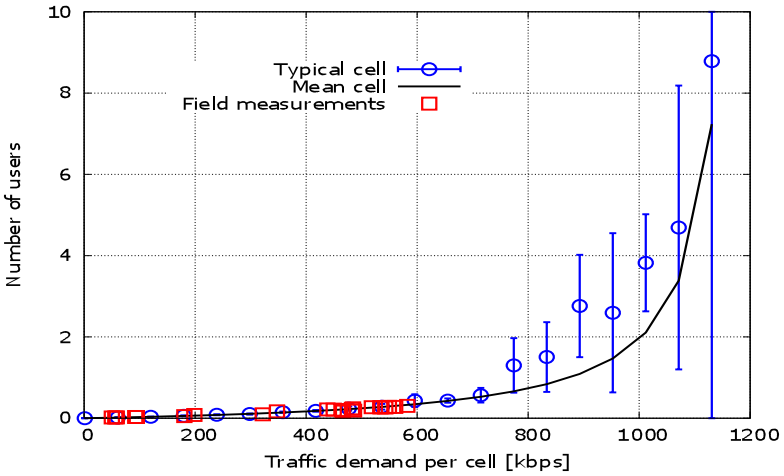


Figure 3.13: Number of users versus traffic demand per cell in the weighted interference model. Also, the same characteristic estimated from the real field measurements.

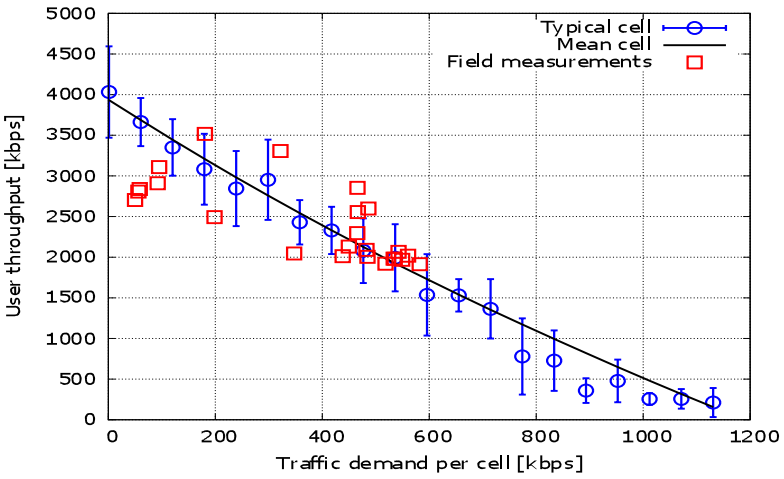


Figure 3.14: Mean user throughput in the network versus traffic demand per cell in the weighted interference model. Also, the same characteristic estimated from the real field measurements.

**Remark 25 (Measurement methodology)** *Measurement points in Figure 3.12 show the fraction of time, within a given hour, when the considered base stations were idle, averaged over the base stations, as function of the average traffic demand during this hour. Similarly, measurement points in Figure 3.13 show the spatial average of the mean number of users reported by the considered base stations within a given hour, as function of the average traffic demand during this hour. Finally, measurements in Figure 3.14 give the ratio of the total number of bits transmitted by all the base stations during a given hour, to the total number of users they served during this hour as function of the average traffic demand during this hour.*

Remark that Figure 3.14 makes evident a macroscopic relation between the traffic demand and the mean user throughput in the network zone already considered on Figure 3.8. This relation, we are primarily looking for here, is not visible without the spatial averaging of the network measurements described in Remark 25. In order to assure the reader that a relatively good matching between the measurements and the analytic prediction is not a coincidence, we present in Figure 3.15 similar results for an *urban zone* of a different European city, where the spatial homogeneous Poissonianity of the base station locations can also be retained; cf. Figure 3.16. The only engineering difference of this network zone with respect to the previously considered dense urban zone is roughly four times smaller density of base stations, more precisely 1.15 stations per $km^2$.

**Remark 26 (Day and night hours)** *Let us make a final remark regarding the empirical relation between the mean user throughput and the mean traffic demand revealed in Figures 3.14 and 3.15. Recall that different points in these plots correspond to different hours of some given day. In fact, the points below the mean curve correspond to day hours while the points above the mean curve correspond to night hours. This "circulation" of the measured values around the theoretical mean curve, indicated in Figure 3.15 and visible in both presented plots of the throughput, seems to be a more general rule, which escapes from the analysis presented in this thesis and remains an open question. A possible explanation can lie in a different space-time structure of the traffic during the day and night, with the former one being much more clustered (fewer users, requesting larger volumes, generating less interference and overhead traffic).*

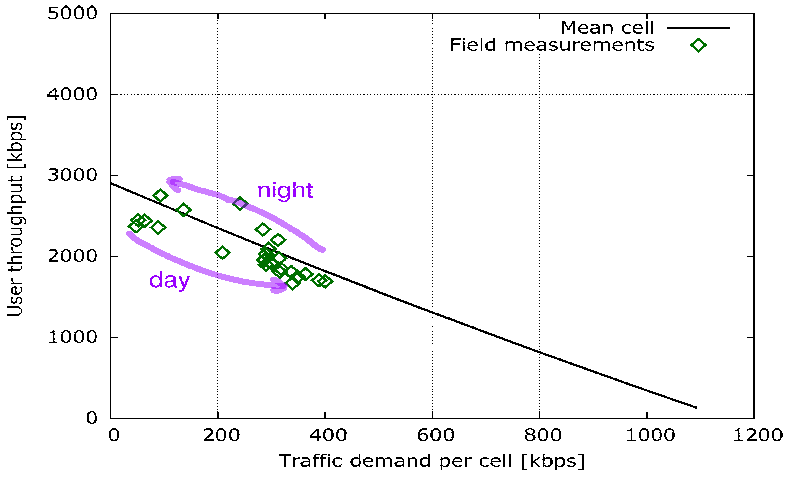


Figure 3.15: Mean user throughput in the network versus traffic demand per area for an urban zone of a big city in Europe. (The density of base stations is 4 times smaller than in the dense urban zone considered in Figure 3.14).

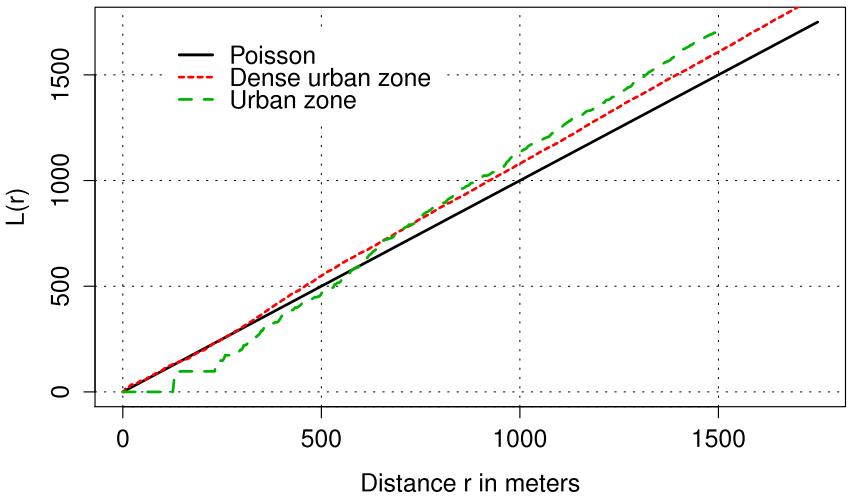


Figure 3.16: Ripley's $L$-function calculated for the considered dense urban and urban network zones. ($L$ function is the square root of the sample-based estimator of the expected number of neigbours of the typical point within a given distance, normalized by the mean number of points in the disk of the same radius. Slinvyak's theorem allows to calculate the theoretical value of this function for a homogeneous Poisson process, which is $L(r) = r$.) In fact, in large cities spatial, homogeneous "Poissonianity" of base-station locations is often satisfied "per zone" (city center, residential zone, suburbs, etc.). Moreover, log-normal shadowing further justifies the Poisson assumption, cf. [20, 29] .

### 3.7.4 Numerical results for heterogeneous networks

We consider the following numerical setting representative of an operational network in some big city in Europe comprising macro and micro base stations. The performance of each category of BS calculated using the approach proposed in the present Chapter is compared to the real field measurements.

#### Model specification

The network comprises *macro* and *micro* base stations indexed by 1 and 2 respectively. The BS locations are generated as a realization of a Poisson point process of intensity $\gamma = \gamma_1 + \gamma_2 = 4.62\text{km}^{-2}$ (which corresponds to an average distance between two base stations of 0.5km) over a sufficiently large observation window which is taken to be the disc of radius 2.63km. The ratio of the micro to macro BS intensities equals $\gamma_2/\gamma_1 = 0.039$.

The powers transmitted by macro and micro BS equal $P_1 = 58.26\text{dBm}$, $P_2 = 47.42\text{dBm}$ respectively. The global average power calculated by (3.47) equals $P = 58.03\text{dBm}$.

The propagation loss due to distance is $l(x) = (K|x|)^\beta$ where $K = 7117\text{km}^{-1}$ and the path loss exponent $\beta = 3.8$. Shadowing is assumed log-normally distributed with standard deviation $\sigma = 10\text{dB}$ and spatial correlation 0.05km.

The technology is HSDPA (High-Speed Downlink Packet Access) with MMSE (Minimum Mean Square Error) receiver in the downlink. The peak bit-rate equals to 30% of the information theoretic capacity of the Rayleigh fading channel with AWGN; that is

$$R\left(\mathrm{SINR}\right) = 0.3WE\left[\log_2\left(1+\left|H\right|^2\mathrm{SINR}\right)\right]$$

where the expectation $E\left[\cdot\right]$ is with respect to the Rayleigh fading $H$ of mean power $E[\left|H\right|^2] = 1$, and $W = 5$MHz is the frequency bandwidth.

A fraction $\epsilon = 10\%$ of the transmitted power is used by the pilot channel (which is always transmitted whether the BS serve users or not). The antenna pattern is described in [3, Table A.2.1.1-2]. The noise power is $-96$dBm.

We study now our model using the typical and mean cell approaches, assuming first the full interference model and then the weighted one. We shall give the performance results both globally for all the categories of BS and separately for macro and micro categories.

The results for the typical cell approach are obtained either by simulation or from measurements, whereas the results for the mean cell approach are analytic. Note that the mean load is known analytically (3.19) for both the typical and mean cells; nevertheless, we will associate this analytic expression to the mean cell approach in the legends of the subsequent curves.

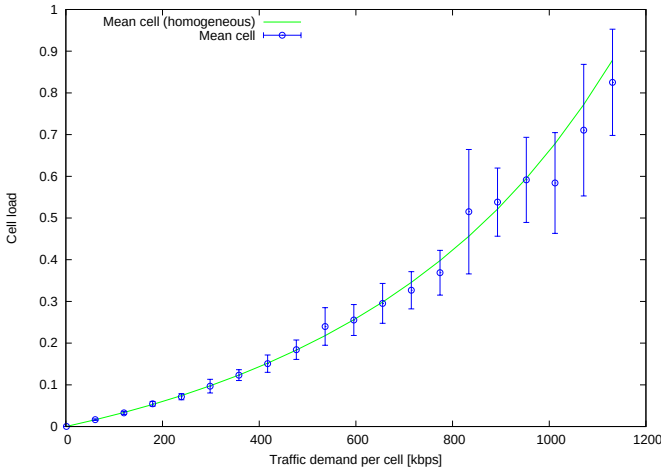


Figure 3.17: Accuracy of the homogeneous approximation of the mean cell

### Full interference

We consider first the full interference model (i.e. each BS always transmits at its maximal power even when it has no user to serve). This model is analyzed by simulation with the typical cell approach or analytically with the mean cell approach (no measurements are available in the full interference case).

Figure 3.18 shows the mean cell load of the typical cell $\bar{\theta}$ and the stable fraction of the network $\pi$ obtained from simulations, as well as the load of the mean cell $\tilde{\theta}$ calculated analytically, versus mean traffic demand per cell $\rho/\gamma$. This figure confirms that the typical cell and the mean cell models have the same load both globally and for each category of BS.

Figure 3.19 shows the mean number of users per cell $\bar{N}$ (obtained from simulations) and the analytically calculated number of users in the mean cell $\tilde{N}$ versus mean traffic demand per cell.

Again the mean cell reproduces well the results of the typical cell at least for moderate traffic demands; i.e. as long as the stable fraction of the network $\pi$ remains close to 1 as may be seen in Figure 3.18.

Finally, Figure 3.20 presents the dependence of the mean user throughput in the network on the mean traffic demand per cell obtained using the two approaches: $\bar{r}$ and for the typical cell and $\tilde{r}$ for the mean cell. Observe again for moderate traffic demands the good fit between the mean and typical cells both globally and for each BS category.

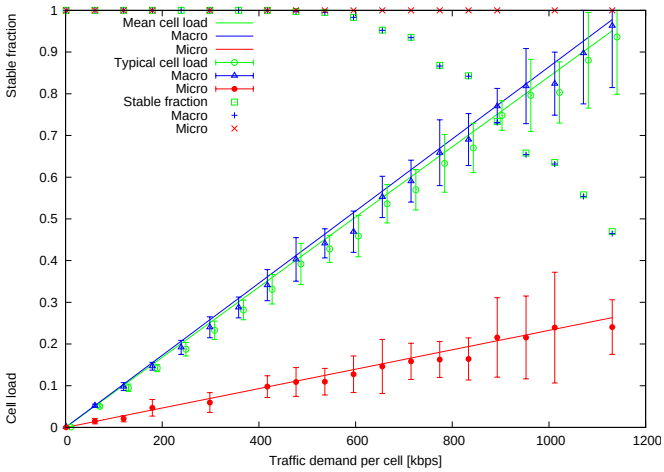


Figure 3.18: Cell load versus traffic demand per cell in the full interference model.

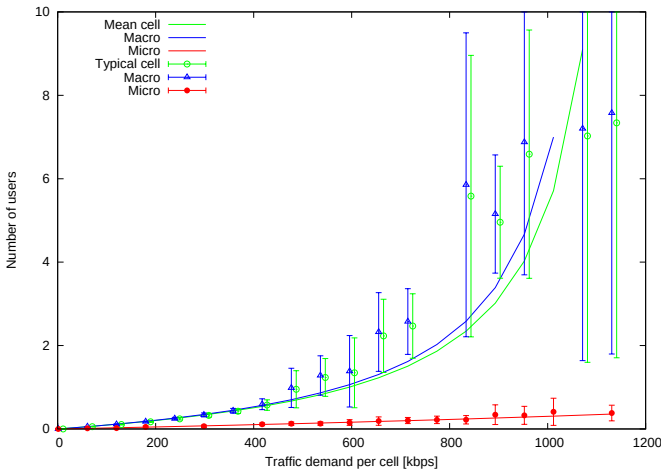


Figure 3.19: Number of users per cell versus traffic demand per cell in the full interference model.

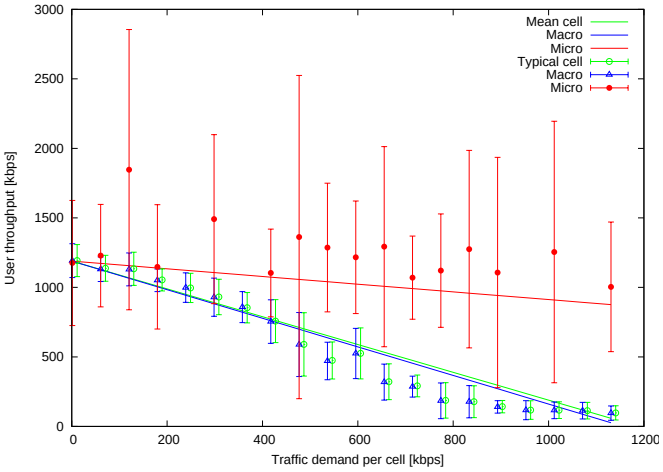


Figure 3.20: Mean user throughput in the network versus traffic demand per cell in the full interference model.

### Weighted interference

We consider now the load-weighted interference model accounting for idle periods of the interfering BS.

Figure 3.21 shows the mean cell load and the stable fraction of the network, Figure 3.22 presents the mean number of users per cell, and Figure 3.23 shows the mean user throughput. Besides the results of the analytic mean cell and the simulated typical cell, these figures give also the typical cell characteristics deduced from measurements in the operational network. As expected, the performance is improved compared to the full interference model.

Moreover, observe that the analytic mean cell load and number of users fits well with both the simulated and measured typical cell; particularly in the range of traffic demands for which measurements are available. This agreement holds both globally and for each category of BS; in particular the dependence of the average load of each category of BS with its power in the operational network is well explained by the simple law (3.60).

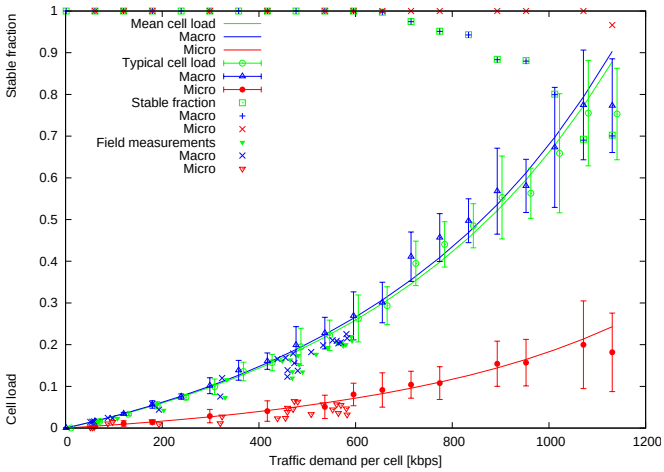


Figure 3.21: Cell load versus traffic demand per cell in the weighted interference model.

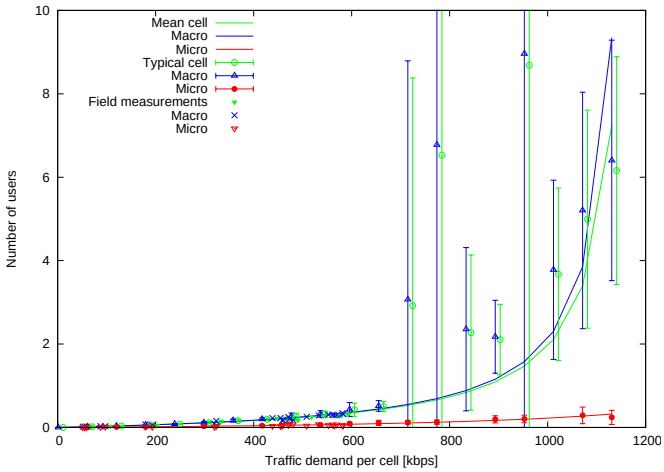


Figure 3.22: Number of users per cell versus traffic demand per cell in the weighted interference model.

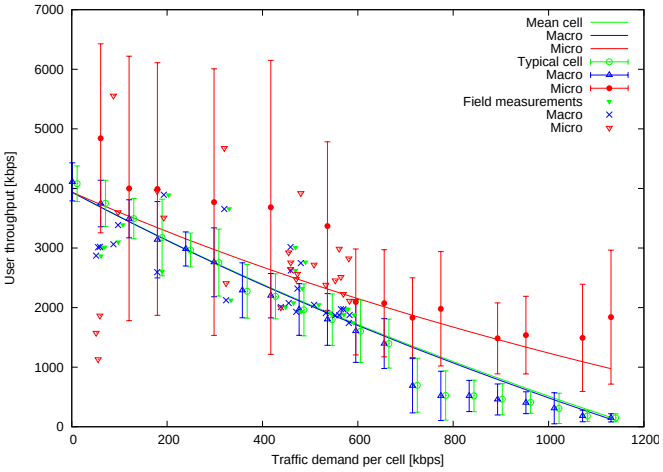


Figure 3.23: Mean user throughput in the network versus traffic demand per cell in the weighted interference model.

### 3.7.5 Spatial distribution of QoS parameters averaged over many cells in the network

In Section 3.7.3 we demonstrated the results for estimation of the QoS metrics as function of the traffic demand. Also, we assumed that all base stations transmit at the same power. In this section we are interested in the spatial distribution of cell load, number of users and mean user throughput considering that different base stations transmit at different powers. We consider only the weighted interference case. Each BS $v$ is characterized by a transmitting power $P_v \in \mathbb{R}_+^*$. We shall assume that $P_v$ are *i.i.d.* (independent and identically distributed) *marks* of the point process $\Phi$ (i.e., given $\Phi$, the transmitting powers $P_1, P_2, \ldots$ are i.i.d. random variable with some fixed distribution).

The received power at location $y \in \mathbb{R}^2$ from BS $u$ equals now

$$L_u^{-1}(y) = \frac{P_u \mathbf{S}_u (y-u)}{\ell (y-u)} \tag{3.63}$$

The cell is now given by the follwing formula:

$$V(u) = \left\{ y \in \mathbb{R}^2 : L_u(y) \leq \min_{v \in \Phi \setminus \{u\}} L_v(y) \right\} \tag{3.64}$$

The SINR formula given in (3.35) is now

$$\mathrm{SINR}(y, \Phi) := \frac{P_u \mathbf{S}_u (y-u) / l(|y-u|)}{\mathcal{N} + \sum\limits_{v \in \Phi \setminus \{u\}} P_v \mathbf{S}_v (y-v) p(v) / l(|y-v|)} \tag{3.65}$$

for a user at position $y$ served by the base station $u$ as a base station offering him the strongest signal. Considering the different transmitting powers does not change the model itself developed in Section 3.5 and consequently we obtain the same system of cell-load equations as presented in Section 3.5.3. Solving this system again we obtain the cell loads of all cells in a considered area (network). It is obvious from the results in Section 3.5.2 and Theorem 1 that solving the (3.36) one can obtain the cell load and further deduce the number of users and the mean user throughput for all cells in the network. In such a way we obtain the cumulative distribution function of aforementioned parameters. The ultimate goal is to compare the results to the real-field measurements.

#### Real-field measurements

Now we describe the real-field measurements. The raw data are collected using a specialized tool which is used by operational engineers for network maintenance. This tool measures several parameters for every base station 24 hours a day. In particular, one can get the cell load, traffic demand, number of users, mean user throughput for each cell in each hour. We have also the BS coordinates which permits to estimate the intensity $\gamma$ of BS per unit area.

We choose one hour during the day and estimate the corresponding empirical CDF of the QoS parameters.

#### Numerical setup for simulation

The numerical setup is the same as in Section 3.7.3 except that the transmitting power is not constant over the network. We assume that the transmitting power $P_v$ has a log-normal distribution of logarithmic-standard deviation $\sigma_P$. In order to justify this model, we give the empirical

CDF of transmitting powers in dB estimated from measurements in the operational network in Figure 3.24. This figure shows that this CDF may be approximated by a normal distribution with standard deviation $\sigma_P = 5.3$dB. The mean transmitting power of each BS including a global antenna gain equals $E[P_n] = 60$dBm.

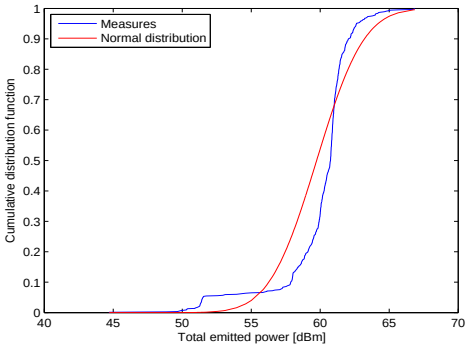


Figure 3.24: CDF of BS powers in the operational network in the downtown of a big city (blue) and normal distribution approximation (red).

With the constant power described in Section 3.5.1, each BS transmits at a constant power $\tilde{P}_n = 60$dBm and the shadowing (3.5.3) has a log-normal distribution of standard deviation

$$\sigma_{\tilde{S}} = \sqrt{\sigma_S^2 + \sigma_P^2} \simeq 9.6\text{dB}$$

The other sources of irregularities, as for example non-uniform traffic demand, are not considered. Consequently, we will consider networks or parts of a network where we can assume uniform spatial traffic demand (e.g. downtown of a big city or a typical rural area).

| **MEASURES** | mean | standard deviation |
|---|---|---|
| cell load | 0.1854 | 0.1337 |
| mean number of users | 0.2608 | 0.2511 |
| mean user throughput[kbit/s] | 2135 | 698 |
| **SIMULATIONS** | mean | standard deviation |
| cell load | 0.1845 | 0.1059 |
| mean number of users | 0.2531 | 0.2147 |
| mean user throughput[kbit/s] | 2054 | 425 |

Table 3.1: Mean and standard deviation of spatial distribution of QoS metrics for the downtown of a big city

### Results

Figures 3.25, 3.26 and 3.27 show the spatial distribution (across different cells) of the cell load, mean number of users per cell and the mean user throughput in the network deployed in

| **MEASURES** | mean | standard deviation |
|---|---|---|
| cell load | 0.1190 | 0.1035 |
| mean number of users | 0.1530 | 0.1734 |
| mean user throughput[kbit/s] | 1975 | 733 |
| **SIMULATIONS** | mean | standard deviation |
| cell load | 0.1321 | 0.0937 |
| mean number of users | 0.1774 | 0.2861 |
| mean user throughput[kbit/s] | 2051 | 601 |

Table 3.2: Mean and standard deviation of spatial distribution of QoS metrics for the mid-size city

the downtown of a big city. Recall that these metrics represent, themselves, the steady-state (averaged over time) performance characteristics of individual cells.

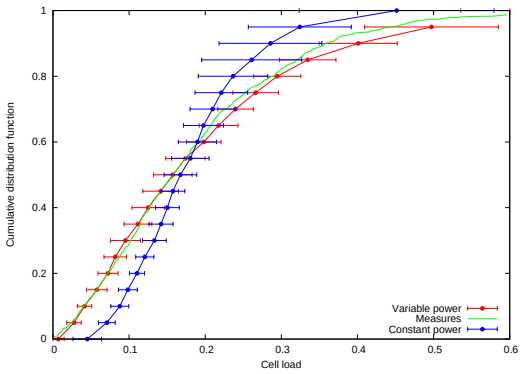


Figure 3.25: CDF of cell load for the downtown of a big city obtained either from the variable power model, from real-field measurements, or from the model where the transmitted powers are assumed constant.

Analogous characteristics regarding the network in a mid-size city are presented in Figures 3.28, 3.29 and 3.30. Tables 3.1 and 3.2 show means and standard deviations of these spatial distributions.

All figures and tables present the distributions estimates in our model as well as the real-field measurements. For sake of comparison, we present also in the figures the results obtained in the model described in Section3.5.1 where the transmitted powers are assumed constant. The simulation curves represent the means over ten repeated network simulations, with the horizontal bars giving the standard deviation of this averaging. In what follows we discuss the presented results in more detail.

The estimated network density and the traffic demand in the downtown of the big city are, respectively, $\gamma = 4.62\text{km}^{-2}$ and $\rho = 483\text{kbit/s/cell}$. Analogous values for the mid-size city are $\gamma = 1.27\text{km}^{-2}$ and $\rho = 284\text{kbit/s/cell}$. Note that in the latter scenario the traffic demand is smaller, but the network is less dense and also less regular (cf Figure 3.16). We use these values as input parameters for our model.

In general we see a good agreement between real field measures and the model analysis with

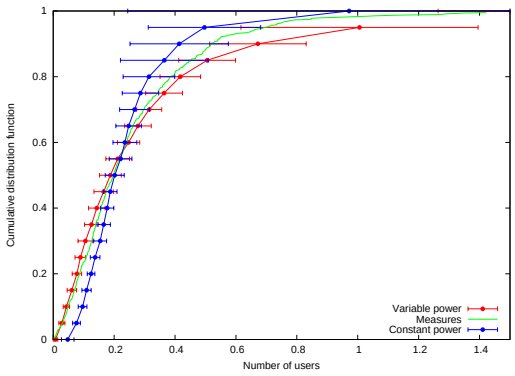


Figure 3.26: CDF of the mean users number for the downtown of a big city.

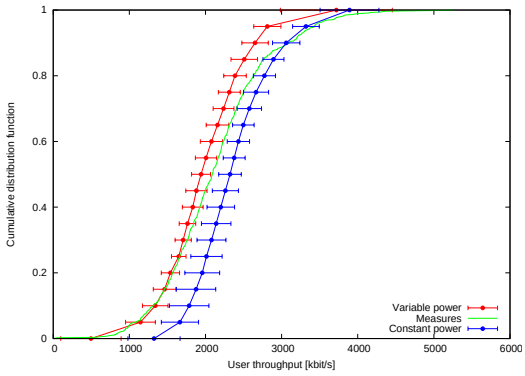


Figure 3.27: CDF of the throughput for the downtown of a big city.

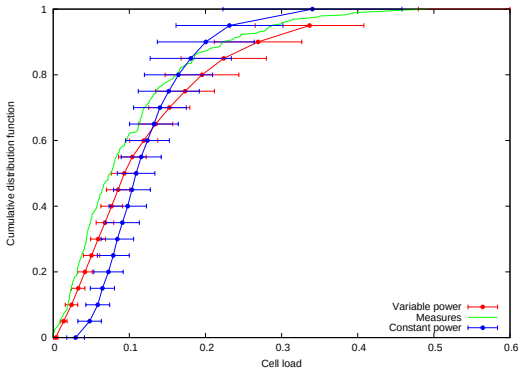


Figure 3.28: CDF of cell load for a mid-size city obtained either from the variable power model, from real-field measurements, or from the model where the transmitted powers are assumed constant.

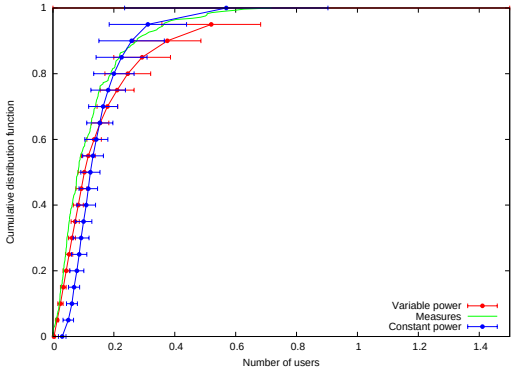


Figure 3.29: CDF of the mean number of users for the mid-size city.

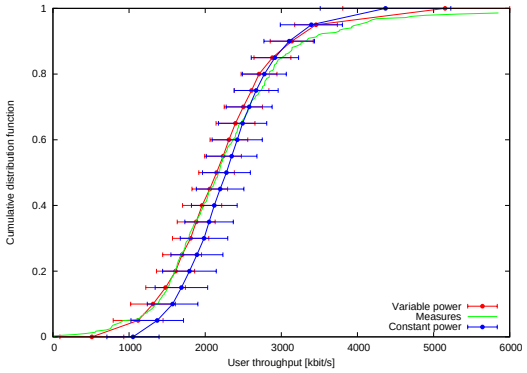


Figure 3.30: CDF of the throughput for the mid-size city.

randomized transmitted power. Under the constant power assumption the model predicts well the median of the cell load and the mean number of users but fails to match the spatial distribution of these characteristics. Clearly, the spatial variability of power creates more spatial heterogeneity of these characteristics in the network. Regarding the mean user throughput the constant power assumption fails to predict even the median. Extensions of the model, e.g. letting it account for further sources of disparity in the deployed networks (e.g. different heights of antennas) could perhaps improve the accuracy of prediction.

# Appendix

## 3.A Proof of Proposition 4 in the Markovian case

Assume that the transmitted volumes are *exponentially distributed*. In this particular case, the process $\{X(t); t \geq 0\}$ describing the number of users of the different classes is a continuous-time Markov process with discrete state space $\mathbb{N}^{\mathbb{D}}$ and admits the following generator

$$\begin{cases} q(x, x+\varepsilon_j) = \lambda_j, & x \in \mathbb{N}^{\mathbb{D}} \\ q(x, x-\varepsilon_j) = \mu_j R_j \frac{x_j}{x_{\mathbb{D}}}, & x \in \mathbb{N}^{\mathbb{D}}, x_j > 0, \end{cases} \tag{3.66}$$

where $\varepsilon_j$ designates the vector of $\mathbb{N}^{\mathbb{D}}$ having coordinate 1 at position $j$ and 0 elsewhere and $x_{\mathbb{D}} := \sum_{j\in\mathbb{D}} x_j$ the total number of users in the queue. It is easy to see that the process $\{X(t); t \geq 0\}$ is regular [27, p.337] and irreducible [27, p.357] and that it admits as invariant measure

$$\alpha(x) = x_{\mathbb{D}}! \prod_{j\in\mathbb{D}} \frac{(\rho'_j)^{x_j}}{x_j!}, \quad x \in \mathbb{N}^{\mathbb{D}}, \tag{3.67}$$

where $\rho'_j := \lambda_j / (\mu_j R_j) = \rho_j / R_j$. If $\rho' := \sum_{j=1}^{J} \rho'_j < 1$ then $\sum_{x\in\mathbb{N}^{\mathbb{D}}} \alpha(x) = \frac{1}{1-\rho'}$, indeed

$$\begin{aligned} \frac{1}{1-\rho'} &= \sum_{n=0}^{\infty} \rho'^n \\ &= \sum_{n=0}^{\infty} \left( \sum_{j\in\mathbb{D}} \rho'_j \right)^n \\ &= \sum_{n=0}^{\infty} \sum_{x\in\mathbb{N}^{\mathbb{D}}: x_{\mathbb{D}}=n} n! \prod_{j\in\mathbb{D}} \frac{\rho'^{x_j}_j}{x_j!} \\ &= \sum_{x\in\mathbb{N}^{\mathbb{D}}} x_{\mathbb{D}}! \prod_{j\in\mathbb{D}} \frac{(\rho'_j)^{x_j}}{x_j!} = \sum_{x\in\mathbb{N}^{\mathbb{D}}} \alpha(x) \end{aligned}$$

We deduce that if $\rho' < 1$ then the process $\{X(t); t \geq 0\}$ admits $\pi = (1-\rho)\alpha$ as invariant distribution; and hence this process is t-positive recurrent [27, p.357]. We deduce from (3.67) that the invariant distribution is

$$\pi(x) = (1-\rho') x_{\mathbb{D}}! \prod_{j\in\mathbb{D}} \frac{(\rho'_j)^{x_j}}{x_j!}, \quad x \in \mathbb{N}^{\mathbb{D}}$$

Let $X = (X_1, X_2, \ldots, X_J)$ be the vector counting the number of users of each class in the steady state, and let $X_{\mathbb{D}} := \sum_{j\in\mathbb{D}} X_j$ be the total number of users in the queue. The vector $X$ has $\pi$ as

distribution, then, for $n \in \mathbb{N}$,

$$\begin{aligned} P\left(X_{\mathbb{D}} = n\right) &= \sum_{x \in \mathbb{N}^{\mathbb{D}} : x_{\mathbb{D}} = n} \pi\left(x\right) \\ &= \left(1 - \rho'\right) \sum_{x \in \mathbb{N}^{\mathbb{D}} : x_{\mathbb{D}} = n} n! \prod_{j \in \mathbb{D}} \frac{\left(\rho'_j\right)^{x_j}}{x_j!} \\ &= \left(1 - \rho'\right) \rho'^n, \end{aligned}$$

which is the geometric distribution on $\mathbb{N}$ with parameter $1 - \rho' = 1 - \rho/\rho_{\mathrm{c}}$ where $\rho_{\mathrm{c}}$ is given by (3.4). The mean number of users is

$$\bar{N} := E\left[X_{\mathbb{D}}\right] = \frac{\rho'}{1 - \rho'} = \frac{\rho}{\rho_{\mathrm{c}} - \rho}$$

From Little's formula [11] the expected delay, denoted $\bar{T}$, equals

$$\bar{T} = \frac{E\left[X_{\mathbb{D}}\right]}{\lambda} = \frac{\rho}{\left(\rho_{\mathrm{c}} - \rho\right)\lambda}$$

In the steady state the queue throughput equals the traffic demand $\rho$. The throughput per user is *defined* as the ratio of the above queue throughput by the average number of users; that is

$$\bar{r} = \frac{\rho}{E\left[X_{\mathbb{D}}\right]} = \rho_{\mathrm{c}} - \rho$$

For a given class $j \in \mathbb{D}$,

$$\begin{aligned} \bar{N}_j &:= E\left[X_j\right] \\ &= \sum_x x_j \pi\left(x\right) \\ &= \sum_{x : x_j \neq 0} \left(1 - \rho'\right) x_{\mathbb{D}}! x_j \prod_{i \in \mathbb{D}} \frac{\left(\rho'_i\right)^{x_i}}{x_i!} \\ &= \sum_{x'} \left(1 - \rho'\right) \left(x'_{\mathbb{D}} + 1\right) x'_{\mathbb{D}}! \prod_{i \in \mathbb{D}} \frac{\left(\rho'_i\right)^{x_i}}{x'_i!} \\ &= \sum_{x'} \left(x'_{\mathbb{D}} + 1\right) \pi\left(x'\right) \\ &= \rho'_j E\left[X_{\mathbb{D}} + 1\right] = \frac{\rho_j}{\left(1 - \frac{\rho}{\rho_{\mathrm{c}}}\right) R_j} \end{aligned}$$

where for the fourth equality we introduce the vector $x'$ related to $x$ as follows

$$x'_i = \begin{cases} x_i & i \neq j \\ x_i - 1 & i = j \end{cases}$$

From Little's formula the expected delay, denoted $\bar{T}_j$, equals

$$\bar{T}_j = \frac{E\left[X_j\right]}{\lambda_j} = \frac{1}{\left(1 - \frac{\rho}{\rho_{\mathrm{c}}}\right) R_j \mu_j}$$

The expected throughput of class $j$, denoted $\bar{r}_j$, is the average required volume $\mu_j^{-1}$ divided by the expected delay, that is

$$\bar{r}_j = \frac{\mu_j^{-1}}{\bar{T}_j} = \left(1 - \frac{\rho}{\rho_c}\right) R_j$$

# Chapter 4

# Quality of service in real-time streaming

## 4.1 Introduction

Wireless cellular networks offer nowadays possibility to watch TV on mobile devices, which is an example of the real-time content streaming. This type of traffic demand is expected to increase significantly in the future. In order to cope with this process, network operators need to implement in their dimensioning tools efficient methods allowing to predict the quality of this type of service. The quality of real-time streaming (RTS) is principally related to the number and duration of *outage incidents* — (hopefully short) periods when the network cannot deliver to a given user in real-time the requested content of the required quality. In this Chapter we propose a stochastic model allowing for an analytic evaluation of such metrics. It assumes a traffic demand with different radio conditions of calls, and can be specified to take into account the parameters of a given wireless cellular technology. We develop expressions for several important performance characteristics of this model, including the mean time spent in outage and the mean number of outage incidents for a typical streaming call as function of its radio conditions. These expressions involve only stationary probabilities of the (free) traffic demand process, which is a vector of independent Poisson random variables describing the number of users in different radio conditions.

We use this model to analyze RTS in a typical cell of a 3GPP Long Term Evolution (LTE) cellular network assuming orthogonal intra-cell user channels with the peak bit-rates (achievable when there are no other users in the same cell) close to the theoretical Shannon bound in the additive white Gaussian noise (AWGN) channel, with the extra-cell interference treated as noise. These assumptions lead to a radio resource constraint in a multi-rate linear form. Namely, each user experiencing a given signal-to-(extra-cell)-interference-and-noise ratio (SINR) requires a fixed fraction of the normalized radio capacity, related to the ratio between its requested and peak bit-rates. All users of a given configuration (experiencing different SINR values) can be entirely satisfied if and only if the total required capacity is not larger than one.[1]

In the above context of a multi-rate linear radio resource constraint, we analyze some natural

[1]Recall that in the case of voice calls and, more generally, constant bit-rate (CBR) calls the multi-rate linear form of the resource constraints has already proved to lead to efficient model evaluation methods, via e.g. Kaufman-Roberts algorithm [65, 81]. Despite some fundamental similarities to CBR service, the RTS gives rise to a new model, due to the fact that the service denials are not definitive for a given call, but have a form of temporal interruptions (outage) periods.

parametric class of *least-effort-served-first* (LESF) service policies, which assign service to users in order of their increasing radio capacity demand, until the full capacity (possibly with some margin) is reached. The capacity margin may be used to offer some "lower quality" service to users temporarily in outage thus realizing some type of fairness with respect to unequal user radio-channel conditions. This class contains an optimal and a fair policy, the latter being suggested by LTE implementations.

In order to evaluate explicitly the quality of service metrics induced by the LESF policies, we relate the mean time spent in outage and the mean number of outage incidents for a typical streaming call in given radio conditions to the distribution functions of some linear functionals of the Poisson vector describing the steady state of the system. We calculate the Fourier transforms of these functions and use a well-known Fourier transform inversion method to obtain numerical values of the quantities of interest. We also study the mean throughput during a typical streaming call evaluating the expectations of the corresponding non-linear functionals of the Poisson vector describing the steady state of the system via the Monte Carlo method.

Using this approach, we present a thorough study of the quality of RTS with LESF policies in the aforementioned Markovian setting. For completeness we present also some pure-simulation results illustrating the impact of a non Poisson-arrival assumption.

## 4.2 Related work

Let us now recollect a few *related works* on the performance evaluation of cellular networks. In the early 80's, wireless cellular networks were carrying essentially voice calls, which require constant bit-rates (CBR) and are subject to admission control policies with blocking (at the arrival epoch) to guarantee these rates for calls already in service. An important amount of work has been done to propose efficient call admission policies [83, 100, 102]. Policies with admission conditions in the multi-rate linear form have been considered e.g. in [10, 42, 59].

Progressively, cellular networks started carrying also calls with variable bit-rates (VBR), used to transmit data files. The available resources are (fairly) shared between such calls and when the traffic demand increases, the file transfer delays increase as well, but (in principle) no call is ever blocked. These delays may be evaluated analytically using multi-rate linear resource constraint in conjunction with multi-class processor sharing models; cf e.g. [25, 59].

Recently, users may access multimedia streaming services through their mobile devices [45]. They are provided via CBR connections, essentially without admission control, but they tolerate temporary interruptions, when network congestion occurs. One may distinguish two types of streaming traffic. In *real-time streaming (RTS)* (as e.g. in mobile TV), considered in this thesis, the portions of the streaming content emitted during the time when the transmission to a given user is interrupted (is in outage) are definitely lost for him (unless a "secondary", lower-rate streaming is provided during these periods). In *non-real-time streaming (NRTS)* (like e.g., video-on-demand, YouTube, Dailymotion, etc), a user starts playing back the requested multimedia content after some initial delay, required to deliver and buffer on the user device some initial portion of it. If further transmission is interrupted for some time making the user buffer content drop to zero (buffer starvation) then the play-back is stopped until some new required portion of the content is delivered. Several papers study the effect of the variability of the wireless channel on the performance of a single streaming call; see for e.g. [68], [75]. In [82] VBR transmissions and RTS are considered jointly in some analytical model, however the number and duration of outage periods are not evaluated. In [99] the tradeoff between the start-up delay and the probability of buffer starvation is analyzed in a Markovian queuing framework for NRTS streaming.

We do not consider any cell-load balancing; see [14] for some recent work on this problem in

the video streaming context. Also, [67, 98] consider some admission control policies to guarantee non-dropping of multimedia calls due to caller impatience and/or handoffs.

## 4.3 Streaming in wireless cellular networks

In this section we present a new stochastic model of RTS in cellular networks.

### 4.3.1 System assumptions

We consider the following scenario of multi-user streaming in a cellular network.

#### Network layer

Geographically distributed users wish to obtain down-link wireless streaming of some (typically video) content, contacting base stations of a network at random times, for random durations, requesting some fixed streaming bit-rates. We consider a uni-cast traffic (as opposed to the broadcast or multi-cast case), i.e.; the content is delivered to all users via private connections. Different classes of users (calls) need to be distinguished, regarding their radio channel conditions, requested streaming bit-rates and mean streaming times. Each user chooses one base station, the one with the smallest path-loss, independently of the configuration of users served by this base station. Thus, we do not consider any load-balancing policy.

#### Data layer — streaming policies

If a given base station cannot serve all the users present at a given time, it temporarily stops streaming the requested content at the requested rate to users of some classes, according to some given policy (to be described), which is supposed to preserve a maximal subset of served users. We call these (classes of) users with the requested bit-rate temporarily denied *in outage*. The users in outage will not receive the part of the content which is transmitted during their outage times (this is the principle of the RTS). We will also consider policies, which offer some "best-effort" streaming bit-rates for some classes of users in outage, thus allowing for example to keep receiving the requested content but of a lower quality. Users, which are (temporarily) denied even this lower quality of service are called *in deep outage*.

#### Medium access

In this thesis we assume that users are connected to the serving antennas via orthogonal single-input-single-output (SISO) channels allowing for a peak-rate close to the theoretical Shannon bound in the additive white Gaussian noise (AWGN) model, with the (extra-cell) interference treated as noise.[2] We will also comment on how to model multiple-input-multiple-output (MIMO) and broadcast channels.

#### Physical layer

The quality of channel of a given user depends on the path-loss of the signal with respect to its serving base station, a constant noise, and the interference from other (non-serving) base stations. These three components determine its signal-to-interference-and-noise ratio (SINR).

[2]Orthogonality of channels is an appropriate assumption for current LTE (Long Term Evolution) norm for cellular networks based on OFDMA, as well as for other multiple access techniques as FDA, TDMA, CDMA assuming perfect in-cell orthogonality, and even HDR neglecting the scheduler gain.

Both path-loss from the serving station and interference account for the distance and random propagation effects (shadowing). Our main motivation for considering a multi-class model is to distinguish users with different SINR values. In other words, even if we assume that all users require the same streaming times and rates, we still need a multi-class model due to (typically) different SINR's values of users in wireless cellular networks.

**Performance characteristics**

We will present and analytically evaluate performance of some (realistic) streaming policies in the context described above. We will be particularly interested in the following characteristics:

- fraction of time spent in outage and in deep outage during the typical call of a given class,
- number of outage incidents occurring during this call,
- mean throughput (average bit-rate) during such call, accounting for the requested bit-rates and for the "best-effort" bit-rate obtained during the outage periods.

### 4.3.2 Model description

In what follows we describe a mathematical model of the RTS that is an incarnation of a new, more general, stochastic service model with capacity sharing and interruptions presented and analyzed in Appendix 4.4.2. This is a single server model which allows to study the performance of one tagged base station of a multi-cellular network satisfying the above system assumptions. More details on how this model fits the multi-cell scenario will be presented in Section 4.4.

**Traffic demand**

Consider $J \geq 1$ *classes of calls* (or, equivalently, users) characterized by different *requested streaming bit-rates* $r_k$, *wireless channel conditions* described by the signal-to-(extra-cell)-interference-and-noise ratio $\mathrm{SINR}_k$ with respect to the serving base-station [3] and *mean requested streaming times* $1/\mu_k$, $k = 1, \ldots, J$.

We assume that calls of class $k \in \{1, \ldots, J\}$ arrive in time according to a Poisson process with intensity $\lambda_k > 0$ (number of call arrivals per unit of time, per base station) and stay in the system (keep requesting streaming) for independent times, having some *general distribution* with mean $1/\mu_k < \infty$. [4] Different classes of calls are independent from each other. We denote by $X_k(t)$ the number of calls of a given class requesting streaming from a given BS at time $t$; see Section 4.1.1 in the Appendix for a formal definitions of these variables in terms of arrival process and service times. Let $\mathbf{X}(t) = (X_1(t), \ldots, X_J(t))$; we call it the (vector of) user configuration at time $t$. The stationary distribution $\pi$ of $\mathbf{X}(t)$ coincides with the distribution of the vector $(X_1, \ldots, X_J)$ of independent Poisson random variables with means $\mathbf{E}[X_k] := \rho_k = \lambda_k/\mu_k$, $k = 1, 2, \ldots, J$. We call $\rho_k$ the *traffic demand* (per base station) of class $k$.

**Wireless resource constraints**

Users are supposed to be offered the requested streaming rates for the whole requested streaming times. However, due to limited wireless resources, for some configuration of users $\mathbf{X}(t)$, the requested streaming rates $\mathbf{r} = (r_1, \ldots, r_J)$ may not be achievable. Following the assumption of

[3] In this thesis the interference is always caused only by non-serving base stations.

[4] All the results presented in this Chapter do not depend on the particular choice of the streaming time distributions. This property is often referred to in the queuing context as the insensitivity property.

orthogonal AWGN SISO wireless channels (with the (extra-cell) interference treated as noise) available for users of a given base station, we assume that the requested rates are achievable for all calls present at time $t$ if

$$X_k(t)r_k = \nu_k R_k, \quad k = 1, \ldots, J, \tag{4.1}$$

for some non-negative vector $(\nu_1, \ldots, \nu_J)$, such that $\sum_{k=1}^{J} \nu_k \leq 1$, where

$$R_k = cW \log(1 + \mathrm{SINR}_k) \tag{4.2}$$

is the maximal (peak) bit-rate of a user of class $k$, whose channel conditions are characterized by $\mathrm{SINR}_k$. (The rate $R_k$ is available to a user of class $k$ if it is the only user served by the base station.) Here $W$ is the frequency bandwidth and $c$ (with $0 < c \leq 1$) is a coefficient capturing how close a given coding scheme approaches the theoretical Shannon bound (corresponding to $c = 1$); cf [34, Th .9.1.1]. [5] Note that the assumption (4.1) corresponds to the situation, when users neither hamper nor assist each other's transmission. They use channels which are perfectly separated in time, frequency or by orthogonal codes, nevertheless sharing these resources. [6]

We can interpret the ratio between the requested and maximal bit-rates $\varphi_k = r_k/R_k$ as the *resource demand* of a user of class $k$. Note that the configuration of users $\mathbf{X}(t)$ can be entirely served if and only if the total resource demand satisfies the constraint

$$\sum_{k=1}^{J} \varphi_k X_k(t) \leq 1\,. \tag{4.3}$$

This is a *multi-rate linear resource constraint.*

### Service policy

If the requested streaming rates are not achievable for a given configuration of users $\mathbf{X}(t)$ present at time $t$, then some classes of users will be temporarily put in outage at time $t$, meaning that they will receive some smaller bit-rates (whose values are not guaranteed and may depend on the configuration $\mathbf{X}(t)$). These smaller, "best-effort" bit-rates may drop to 0, in which case we say that users are in deep-outage. Let us recall that the times at which users are in outage and deep outage do *not* alter the original streaming times; i.e. the streaming content is not buffered, nor delayed during the outage periods.

We will define now a parametric family of service polices for which *classes with smaller resource demands have higher service priority.* In this regard, in the remaining part of this Section we assume (without loss of generality) that the resource demands of users from different classes are ordered $\varphi_1 < \varphi_2 < \ldots < \varphi_J$.

[5] It was also shown in [61] that the performance of AWGN *multiple input multiple output* (MIMO) channel can be approximated by taking values of $\gamma \geq 1$. Another possibility to consider MIMO channel is to use the exact capacity formula given in [90].

[6] From an information theory point of view, the orthogonality assumption is not optimal. In fact, the theoretically optimal performance is offered by the *broadcast channel* model. It is known that in the case of AWGN broadcast channel the rates $\boldsymbol{r}$ are (theoretically) achievable for the configuration $\boldsymbol{X}$ if (and only if) there exists a vector $(\nu_1, \ldots, \nu_J)$, such that $\sum_{k=1}^{J} \nu_k \leq 1$ and

$$X_k r_k = W \log\left(1 + \frac{\nu_k}{1/\mathrm{SINR}_k + \sum_{i=1}^{k-1} \nu_i}\right) \quad k = 1, \ldots, J,$$

where the classes of users are numbered such that $\mathrm{SINR}_1 \geq \mathrm{SINR}_2 \geq \ldots \geq \mathrm{SINR}_J$; cf [92, Eq. 6.29].

**Least-effort-served-first policy** For a given configuration of users $\mathbf{X} = \mathbf{X}(t)$ requesting streaming at time $t$, *least-effort-served-first policy with $\delta$-margin* (*LESF($\delta$)* for short) attributes the requested bit-rates to all users in classes $k = 1, \ldots, K$, where

$$K = K^{\delta}(\mathbf{X}) = \max\{k \in \{1, \ldots, J\} : \sum_{j=1}^{k-1} \varphi_j X_j + \varphi_k \sum_{j=k}^{J} X_j \mathbf{1}(\varphi_j \leq \varphi_k(1+\delta)) \leq 1\}, \tag{4.4}$$

where $\mathbf{1}_A(x) = 1$ is the indicator function of set $A$ and $\delta$ is a constant satisfying $0 \leq \delta \leq \infty$.

**Remark 27** *The LESF(0) policy is* optimal *in the following sense: given constraint (4.3) and the assumption that the classes with smaller resource demands have higher priority, this policy allows to serve the maximal subset of users present in the system. For the same reason any LESF($\delta$) policy with $\delta > 0$ is clearly sub-optimal. In order to explain the motivation for considering such policies, one needs to extend the model and explain what actually happens with classes of users which experience outage. In this regard, note that $C = \sum_{j=1}^{K} \varphi_j X_j \leq 1$ is the actual fraction of the server capacity consumed by the users which are not in outage. The remaining server capacity $1 - C$ (which is not needed to serve users in classes $1, \ldots, K$) can be used to offer some "lower quality" service (e.g. streaming with lower video resolution, etc) to the users in classes $K+1, \ldots, J$ which are in outage. Note by (4.4) that the remaining server capacity under the policy LESF($\delta$) is at least*

$$1 - C \geq \varphi_K \sum_{j=K+1}^{J} X_j \mathbf{1}(\varphi_j \leq \varphi_K(1+\delta)).$$

*Hence, the server accepting the class $K$ as the least-priority class being "fully" served, leaves enough remaining capacity to be able to make the same effort (allocate service capacity $\varphi_K$) for all users in outage in classes whose service demand exceeds $\varphi_K$ by no more than $\delta \times 100\%$. These latter users will not have "full" required service (since this requires more resources, $\varphi_j > \varphi_K$, for the full service) but only some "lower quality" service (to be specified in what follows). Consequently, one can conclude that policies LESF($\delta$) with $\delta > 0$, being sub-optimal, ensure some* fairness, *in the sense explained above. Clearly the policy LESF($\infty$) (i.e., with $\delta = \infty$) is the most fair, in the sense that it reserves enough remaining capacity to offer the "lower quality" service for* all *users in outage (no deep outage). Thus, we will call LESF($\infty$) the* LESF fair *policy.*

**Best-effort service for users in outage** We will specify now a natural model for the "best-effort" streaming bit-rates that can be offered for users in outage in association with a given LESF($\delta$) policy. For $k > K = K^{\delta}(\mathbf{X})$ denote

$$r'_k = r'^{\delta}_k(\mathbf{X}) = R_k \frac{1 - \sum_{j=1}^{K} X_j \varphi_j}{\sum_{j=K+1}^{J} X_j \mathbf{1}(\varphi_j \leq (1+\delta)\varphi_K)} \tag{4.5}$$

if $\varphi_k \leq (1+\delta)\varphi_K$ and 0 otherwise.

The rates $(r_1, \ldots, r_K, r'_{K+1}, \ldots, r'_J)$ are achievable for the configuration $\mathbf{X}$ under resource constraint (4.3). Note that users in classes $j$ such that $\varphi_j > (1+\delta)\varphi_K$ do not receive any positive bit-rate. We say, they are in *deep outage*. Finally, we remark that the service (4.5) is "resource fair" among users in outage but not in deep outage.

**Performance metrics**

The configuration of users $\mathbf{X}(t)$ evolves in time, it changes at arrival and departure times of users. At each arrival or departure epoch the base station applies the outage policy to the new configuration of users to decide which classes of users receive requested streaming rates and which are in outage (or deep outage).

Let us introduce the following characteristics of the *typical call (user)* of class $k = 1, \ldots, J$.

- $\mathcal{P}_k$ denotes the *probability of outage at the arrival epoch for class k*. This is the probability that the typical call of this class is put in outage immediately at its arrival epoch.
- $D_k$ denotes the *mean total time spent in outage during the typical call of class k*.
- $M_k$ denotes the *mean number of outage incidents experienced during the typical call of class k*.

More formal definitions of these characteristics, as well as other *system* characteristics (as e.g. the intensity of outage incidents) are given in the Appendix. We also introduce two further characteristics related to the mean *throughput* obtained during the typical call of class $k = 1, \ldots, J$.

- Denote by $\Upsilon_k$ the *mean throughput during the typical call of class k*. This is the mean bit-rate obtained during such a call, taking into account the bit-rate $r_k$ when the call is not in outage and the best-effort bit rate $r'_k$ obtained during the outage periods, averaged over the call duration.
- Let $\Upsilon'_k$ be the *part of the throughput obtained during the outage periods of the typical call of class k*. This is the mean best-effort bit-rate of such call averaged over outage periods.

### 4.3.3 Model evaluation

**Results**

We will show how the performance metrics regarding outage incidents and duration, introduced in Section 4.3.2, can be expressed using probability distribution functions of some *linear* functionals of the random vector $X_1, \ldots, X_J$ of independent Poisson random variables with parameters $\rho_j$, respectively. Recall that these random variables correspond to the number of calls of different classes present in the stationary regime of our streaming model.

Specifically, for given $\delta > 0$, $k = 1, \ldots, J$ and $t \geq 0$ denote

$$F_k^{\delta}(t) := \Pr\left\{ \sum_{j=1}^{k} X_j^{\delta,k} \varphi_j \leq t \right\}, \tag{4.6}$$

where $X_j^{\delta,k} = X_j$ for $j = 1, \ldots, k-1$ and $X_k^{\delta,k} = \sum_{j=k}^{J} X_j \mathbf{1}(\varphi_j \leq \varphi_k(1+\delta))$.

The following results follow from the analysis of a more general model presented in the Appendix.

**Proposition 11** *The probability of outage at the arrival epoch for user of class k is equal to*

$$\mathcal{P}_k = 1 - F_k^{\delta}(1 - \varphi_k) \qquad k = 1, \ldots, J\,. \tag{4.7}$$

*The mean total time spent in outage during the typical call of class $k$ is equal to*

$$D_k = \frac{\mathcal{P}_k}{\mu_k} = \frac{1 - F_k^\delta(1-\varphi_k)}{\mu_k} \qquad k = 1, \ldots, J\,. \tag{4.8}$$

*The mean number of outage incidents experienced during the typical call of class $k$ (after its arrival) is equal to*

$$M_k = \frac{1}{\mu_k} \sum_{j=1}^{J} \lambda_j \left( F_k^\delta(1-\varphi_k) - F_k^\delta(1-\varphi_k-\varphi_j) \right) \;\; k = 1, \ldots, J\,. \tag{4.9}$$

**Proof.** Note first that the functions $F_k^\delta(t)$ defined in (4.6) allow one to represent the stationary probability that the configuration of users is in a state in which the LESF($\delta$) policy serves users of class $k$

$$F_k^\delta(1) = \Pr\left\{ \sum_{j=1}^{k} X_j^{\delta,k} \varphi_j \leq 1 \right\}\,.$$

In the general model described in the Appendix we denote this state by $\mathcal{F}_k$ and its probability by $\pi(\mathcal{F}_k)$. Thus $\pi(\mathcal{F}_k) = F_k^\delta(1)$. Moreover,

$$1 - F_k^\delta(1-\varphi_k) = \Pr\left\{ \sum_{j=1}^{k} X_j^{\delta,k} \varphi_j > 1 - \varphi_k \right\}$$

is the probability that the steady state configuration of users appended with one user of class $k$ is in the complement $\mathcal{F}'_k$ of the state $\mathcal{F}_k$, i.e., all users of class $k$ are in outage (meaning $k > K^\delta(\mathbf{X}')$, where $\mathbf{X}' = (X_1, \ldots, X_k + 1, \ldots, X_J)$). Thus the expression (4.7) follows from Proposition 13. Similarly (4.8) follows from Proposition 14 and (4.9) follows from Proposition 15. ■

Regarding the throughput characteristics, we have the following result.

**Proposition 12** *The mean throughput during the typical call of class $k$ is equal to*

$$\Upsilon_k = r_k(1 - \mathcal{P}_k) + \Upsilon'_k = r_k F_k^\delta(1-\varphi_k) + \Upsilon'_k\,,$$

*where*

$$\begin{aligned} \Upsilon'_k = \mathbf{E}\Big[ & r_k'^{\delta}(X_1, \ldots, X_k + 1, \ldots, X_J) \\ & \mathbf{1}\left(K^\delta(X_1, \ldots, X_k + 1, \ldots, X_J) < k\right)\Big]\,, \end{aligned} \tag{4.10}$$

*with the best-effort rate $r'_k(\cdot)$ given by (4.5) and the least-priority class $K^\delta(\cdot)$ begin served by the LESF($\delta$) policy given by (4.4), is the part of the throughput obtained during the outage periods.*

*Proof* of this proposition is given in the Appendix.

**Remark 28** *Recall from (4.5) that the variable rates $r'_k$ are obtained by the user of class $k$ when he is in outage, i.e., $k > K$. They are non-null, $r'_k > 0$, only if $\varphi_k \leq (1+\delta)\varphi_K$. In the case of equal requested rates $r_k$, the intersection of the two conditions $0 < r'_k$ and $k > K$ is equivalent to*

$$(1 + SINR_K)^{1/(1+\delta)} - 1 \leq SINR_k \leq SINR_K\,. \tag{4.11}$$

**Remarks on numerical evaluation**

In order to be able to use the expressions given in (11) we need to evaluate the distribution functions $F_k^\delta(t)$. In what follows we show how this can be done using Laplace transforms. Regarding the throughput in outage $\Upsilon'_k$, expressed in (4.10) as the expectation of a *non-linear* functional of the vector $(X_1, \ldots, X_J)$, we will use Monte Carlo simulations to obtain numerical values for this expectation.

Denote by $\mathcal{L}_k^\delta(\theta) := \int_0^\infty e^{-\theta s} F_k^\delta(s)\mathrm{d}s$ the Laplace transform of the function $F_k^\delta(t)$.

**Fact 2** *We have*

$$\mathcal{L}_k^\delta(\theta) = \frac{1}{\theta} \exp\left[\sum_{j=1}^{k} \rho_j^{\delta,k} \left(e^{-\theta\varphi_j} - 1\right)\right],$$

*where* $\rho_j^{\delta,k} = \rho_j$ *for* $j = 1, \ldots, k-1$ *and* $\rho_k^{\delta,k} = \sum_{j=k}^{J} \rho_j \mathbf{1}(\varphi_j \le \varphi_k(1+\delta))$.

**Proof.** Note that for given $\delta > 0$, $k = 1, \ldots, J$ the random variables $X_1^{\delta,k}, \ldots, X_k^{\delta,k}$ are independent, of Poisson distribution, with parameters $\rho_1^{\delta,k}, \ldots, \rho_k^{\delta,k}$, respectively. The result follows from [9, Proposition 1.2.2] and a general relation $\int_0^\infty e^{-\theta s} F(s)\,\mathrm{d}s = \frac{1}{\theta}\int_0^\infty e^{-\theta s} F(\mathrm{d}s)$. ■

The probabilities $F_k^\delta(\cdot)$ may be retrieved from $\mathcal{L}_k^\delta(\cdot)$ using standard techniques. For example with the algorithm implemented by [55] in Matlab [38]. In what follows we present a more explicit result based on the Bromwich contour inversion integral. In this regard, denote $\overline{\mathcal{L}}_k^\delta(\theta) = 1/\theta - \mathcal{L}_k^\delta(\theta)$ (which is the Laplace transform of complementary distribution function $1 - F_k^\delta(t)$). Also, denote by $\mathcal{R}(z)$ the real part of the complex number $z$.

**Fact 3** *We have*

$$F_k^\delta(t) = 1 - \frac{2e^{at}}{\pi} \int_0^\infty \mathcal{R}\left(\overline{\mathcal{L}}_k^\delta(a+iu)\right) \cos ut\, du\,, \tag{4.12}$$

*where* $a > 0$ *is an arbitrary constant.*

**Proof.** See [6]. ■

**Remark 29** *As shown in [6], the integral in (4.12) can be numerically evaluated using the trapezoidal rule, with the parameter a allowing to control the approximation error. Specifically, for* $n = 0, 1, \ldots$ *define*

$$h_n(t) = h_n(t; a, k, \delta) := \frac{(-1)^n e^{a/2}}{t} \mathcal{R}\left(\overline{\mathcal{L}}_k^\delta\left(\frac{a + 2n\pi i}{2t}\right)\right),$$

$S_n(t) := \frac{h_0(t)}{2} + \sum_{i=1}^{n} h_i(t)$, *and* $S(t) = \lim_{n\to\infty} S_n(t)$. *Then* $\left|F_k^\delta(t) - (1 - S(t))\right| \le e^{-a}$. *Finally, the (alternating) infinite series* $S(t)$ *can be efficiently approximated using for example the Euler summation rule*

$$S(t) \approx \sum_{i=0}^{M} \binom{M}{i} 2^{-M} S_{N+i}(t)$$

*with a typical choice* $N = 15$, $M = 11$.

**Remark 30** *The expression (4.9) for the mean number of outage incidents involves a sum of a potentially big number of terms* $F_k^\delta(1-\varphi_k) - F_k^\delta(1-\varphi_k-\varphi_j)$, $j = 1, \ldots, J$, *which are typically small, and which are evaluated via the inversion of the Laplace transform. Consequently the sum*

*may accumulate precision errors. In order to avoid this problem we propose another numerical approach for calculating $M_k$. It consists in representing $M_k$ equivalently to (4.9) as*

$$M_k = \frac{F_k^{\delta}(1-\varphi_k)}{\mu_k} \sum_{j=1}^{J} \lambda_j b_k(j) \qquad k = 1, \ldots, J \tag{4.13}$$

*where*

$$b_k(j) = \frac{F_k^{\delta}(1-\varphi_k) - F_k^{\delta}(1-\varphi_k-\varphi_j)}{F_k^{\delta}(1-\varphi_k)} \tag{4.14}$$

*Let $k$ and $\delta$ be fixed. Recall the definition of $F_k^{\delta}(t)$ in (4.6) and note that the expression (4.14) may be written as*

$$b_k(j) = \frac{\Pr(X \in \mathcal{F}, X + \varepsilon_j \notin \mathcal{F})}{\Pr(X \in \mathcal{F})}$$

*where $\mathcal{F} = \mathcal{F}(k) = \left\{ X \in \mathbb{R}^J : \sum_{j=1}^{k} X_j^{\delta,k} \varphi_j \leq 1 - \varphi_k \right\}$. The above expression may be seen as the blocking probability for class $j$ in a classical multi-class Erlang loss system with the admission condition $X \in \mathcal{F}$. Consequently, $b_k(\cdot)$ may be calculated by using the* Kaufman-Roberts algorithm *[65, 81] and plugged into (4.13). Note that by doing this we still need to calculate $F_k^{\delta}(1-\varphi_k)$ however avoid summing of $O$ differences of these functions as in (4.9).*

## 4.4 Quality of real-time streaming in LTE

In this section we will use the model developed in Section 4.3 to evaluate the quality of RTS in LTE symmetric networks. This single-server (base station) model will be used to study the performance of one tagged base station of a multi-cellular network under the following assumptions:

- We assume a regular hexagonal lattice of base stations on a torus. This allows us to consider the tagged base station of the network as a typical one.
- Homogeneous (in space and time) Poisson arrivals on the torus are marked by i.i.d. (across users and base stations) variables representing their shadowing with respect to different base stations. These variables, together with independent user locations determine their serving (strongest) base stations. A consequence of the independence of users locations and shadowing variables is that the arrivals served by the tagged base station form an independent thinning of the total Poisson arrival process to the torus and thus a Poisson process too. Uniform distribution of user locations and identical distribution of the their shadowing variables imply that the intensity of the arrival process to the tagged base station is equal to the total arrival intensity to the torus divided by the number of stations. Moreover, the distribution of the SINR of the typical user of the tagged base station coincides with the distribution of the typical user of the whole network.
- The intensity of arrivals of some particular (SINR)-class to the tagged base station is equal to the total intensity of arrivals to the tagged cell times the probability of the random SINR of the typical user being in the SINR-interval corresponding to this class.
- We consider the "full interference" scenario, i.e. all base stations transmit the signal at the constant power, regardless of the number of users they serve (this number can be zero). This makes the interference, and hence the service rates, of users of a given base station independent of the service of other base stations (decouples the service processes of different base stations).

### 4.4.1 LTE model and traffic specification

#### SINR distribution

Recall that the main motivation for considering a multi-class model was the necessity to distinguish users with different radio conditions, related to different values of the SINR they have with respect to the serving base stations. In order to choose representative values of SINR in a given network and to know what fraction of users experience a given value, we need to know the *(spatial) distribution of the SINR* (with respect to the serving base station) experienced in this network (possibly biased by the spatial repartition of arrivals of streaming calls). This distribution can be obtained from real-network measurements, simulations or analytic evaluation of an appropriate spatial, stochastic model.[7] In this Section we will use the distribution of SINR obtained from the simulation compliant with the 3GPP recommendation in the so-called calibration case (to be explained in what follows). At present, assume simply, that we are given a cumulative distribution function (CDF) of the SINR expressed in dB, $F(x) := \Pr\{10\log_{10}(\text{SINR}) \leq x\}$, obtained from either of these methods. In other words, $F(x)$ represents the fraction of mobile users in the given network which experience the SINR (expressed in dB) not larger than $x$.

Consider a discrete probability mass function

$$p_k := F\left(\frac{x_{k+1}+x_k}{2}\right) - F\left(\frac{x_k+x_{k-1}}{2}\right) \qquad k = 1, 2, \ldots, J\,, \tag{4.15}$$

with $x_0 = -\infty$, $x_{J+1} = \infty$. We define the class $k = 1, \ldots, J$ of users as all users having the SINR expressed in dB in the interval $\left(\frac{(x_k+x_{k-1})}{2}, \frac{(x_{k+1}+x_k)}{2}\right)$, and approximate their SINR by the common value $\text{SINR}_k = 10^{x_k/10}$. Clearly $p_k$ is the fraction of mobile users in the given network which experience the SINR close to $\text{SINR}_k$. Hence, in the case of homogeneous streaming traffic (the same requested streaming rates and mean streaming times, which will be our default assumption in the numerical examples) we can assume the intensity of arrivals $\lambda_k$ of users of class $k$ to be equal to $\lambda_k = p_k\lambda$ where $\lambda = \sum_{i=k}^{J} \lambda_k$ is the total arrival intensity (per unit of time per serving base station) to be specified together with the CDF $F$ of the SINR.

**CDF of the SINR for 3GPP recommendation** We obtain the CDF $F$ of the SINR from the simulation compliant with the 3GPP recommendation in the so-called calibration case, (compare to [3]). More precisely, we consider the geometric pattern of BS placed on the $6 \times 6$ hexagonal lattice. In the middle of each hexagon there are three symmetrically oriented BS antennas, which gives a total of 108 BS antennas. The distance between the centers of two neighboring hexagons is 0.5 km. Each BS antenna is characterized by the following horizontal pattern $A(\phi) = -\min(12(\phi/\zeta)^2, A_m)$, where $\phi$ is the angle in degrees, with $\zeta = 70°$, $A_m = 20dB$, and uses transmission power $P = 60$dBm (including omnidirectional gain of 14dBi). The distance-loss model (corresponding to the frequency carrier 2GHz) is $L(r) = 128.1 + 37.6 \times \log_{10}(r)$[dB] where $r$ is the distance in km. A supplementary penetration loss of 20dB is added. The shadowing is modeled as a log-normal random variable of mean one and logarithmic standard deviation of deviation 8dB, cf [18]. The noise power equals $-95$dBm (which corresponds to a system bandwidth of 10MHz, a noise floor of $-174$dBm/Hz and a noise figure of 9dB). In order to obtain the empirical CDF of the SINR we generate 3600 random user locations uniformly in the network (100 user locations per hexagon on average). Each user is connected to the antenna with the strongest received signal (smallest propagation-loss including distance, shadowing and

[7] For this latter possibility, we refer the reader to a recent paper on Poisson modeling of real cellular networks subject to shadowing [20], as well as to [39], completed in [19], where the distribution of the the SINR in Poisson networks is evaluated explicitly.

antenna pattern) and the SINR is calculated. The obtained empirical CDF $F$ of the SINR is shown in Figure 4.1.

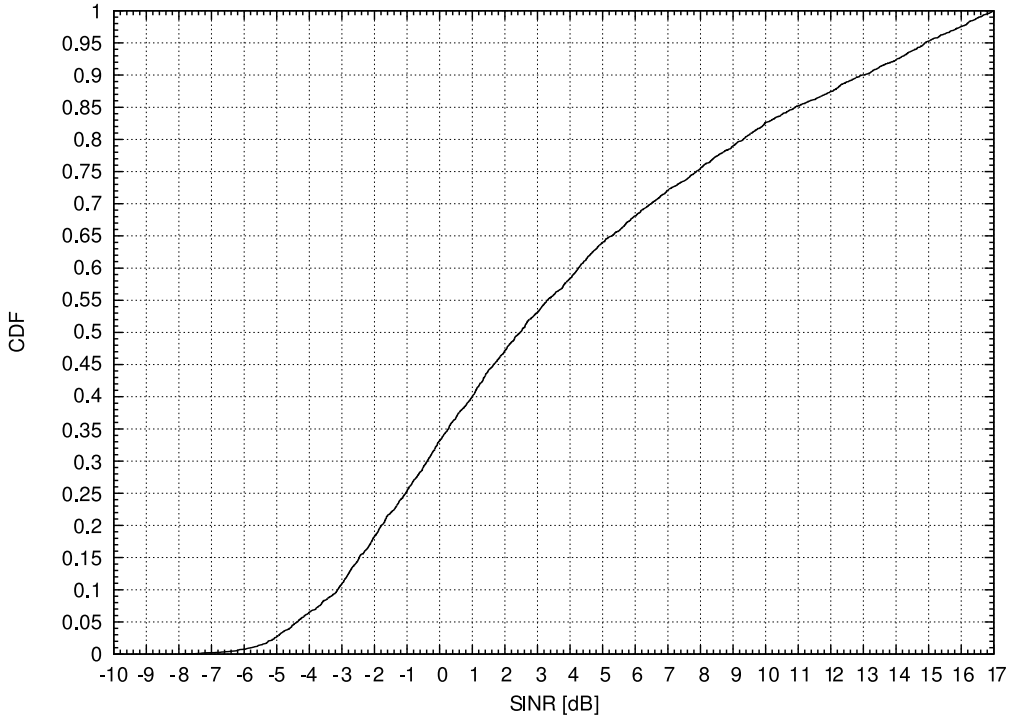


Figure 4.1: Cumulative distribution function of the SINR obtained according to 3GPP specification; see Section 4.4.1. An abrupt transition of the CDF to 1 at SINR = 17dB is due to the cell sectorization: each mobile is interfered by each of the two antennas co-located with its serving antenna on the same site (and serving the different sectors) with the power equal to at least 1% of the power received from the serving BS. Therefore the signal to interference ratio is at most $0.5 \times 10^{-2} = 17$dB.

**Link characteristics**

3GPP shows in [4] that there is a 25% gap between the practical coding schemes and the Shannon limit for the AWGN channel. Moreover, some of the transmitted bits are used for signaling, which induces a supplementary capacity loss of about 30% (see [2]). This made us assume $c = 0.5 (\approx 0.75(1 - 0.3))$ in (4.2). The system bandwidth is $W = 10$MHz.

**Streaming traffic**

We assume that all calls require the same streaming rate $r_k = 256\,\text{kbit/s}$ and have the same streaming call time distribution. We split them into $J = 100$ user classes characterized by values of the SINR falling into different intervals regularly approximating the SINR domain from $x_1 = -10$dB to $x_J = 17$dB as explained in Section 4.4.1. In our performance evaluation we will consider two values of the spatially uniform traffic demand: 900 and 600 Erlang/km$^2$. (All results presented in what follows do not depend on the mean streaming time but only on

the traffic demand). Consequently, the $k$ th class traffic demand per unit of area is equal to, respectively, $p_k \times 900$ and $p_k \times 600\,\text{Erlang/km}^2$, where $p_k$ are given by (4.15). Multiplying by the area served by one base station equal to $\sqrt{3} \cdot (0.5\,\text{km})^2/6 \approx 0.0722\,\text{km}^2$ we obtain the traffic demand per cell, per class, equal to $\rho_k = p_k \times 900 \times 0.0722 \approx p_k \times 64.9$ Erlang and $\rho_k = p_k \times 600 \times 0.0722 \approx p_k \times 43.3$ Erlang, respectively, for the two studied scenarios.

### 4.4.2 Performance evaluation

Assuming the LTE and traffic model described above, we consider now streaming policies LESF($\delta$) defined in Section 4.3.2. Recall that in doing so, we assume that users are served by the antenna offering the smallest path-loss, and dispose orthogonal down-link channels, with the maximal rates $R_k$ depending on the value of the SINR (interference comes from non-serving BS) characterizing class $k$. Roughly speaking, LESF($\delta$) policy assigns the total requested streaming rate $r_k = 256$kbit/s for the maximal possible subset of classes in the order of decreasing SINR, leaving some capacity margin to offer some "best-effort" streaming rates for (some) users remaining in outage. These streaming rates $r'_k$ given by (4.5) depend on the current configuration of users and are non-zero for users with SINR within the interval $(1 + \text{SINR}_K)^{1/(1+\delta)} - 1 \leq \text{SINR} \leq \text{SINR}_K$, where $\text{SINR}_K$ is the minimal value of SINR for which users are assigned the total requested streaming rate; cf Remark 28. In particular, LESF(0), called the *optimal* policy, leaves no capacity margin for users in outage, while LESF($\infty$), called the *fair* one, offers a "best-effort" streaming rate for all users in outage at the price of assigning the full requested rate 256kbit/s to a smaller number of classes (higher value of the $\text{SINR}_K$) [8]. In what follows, we use our results of Section 4.3.3 to evaluate performance of these streaming policies in the LTE network model.

#### Outage time

Figure 4.2 shows the mean time of the streaming call spent in outage normalized by call duration, $\mu_k D_k$, evaluated using (4.8), as function of the SINR value characterizing class $k$, for the traffic 900 Erlang/km$^2$ and different policies LESF($\delta$). Figure 4.3 shows the analogous results assuming a traffic load of 600 Erlang/km$^2$. The main observations are as follows:

- All LESF policies exhibit a cut-off behaviour: the fraction of time in outage drops rapidly from 100% to 0% when the SINR exceeds some critical values. This cut-off is more strict for the optimal policy.
- For the traffic of 900 Erlang/km$^2$, users with the SINR$\geq$ 3dB are practically never in outage, when the optimal policy is used. The same holds true for users with SINR$\geq$ 13dB, when the fair policy is used.
- When the traffic drops to 600 Erlang/km$^2$, these critical values of SINR decrease by 2dB and 5dB, respectively, for the optimal and the fair policy. Note that the fair policy is more sensitive to higher traffic load.

#### Number of outage incidents

Figure 4.4 shows the mean number of outage incidents per streaming call, $M_k$ evaluated using (4.9), as function of the SINR value characterizing class $k$, for the traffic 900 Erlang/km$^2$ and different policies LESF($\delta$). (Recall that we assume the same streaming time distribution for all users, and hence $\lambda_j/\mu_k = \rho_j$ making the expression in (4.9) depend only on the vector of

[8] The LESF fair policy seems to be adopted in some implementations of the LTE.

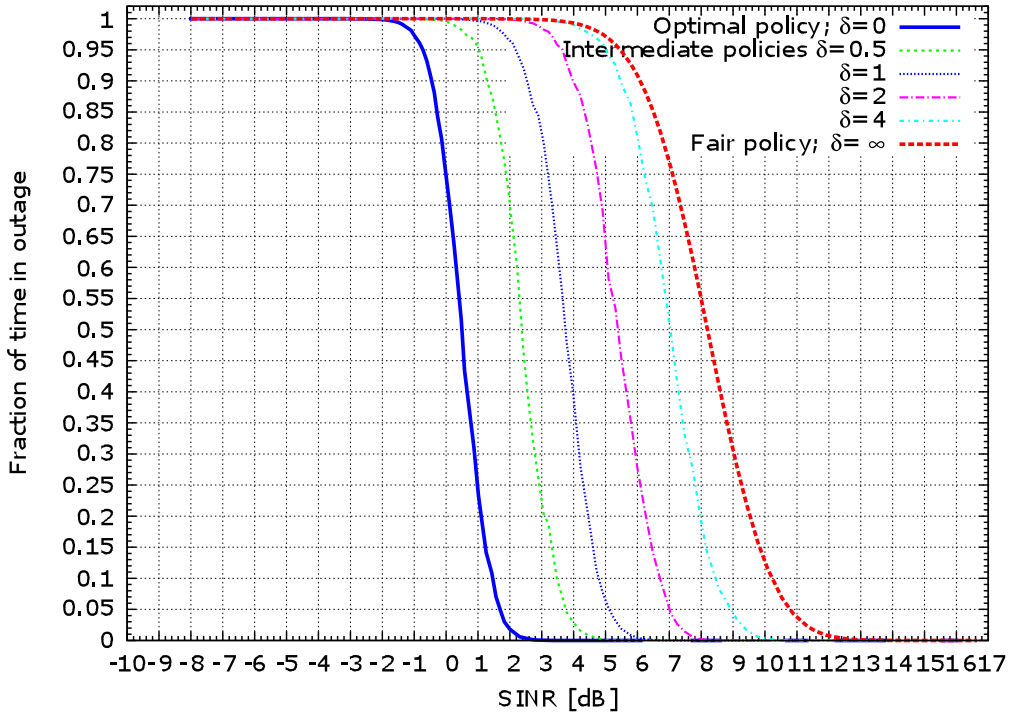


Figure 4.2: Mean fraction of the requested streaming time in outage, as function of the user SINR for different policies LESF($\delta$); traffic 900 Erlang/km$^2$.

traffic demand per class.) Figure 4.5 shows the analogous results assuming a traffic demand of 600 Erlang/km$^2$. The main observations are as follows:

- For all policies, the number of outage incidents (during the service) is non-zero only for users with the SINR close to the critical values revealed by the analysis of the outage times. Users with SINR below these values are constantly in outage while users with SINR above them are never in outage.
- More fair policies generate slightly more outage incidents. The worst values are 2 to 2.2 interruptions per call for the optimal policy, depending on the traffic value, and 2.4 to 3 interruptions per call for the fair policy.

Studying outage times and outage incidents we do not see apparent reasons for considering fair policies. This motivates our study of the best-effort service in outage.

### The role of the "best effort" service

Figure 4.6 shows the fraction of time spent in deep outage as function of the SINR, assuming traffic 900 Erlang/km$^2$. These values should be compared to the fraction of time spent in outage (for convenience copied in Figure 4.6 from Figure 4.2). Recall, users in outage do not receive the full requested streaming rate (assumed 256kbit/s in our example), however they do receive some non-null "best effort" rates given by (4.5), unless they are in deep outage — have SINR too small; cf Remark 28. Considering users in outage but not in deep outage as "partially satisfied",

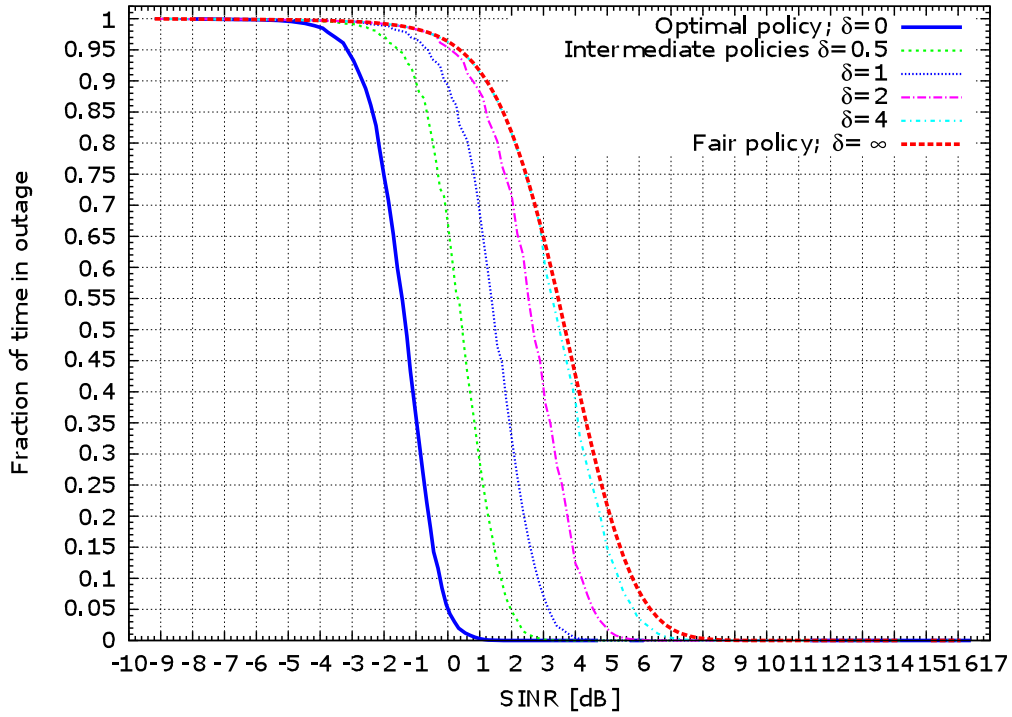


Figure 4.3: Fraction of time in outage as on Figure 4.2 for traffic 600 Erlang/km$^2$.

increasing fairness margin $\delta$ allows to (at least) partially satisfy users with decreasing SINR values. Obviously the level of the "partial satisfaction" depends on the throughput obtained in outage periods, which is our quantity of interest in Figure 4.7. It shows also two curves for all policies LESF($\delta$) assuming traffic 900 Erlang/km$^2$. The upper ones represent the mean total throughput realized during the service, normalized to its maximal value; i.e., $\Upsilon_k/(256\text{kbit/s})$, in function of the SINR value characterizing class $k$. The fractions of this throughput realized during outage periods, $\Upsilon'_k/(256\text{kbit/s})$, are represented by the lower curves.

Figures 4.7 and 4.6 teach us that the role of the LESF($\delta$) policies with $\delta > 0$ may be two-fold.

- LESF($\delta$) policies with small values of $\delta$, e.g. $\delta = 0.5$, *improve "temporal homogeneity" of service with respect to the optimal policy, for users having SINR near the critical value.* For example, a user having SINR equal to 1dB is served by the optimal policy during 80% of the time with the full requested streaming rate (cf. Figure 4.6). However, for the remaining 20% of the time it does not receive any service (deep outage, rate 0bits/s). The policy LESF(0.5) offers to such a user 80% of the requested streaming rate during the whole streaming time (cf. Figure 4.7), with no deep outage periods (cf. Figure 4.6). The price for this is that a slightly higher SINR is required to receive the full requested streaming rate (at least 5dB, instead of 3dB for the optimal policy).

- The fair policy LESF($\infty$) *improves the spatial homogeneity of service.* It leaves no user in deep outage, however a much larger SINR= 13dB is required for not to be in outage (cf. Figure 4.6). Moreover, the throughput of all users in outage but not in deep outage is substantially reduced e.g. from 80% to 40% for SINR= 1dB, with respect to some

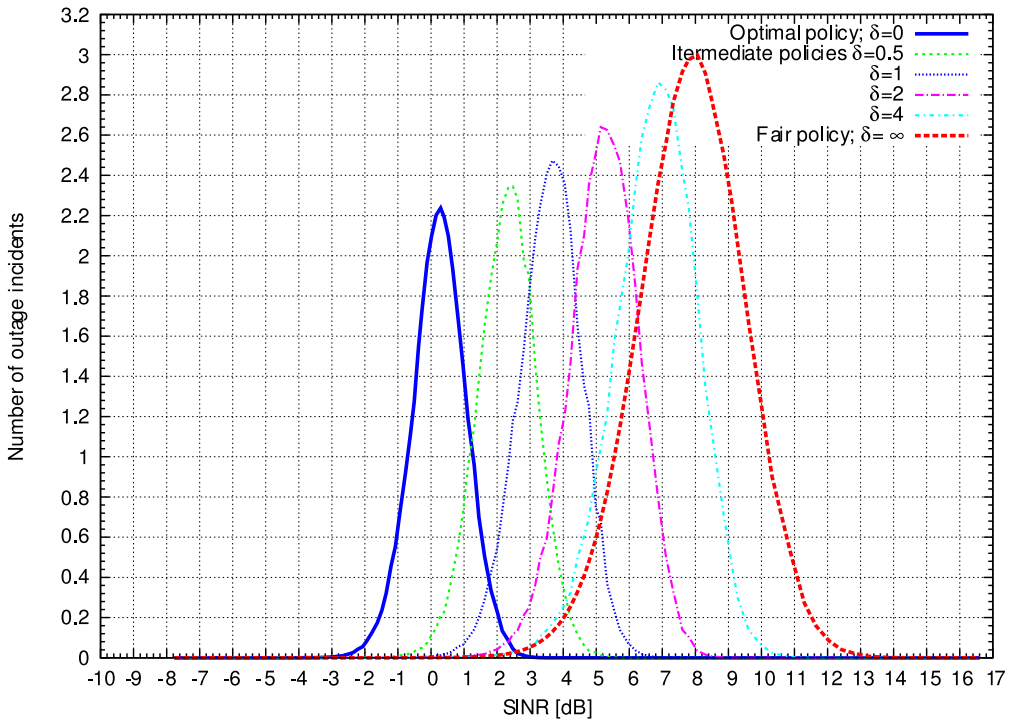


Figure 4.4: Number of outage incidents during the requested streaming time, as function of the user SINR for different policies LESF($\delta$); traffic 900 Erlang/km$^2$.

intermediate LESF($\delta$) policies (with $0 < \delta < \infty$). These intermediate policies can offer an interesting compromise between the optimality and fairness.

### Impact of non-Poisson-arrivals

Recall that the performance analysis of the model presented in this Chapter is insensitive to the distribution of the requested streaming times. In this section we will briefly study the impact of a non Poisson-arrival assumption. In this regard we simulate the dynamics of the model with deterministic inter-arrival times (with all other model assumptions as before) and estimate the mean fraction of time in outage $\mu_k D_k$ and mean number of outage incidents $M_k$ for each class $k$. For the comparison, as well as for the validation of the theoretical work, we perform also the simulation of the model with Poisson arrivals. The results are plotted in Figures 4.8, 4.9 and 4.10, 4.11. Observe first that the simulations of the Poisson model confirm the results of the theoretical analysis. Regarding the impact of the deterministic inter-arrival times a (somewhat expected) fact is that the optimal policy remains optimal regarding the fraction of time spent in the outage and the number of outage incidents. Another, less evident, observation is that the deterministic inter-arrivals (more regular than in the Poisson case) do *not* improve the situation for *all* classes of users. In fact, users with small values of the SINR have a smaller fraction of time in outage under Poisson arrival assumption than in the deterministic one! This is different from what we can observe for the blocking probability for the classical Erlang's loss model; cf e.g. [101, Figure 8]. Moreover, the deterministic arrivals increase the number of outage incidents

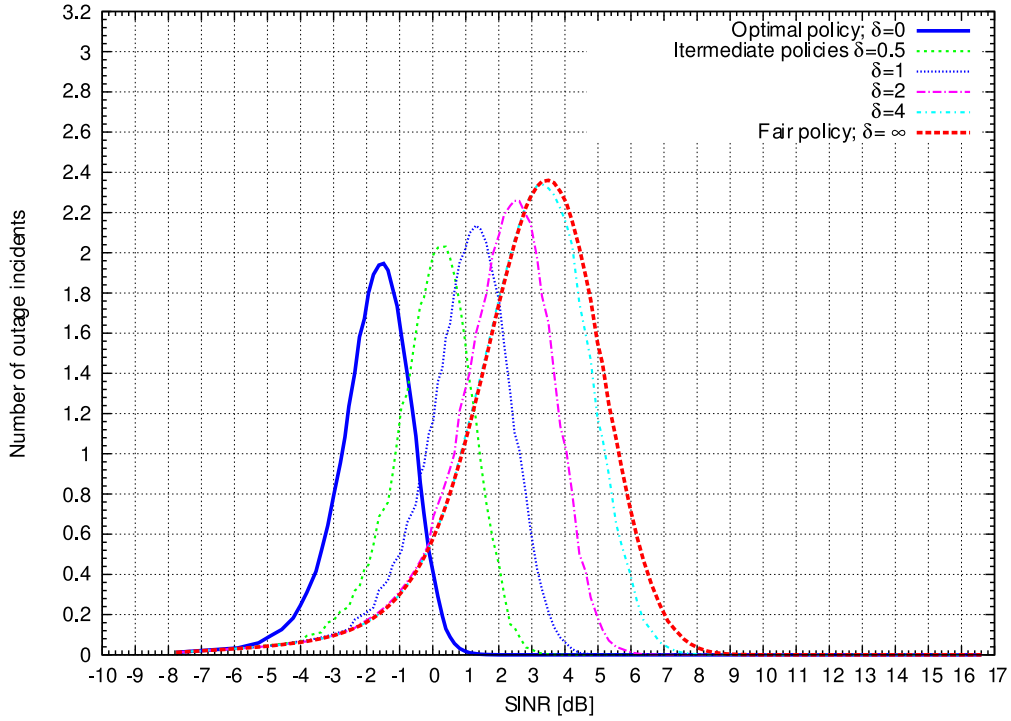


Figure 4.5: Number of outage incidents as in Figure 4.4 for traffic 600 Erlang/km$^2$.

for intermediate values of the SINR and decrease for extreme ones, especially with the fair policy. Concluding these observations one can say however, that the differences between Poisson and deterministic are not very significant and hence the Poisson model can be used to approximate a more realistic traffic model.

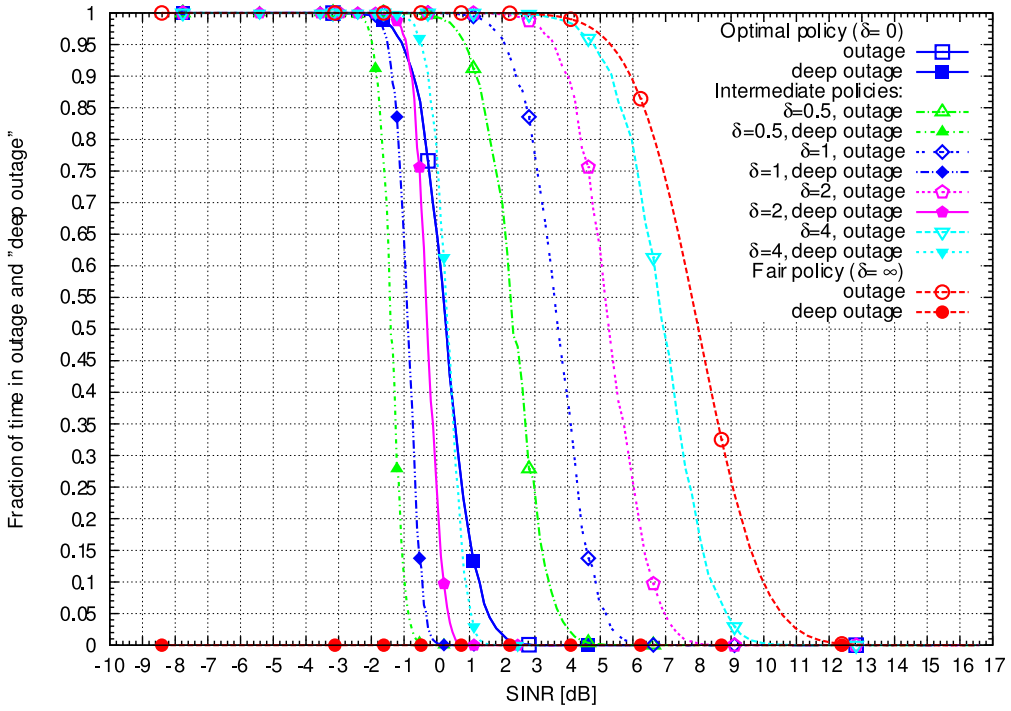


Figure 4.6: Deep outage versus outage time. For any policy LESF($\delta$), with $0 < \delta < \infty$, the left curve of a given style represents the fraction of time spent in deep outage. The right curve of a given style recalls the fraction of time spent in outage (already plotted on Figure 4.2). The optimal policy ($\delta = 0$) does not offer any "best effort" service. The fair policy ($\delta = \infty$) offers this service for all users in outage.

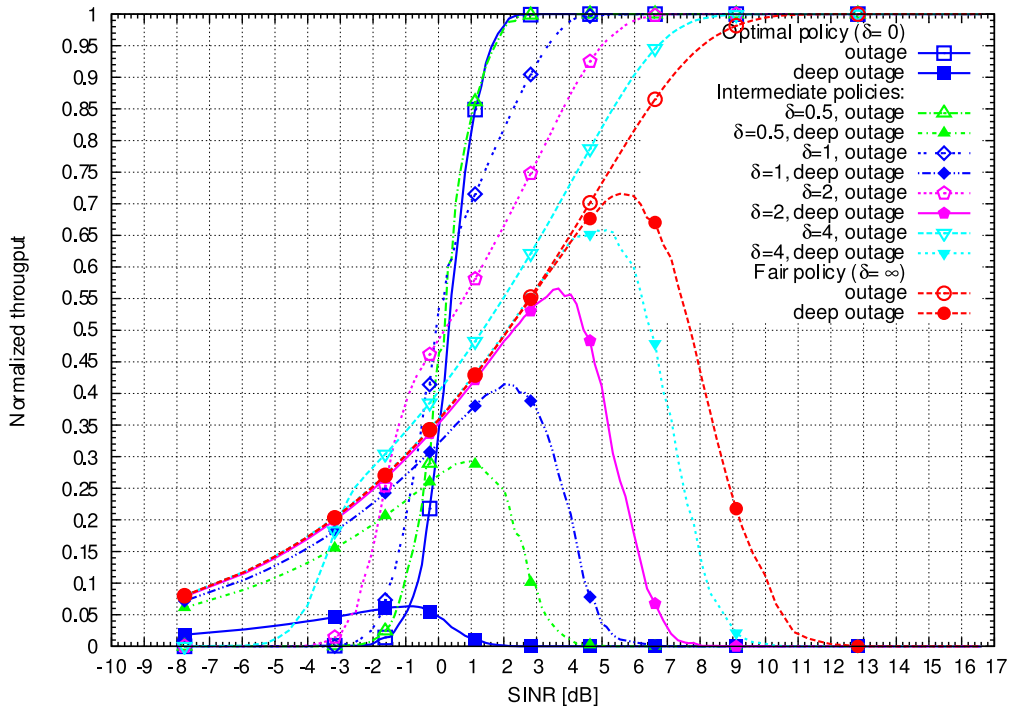


Figure 4.7: Mean total throughput normalized to its maximal value 256kbit/s obtained during the service time (upper curves) and its fraction obtained when a user is in outage (lower curves) for different policies LESF($\delta$) traffic 900 Erlang/km$^2$.

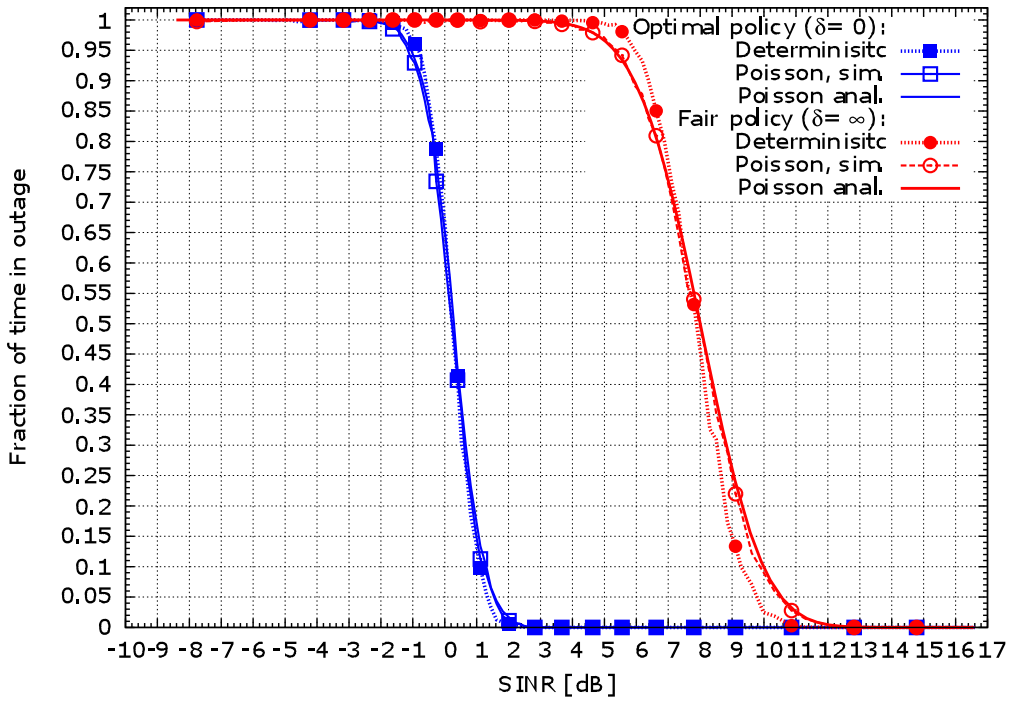


Figure 4.8: Impact of the deterministic arrival process (as compared to the Poisson one) on the mean fraction of the requested streaming time in outage, for the optimal and fair policy; traffic 900 Erlang/km$^2$.

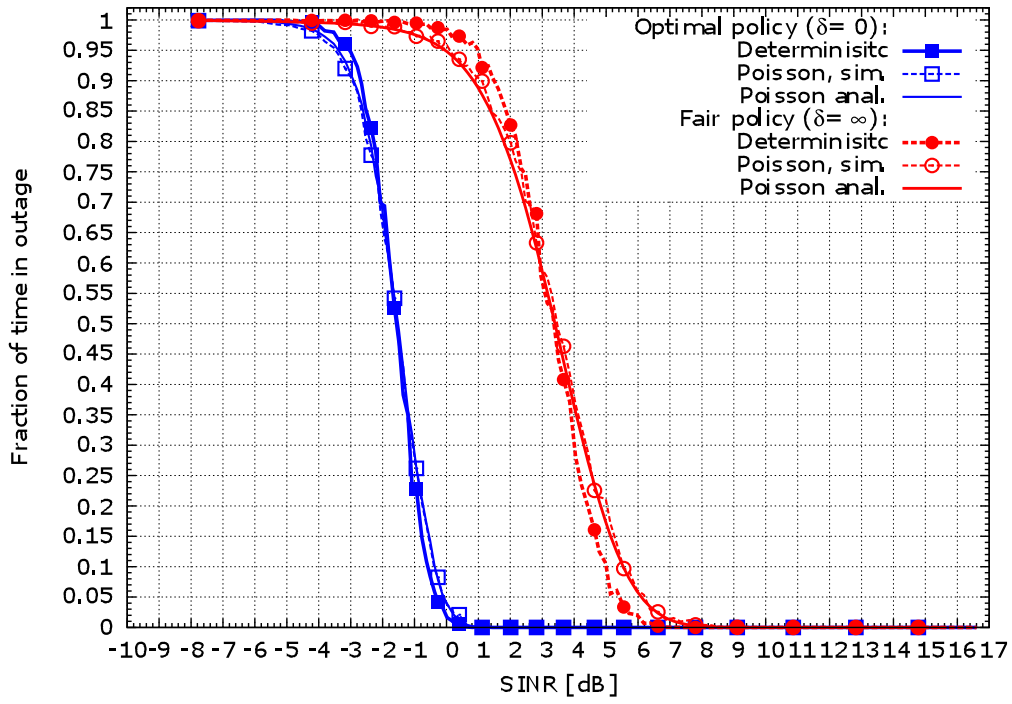


Figure 4.9: Impact of the deterministic arrival process (as compared to the Poisson one) on the mean fraction of the requested streaming time in outage, for the optimal and fair policy; traffic 600 Erlang/km$^2$.

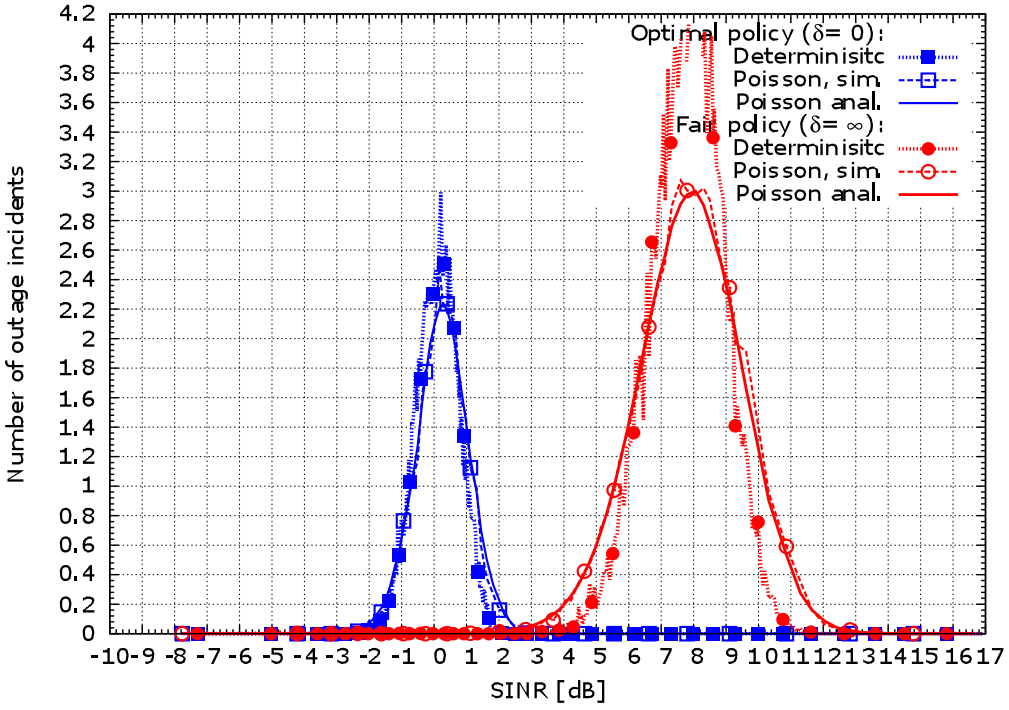


Figure 4.10: Impact of the deterministic arrival process (as compared to the Poisson one) on the mean number of outage incidents for the optimal and fair policy; traffic 900 Erlang/km$^2$.

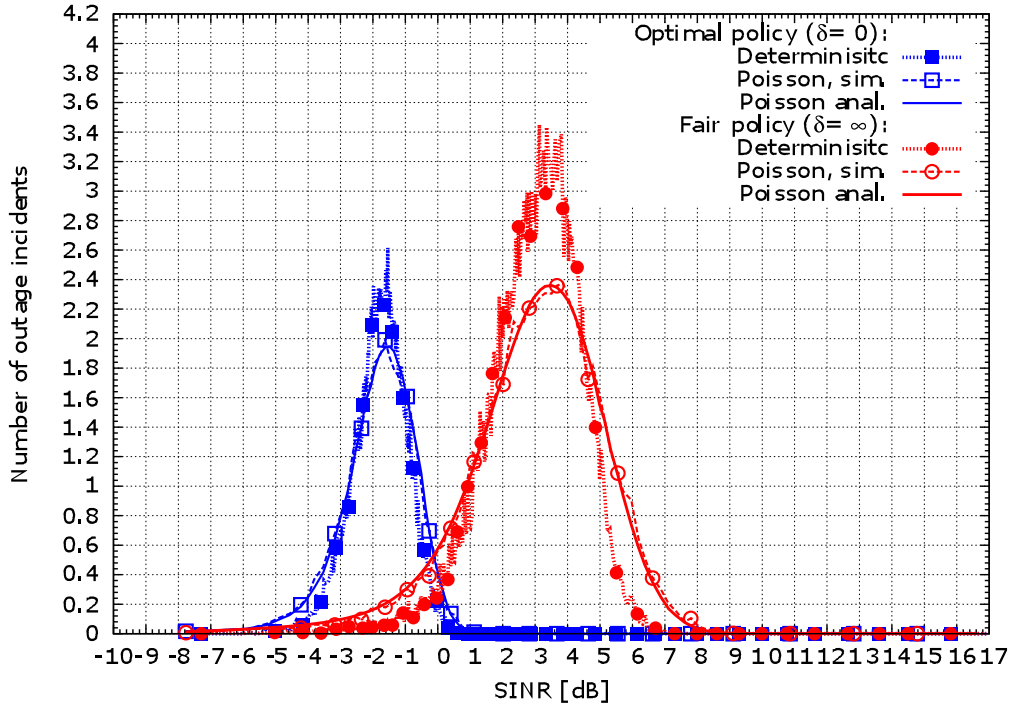


Figure 4.11: Impact of the deterministic arrival process (as compared to the Poisson one) on the mean number of outage incidents for the optimal and fair policy; traffic 600 Erlang/km$^2$.

# Appendix

## 4.A A general real-time streaming (RTS) model

In this section we will present a general stochastic model for real-time streaming. An instantiation of this model was used in the main body of this Chapter to evaluate the real-time streaming in wireless cellular networks. This model comprises a Markovian, multi-class process of call arrivals and their independent, arbitrarily distributed streaming times. These calls are served by a server whose service capacity is limited. Depending on the numbers of calls of different classes present in the system, the server may not be able to serve some classes of users. If such a congestion occurs, these classes are temporarily denied the service, until the next call arrival or departure, when the situation is reevaluated. These service denial periods, called outage periods, do not alter the call sojourn times in the system. Our model allows for a very general service (outage) policy saying which classes of users are temporarily denied the service due to insufficient service capacity. We will evaluate key characteristics of this model using the formalism of point processes and their Palm theory, often used in the modern approach to stochastic networking [11]. Specifically, we are interested in the intensity of outage incidents, the mean inter-outage times and the outage durations of a given class, seen from the server perspective, as well as the probability of outage at the arrival epoch, mean total time in outage and mean number of outage incidents experienced by a typical user of a given class. The expressions developed for these characteristics involve only stationary probabilities of the (free) traffic demand process, which in our case is a vector of independent Poisson random variables. Recall that such a representation is possible e.g. for the well known Erlang-B formula, giving the blocking probability in the classical (possibly multi-class) Erlang's loss model. Indeed, our model can be seen as an extension of the classical loss model, where the losses (i.e., service denials) are not definitive for a given call, but only temporal — having the form of outage periods.

### 4.1.1 Traffic demand

Consider $J \geq 1$ classes of users identified with calls. We assume that users of class $k \in \{1, \ldots, J\}$ arrive in time according to a Poisson process $N_k = \{T_n^k : n\}$ [9] with intensity $\lambda_k > 0$ and stay in the system for independent requested streaming times $W_n^k$ having some general distribution with mean $1/\mu_k < \infty$. All the results presented in what follows do not depend on the particular choice of the streaming time distributions — the property called in the queueing-theoretic context *insensitivity property*. Denote by $\tilde{N}_k = \{(T_n^k, W_n^k) : n\}$ the process of arrival epochs and streaming times (call durations) of users of class $k$. We assume that $\tilde{N}_k$ are independent across $k = 1, \ldots, J$. Denote by $X_k(t) = \sum_n \mathbf{1}_{[T_n^k, T_n^k + W_n^k)}(t)$ the number of users of class $k$ present in the system at

[9] The time instants $T_n^k$ are used only in the Appendix and should not be confused with $T_k$ denoting in the main stream of (and in the proof of Proposition 12 at the end of the Appendix) the mean throughput of a user in class $k$.

time $t$ and let $\mathbf{X}(t) = (X_1(t), \ldots, X_J(t))$; we call it the (vector of) user configuration at time $t$. The stationary distribution $\pi$ of $\mathbf{X}(t)$ coincides with the distribution of a vector of independent Poisson random variables $(X_1, \ldots, X_J)$ with means $\mathbf{E}[X_k] := \rho_k = \lambda_k/\mu_k$, $k = 1, 2, \ldots, J$. We call $\rho_k$ the *traffic demand* of class $k$.

We adopt the usual convention for the numbering of the arrival epochs $T_0^k \leq 0 < T_1^k$. The same convention is used with respect to all point processes denoting some time epochs.

### 4.1.2 Resource constraints and outage policy

For class $k = 1, \ldots, J$, let a subset of user configurations $\mathcal{F}_k \subset \bar{\mathbb{N}}^J$ be given, where $\bar{\mathbb{N}} = \{0, 1, \ldots\}$, such that all $X_k$ users of class $k$ present in the configuration $\mathbf{X} = (X_1 \ \ldots, X_k, \ldots, X_J)$ are served if and only if $\mathbf{X} \in \mathcal{F}_k$ and no user of class $k$ is served (we say it is in *outage*) if $\mathbf{X} \notin \mathcal{F}_k$. We call $\mathcal{F}_k$ the *k th class (service) feasibility set*. Denote by $\pi_k = \pi(\mathcal{F}_k)$ the probability that the stationary configuration of users is in $k$ th class feasibility set.

We assume that, upon each arrival or departure of a user, the system updates its decision and, for any class $k$, it assigns the service to all users of class $k$ if the updated configuration of users is in $\mathcal{F}_k$. All users of any class $j$ for which the updated configuration is in $\mathcal{F}_k' = \bar{\mathbb{N}}^J \setminus \mathcal{F}_k$ will be placed in outage (at least) until the next user arrival or departure.

In what follows we will assume that no user departure can cause outage of any class of users i.e., switch a given configuration from $\mathcal{F}_k$ to $\mathcal{F}_k'$. (However a user departure may make some class $j$ switch from $\mathcal{F}_j'$ to $\mathcal{F}_j$.)

Denote by $\tilde{X}_k(t) := X_k(t)\mathbf{1}_{\mathcal{F}_k}(\mathbf{X}(t))$ the number of users of class $k$ *not in outage* at time $t$. Denote by $\tilde{\mathbf{X}}(t) = (\tilde{X}_i(t), \ldots, \tilde{X}_J(t))$ the configuration of users *not* in outage at time $t$.

### 4.1.3 Performance metrics

In what follows we will be interested in the following characteristics of the model.

#### Virtual system metrics

During its time evolution, the user configuration $\mathbf{X}(t)$ alternates visits in the feasibility set $\mathcal{F}_k$ and its complement $\mathcal{F}_k'$, for each class $k = 1, \ldots, J$. We are interested in the expected visit durations in these sets as well as the intensities (frequencies) of the alternations. More formally, for each given $k = 1, \ldots, J$, we define the point process $B_k := \{\tau_n^k : n\}$ of exit epochs of $\mathbf{X}(t)$ from $\mathcal{F}_k$; i.e., all epochs $t$ such that $(\mathbf{X}(t-), \mathbf{X}(t)) \in \mathcal{F}_k \times \mathcal{F}_k'$ (with the convention $\tau_0^k \leq 0 < \tau_1^k$). These are epochs when all users of class $k$ present in the system (if any) have their service interrupted.

Denote by $\sigma_n'^k := \sup\{t - \tau_n^k : \mathbf{X}(s) \in \mathcal{F}_k' \ \forall s \in [\tau_n^k, t)\}$ the duration of the $n$ th visit of the process $\mathbf{X}(t)$ in $\mathcal{F}_k'$ and by $\sigma_n^k := \tau_{n+1}^k - \tau_n^k - \sigma_n'^k$ the duration of the $n$ th visit of the process $\mathbf{X}(t)$ in $\mathcal{F}_k$. We define for each class $k = 1, \ldots, J$:

- *The intensity of outage incidents of class $k$*, i.e., the mean number of outage incidents of this class per unit of time

$$\Lambda_k := \lim_{T \to \infty} \frac{1}{T} \sum_n \mathbf{1}_{[0,T)}(\tau_n^k)\,.$$

  Obviously $\Lambda_k$ is also the intensity of entrance to the $k$ th class feasibility set $\mathcal{F}_k$.

- *The mean service time between two outage incidents of class $k$*

$$\bar{\sigma}_k := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \sigma_n^k \,.$$

- *The mean outage duration of class $k$*

$$\bar{\sigma}'_k := \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \sigma_n'^k \,.$$

Note that the above metrics characterize a "virtual" quality of the service, since some visits in $\mathcal{F}_k$ and $\mathcal{F}'_k$ may occur when there is no $k$ th class user in the system (in the latter case the outage of this class is not experienced by any user).

**User metrics**

We adopt now a user point of view on the system. We define for each class $k = 1, \ldots, J$:

- *The probability of outage at the arrival epoch for user of class $k$*

$$\mathcal{P}_k = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{\mathcal{F}'_k}(\mathbf{X}(T_n^k)) \,.$$

- *The mean total time in outage of user of class $k$*

$$D_k = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \int_{[T_n^k, T_n^k + W_n^k)} \mathbf{1}_{\mathcal{F}'_k}(\mathbf{X}(t)) \, \mathrm{d}t \,.$$

- *The mean number of outage incidents experienced by user of class $k$ after its arrival*

$$M_k = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \sum_{m} \mathbf{1}_{(T_n^k, T_n^k + W_n^k)}(\tau_m^k) \,.$$

  Note that possible outage experienced at the arrival of a given user is not counted in $M_k$. The mean total number of outage incidents (including possibly at the arrival epoch) is hence $\mathcal{P}_k + M_k$.

### 4.1.4 Mathematical results

For a given class $k = 1, \ldots, J$, denote by $\varepsilon_k = (0, \ldots, 1, \ldots, 0) \in \bar{\mathbb{N}}^J$ the unit vector having its $k$ th component equal to 1. Hence $\mathbf{x} + \varepsilon_k$ represents adding one user of class $k$ to the configuration of users $\mathbf{x} \in \bar{\mathbb{N}}^J$. Denote by Pr the probability under which $\{\mathbf{X}(t) : t\}$ is stationary and by $\mathbf{E}$ the corresponding expectation. Recall that $\pi\{\mathbf{x} \in \cdot\} = \Pr\{\mathbf{X}(t) \in \cdot\}$ is the distribution of the stationary configuration of users $\mathbf{X}(t)$ (it corresponds to independent Poisson variables of mean $\rho_k$).

**General results**

We present first results regarding the virtual system metrics. These results will be next used to evaluate the user metrics.

**Lemma 8** *The intensity of outage incidents of class $k$ is* Pr*-almost surely equal to*

$$\Lambda_k = \sum_{j=1}^{J} \lambda_j \pi \left\{ \mathbf{x} \in \mathcal{F}_k, \mathbf{x} + \varepsilon_j \in \mathcal{F}'_k \right\} \qquad k = 1, \ldots, J.$$

**Proof.** Let $N = \sum_{j=1}^{J} N_j$ be the point process counting the arrival times of users of all classes. By independence, $N$ is the Poisson point process of intensity $\lambda = \sum_{j=1}^{J} \lambda_j$. Then, by the ergodicity of the process $\mathbf{X}(t)$ and the fact that the exits from $\mathcal{F}_k$ can take place only at some user arrival epoch we have by the Campbell's formula [10] [11, Equation (1.2.19)],

$$\begin{aligned} \Lambda_k &= \mathbf{E}\left[ \int_{[0,1)} \mathbf{1}_{\mathcal{F}_k \times \mathcal{F}'_k} \left( \mathbf{X}\left(t-\right), \mathbf{X}\left(t\right) \right) N\left(\mathrm{d}t\right) \right] \\ &= \lambda \overset{0}{\Pr_N} \left\{ \mathbf{X}(0-) \in \mathcal{F}_k, \mathbf{X}(0) \in \mathcal{F}'_k \right\} , \end{aligned}$$

where $\Pr_N^0$ designates the Palm probability associated to $N$ (which is, roughly speaking, the conditional probability given an arrival at time 0). By the PASTA (Poisson Arrivals See Time Averages) property [11, Equation (3.3.4)] the configuration of users $\mathbf{X}(0-)$ under $\Pr_N^0$ has distribution $\pi$. Moreover, $\mathbf{X}(0) = \mathbf{X}(0-) + \varepsilon_\xi$ where $\xi \in \{1, \ldots, J\}$ is under $\Pr_N^0$ independent of $X(0-)$ and takes value $j$ with probability $\lambda_j/\lambda$. This completes the proof. ■

**Lemma 9** *The mean service time between two outage incidents and the mean outage duration of class $k$ are* Pr*-almost surely equal to, respectively,*

$$\bar{\sigma}_k = \frac{\pi\left(\mathcal{F}_k\right)}{\Lambda_k}, \quad \bar{\sigma}'_k := \frac{\pi(\mathcal{F}'_k)}{\Lambda_k} \qquad k = 1, \ldots, J,$$

*where $\Lambda_k$ is given in Lemma 8.*

**Proof.** First we prove the expression for $\bar{\sigma}_k$. By ergodicity $\bar{\sigma}_k = \mathbf{E}^0_{B_k}\left[\sigma_0^k\right]$ Pr-almost surely, where $\mathbf{E}^0_{B_k}$ designates the expectation with respect to the Palm probability associated to $B_k$, and $\mathbf{E}^0_{B_k}\left[\tau_0^k\right] = 1/\Lambda_k$; [11, see e.g. Equation (1.6.8) and Equation (1.2.27)]. Applying the mean value formula [11, Equation (1.3.2)][11] we get $\pi(\mathcal{F}_k) = \Lambda_k \mathbf{E}^0_{B_k}\left[\sigma_0^k\right]$, which completes the proof of the expression for $\bar{\sigma}_k$. For the other expression, note by the definition of the sequence $\sigma_n^k, \sigma_n'^k$ and $\tau_n^k$ that Pr-almost surely,

$$\bar{\sigma}'_k = \mathbf{E}^0_{B_k}\left[\sigma_0'^k\right] = \mathbf{E}^0_{B_k}\left[\tau_1^k - \sigma_0^k\right] = \frac{1}{\Lambda_k} - \frac{\pi(\mathcal{F}_k)}{\Lambda_k} = \frac{\pi(\mathcal{F}'_k)}{\Lambda_k},$$

which completes the proof. ■

**Proposition 13** *The probability of outage at the arrival epoch for a user of class $k$ is equal to*

$$\mathcal{P}_k = \pi\left\{ \mathbf{x} + \varepsilon_k \in \mathcal{F}'_k \right\} \qquad k = 1, \ldots, J \tag{4.16}$$

Pr*-almost surely.*

[10] With $Z_n := (\mathbf{X}(T_n-), \mathbf{X}(T_n))$ and $f(t,z) = \mathbf{1}_{[0,1)}(t)\mathbf{1}_{\mathcal{F}_k \times \mathcal{F}'_k}(z)$

[11] with $Z_k\left(t\right) = \mathbf{1}_{\mathcal{F}_k}\left(\mathbf{X}\left(t\right)\right)$

**Proof.** By ergodicity we have $\mathcal{P}_k = \Pr^0_{N_k}\{\mathbf{X}(0) \in \mathcal{F}'_k\}$, where $\Pr^0_{N_k}$ designates the Palm probability associated to $N_k$ (arrival process of the users of class $k$). By the PASTA property the configuration of users $\mathbf{X}(0-)$, just before arrival of the user of class $k$ at time 0, has distribution $\pi$. Once the user enters the system, the user configuration becomes $\mathbf{X}(0-) + \varepsilon_k$, whence the result. ■

**Proposition 14** *The mean total time in outage of a user of class $k$ is* Pr*-almost surely equal to*

$$D_k = \frac{1}{\mu_k}\pi\{\mathbf{x} + \varepsilon_k \in \mathcal{F}'_k\} \qquad k = 1, \ldots, J\,. \tag{4.17}$$

**Proof.** Again using the ergodicity of $\{\mathbf{X}(t)\}$ we can write

$$D_k = \mathbf{E}^0_{N_k}\left[\int_{[0,W^k_0)} \mathbf{1}_{\mathcal{F}'_k}(\mathbf{X}(t))\,\mathrm{d}t\right].$$

Denote by $\mathbf{Y}(t) := \mathbf{X}(t) - \varepsilon_k \mathbf{1}_{[T^k_0, T^k_0 + W^k_0)}(t)$ the process of configurations of users other than the user number 0 of class $k$ (which arrives at time 0 under $\mathbf{E}^0_{N_k}$). By Slivnyak theorem [9, see e.g. Theorem 1.13] the distribution of the process $\{\mathbf{Y}(t) : t\}$ under $\Pr^0_{N_k}$ is the same as this of $\{\mathbf{X}(t) : t\}$ under Pr. Using the fact that $W^k_0$ and $\mathbf{Y}(t)$ are independent under $\Pr^0_{N_k}$ with $\mathbf{E}^0_{N_k}[W^k_0] = 1/\mu_k$ we obtain

$$\begin{aligned} D_k &= \int_0^\infty \mathbf{E}^0_{N_k}\left[\mathbf{1}_{[0,W^k_0)}(t)\mathbf{1}_{\mathcal{F}'_k}(\mathbf{Y}(t) + \varepsilon_k)\right]\,\mathrm{d}t \\ &= \frac{1}{\mu_k}\,\pi\{\mathbf{x} + \varepsilon_k \in \mathcal{F}'_k)]\,, \end{aligned}$$

which completes the proof. ■

**Proposition 15** *The mean number of outage incidents experienced by user of class $k$ after its arrival is* Pr*-almost surely equal to*

$$\begin{aligned} M_k =& \frac{1}{\mu_k}\sum_{j=1}^{J} \lambda_j \pi\{\mathbf{x} + \varepsilon_k \in \mathcal{F}_k, \mathbf{x} + \varepsilon_k + \varepsilon_j \in \mathcal{F}'_k\}\,, \\ & k = 1, \ldots, J\,. \end{aligned} \tag{4.18}$$

**Proof.** Again using the ergodicity of $\{X(t)\}$ we know that, Pr-almost surely,

$$M_k = \mathbf{E}^0_{N_k}\left[\int_{(0,W^k_0)} B_k(\mathrm{d}t)\right].$$

Using the fact that $W^k_0$ and $\mathbf{Y}(t)$ are independent under $\Pr^0_{N_k}$ with $\mathbf{E}^0_{N_k}[W^k_0] = 1/\mu_k$ we obtain

$$M_k = \mathbf{E}^0_{N_k}\left[B^*_k(0, W^k_0)\right] = \frac{\Lambda^*_k}{\mu_k}\,,$$

where $B^*_k =: \{\tau^{*k}_n : n\}$ is the point process of exit epochs of $\mathbf{X}(t)$ from $\mathcal{F}^*_k = \{\mathbf{x} : \mathbf{x} + \varepsilon_k \in \mathcal{F}_k\}$ and $\Lambda^*_k$ its intensity. Using Lemma 8 with $\mathcal{F}_k$ replaced by $\mathcal{F}^*_k$ concludes the proof. ■

We will now prove the result regarding the throughput of the typical call of class $k$.

**Proof of Proposition 12.** We have

$$\Upsilon_k = \Upsilon_k^\delta = \mu_k \mathbf{E}_{N_k}^0 \left[ \int_{[0,W_0^k)} r_k \mathbf{1}(\mathbf{X}(t) \in \mathcal{F}_k^\delta) + r_k^{'\delta}(\mathbf{X}(t)) \mathbf{1}(\mathbf{X}(t) \notin \mathcal{F}_k^\delta) \, \mathrm{d}t \right] .$$

It is easy to see, as in the proof of Proposition 14, that $\Upsilon_k = r_k \pi \left\{ \mathbf{x} + \varepsilon_k \in \mathcal{F}_k^\delta \right\} + \Upsilon_k'$, where

$$\Upsilon_k' = \mathbf{E} \left[ r_k^{'\delta}(\mathbf{X}(t) + \varepsilon_k) \mathbf{1}((\mathbf{X}(t) + \varepsilon_k) \notin \mathcal{F}_k^\delta) \right]$$

is the part of the throughput obtained by a user of class $k$ during its outage time. ■

# Chapter 5

# Conclusion and future work

This thesis aimed at developing the methods for the examination of the QoS perceived by users in cellular networks. QoS for services such as data transmission and real-time streaming are examined. The approach consists in decomposing the problem into three levels corresponding to three time scales. Firstly, the single link capacity is studied on the ground of information theory, after that users' arrivals and departures are considered using queuing theory. Finally, the spatial patterns of network resources are taken into account using stochastic geometry.

More specifically, in Chapter 2 we describe a simple model of a MIMO cellular network which permits to obtain an analytical lower bound for user bit-rates which are feasible from the information theory point of view. This expression accounts for the variety of MIMO configurations (numbers of transmitting and receiving antennas) and radio conditions (SINR). We validate the analytical lower bound by comparison to the results of 3GPP simulations and to measurements in the field.

In Chapter 3 we establish the dependence relation between the traffic demand and mean user throughput for large wireless cellular networks serving variable bit-rate calls.

Further, we evaluate the user QoS (spatial CDFs of mean user throughput per cell, mean number of users per cell and cell loads. We develop two approaches: the typical cell approach corresponds to spatial averages of the characteristics of the cells in a large network. Since the averages of some crucial characteristics do not have explicit expressions, we propose the alternative mean cell approach. It permits an explicit expression of the major characteristics and approximates well the typical cell. Also a heterogeneous cellular network model allowing for different BS types (having different transmission powers) is proposed, aiming to help in performance evaluation and dimensioning of real (large, irregular) operational networks. It allows one to identify key laws relating the performance of the different base station types.We validate the proposed approach by comparing its results to real field measurements.

The dimensioning for streaming traffic as well as mixing such traffic with variable bit-rate calls are important axes for future work. These studies raised also open theoretical questions regarding the stability of spatially and, more difficult, space-time dependent processor sharing queues modeling the performance of individual network cells (cf Section 3.5.3). More work is also required to understand the problem of different performance of the network during day and night hours (cf Figure 3.15).

In Chapter 4, a real-time streaming (RTS) traffic, as e.g. mobile TV, is analyzed in the context of wireless cellular networks. An adequate stochastic model is proposed to evaluate user performance metrics, such as frequency and number of interruptions during RTS calls as function of user radio conditions. Despite some fundamental similarities to the classical Erlang loss model,

a new model was required for this type of service, where the service denials are not definitive for a given call, but only temporal – having the form of, hopefully short, interruptions (outage) periods. Our model allows one to take into account realistic implementations of the RTS service, e.g. in the LTE networks. In this latter context, several numerical demonstrations are given, presenting the quality of service metrics as function of user radio conditions.

# Bibliography


[1] 3GPP. TR 25.814-V710 Physical layer aspects for evolved Universal Terrestrial Radio Access (UTRA). In *3GPP Ftp Server*, 2006.

[2] 3GPP. TR 36.211-V910 Evolved Universal Terrestrial Radio Access (E-UTRA) - Physical Channels and Modulation. In *3GPP Ftp Server*, March 2010.

[3] 3GPP. TR 36.814-V900 Further advancements for E-UTRA - Physical Layer Aspects. In *3GPP Ftp Server*, 2010.

[4] 3GPP. TR 36.942-V830 Evolved Universal Terrestrial Radio Access (E-UTRA) - Radio Frequency (RF) system scenarios. In *3GPP Ftp Server*, September 2010.

[5] 3GPP. TS 36.101-V8.16.0 Evolved Universal Terrestrial Radio Access (E-UTRA) - User Equipment (UE) radio transmission and reception. In *3GPP Ftp Server*, January 2012.

[6] J. Abate and W. Whitt. Numerical inversion of Laplace transforms of probability distributions. *ORSA Journal on Computing*, 7(1):38–43, 1995.

[7] M.-S. Alouini and A. J. Goldsmith. Area spectral efficiency of cellular mobile radio systems. *IEEE Trans. Veh. Technol.*, 48:1047–1066, 1999.

[8] J.G. Andrews, F. Baccelli, and R.K. Ganti. A Tractable Approach to Coverage and Rate in Cellular Networks. *IEEE Trans. Commun.*, 59(11):3122 –3134, november 2011.

[9] F. Baccelli and B. Błaszczyszyn. *Stochastic Geometry and Wireless Networks, Volume I — Theory*, volume 3, No 3–4 of *Foundations and Trends in Networking*. NoW Publishers, 2009.

[10] F. Baccelli, B. Błaszczyszyn, and F. Tournois. Downlink admission/congestion control and maximal load in CDMA networks. In *Proc. of IEEE Infocom*, 2003.

[11] F. Baccelli and P. Brémaud. *Elements of queueing theory. Palm martingale calculus and stochastic recurrences*. Springer, 2003.

[12] N. Baldo, M. Miozzo, M. Requena, and J. N. Guerrero. An open source product-oriented LTE network simulator based on ns-3. In *Proc. of MSWIM*, 2011.

[13] N. Baldo, M. Requena, J. Nin, and M. Miozzo. A new model for the simulation of the LTE-EPC data plane. In *Proc. of WNS3*, 2012.

[14] Dilip Bethanabhotla, Giuseppe Caire, and Michael J Neely. Joint transmission scheduling and congestion control for adaptive video streaming in small-cell networks. *arXiv preprint arXiv:1304.8083*, 2013.

[15] B. Błaszczyszyn, M. Jovanovic, and M. K. Karray. How user throughput depends on the traffic demand in large cellular networks. In *In Proc. of WiOpt/SpaSWiN*, 2014.

[16] B. Błaszczyszyn, M. Jovanovic, and M. K. Karray. Performance laws of large heterogeneous cellular networks. In *arXiv:1411.7785 submitted*, 2014.

[17] B. Błaszczyszyn, M. Jovanovic, and M. K. Karray. Quality of real-time streaming in wireless cellular networks - stochastic modeling and analysis. *IEEE Trans. on Wireless Commun.*, 13(6), 2014.

[18] B. Błaszczyszyn and M. K. Karray. Effect of opportunistic scheduling on the quality of service perceived by the users in ofdma cellular networks. *Annals of telecommunications*, 67(3):203–213, April 2012.

[19] B. Błaszczyszyn, M. K. Karray, and H. P. Keeler. SINR-based coverage probability in cellular networks under multiple connections. In *Proc. of ISIT*, 2013.

[20] B. Błaszczyszyn, M. K. Karray, and H. P. Keeler. Using Poisson processes to model lattice cellular networks. In *Proc. of Infocom*, 2013.

[21] B. Błaszczyszyn and Holger Paul Keeler. Equivalence and comparison of heterogeneous cellular networks. In *Proc. of PIMRC'13 - WDN-CN2013*, 2013.

[22] R. S. Blum, J. H. Winters, and N. R. Sollenberger. On the capacity of cellular systems with MIMO. *Communications Letters, IEEE*, 6(6):242 –244, jun 2002.

[23] H. Bolcskei, D. Gesbert, and A. J. Paulraj. On the capacity of OFDM-based spatial multiplexing systems. *IEEE Trans. Commun.*, 50(2), February 2002.

[24] T. Bonald, S. C. Borst, N. Hegde, M. Jonckheere, and A. Proutière. Flow-level performance and capacity of wireless networks with user mobility. *Queueing Systems*, 63(1-4):131–164, 2009.

[25] T. Bonald and A. Proutière. Wireless downlink data channels: user performance and cell dimensioning. In *Proc. of Mobicom*, September 2003.

[26] S. Borst. User-level performance of channel-aware scheduling algorithms in wireless data networks. In *Proc. of IEEE Infocom*, 2003.

[27] P. Brémaud. *Markov chains. Gibbs fields, Monte Carlo simulation, and queues.* Springer, 1999.

[28] P. Brémaud. *Initiation aux probabilités et aux chaînes de Markov.* Springer, France, 2009.

[29] Timothy X Brown. Cellular performance bounds via shotgun cellular systems. *Selected Areas in Communications, IEEE Journal on*, 18(11):2443–2455, 2000.

[30] G. Caire, G. Taricco, and E. Biglieri. Optimum power control over fading channels. *IEEE Trans. Inf. Theory*, 45(5):1468–1489, July 1999.

[31] V. Chandrasekhar and J. G. Andrews. Spectrum Allocation in Tiered Cellular Networks. *IEEE Trans. on Comm.*, 57(10), 2009.

[32] A. Clark, P. J. Smith, and D. P. Taylor. Instantaneous capacity of OFDM on Rayleigh-fading channels. *IEEE Trans. Inf. Theory*, 53(1):355–361, 2007.

[33] J. W. Cohen. *On regenerative processes in queueing theory*, volume 121 of *Lecture Notes in Economics and Mathematical Systems*. Springer Berlin Heidelberg, 1976.

[34] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 2006.

[35] E. Dahlman, S. Parkvall, and J. Skold. *4g: LTE/LTE-advanced for mobile broadband*. Academic Press Inc, 2011.

[36] D. J. Daley and D. VereJones. *An introduction to the theory of point processes*. Springer, New York, 1988.

[37] D. J. Daley and D. VereJones. *An introduction to the theory of point processes*. Springer, New York, 2 edition, 2003.

[38] F. R. de Hoog, J. H. Knight, and A. N. Stokes. An improved method for numerical inversion of laplace transforms. *SIAM Journal of Scientific and Statistical Computation*, 3(3):357–366, 1982.

[39] H. S. Dhillon, R. Krishna Ganti, F. Baccelli, and J. G. Andrews. Modeling and analysis of K-tier downlink heterogeneous cellular networks. *IEEE Journal on Selected Areas in Communications*, 30(3):550–560, 2012.

[40] S.N. Diggavi and T.M. Cover. The worst additive noise under a covariance constraint. *IEEE Trans. Inf. Theory*, 47(7):3072–3081, 2001.

[41] L. Du, J. Bigham, and L. Cuthbert. Geographic Load Balancing for WCDMA Mobile Networks Using a Bubble Oscillation Algorithm. In *Wireless Communications and Networking Conference*, 2005.

[42] S-E. Elayoubi, O. Ben Haddada, and B. Fourestié. Performance evaluation of frequency planning schemes in OFDMA-based networks. *IEEE Trans. Wireless Commun.*, 7(5-1):1623–1633, 2008.

[43] J. Evans and D. N. C. Tse. Large system performance of linear multiuser receivers in multipath fading channels. *IEEE Trans. Inf. Theory*, 46(6), September 2000.

[44] J. S. Evans and D. Everitt. On the teletraffic capacity of CDMA cellular networks. *IEEE Trans. Veh. Technol.*, 48, January 1999.

[45] F.H.P. Fitzek, S. Hendrata, P. Seeling, and M. Reisslein. *Wireless Internet – Video Streaming in Wireless Internet*, chapter 11, pages 1–41. Electrical Engineering & Applied Signal Processing Series. CRC Press, March 2004.

[46] G. J. Foschini and M. J. Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, 6(3):311–335, March 1998.

[47] R.G. Gallager. *Information theory and reliable communication*. Wiley, New York, 1968.

[48] A. Gember, A. Akella, A. Varshavsky J. Pang, and R. Caceres. Obtaining In-Context Measurements of Cellular Network Performance. In *Internet Measurement Conference (IMC)*, 2012.

[49] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu. Multi-cell mimo cooperative networks : A new look at interference. *IEEE J. Select. Areas Commun.*, 28(9), 2010.

[50] A. Ghosh, J. G. Andrews, and al. Heterogeneous Cellular Networks: From Theory to Practice. *IEEE Communications Magazine*, june 2012.

[51] M. Girnyk, M. Vehkaperä, and L. Rasmussen. On the asymptotic sum-rate of uplink mimo cellular systems in the presence of non-gaussian inter-cell interference. In *Proc. of Globecom*, 2012.

[52] R. Giuliano and F. Mazzenga. Dimensioning of OFDM OFDMA-Based Cellular Networks Using Exponential Effective SINR. *IEEE Trans. Veh. Technol.*, 58(8), 2009.

[53] A. J. Goldsmith and S.-G Chua. Variable-rate variable-power MQAM for fading channels. *IEEE Trans. Commun.*, 45:1218–1230, 1997.

[54] N. Hegde and E. Altman. Capacity of multiservice WCDMA Networks with variable GoS. In *Proc. of IEEE WCNC*, 2003.

[55] K. J. Hollenbeck. Invlap.m: A Matlab function for numerical inversion of Laplace transforms by the de Hoog algorithm. unpublished work; downloadable from www.isva.dtu.dk/staff/karl/invlap.htm, 1998.

[56] S. Ihara. *Information theory for continuous systems.* World Scientific, 1993.

[57] M. Jovanovic, M. K. Karray, and B. Błaszczyszyn. QoS and network performance estimation in heterogeneous cellular networks validated by real-field measurements. In *In Proc. of PE-WASUN/MSWiM*, 2014.

[58] M. K. Karray. *Analytic evaluation of wireless cellular networks performance by a spatial Markov process accounting for their geometry, dynamics and control schemes.* PhD thesis, Ecole Nationale Supérieure des Télécommunications, 2007.

[59] M. K. Karray. Analytical evaluation of QoS in the downlink of OFDMA wireless cellular networks serving streaming and elastic traffic. *IEEE Trans. Wireless Commun.*, 9(5), May 2010.

[60] M. K. Karray. User's mobility effect on the performance of wireless cellular networks serving elastic traffic. *Wireless Networks (Springer)*, 17(1), January 2011.

[61] M. K. Karray and M. Jovanovic. Theoretically Feasible QoS in a MIMO Cellular Network Compared to the Practical LTE Performance. In *Proc. of ICWMC*, 2012.

[62] M. K. Karray and M. Jovanovic. A queueing theoretic approach to the dimensioning of wireless cellular networks serving variable bit-rate calls. *IEEE Trans. Veh. Technol.*, 62(6), July 2013.

[63] M. K. Karray, M. Jovanovic, and B. Błaszczyszyn. Theoretical expression of link performance in OFDM cellular networks with MIMO compared to simulation and measurements. In *submitted*, 2014.

[64] A. Kashyap, T. Basar, and R. Srikant. Correlated jamming on MIMO Gaussian fading channels. *IEEE Trans. Inf. Theory*, 50(9):2119–2123, 2004.

[65] J.S. Kaufman. Blocking in a shared resource environment. *IEEE Trans. Commun.*, 29(10):1474–1481, 1981.

[66] J. M. Kelif, M. Coupechoux, and P. Godlewski. On the Dimensioning of Cellular OFDMA Networks. *Elsevier Physical Communication*, 5(1), 2012.

[67] Bo Li, Samuel T Chanson, and Chuang Lin. Analysis of a hybrid cutoff priority scheme for multiple classes of traffic in multimedia wireless networks. *Wireless Networks*, 4(4):279–290, 1998.

[68] G. Liang and B. Liang. Effect of delay and buffering on jitter-free streaming over random VBR channels. *IEEE transactions on multimedia*, 10(6):1128–1141, 2008.

[69] A. Lozano and A. M. Tulino. Capacity of multiple-transmit multiple receive antenna architectures. *IEEE Trans. Inf. Theory*, 48(12), December 2002.

[70] K. Majewski and M. Koonert. Conservative Cell Load Approximation for Radio Networks with Shannon Channels and its Application to LTE Network Planning. In *Sixth Advanced International Conference on Telecommunications*, 2010.

[71] M. Medard. Capacity of correlated jamming channels. In *Allerton Conference on Communications, Computing and Control*, 1997.

[72] C. Mehlführer, J. C. Ikuno, M. Simko, S. Schwarz, M. Wrulich, and M. Rupp. The Vienna LTE simulators - Enabling reproducibility in wireless communications research. *EURASIP Journal on Advances in Signal Processing*, 2011.

[73] P. E. Mogensen, W. Na, I. Z. Kovács, F. Frederiksen, A. Pokhariyal, K. I. Pedersen, T. E. Kolding, K. Hugl, and M. Kuusela. LTE Capacity Compared to the Shannon Bound. In *Proc. of VTC Spring*, pages 1234–1238, 2007.

[74] Z. Niu, Y. Wu, J. Gong, and Z. Yang. Cell Zooming for Cost-Efficient Green Cellular Networks. *IEEE Communications Magazine*, november 2010.

[75] A. ParandehGheibi, M. Médard, S. Shakkottai, and A. Ozdaglar. Avoiding interruptions - qoe trade-offs in block-coded streaming media applications. In *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pages 1778 –1782, june 2010.

[76] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das. Understanding Traffic Dynamics in Cellular Data Networks. In *INFOCOM*, 2011.

[77] G. Piro, L. A. Grieco, G. Boggia, F. Capozzi, and P. Camarda. Simulating LTE cellular systems: an open source framework. *IEEE Trans. Veh. Technol.*, 60(2), 2011.

[78] S. Ramanath, E. Altman, V. Kumar, and M. Debbah. Optimizing cell size in Pico-cell networks. In *Proc. of RAWNET/WCNC-3*, 2009.

[79] P. B. Rapajic and D. Popescu. Information capacity of a random signature multiple-input multiple-output channel. *IEEE Trans. on Comm.*, 48(8), 2000.

[80] Ph. Robert. *Stochastic networks and queues.* Springer, 2003.

[81] J.W. Roberts. A service system with heterogeneous user requirements. In *Performance of Data Communications Systems and their Applications (edited by G. Pujolle)*, 1981.

[82] L. Rong, S. E. Elayoubi, and O. B. Haddada. Performance evaluation of cellular networks offering TV services. *IEEE Trans. Veh. Technol.*, 60(2):644 –655, feb. 2011.

[83] A. Sampath, P. S. Kumar, and J. Holtzmann. Power control and resource management for a multimedia CDMA wireless system. In *Proc. of IEEE PIMRC*, volume 1, September 1995.

[84] C. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948.

[85] I. Shomorony and A. S. Avestimehr. Worst-case additive noise in wireless networks. *CoRR*, abs/1202.2687, 2012.

[86] M. Simko, Q. Wang, and M. Rupp. Optimal pilot symbol power allocation under time-variant channels. *EURASIP Journal on Wireless Communications and Networking*, 225(25), 2012.

[87] I. Siomina and D. Yuan. Analysis of Cell Load Coupling for LTE Network Planning and Optimization. *IEEE Trans. Wireless Commun*, 11(6), June 2012.

[88] Dietrich Stoyan, Wilfrid S. Kendall, and Joseph Mecke. *Stochastic geometry and its applications*. Wiley Chichester, 1995.

[89] W. Lum Tan, F. Lam, and W. Cheong Lau. An Empirical Study on the Capacity and Performance of 3G Networks. *IEEE Trans. Mob. Computing*, 7(6), 2008.

[90] E. Telatar. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, 10(6):585–596, November 1999.

[91] I. E. Telatar. Capacity of multi-antenna Gaussian channels. *AT&T Technical Memorandum*, June 1995.

[92] D. Tse and P. Viswanath. *Fundamentals of wireless communication*. Cambridge University Press, 2005.

[93] D. N. C. Tse and S. V. Hanly. Linear multiuser receivers: Effective interference effective bandwidth and user capacity. *IEEE Trans. Inf. Theory*, 45(2), March 1999.

[94] A. M. Tulino and S. Verdú. Random matrix theory and wireless communications. *Foundations and Trends in Communications and Information Theory*, 1(1), 2004.

[95] S. Verdu and S. Shamai. Spectral efficiency of CDMA with random spreading. *IEEE Trans. Inf. Theory*, 45(2), March 1999.

[96] H. Wang, L. Ding, P. Wu, Z. Pan, N. Liu, and X. You. Dynamic load balancing in 3GPP LTE multi-cell networks with heterogenous services. In *Communications and Networking in China (CHINACOM)*, 2010.

[97] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz. Primary Users in Cellular Networks: A Large-Scale Measurement Study. In *3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN)*, 2008.

[98] Si Wu, KY Michael Wong, and Bo Li. A dynamic call admission policy with precision QoS guarantee using stochastic control for mobile wireless networks. *Networking, IEEE/ACM Transactions on*, 10(2):257–271, 2002.

[99] Yuedong Xu, Eitan Altman, Rachid El Azouzi, Majed Haddad, Salah-Eddine Elayoubi, and Tania Jiménez. Probabilistic analysis of buffer starvation in Markovian queues. In *in Proc. of Infocom'12*, Orlando, FL USA, 2012.

[100] R. Yates. A framework for uplink power control in cellular radio systems. *IEEE J. Select. Areas Commun.*, 13(7), September 1995.

[101] James Yu, Mohamed Amezziane, et al. G/g/c/c simulation model for voip traffic engineering with non-parametric validation. In *ICDT 2013, The Eighth International Conference on Digital Telecommunications*, pages 44–49, 2013.

[102] J. Zander. Distributed co-channel interference control in cellular radio systems. *IEEE Trans. Veh. Technol.*, 41, 1992.

[103] H. Zhang, X. Qiu, L. Meng, and X. Zhang. Design of distributed and autonomic load balancing for self-organization LTE. In *Vehicular Technology Conference Fall (VTC 2010-Fall)*, 2010.