Διδακτορική Διατριβή

Ανάλυση και Σχεδίαση Πυκνά Δομημένων Ασυρμάτων Δικτύων

Στέλιος Στεφανάτος

Ιούνιος 2017

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

Πανεπιστήμιο Πειραιώς,
Σχολή Τεχνολογιών Πληροφορικής και Επικοινωνιών,
Τμήμα Ψηφιακών Συστημάτων

PhD Thesis

Analysis and Design of Ultra-Dense Wireless Networks

Stelios Stefanatos

June 2017

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

UNIVERSITY OF PIRAEUS

University of Piraeus,
School of Information and Communication Technologies,
Department of Digital Systems

Στη μητέρα μου, Ελπινίκη

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά την επιβλέποντα μου, κα. Αγγελική Αλεξίου, Αναπλ. Καθηγήτρια Πανεπιστημίου Πειραιώς, για την καθοδήγηση και υποστήριξη που μου παρείχε καθόλη την διάρκεια αυτής της διατριβής. Χωρίς την βοήθειά της, το αποτέλεσμα αυτής της διατριβής δεν θα ήταν εφικτό.

Θα ήθελα επίσης να ευχαριστήσω τα μέλη της εξεταστικής μου επιτροπής για την τιμή που μου έκαναν να αξιολογήσουν την εργασία μου, κ. Αθανάσιο Κανάτα, Καθηγητή Πανεπιστημίου Πειραιώς, κ. Άγγελο Ρούσκα, Αναπλ. Καθηγητή Πανεπιστημίου Πειραιώς, κ. Μιλτιάδη Αναγνώστου, Καθηγητή Εθνικού Μετσόβειου Πολυτεχνείου, κ. Γεώργιο Ευθύμογλου, Αναπλ. Καθηγητή Πανεπιστημίου Πειραιώς, κ. Μιχαήλ Φιλιππάκη, Επικ. Καθηγητή, Πανεπιστημίου Πειραιώς, και κ. Στάυρο Τουμπή, Επικ. Καθηγητή Οικονομικού Πανεπιστημίου Αθηνών.

Περίληψη

Η ραγδαία αύξηση ασυρμάτων συσκευών καθώς και η εισαγωγή νέων ασύρματων εφαρμογών με στόχο την υλοποίηση του "δικτύου των πραγμάτων" εισάγουν σημαντικές προκλήσεις στην λειτουργία του κυψελλωτού δικτύου. Είναι ευρέως αποδεκτό ότι η αντιμετώπιση αυτών των προκλήσεων απαιτεί ραγδαίες αλλαγές στην λειτορυγία του δικτύου όπως το γνωρίζουμε σήμερα. Η πιο σημαντική αλλαγή αφορά την δημιουργία δικτύων *πυκνής δόμησης* με στόχο την μεγαλύτερη χωρική επαναχρησιμοποίηση φάσματος και, εν τέλει, την εξυπηρέτηση του τεραστίου αριθμού συσκευών που αναμένεται να υπάρχει στο μέλλον. Ταυτόχρονα, αυτός ο μεγάλος αριθμός ασυρμάτων συσκευών ενθαρρύνει την εισαγωγή *νέων μεθόδων πρόσβασης* στο δίκτυο - επιπλέον των παραδοσιακών ανοδικών/καθοδικών ζεύξεων - με τις συσκευές να είναι σε θέση να επικοινωνήσουν απευθείας χωρίς την διαμεσολάβηση ενός ή περισσοτέρων σταθμών βάσης. Αυτά τα νέα χαρακτηριστικά του μελλοντικού δικτύου, αν και θεωρητικά προσφέρουν την ζητούμενη αύξηση των επιδόσεων, στην πράξη εισάγουν σημαντικές προκλήσεις που πρέπει να αντιμετωπιστούν και οι οποίες δεν ήταν παρούσες στα δίκτυα προηγούμενων γενεών.

Στην παρούσα διδακτορική διατριβή μελετώνται διάφορα χαρακτηριστικά της λειτουργίας μελλοντικών, πυκνά δομημένων κυψελλωτών δικτύων με χρήση εργαλείων στοχαστικής γεωμετρίας. Στο Κεφάλαιο 2 μελετάται η επίδραση που έχει η ολοένα αυξανόμενη πυκνότητα σταθμών βάσης στην ποιότητα υπηρεσιών που αντιλαμβάνεται ο χρήστης. Αποδεικνύεται ότι η αύξηση της πυκνότητας, αν και είναι σε θέση από μόνη της να αυξήσει τις επιδόσεις, πρέπει απαραίτητα να συνοδεύεται και από ένα προσεκτικά σχεδιασμό μηχανισμό κατανομής πόρων μεταξύ των χρηστών της κάθε κυψέλης. Αν και η επίδοση αυξάνει μονοτονικά με την αύξηση της πυκνότητας, αύξηση πέραν του σημείου όπου είναι συγκρίσιμη με την πυκνότητα των συσκευών παρέχει λίγα οφέλη και, συνεπώς, δεν έχει πρακτικό ενδιαφέρον.

Στο Κεφάλαιο 3 εξετάζεται κατά πόσο η απευθείας επικοινωνία μεταξύ συσκευών (device-to-device/D2D επικοινωνία) είναι ωφέλιμη για την συνολική λειτουργία του δικτύου. Για αυτόν τον σκοπό εισάγεται η έννοια της "περιοχής λειτουργίας" που καθορίζει τα σημεία λειτουργίας του δικτύου (χαρακτηριζόμενα από παραμέτρους όπως πυκνότητα σταθμών βάσης και χρηστών) για τα οποία η εισαγωγή D2D επικοινωνίας ωφελεί. Υπό παραδοχές, η περιοχή λειτουργίας προσδιορίζεται σε απλές κλειστές μορφές για διάφορες εναλλακτικές όσων αφορά την μέθοδο επιλογής μεθόδου πρόσβασης και τον τρόπο πρόσβασης στο κανάλι. Από την ανάλυση προκύπτει ότι με επιλογή μεθόδου πρόσβασης βασιζόμενη στην απόσταση των συσκευών, η εισαγωγή D2D επικοινωνιών είναι ωφέλιμη για όλα τα δυνατά σημεία λειτουργίας. Επιπλέον, η εγκαθίδρυση

D2D ζεύξεων δεν είναι απαραίτητο να περιορίζεται σε πολύ μικρές αποστάσεις μεταξύ των συσκευών, μιας και, υπό συνθήκες, D2D ζεύξεις απόστασης της τάξης της απόστασης από τον κοντινότερο σταθμό βάσης είναι ωφέλιμες.

Στο Κεφάλαιο 4 γίνεται χαρακτηρισμός της στατιστικής της παρεμβολής που προέρχεται από μεταδόσεις διαφορετικής μεθόδου πρόσβασης. Ένα τέτοιο παράδειγμα παρεμβολής είναι αυτή που αντιλαμβάνεται ένας δέκτης που λαμβάνει δεδομένα από ένα σταθμό βάσης λόγω μεταδόσεων άλλων συσκευών (και όχι σταθμών βάσης) που βρίσκονται σε D2D επικοινωνία. Κάνοντας χρήση ενός γενικευμένου μοντέλου συστήματος η στατιστική περιγραφή της παρεμβολής προσδιορίστηκε σε απλές προσεγγιστικές εκφράσεις, οι οποίες έχουν εφαρμογή σε πολλαπλά σενάρια λειτουργίας, όπως οι D2D επικοινωνίες μεταξύ συσκευών που ενδεχομένως να βρίσκονται σε διαφορετικές κυψέλες.

Στο Κεφάλαιο 5 παρουσιάζεται μία τεχνο-οικονομική μελέτη της λειτουργίας του δικτύου υπό την υπόθεση ότι η διαχείρισή του γίνεται από πολλούς ανταγωνιστικούς παρόχους. Μία τέτοια ανάλυση έχει ενδιαφέρον μιας και η μεγάλη πυκνότητα του μελλοντικού δικτύου καθώς και η ευκολία δημιουργίας υπο-δικτύων μεσω λογισμικού, ενθαρρύνουν την είσοδο νέων παρόχων. Με συνδυασμό εργαλείων στοχαστικής γεωμετρίας και θεωρίας παιγνίων προτάθηκε ένα απλό μοντέλο της αλληλεπίδρασης παρόχων και χρηστών. Τα αποτελέσματα της ανάλυσης δείχνουν ότι ο μέσος χρήστης ωφελείται από την παρουσία πολλαπλών ανταγωνιστικών παρόχων της ίδιας αγοραστικής δύναμης, με την τελευταία να προσδιορίζεται από την πυκνότητα και την τεχνολογία των δικτύων. Εν γένει, η παρουσία παρόχων που έχουν στην κατοχή του πυκνά δομημένα δίκτυα ή/και προσφέρουν τις υπηρεσίες του σε χαμηλές τιμές είναι ωφέλιμη για τον μέσο χρήστη. Η ανάλυση επιπλέον δείχνει οτι η αύξηση της πυκνότητας σταθμών βάσης από ένα σημείο και μετά ενδέχεται να μην οδηγεί την αγορά σε σταθερή κατάσταση, ενθαρρύνοντας έτσι την χρησιμοποίηση άλλων, π.χ., συνεργατικών, μοντέλων αλληλεπίδρασης μεταξύ των παρόχων.

Abstract

The proliferation of wireless devices and the introduction of novel applications and services towards the realization of the internet of things (IoT) impose serious challenges for the cellular network operation. It is widely recognized that addressing these challenges will require a revolutionary (rather than evolutionary) upgrade of current cellular networks. The major enabler of this upgrade will be *ultra dense networking*, i.e., operation under extreme infrastructure densification levels, motivated by the need for an aggressive spatial frequency reuse towards providing the high quality of service requested by an extremely large number of user devices. At the same time, this extreme density of devices strongly motivates the introduction of *new channel access methods (modes),* in addition to the traditional uplink/downlink transmissions, with the most prominent example being that of direct device-to-device (D2D) communications. These two features of the future network (densification and new access methods), although, in principle, able to provide performance benefits, also introduce significant challenges that were not present in the previous cellular network generations.

This thesis investigates various aspects of the operation of future ultra dense cellular networks using analytical tools from stochastic geometry. Regarding conventional downlink transmissions, the impact of an increasing infrastructure density on user performance is investigated. It is shown that *infrastructure densification* alone is able to provide considerable gains, however, when supported by a carefully designed intra-cell resource allocation (scheduling) scheme. Even though performance is monotonically increasing with increasing infrastructure density, densification beyond the level where the number of access points is comparable to the number of users provides only moderate gains and, therefore, may not be of interest in practice.

The potential of D2D communications as a technique that is beneficial to the overall system performance is investigated based on the new notion of *operational region of D2D communications*, which identifies the range of system conditions (in terms of parameters such as density of access points and user devices) where D2D communications are beneficial. The operational region is identified in closed form for various combinations of mode selection, spectrum sharing, and channel access schemes. Analysis reveals that the introduction of D2D communications enhances the performance of the conventional (non-D2D-enabled) cellular network under all operational conditions as long as a distance-based mode selection scheme is adopted and D2D transmissions are performed on the same bandwidth as cellular transmissions with appropriate power control.

The critical issue for future cellular networks of characterization of *cross-mode interference*, i.e. interference experienced by a receiver operating on a certain access mode (e.g., cellular downlink) due to transmissions corresponding to other modes (e.g., D2D) is examined. By adopting a general system model, a characterization of *intercell* cross-mode interference statistics is analytically obtained that is applicable for many operational scenarios of interest, ranging from conventional uplink/downlink to cross-cell D2D transmissions, with interference generated by transmissions of different mode(s). The analysis allows for an efficient computation of performance measures of interest such as coverage probability and can serve as an analytical baseline for sophisticated design of, e.g., resource allocation algorithms.

As the extreme infrastructure densification motivates the management of the network by multiple competing operators, a *techno-economic investigation* of the system is performed under the assumption of network partitioning over multiple operators offering their services to self-interested users. By combining tools from stochastic geometry and game theory, it is shown, among others, that users enjoy maximum quality of experience in a market where multiple operators are competing with comparable networks in terms of density and transmission technology. Although increasing infrastructure density is always beneficial from an engineering perspective (e.g., higher user rates), it is shown for the tractable case of two competing operators that stable market conditions cannot be achieved above a certain density threshold, suggesting a different market model (e.g., based on partial cooperation among operators).

List of Publications

The results of this thesis appeared in the following publications:

Journals:

1. S. Stefanatos, A. G. Gotsis, and A. Alexiou, "Cross-mode interference characterization in cellular networks with Voronoi guard regions," *IEEE Transactions on Wireless Communications*, Feb. 2017, submitted.
2. S. Stefanatos and A. Alexiou, "On the competition among small cell wireless operators with large scale deployments," *IEEE Transactions on Mobile Computing*, 2017, to appear.
3. A. G. Gotsis, S. Stefanatos, and A. Alexiou, "Ultra dense networks: The new wireless frontier for enabling 5G access," *IEEE Vehicular Technology Magazine*, vol. 11, no. 2, pp. 71–78, Apr. 2016.
4. S. Stefanatos, A. G. Gotsis, and A. Alexiou, "Operational region of D2D communications for enhancing cellular network performance," *IEEE Transactions on Wireless Communications*, vol. 14 , no. 11, pp. 5984–5997, Nov. 2015.
5. S. Stefanatos and A. Alexiou, "Access point density and bandwidth partitioning in ultra dense wireless networks," *IEEE Transactions on Communications*, vol. 62, no. 9, pp. 3376–3384, Sep. 2014.

Conferences:

1. S. Stefanatos and A. Alexiou, "Duopoly competition between small cell operators with large scale deployments," *IEEE Signal Proces. Adv. In Wireless Commun. (SPAWC)*, UK, Jul. 2016.
2. S. Stefanatos, A. G. Gotsis, and A. Alexiou, "Operational region of overlay D2D communications," *IEEE International Conference on Communications (ICC)*, London, UK, Jun. 2015.
3. S. Stefanatos, A. G. Gotsis, and A. Alexiou, "Analytical assessment of coordinated overlay D2D communications," *European Wireless (EW) conference 2014*, Barcelona, Spain, 14-16 May 2014.
4. S. Stefanatos and A. Alexiou, "Exploiting frequency and spatial dimensions in small cell wireless networks," *IEEE Wireless Communications and Networking Conference (WCNC '14)*, Istanbul, Turkey, Apr. 2014.

Contents

1. Introduction

1.1. Densification as the Major Feature of the Future Cellular Network

Wireless data traffic has increased exponentially in recent years, aided by the introduction of affordable, powerful devices (e.g., laptops, smartphones), as well as new wireless applications towards implementing the Internet of Things (IoT). This trend is expected to continue in the forthcoming years [1], resulting in a highly challenging operational environment for the cellular network. Accordingly, the wireless industry and research community are working towards defining and eventually implementing a new (fifth) generation of cellular networks that is able to support the extremely large data rates per link as well as the extremely large number of concurrent links expected in the future [75]. Towards this end, various technological enablers can be identified [52], which can be roughly partitioned into three major categories.

1. Advanced transmission techniques, e.g., massive MIMO [2] and non-orthogonal multiple access [3];
2. Spectrum increase, e.g., exploitation of mmWave frequencies [4];
3. Extreme densification of infrastructure [6, 5].

The first two categories have been present in the evolution of cellular networks since their beginning and may be viewed as the "standard" approaches towards performance enhancement (of course, this does not imply that there are no significant technical challenges associated with them). However, the advances in transmission techniques and spectrum usage, although necessary, are not sufficient to accommodate for the high demands of the future network. In particular, densification of the infrastructure appears to be the most important aspect of the future architecture towards achieving its goals [6].

The importance of infrastructure densification has already been recognized in current network deployments where the introduction of small-size, small-range access points (APs) is the main method to cope with the high demands in certain geographical areas. This approach has resulted in the introduction of the new cellular networking notions of *small cells* and *heterogeneous networks* [22]. As the network demands continue to increase, it is expected that the future network will consist exclusively of small cells over its whole geographical coverage, thus achieving, in principle,

1. Increased spatial reuse allowing for more transmissions per unit area in the same bandwidth;
2. Reduced AP-to-user distance, therefore, reduced path loss resulting in increased per-link spectral efficiency and/or energy efficiency;
3. Reduced number of users served by the same AP, therefore, increased per-user long-term throughput.

Clearly, network *densification* is the main theme of the future network, both regarding the number of APs as well as user equipments (UEs) per unit area. Although densification of the cellular network was observed in all generations of cellular networks, its manifestation in the future network is expected to have such an impact that will result in a revolutionary (instead of evolutionary) change of a fundamentally stable attribute of the cellular system, namely, its architecture. In accordance, much of the current research on cellular networking focuses on the new architectural features due to (extreme) densification with terms like "ultra dense" or "hyper dense" networks already well-established in the literature (e.g., [7, 8, 51]).

The obvious system attribute affected by the densification is the infrastructure (AP) *density* λ_a, i.e., average number of APs per unit area, assuming a uniform distribution of APs over the area of interest. Not only is the value of λ_a expected to significantly increase in future network deployment but also the previously absurd scenario where λ_a may as well be of the same order as λ_u, the density of user equipments (UEs), is envisaged. The realization of such extreme infrastructure densification is promised not only by the increase of physical APs per unit area, but also by the introduction of new roles to the UEs as well. Indeed, the increasing density of powerful UEs will effectively result in an operational environment where having UEs acting as *virtual* APs themselves can be beneficial to the overall system performance. This new role of UEs is driven by applications/use cases where communicating UEs are in close proximity [53, 55], as well as by the potential exploitation of UE caches (memories) for storing and directly transmitting popular data contents [9, 10]. Towards realizing this new role of the UEs, a completely new concept in cellular networking has recently emerged, namely, direct *device-to-device (D2D) communications*, which has attracted a lot of attention from both industry and academia.

D2D communications constitute one of the major *disruptive* technology directions of the future network [52] as they allow for the establishment of direct links among UEs bypassing the conventional (uplink/downlink) access methods of the cellular network. When D2D communications are *inband*, i.e., performed on the same spectrum as uplink and/or downlink transmissions, the D2D communicating nodes may be viewed as virtual APs, effectively increasing the infrastructure density of the future network thus providing an ingenius means to further harvest the gains offered by densification.

It is clear that the new operational conditions and features of the future wireless network render its performance analysis of critical importance towards identifying

parameters that affect performance and, ultimately, achieving an optimal system design. This is a difficult task as the unprecedented infrastructure densification and new envisioned access methods introduce many new challenges in the system operation that were not relevant in previous cellular generations. It is clear that in order to identify, understand and overcome these challenges, towards an optimal system design, a radically different analysis methodology is required compared to the ones employed previously. The goal of this thesis is exactly that:

Identify and analyze the benefits and challenges of the new architectural and operational features of the future, ultra-dense network, towards achieving an optimal system design.

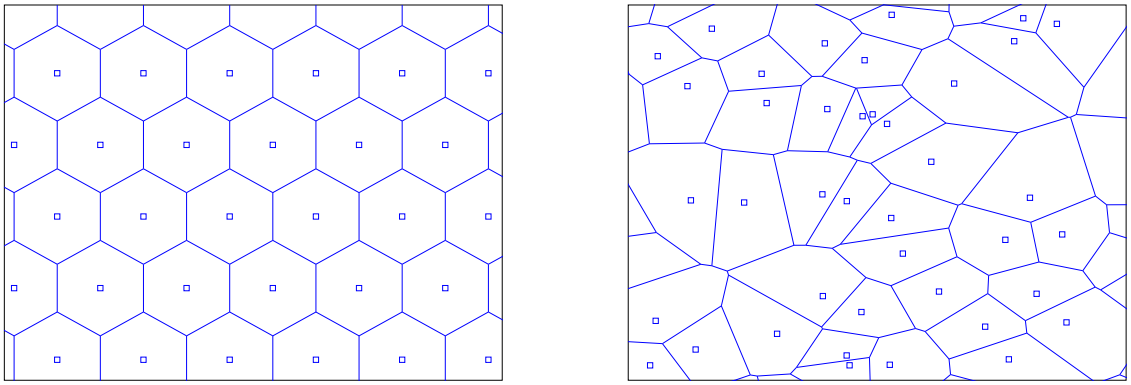
1.2. Challenges in Ultra-Dense Cellular Networks

The new characteristics and features of the future cellular network have a strong impact on all its aspects such as topology, architecture, operation, performance, and even management. This section provides a description of those major challenges of the future network that are the focus of this thesis.

1.2.1. Randomness of AP Positions and Inter-cell Interference

One may argue that harvesting the gains of densification is straightforward, if not trivial: simply increase the density of APs and have each small cell operate exactly as in the case of conventional ("macro") cells with a proportional decrease of transmit powers. By this argument, densification is nothing more than a scaling ("miniaturization") of the network which has no effect on performance apart from offering the advantage of higher and potentially unlimited spatial re-use. Indeed, this was the reasoning that motivated the infrastructure densfication experienced in previous generations of cellular networks. Unfortunately, this argument does not apply for the future network since the *extreme* densification of infrastructure does not only mean increase of the AP density (λ_a) but also *(quasi-)randomness of the AP positions*.

As the density of APs increases, it becomes very difficult to place APs under a well-planed, regular pattern, typically represented by the well-known honeycomb model [11] in cellular engineering (see Fig. 1.1a). This is due to (a) the difficulty of operators installing APs at arbitrary positions (e.g., due to regulatory restrictions) and (b) the continuing introduction of additional APs "on top" of the existing cellular deployment, many of which are user-installed. This results in a highly non-regular spatial pattern for the APs, with minimal to no guarantees on the minimum distance among them, i.e., two APs may be located arbitrarily close to each other (see Fig. 1.1b). The irregularity of the AP placement pattern is further aggravated by the presence of randomly located mobile UEs acting as virtual APs.



(a) **(b)**

Figure 1.1.: Comparison of infrastructure deployments: (a) regular distribution of APs, modeling traditional cellular networks, and (b) random distribution of APs, modeling future dense networks. Marks depict APs (base stations) and lines depict the (geometrical) boundaries of the cells.

It is therefore obvious that extreme densification of infrastructure is *not* equivalent to a scaling of the network geometry. On the contrary, the future network will have a fundamentally different topology/architecture, where the conventional notion of (hexagonal) cell becomes irrelevant. This, in turn, results in a fundamentally different operational environment with respect to the perceived inter-cell interference power, which depicts a (quasi-)random spatial distribution of very large dynamic range [60], even without any random fading effects present. It is intuitively expected, and was indeed verified in [29], that the effect of interference is more severe with an irregular rather than a regular AP deployment (under the same AP density). This can be viewed as the price to pay for the inherent spatial reuse gain due to the increase of λ_a in future networks.

Furthermore, the irregularity of AP positions introduces an additional level of randomness to the network regarding the number of users served by the same AP. For example, with a regular AP position pattern and a uniform spatial distribution of the users, each AP serves more or less the same number of users as every other AP in the network. Since the randomness of AP positions of the future network results in a highly irregular cell pattern (see Fig. 1.1b), the number of users an AP will serve is highly uncertain, depending not only on the user positions but on the positions of its neighbor APs as well.

Clearly, a careful analysis/design of the future network should be based on (new) methodologies and tools that take into account the irregularity of the AP positions and its effect on interference statistics and intra-cell resource sharing.

1.2.2. Coexistence and Selection of Alternative Access Modes

The extreme density of UEs and the envisioned proliferation of IoT applications strongly motivate the incorporation of D2D communications as a means for enhancing link performance as well as system offloading. However, as stated above, this approach transforms UEs into virtual APs, which contribute to the total interference level as well as to the (random) spatial profile of the interference. Even more important, in D2D-enabled cellular networks, the new cellular networking notion of *cross-mode* interference becomes relevant, i.e., interference experienced by a receiver operating on a certain access mode due to transmissions corresponding to different access modes. For example, a downlink transmission is no longer affected by inter-cell interference due to other transmitting APs but also by interference generated by inband D2D transmissions, which may in principle be both inter-cell *and* intra-cell. This implies that a conventional downlink (uplink) transmission may fail due to the presence of a D2D transmitter in the close vicinity of the downlink (uplink) receiver.

It is clear that there exists a trade-off between the benefits and costs of allowing D2D transmissions within the cellular network. A sophisticated system design is therefore required so that D2D transmissions are optimally performed in terms of maximizing both user-level as well as network-level performance. One critical aspect of this design is the specification of the mode selection procedure, i.e., how a UE chooses between establishing a (conventional) cellular link (uplink/downlink) or a D2D link (in case the latter is possible). The mode selection procedure should ideally take into consideration performance aspects of not only the (potential) D2D link involved but also of the cellular links established in the same geographical area. Another important issue to be considered is spectrum sharing between D2D and cellular transmissions, either via means of spectrum partitioning or appropriate power control.

Note that similar remarks as above can be made for the optimization of other access modes proposed for the future cellular network such relay-assisted downlink/uplink transmissions [12] as well operation with full-duplex radios [54]. Insights regarding the co-existance of a D2D mode with conventional uplink/downlink transmissions are also applicable/relevant to these access schemes as the nature of the generated (cross-mode) interference is similar.

1.2.3. Techno-economical Aspects of Large-scale Network Operation

Apart from purely technical aspects, the extreme network densification of the future network will certainly have an impact on how it is operated and managed [76]. In particular, the continuation of the long-standing network management model, where all the infrastructure and spectrum is owned/exploited by a handful of major operators/providers appears highly unlikely. This is due to the high costs related

to the installment and operation of ultra high density infrastructure over large scale areas, as well as the inability of a single network to optimally support the diverse type of future wireless applications. For example, a network designed/optimized towards "conventional", high-rate and latency-tolerant applications (such as file downloading) is very difficult to also efficiently (optimally) support low-rate and latency-sensitive transmissions (tactile internet [13]) requested by a massive number of nodes/machines (IoT). Network fragmentation over multiple operators is further promoted by the recent emergence of the concept of virtualization in wireless networking [77], which transforms infrastructure and spectrum as commodities available to anyone interested in setting up a network from scratch.

An interesting issue that naturally arises in this setting is the determination of the *techno-economic* aspects of the future network in order to understand the profit potential for the multiple operators that are expected to enter the market, as well as how (if at all) competition among operators improves the end-user experience. This analysis is challenging as it has to take into account multiple aspects of the network operation including

- technical (engineering) performance aspects of each subnetwork which, as stated above, are critically affected by the randomness of AP positions;
- economic competition among operators, as reflected by the quality of their offered services and their offered price;
- the implicit correlation among UEs on their choice of "best" operator/provider, since the quality of services offered by each operator largely depends on the load (number of users) it serves.

1.3. Fundamental Questions Regarding Future Network Operation

The previous section identified (some of) the features and challenges of the future cellular network, which are the focus of this thesis. In particular, the work of this thesis attempts to provide answers and insights on the following fundamental questions regarding the future network operation:

1. *How do the benefits of extreme infrastructure density outweigh the costs? Is infrastructure densification alone sufficient to achieve potentially unlimited performance gains?*
2. *Under which system operational conditions, specified in terms of, e.g., user and infrastructure density, is the introduction of D2D communications beneficial for the overall system performance? How should mode (cellular or D2D) selection be performed?*

3. *What is the statistical characterization of cross-mode inter-cell interference in an ultra-dense network? How doe the statistics of the AP-generated interference differ from the UE-generated interference?*
4. *What is the techno-economic network performance under multiple operators management? How beneficial is operator competition to user experience and what are the conditions, in terms of, e.g., density and transmission technology of the competing subnetworks, that lead to stable market conditions where multiple players can achieve profits?*

1.4. Stochastic Geometry as a Basic Tool for Future Network Analysis and Design

Questions such as the above regarding system-level performance analysis and design are traditionally addressed via two methodologies [14]:

1. System-level simulation
2. Optimization theory

System-level simulation is the only method that can provide reliable quantitative insights on the performance of the network, taking into account the multiple complex aspects involved with its operation. However, as the infrastructure density of the future network increases and, along with it, its various design parameters, the simulation approach for system design becomes impractical. This is due to the unacceptably long time it would take to simulate the essentially unlimited possible system configurations. Therefore, in practice, simulations are performed over a limited subset of possible system configurations and mostly towards obtaining insights on system performance rather than design.

Towards obtaining insights on system design, optimization theory is the natural tool. In particular, critical system operational aspects such as, e.g., power/admission control, resource allocation and spectrum sharing, have long been addressed by solving appropriate optimization problems. Unfortunately, the inherent issue with the optimization theory approach is that safe conclusions regarding system operation cannot be obtained. This is because a tractable problem formulation will necessarily consider a limited number of nodes with the interference generated by the rest of the network typically abstracted as a common noise term. As interference in dense wireless networks exhibits significant spatial variability, the validity of the conclusions of the optimization theory approach to a large scale network setting becomes highly questionable [14, 30].

The above observations have resulted in the recent emergence of a third, complementary to the above, methodology for the analysis and design of wireless networks, based on *stochastic geometry* [19, 20, 45, 56, 18]. Stochastic geometry is a broad term referring to the study of phenomena involving *point processes,* i.e., processes

consisting of a countable (and possibly of unbounded cardinality) set of points that are randomly distributed in space according to some probabilistic law [17][1]. Clearly, modeling the positions of mobile users over a large geographical area as a realization of a point process is a standard and well-accepted approach that has been routinely applied in studies on traditional cellular networks [16]. However, as discussed above, future cellular network deployment will consist of irregularly placed APs, which provides a strong motivation for modeling their positions also as a realization of a point process [60]. As the point processes typically considered in stochastic geometry are assumed to be distributed over the whole $\mathbb{R}^2$, their incorporation for modeling AP and user positions also provides a model for deployments extending over large geographical areas and whose size does not allow for investigation via simulation and/or optimization approaches.

Of course, this modeling would be useless if it would not allow for a tractable system analysis and design. Towards obtaining tractable expressions for system performance that, in turn, allow for efficient optimization of their parameters, consideration of *Poisson point processes* as models for the AP and user positions are typically considered in the literature. Poisson point processes are attractive as they are characterized by a single parameter, their *density function* $\lambda : \mathbb{R}^2 \to [0, \infty)$. A typical realization of a Poisson point process with constant density function $\lambda(x) = \lambda, x \in \mathbb{R}^2$, is shown in Fig. 1.1b, which can be generated as follows [18]:

1. For any bounded area $\mathcal{A} \subset \mathbb{R}^2$, the number N of points distributed over $\mathcal{A}$ is a Poisson distributed random variable of mean $\int_{\mathcal{A}} \lambda(x)dx$.
2. The N points are uniformly distributed over $\mathcal{A}$.

The Poisson point process has many convenient analytical properties due to which, it is the "standard" modeling tool of stochastic geometry. The reader is referred to the excellent texts of [17, 18] for a detailed treatment of Poisson processes. Regarding their application to wireless networking, it is clear the the Poisson process is a reasonable model for mobile users positions where no correlation among them exists. As was discussed in [29, 60, 88], a Poisson point process is also a reasonable model for the spatial distribution of APs observed in actual cellular networks deployments.

More importantly, modeling the positions of intereferers as Poisson point distributed allows for a tractable statistical characterization of their generated interference when the latter follows the standard model [45, 14, 18]

$$i_{x_0} = \sum_{x \in \Phi} s_x \sqrt{P_x g_x} \|x - x_0\|^{-\alpha/2},$$

where $\Phi = \{x_1, x_2, \ldots\} \subset \mathbb{R}^2$ is the point process of the transmitting (interfering) nodes (e.g., APs or users), $x_0 \in \mathbb{R}^2 \setminus \Phi$ is the point in space where the interference

[1]Throughout this thesis, APs and UEs will be considered to be distributed over the two-dimensional space (plane) $\mathbb{R}^2$, as is the case in the vast majority of the related literature.

is measured, $s_x \in \mathbb{C}, P_x > 0$ and $g_x \geq 0$ is the unit-power transmitted symbol, transmit power and channel fading gain corresponding to the node positioned at $x \in \Phi$, respectively, $\alpha > 2$ is the path loss exponent, and $\|\cdot\|$ denotes the Euclidean norm. Note that a tractable characterization of the interference is essential towards obtaining characterization of metrics of interest such as the distribution of the *signal-to-interference-ratio* (SIR) defined as

$$\text{SIR} \triangleq \frac{P^* g^* \|x^* - x_0\|^{-\alpha}}{I_{x_0}},$$

where P^*, g^*, x^* are the transmit power, channel gain, and position of the serving transmitter, respectively, and $I_{x_0} \triangleq \sum_{x \in \Phi} P g_x \|x - x_0\|^{-\alpha}$ is the interference *power,* under the common assumption of zero mean, uncorrelated signaling among interfering nodes.

The texts of [45, 18] provide a detailed treatment of how stochastic geometry has been successfully applied towards the performance characterization of wireless networks, notably, mobile *ad hoc* networks (MANETS). However, it should be noted that despite offering the tools to analytically study large scale wireless networks, the stochastic geometry framework does have its limitations. In particular, stochastic geometry analysis

- is usually employed to obtain insights for the performance of a *single* user or AP as correlation properties of performance among nodes is typically not tractable.
- offers limited flexibility for the analysis of sophisticated transmission and/or resource allocation schemes such as cooperative AP transmission and OFDMA scheduling.

For these reasons, stochastic geometry provides a third, supplementary methodology framework for understanding wireless networks, in addition to simulation and optimization theory. However, note that in many studies, such as those aiming to obtain insights on the effect of (large scale) network density, stochastic geometry is the *only* methodology currently available.

Application of stochastic geometry to cellular networking was first presented in the recent seminal work of [29]. This work initiated a multitude of works applying stochastic geometry for the characterization of cellular network, which have aided to establish stochastic geometry as an essential tool for the cellular network engineer/designer. This thesis contributes to this line of works, by applying *stochastic geometry methods to the study of open topics related to future cellular networking.*

1.5. Contributions of this Thesis

Throughout this thesis, the stochastic geometry framework was applied towards obtaining insights on the analysis/design of critical aspects of future cellular networks that were not sufficiently addressed in the related literature. This section presents a summary of the main contributions of this thesis, which are discussed in detail in the next chapters.

1.5.1. Effect of Access Point Density and Bandwidth Partitioning in Ultra-Dense Downlink Cellular Networks

The seminal paper of [29] was the first to apply stochastic geometry tools for the investigation of the properties of a (downlink) cellular system, which spurred the interest of the cellular research community to stochastic geometry methods. One of the major insights given in [29] was that the Poisson point process (PPP) model for the positions of the APs provides a tractable yet reasonably accurate performance prediction of actual irregular AP deployments. In addition, it was shown that under the PPP model with *constant* density λ_a, *the SIR statistics of a random downlink user served by its nearest AP is independent of* λ_a. This unexpected result provides a strong argument in favor of the deployment of ultra dense networks as it implies that although random densification amplifies the effects of interference, at the same time, it provides an improved direct link transmission.

However, in [29], as well as a multitude of relevant subsequent publications, the following critical assumption is made: Each cell operates with one (downlink) user. Although this assumption is sufficient to obtain valuable insights about the SIR and the spectral efficiency of downlink transmissions under worst-case interference (all APs transmit), it does ignore the following aspects of network operation which are especially relevant in ultra dense AP deployments:

1. It is possible that there exist cells with no users to serve, i.e., not all APs in the network transmit,
2. In general, multiple UEs will be served by a single AP and the effect of resource sharing on UE long-term throughput must be taken into account.

Realizing these shortcoming, a more sophisticated system model was considered in this thesis, taking into account both the effects of inactive cells as well as the user throughput loss due to multiple access in cells serving more than one user. Regarding the latter factor, two multiple access schemes were investigated:

- Users within a cell are served in a fair, time-division-multiple-access (TDMA) fashion (no user priorities). For improved interference protection, the available downlink spectrum is partitioned into $N \geq 1$ partitions of equal bandwidth and each downlink transmission is performed on a single randomly selected partition.

- As a means to reduce the throughput cost due to TDMA, a modified version is considered where all users are still served via a single spectrum partition, however, multiple parallel links are established during a single timeslot. In case multiple UEs are served by the same spectrum partition are served via TDMA. The resulting scheme is referred to as TDMA/FDMA and, in principle, provides improved bandwidth utilization per cell, however, at the cost of increased interference (as the probability of a neighbor AP transmitting on the same bandwidth partition(s) is larger).

For both schemes, the distribution of the downlink throughput experienced by a random (typical) user of the system was analytically obtained. The statistical characterization was made in terms of density of APs, density of users, number of spectrum partitions and path loss factor. Investigation of the analytical expressions reveals the following:

1. Although TDMA offers improved interference protection, FDMA/TDMA is preferable in terms of user throughput when the number of partitions, N, is optimally selected.

2. The optimal value of N that maximizes the probability of a random user achieving a rate r_0 (in b/s/Hz) scales inversely proportional to r_0.

3. Although previous studies showed that the user SIR performance is independent of λ_a, increasing the ratio λ_a/λ_u (density of APs to density of UEs) with optimal selection of N improves the user rate due to the reduced number of users sharing the resource of a single AP. The statistical characterization of the user rate depends only on the value of the ratio λ_a/λ_u and *not* on their absolute values.

4. Given a requested user long-term throughput value r_0, the probability of the rate outage event decreases as the ratio λ_a/λ_u increases. The minimum value of the ratio λ_a/λ_u required to achieve a maximum allowable rate outage probability can be analytically upper bounded. The upper bound is tight and its asymptotic expansion shows that in the regime of $\lambda_a/\lambda_u \ll 1$, linear increase of the ratio λ_a/λ_u results in a linear increase of the user throughput (for a fixed outage probability), whereas in the regime $\lambda_a/\lambda_u \geq 1$, the density ratio must increase exponentially in order to achieve a linear increase of the requested throughput. This observation suggests that unlimited performance gains by means of infrastructure densification alone are practically impossible, i.e., densification should be complemented by advances in spectrum availability and/or spectral efficient techniques.

The detailed study is presented in chapter 2.

1.5.2. Characterization of Operational Region where Introduction of D2D Communications Enhances Cellular Network Performance

It is clear that D2D communications are beneficial at virtually no cost when applied for the establishment of direct links among communicating nodes lying very close to each other. In that scenario, the D2D link provides a much higher spectral efficiency (compared to conventional uplink/downlink) due to the reduced transmission distance. This topological advantage also allows for a minimum transmission power that results in insignificant interference to conventional cellular links performed in close distance. However, a natural question that arises is how often does this communication scenario happen in practice? As the concept of D2D communications is at its infancy and related "killer" applications are not yet available, the investment of enhancing the cellular network and devices to accommodate for D2D transmissions appears questionable.

On the other hand, by extending the range of D2D transmissions, the scenarios where they can be employed increases proportionally. However, this is achieved with the cost of an increased D2D transmit power and, therefore, D2D-generated (cross-mode) interference. It is therefore critical to understand the *operational conditions* where the introduction of D2D communications is beneficial to the cellular system overall. Such a study will indicate if D2D communications should be restricted to only proximal communications with limited use/business cases or they can provide benefits over a broader range of scenarios, which would justify their introduction to the network and promote the introduction of D2D-based applications and new use/business cases.

Realizing that the literature did not provide any clear insights on the above considerations, especially for large scale network deployments, a system analysis based on stochastic geometry was performed in order to identify the operational region where D2D communications are beneficial. In particular, the operational region is characterized by the following set of parameters:

- density of (both traditional cellular and *potential* D2D) users
- density of APs
- propagation conditions (path loss factor)
- signal-to-inteference-ratio (SIR) threshold for successful communications
- maximum allowed D2D link distance

Note that the possible combinations of values of the above parameters can cover a multitude of use/business cases of interest to operators taking into consideration the critical parameters related to D2D communications, namely, density of potential D2D users and maximum allowed D2D link distance.

By considering two simple mode-selection schemes, an Aloha-based D2D channel access scheme, as well as underlay and overlay D2D communications (i.e., performed

inband or outband to the cellular spectrum, respectively), the operational region of D2D communications was characterized in *closed form* . The figure of merit used was the average rate experienced by a random user of the network (that may or may not be able to establish a D2D link). Important outcomes of this analysis are:

1. D2D communications can provide benefits for operational conditions of practical interest including those with D2D link distances of the order of the user-to-closest-AP distance. This observation strongly motivates the introduction of D2D communications and development of D2D-based applications and use/business cases.
2. Employing both a non-trivial mode selection and a non-trivial D2D channel access procedure does not necessarily provide significant (if any) gains compared to employing one of them only.
3. Exploitation of D2D link distance information alone for mode selection purposes is sufficient for D2D communications to be beneficial for *any* operational scenario.

The detailed study is presented in chapter 3.

1.5.3. Statistical Characterization of Cross-Mode Interference in Cellular Networks with Voronoi Guard Regions

When the maximum D2D link distance is very short, the generated interference to the traditional cellular links becomes very small. This allows for the establishment of multiple parallel D2D links over the whole coverage area of the cellular network, essentially resulting in an independent network whose nodes are distributed similarly to an ad hoc network and transmit on the same spectrum as the cellular network. However, as noted above, it may be advantageous to allow for the establishment of D2D links between nodes whose distance is comparable to the user-AP distance. This, in turn, implies a significant level of cross-mode interference. In order to reduce cross-mode interference it is natural to consider a cellular-network-controlled operation, where each AP controls the transmissions generated by all the nodes within its cell, e.g., allowing for only one active link (either D2D or cellular) per time slot.

In this network-controlled operation where the intra-cell interference is controlled, the inter-cell interference becomes the major limiting factor of performance. In order to analyze and design future networks under this approach, it is critical to understand the properties of the inter-cell, cross-mode interference. Note that, in contrast to previous generations of cellular networks, the (Voronoi) cell of each AP is random due to the random infrastructure deployment (see Fig. 1.1b), which makes the analysis complicated.

Towards a tractable analytical characterization of cross-mode inter-cell interference, a stochastic geometry model was considered and the interference power measured at

an arbitrary position due to either AP or user (D2D) transmissions was investigated. The model is general enough to capture many communication scenarios of interest, from the conventional downlink/uplink to *cross-cell* D2D communications where the receiver of a D2D link does not lie in the same link as the corresponding D2D transmitter. The major outcomes of this investigation are the following:

1. For the case of AP-generated intra-cell interference power, its exact statistical characterization is obtained in closed form. For the case of user-generated inter-cell interference power, a lower bound of its Laplace transform is provided, that can be used to obtain a lower bound on the SIR coverage probability under the common Rayleigh fading assumption.
2. The notion of equivalent interferer density is introduced, which allows for a simple and intuitive understanding of the properties of cross-mode interference and how they depend on the position of the receiver of interest (relative to its closest AP) and the type of experienced interference (AP or user generated).
3. AP-generated and user-generated inter-cell interference demonstrate significant qualitative differences. For example, even though the interference experienced by a downlink receiver due to interfering APs decreases when its distance from its serving AP increases (due to increasing cell area), this is not the case for the user-generated interference as the downlink receiver is more likely to be located at the cell edge and becomes more susceptible to user-generated interference from nearby cells.
4. The analytical interference characterization allows for efficient computation of metrics of interest such as the SIR coverage probability for various operational scenarios of interest as the ones described above. This allows, in turn, for a sophisticated, yet efficient, optimization of system parameters in operational scenarios with heavy cross-mode intra-cell interference.

The detailed this study is presented in chapter 4.

1.5.4. Techno-economic Properties of the Future Network with Multiple Competing Operators

A notable common feature of previous works investigating the techno-economic aspects of wireless networks managed by multiple competing operators (OPs) is that they, either explicitly or implicitly, consider only a small geographical area that each OP can cover by means of a single AP. Therefore, generalization of results for the case of deployments over wide geographical areas essentially requires the assumption that all OPs deploy identical networks with respect to both number and position pattern of APs. Clearly, this is not a realistic assumption, even more so in the case of future networks resulting from irregular AP deployments. A more realistic system model and analysis should take into account the effect of both different

deployment densities among competing OP networks as well as their quasi-random spatial distribution.

Towards this end, a mathematically rigorous techno-economic investigation of the system was performed by combining tools from stochastic geometry and game theory. A large-scale network was considered (mathematically modeled as infinite), with networks managed by different OPs having a constant, but otherwise arbitrary density over the plane, and their APs distributed as independent Poisson point processes. The positions of the users requesting service are also modeled as an independent Poisson point process, and each user is served by the closest AP of the OP that is subscribed to. Tractable user utilities were considered, reflecting the level of service (expressed in terms of average throughput) and price paid to OP, whereas the OP utilities depend on the number of subscribers (market share) and the subscription price.

By employing a two-step (leader-follower) game formulation where the OPs first announce their prices/services and then users subscribe to them accordingly, the game-theoretic market equilibrium was investigated. The equilibrium is characterized by the OP and user strategies for which no entity (player) involved in the game has any incentive to change it (in order to achieve higher utility). Analysis focused on how different OP deployment densities affect the equilibrium and under which conditions can an OP monopoly be avoided. An OP monopoly is undesirable as the lack of competition in that case is expected to result in poorer user experience (utility). The main outcomes of this analysis are the following:

1. The distribution (partitioning) of user population to the various competing OPs as well as the utility experienced by users was obtained in tight closed-form bounds. These formulas verify the intuition that users are benefited by the presence of multiple OPs with large deployment densities and/or offering their services for a small price.

2. For the case of two competing OPs, the notion of *duopoly region* was introduced, consisting of the set of deployment density pairs where both OPs are able to achieve profit (given choices on service pricing and transmission technology). Analysis of the duopoly region shows that for scenarios where one OP has significantly larger density than the competitor, a duopoly cannot be established, i.e., the smaller density OP drops from the market as there are no profit margins. Interestingly, a duopoly can be established when both OPs have the same density, however, as long as the common density value is smaller than a certain maximum value above which no market equilibrium can be established. This implies that, although extreme infrastructure densification is desirable from an engineering user-end perspective, it may result in unstable market conditions under the competition model considered. This provides a motivation towards a different market model in case of extreme densification, e.g., based on (partial) cooperation among OPs.

3. User utilities are maximized when competing subnetworks have the same char-

acteristics (density/technology). Interestingly, when the total density of APs from all OPs is small to moderate and under the competition model considered in this work, users can actually enjoy a service level very close to the utopia scenario where OPs are cooperating towards maximizing user experience without regards of their own profit. In other words, competition among equally-powered multiple operators is beneficial to the end users, which motivates, e.g., regulatory entities, towards providing appropriate OP incentives for taking part in this competitive market.

The detailed study is presented in chapter 5.

2. Access Point Density and Bandwidth Partitioning in Ultra Dense Wireless Networks

2.1. Overview

This chapter examines the impact of system parameters such as access point density and bandwidth partitioning on the performance of randomly deployed, interference-limited, dense wireless networks. While much progress has been achieved in analyzing randomly deployed networks via tools from stochastic geometry, most existing works either assume a very large user density compared to that of access points, which does not hold in a dense network, and/or consider only the user signal-to-interference-ratio as the system figure of merit, which provides only partial insight on user rate as the effect of multiple access is ignored. In this chapter, the user rate distribution is obtained analytically, taking into account the effects of multiple access as well as the SIR outage. It is shown that user rate outage probability is dependent on the number of bandwidth partitions (subchannels) and the way they are utilized by the multiple access scheme. The optimal number of partitions is lower bounded for the case of large access point density. In addition, an upper bound of the minimum access point density required to provide an asymptotically small rate outage probability is provided in closed form.

The results of this chapter were published in

1. S. Stefanatos and A. Alexiou, "Access point density and bandwidth partitioning in ultra dense wireless networks," *IEEE Transactions on Communications*, vol. 62, no. 9, pp. 3376–3384, Sep. 2014.

2. A. G. Gotsis, S. Stefanatos, and A. Alexiou, "Ultra dense networks: The new wireless frontier for enabling 5G access," *IEEE Vehicular Technology Magazine*, vol. 11, no. 2, pp. 71–78, Apr. 2016.

3. S. Stefanatos and A. Alexiou, "Exploiting Frequency and Spatial Dimensions in Small Cell Wireless Networks," *IEEE Wireless Communications and Networking Conference (WCNC '14)*, Istanbul, Turkey, Apr. 2014.

2.2. Introduction

Small cell networks have attracted a lot of attention recently as they are considered a promising method to satisfy the ever increasing rate demands of wireless users. Some studies have suggested that, by employing low cost access points (APs), the density λ_a of APs will potentially reach, or even exceed, the density λ_u of user equipments (UEs), therefore introducing the notion of ultra dense wireless networks [15]. With a large number of APs available, a random UE will most probably connect to a strong signal AP, having to share the AP resources with a limited number of co-served UEs, and, ultimately, achieve high rates. However, in order to exploit the full system resources, a universal frequency reuse scheme is employed which inevitably results in significant interference that has to be taken carefully into account in system design and performance analysis.

2.2.1. Related Work and Motivation

With an increasing network density, the task of optimally placing the APs in the Euclidean plane becomes difficult, if not impossible. Therefore, the APs will typically have an irregular, random deployment, which is expected to affect system performance. Recent research has showed that such randomly deployed cellular systems can be successfully analyzed by employing tools from stochastic geometry [18, 21]. While significant results have been achieved, most of these works assume $\lambda_u \gg \lambda_a$, effectively ignoring UE distribution, and/or consider only the user signal-to-interference ratio (SIR). Assumption $\lambda_u \gg \lambda_a$ does not hold in the case of dense networks, whereas SIR provides only partial insight on the achieved user rate as the effect of multiple access is neglected [22].

A few recent works have attempted to address these issues. Specifically, the UE distribution is taken into account in [38, 24] by incorporating in the analysis the probability of an AP being inactive (no UE present within its cell). However, analysis considers only the SIR. In [48, 47] the UE distribution is employed for computation of user rates under time-division-multiple-access (TDMA) without considering the effect of SIR outage. In addition, TDMA may not be the best multiple access scheme under certain scenarios.

Partitioning the available bandwidth and transmitting on one of the resulting sub-channels (SCs), i.e., frequency-division-multiple-access (FDMA), has been shown in [27] to be beneficial for the case of ad-hoc networks assuming a channel access scheme where each node transmits independently on a randomly selected SC. This decentralized scheme was employed in [79] for modeling the uplink of a cellular network with frequency hopping channel access. However, this approach is inappropriate for a practical cellular network where scheduling decisions are made by the AP and transmissions are orthogonalized to eliminate intra-cell interference (no sophisticated processing at receivers is assumed that would allow for non-orthogonal

transmissions). A straightforward modification of the bandwidth partitioning concept for the downlink cellular network was considered in [29] where UEs are multiplexed via TDMA and transmission is performed on one, randomly selected SC. This simple scheme was shown to provide improved SIR performance, however, with no explicit indication of how many partitions should be employed or how performance would change by allowing more that one UEs transmitting at the same time slot on different SCs.

2.2.2. Contributions

In this chapter, the stochastic geometry framework is employed for analyzing the downlink user rate of a dense wireless network under a multiple access scheme that exploits bandwidth partitioning for *both* interference reduction and efficient resource sharing among UEs. The previously mentioned issues are explicitly addressed by considering in the analysis

- the UE distribution,
- a multiple access scheme that allows for parallel orthogonal transmissions in frequency,
- the effect of SIR outage.

Under this framework, the user rate distribution is analytically derived for two instances of multiple access schemes that reveals dependence of performance on the number of bandwidth partitions as well as the way the are utilized. The analytical rate distribution expression, apart from allowing for efficient numerical optimization of system parameters, is employed to derive a closed-form lower bound of the optimal number of partitions for the case of large AP density, as well as a closed-form upper bound of the minimum AP density required to provide a given, asymptotically small, rate outage probability. The latter is of critical importance given the trend of AP densification in future wireless networks. Numerical results demonstrate the merits of increased AP density, as well as efficient use of bandwidth partitions, in enhancing network performance in terms of achieved user rate.

2.3. System Model and Performance Metrics

The downlink of an interference-limited, dense wireless network is considered. Randomly deployed over $\mathbb{R}^2$ APs and UEs are modeled as independent homogeneous Poisson point processes (PPP) Φ_a, Φ_u, with densities λ_a, λ_u, respectively. Full buffer transmissions and Gaussian signaling are assumed, with interference treated as noise at the receivers. Each UE is served by its closest AP resulting in irregular, disjoint cell shapes forming a Voronoi tessellation of the plane [18]. Elimination of intra-cell interference is achieved by an orthogonal FDMA/TDMA scheme with the

total system bandwidth partitioned offline to N equal size SCs. All active APs in the system transmit at the same power over all (active) SCs, with the power selected appropriately large so that the system operates in the interference limited region in order to maximize spectral efficiency [30]. No coordination among APs is assumed, i.e., each AP makes independent scheduling decisions.

Considering a typical UE located at the origin and served by its closest AP of index, say, 0, the SIR achieved at SC $n \in \{1, 2, \ldots, N\}$ is given by

$$\mathrm{SIR}_n = \frac{g_{0,n} r_0^{-\alpha}}{\sum_{i \in \Phi_a \setminus \{0\}} \delta_{i,n} g_{i,n} r_i^{-\alpha}}, \tag{2.1}$$

where $r_0 \geq 0$ is the distance from the serving AP, $\alpha > 2$ the path loss exponent, and $g_{0,n} \geq 0$ an exponentially distributed random variable with unit mean, modeling small scale (Rayleigh) fading. The denominator in (2.1) represents the interference power at the considered SC, where $g_{i,n}$, r_i are the channel fading and distance of AP i w.r.t. the typical UE, respectively, and $\delta_{i,n} \in \{0, 1\}$ is an indicator variable representing whether AP i transmits on SC n. Note that $\delta_{i,n}$ depends on the total number of UEs associated with AP i as well as the multiple access scheme and its presence in (2.1) is to account for APs that do not interfere due to lack of associated UEs and/or scheduling decisions. Channel fadings $\{g_{i,n}\}$ are assumed independent, identically distributed (i.i.d.) w.r.t. AP index i.

SIR_n is a random variable due to the randomness of fading, AP and UE locations, as well as the multiple access scheme, and its statistical characterization is of interest. To this end, the interference term of (2.1) can be viewed as shot-noise generated by a marked PPP [18] of density λ_a outside a ball of radius r_0 centered at the origin, and marks $\{g_{i,n}, \delta_{i,n}\}$. Statistical characterization of a marked PPP can be obtained by standard methods when the following conditions hold [18]:

1. marks are mutually independent given the location of points, and,
2. each mark depends only on the location of its corresponding point.

Channel fadings $\{g_{i,n}\}$ satisfy both conditions by assumption, whereas variables $\{\delta_{i,n}\}$ satisfy only the first due to their dependence on the total number of UEs associated with each AP. By fundamental properties of the PPP, the numbers of UEs associated with different APs are mutually independent since AP cells are disjoint. For each AP, the number of associated UEs is determined by λ_u and its cell area, with the latter depending not only on its own position but also on the position of its neighbours APs as well, rendering condition (2) invalid for $\{\delta_{i,n}\}$. In order to obtain tractable expressions for the SIR distribution the following assumption is adopted:

Assumption 2.1. The number K of UEs associated with a random AP is independent of Φ_a.

Note that this assumption is actually stronger than the second condition but is convenient as it allows for incorporating the averaged-over-Φ_a probability mass function (PMF) of K that will be used later in the analysis, given by the following lemma:

Lemma 2.2. *The PMF of the number K of UEs associated with a randomly chosen AP, averaged over the statistics of Φ_a, is [92]*

$$\mathbb{P}(K) = \frac{3.5^{3.5}\Gamma(K+3.5)\tau^{3.5}}{\Gamma(3.5)K!(1+3.5\tau)^{K+3.5}}, K \geq 0, \tag{2.2}$$

where $\Gamma(\cdot)$ is the Gamma function and $\tau \triangleq \lambda_a/\lambda_u$.

Note that $\mathbb{P}(K)$ is a decreasing function of τ and the mean of K equals $1/\tau$. The above approach was shown in [38, 24] to provide accurate results and will be validated by simulations in Sect. III.

Under Assumption 1, $\{\delta_{i,n}\}$ are i.i.d. over i and characterized by the *activity probability* $p_n \triangleq \mathbb{P}(\delta_{i,n} = 1) \in (0, 1], \forall i$, whose actual value will be investigated in Sect. III for specific multiple access schemes. The cumulative distribution function (CDF) of SIR_n can now be obtained in a simple expression as given by the following lemma:

Lemma 2.3. *The CDF of SIR_n under an activity probability p_n is given by [29]*

$$F_{SIR_n}(\theta) \triangleq \mathbb{P}(SIR_n \leq \theta) = 1 - \frac{1}{1 + p_n\rho(\theta)}, \tag{2.3}$$

for $\theta \geq 0$, where $\rho(\theta) \triangleq \theta^{2/\alpha} \int_{\theta^{-2/\alpha}}^{\infty} 1/\left(1 + u^{\alpha/2}\right) du$.

Note that setting p_n *a-priori* equal to 1, as in, e.g., [48, 47], implies that there is always a UE available to be allocated in every AP of the system, i.e., $\lambda_u \gg \lambda_a$, which is not the case in dense networks. For example, for the case $\lambda_a = \lambda_u$ and noting that $p_n \leq \mathbb{P}(K > 0)$ for any multiple access scheme, it follows from (2.2) that $p_n \leq 0.58$.

Knowledge of (2.3) is of importance as it provides the probability $F_{\mathrm{SIR}_n}(\theta_0)$ of service outage due to inability of UE operation below SIR threshold θ_0 whose value may be dictated by operational requirements, e.g., synchronization, and/or application (QoS) requirements. In addition, $F_{\mathrm{SIR}_n}(\theta)$ can be used to obtain CDFs of other directly related quantities of interest by transformation of variables. One such quantity employed extensively in the related literature, e.g., [24, 29], is the rate $\overline{R}_n$ achieved *per channel use*, i.e, on a single time slot, given by

$$\overline{R}_n \triangleq \frac{1}{N}\log_2(1 + \mathrm{SIR}_n) \text{ (b/s/Hz)}. \tag{2.4}$$

Examining $\overline{R}_n$ is important from the viewpoint of system throughput [24] but provides little insight on the achieved user rate. Note that $\overline{R}_n$ is an upper bound on the actual user rate. In case when the considered SC has to be time-shared among UEs, rate will only be a fraction of $\overline{R}_n$. In an attempt to remedy this issue, $\overline{R}_n$ was divided by the (random) number of UEs sharing the SC in [48, 47] (case of $N = 1$ was only considered). However, this is still a misleading measure of performance

as it does not take into account the probability of an SIR outage and, therefore, provides overconfident results.

In order to avoid these issues, the achieved rate of a typical UE is defined as

$$R_n \triangleq \mathbf{1}\{\mathrm{SIR}_n \geq \theta_0\}\frac{\overline{R}_n}{(L_n+1)} \text{ (b/s/Hz)}, \tag{2.5}$$

where $\mathbf{1}\{\cdot\}$ is the indicator function and integer $L_n \geq 0$ is the number of time slots between two successive transmissions to the typical UE, referred to as *delay* in the following. Clearly, (2.5) takes into account both the effects of multiple access and SIR outage via L_n and the indicator function, respectively. In order to obtain the rate outage probability, i.e., the CDF of R_n, (statistical) evaluation of p_n and L_n is required, both depending, in addition to the UE distribution, on the multiple access scheme that is investigated in the following section.

2.4. Effect of Multiple Access on Achievable User Rates

In this section the effect of multiple access on the achievable user rate is investigated. The multiple access scheme must strive to maximize UE resource utilization, while at the same time minimize inter-cell interference. These are conflicting requirements which, as it will be shown, can be (optimally) balanced by the choice of N. For analytical purposes the two schemes considered below are non-channel aware, with the corresponding performance serving as a lower bound under a channel aware resource assignment scheme, and fair, i.e., there are no priorities among UEs.

2.4.1. TDMA

The following scheme, referred to in the following as TDMA, will serve as a baseline.

Algorithm 2.1 TDMA

- UEs are multiplexed via TDMA.
- Transmission to any UE is performed on one, randomly selected SC out of total N.

This scheme is a straightforward application of the random SC selection scheme employed in adhoc studies [27] to the downlink cellular setting, and can be also viewed as a generalization of conventional TDMA ($N = 1$) that is usually assumed in works on cellular networks. It was first examined in [29], where it was shown that it provides improved SIR coverage by using essentially the same principle as in a frequency hopping scheme.

2.4.2. FDMA/TDMA

The major argument against TDMA is the inability of parallel transmissions in frequency by multiple UEs when $N > 1$, which is expected to be beneficial under certain operational scenarios. In this chapter, a simple modification, described at the top of the next page, is employed, referred to as FDMA/TDMA in the following, that allows for multiple UEs served at a single time slot ($\lfloor \cdot \rfloor$ denotes the largest smallest integer operator).

Algorithm 2.2 FDMA/TDMA

- Define $\mathcal{N} \subseteq \{1, 2, \ldots, N\}$ the set of available SCs for allocation at any given instant.
- Randomly order the K cell users via an index $k \in \{1, 2, \ldots, K\}$.

```
for L = 0 to ⌊K/N⌋ do
    N ← {1, 2, ..., N};
    for k = LN + 1 to min{LN + N, K} do
        Assign UE k a random SC n_k ∈ N;
        N ← N\{n_k};
    end for
end for
```

- UEs sharing the same SC are multiplexed via TDMA.

Note that the FDMA/TDMA scheme also subsumes conventional TDMA as a special case. Two typical realizations of the scheme for $N = 3$ are shown in Fig. 2.1. As can be seen, there will be cases with unused SCs ($K < N$) or SCs that support one additional UE compared to others ($K > N$) with no action taken to compensate for these effects. The inefficient bandwidth utilization is not a real issue since presence of unused SCs is beneficial in terms of reduced interference and also N is variable that can be set to a small enough value so that this event is avoided, if desired. The load imbalance among SCs is irrelevant for rate computations due to averaging.

Having specified the multiple access schemes, the corresponding quantities p_n and L_n will be evaluated in the following subsections. By the symmetry of the system model and the schemes considered, all the performance metrics presented in Sect. II do not depend on n. Therefore, the typical UE will be considered assigned to SC 1 and index n will be dropped from notation in the following.

2.4.3. Computation of Activity Probability

The following lemma holds for the activity probability p of TDMA and FDMA/TDMA.

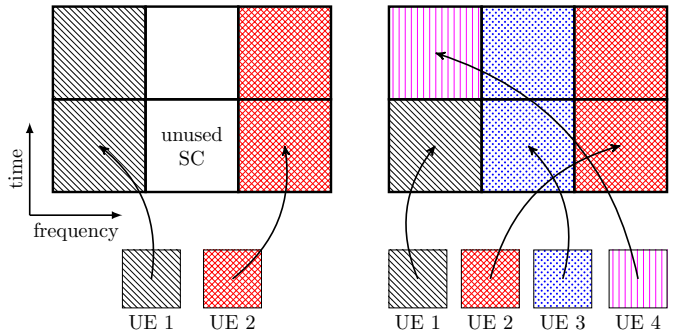


Figure 2.1.: Typical realizations of the FDMA/TDMA scheme, $K < N$ (left) and $K > N$ (right).

Lemma 2.4. *Under assumption 1, the activity probability of any AP in the system, other than* 0*, is*

$$p = \begin{cases} \frac{1}{N}\mathbb{P}(K>0) & \text{, for TDMA,} \\ \frac{1}{N}\sum_{K>0}\mathbb{P}(K>0)\min\{N,K\} & \text{, for FDMA/TDMA,} \end{cases} \tag{2.6}$$

with $\mathbb{P}(K)$ *as given in Lemma 1.*

Proof. See Appendix sec. A.1. □

As expected, p is a decreasing function of N in both cases, and it can be easily shown that, for the same $N > 1$, p of FDMA/TDMA is lower bounded by the corresponding p of TDMA with equality when $\mathbb{P}(K \leq 1) = 1$, i.e., with a dense AP deployment. Note that in [29], p for TDMA was set equal to $1/N$, implying that $\mathbb{P}(K > 0) = 1$, which is (approximately) valid only for $\tau \to 0$. Substituting (2.6) in (2.3) shows that $F_{\text{SIR}}(\theta)$ is decreasing in N, i.e., bandwidth partitioning improves performance in terms of SIR.

Fig. 2.2 shows the behavior of p as a function of N for FDMA/TDMA and TDMA and various values of τ. Conventional TDMA performance corresponds to $N = 1$. It can be seen that both schemes outperform conventional TDMA, with larger values of N required to obtain the same p under heavier system load. TDMA is always better than FDMA/TDMA, especially under heavy system load. For $\tau = 10$, both schemes essentially operate exactly the same and this is reflected on the values of p. For $\tau = 1$, FDMA/TDMA is worse than TDMA but relatively close, with similar dependence on N (inversely proportional).

Remark: According to the previous discussion, the SIR grows unbounded with increasing τ and/or N, which is unrealistic. However, arbitrarily large values of τ are not of interest due to practical considerations, whereas arbitrarily large values of N are not acceptable from a user rate perspective as will be shown in Sect. IV (Lemma 6).

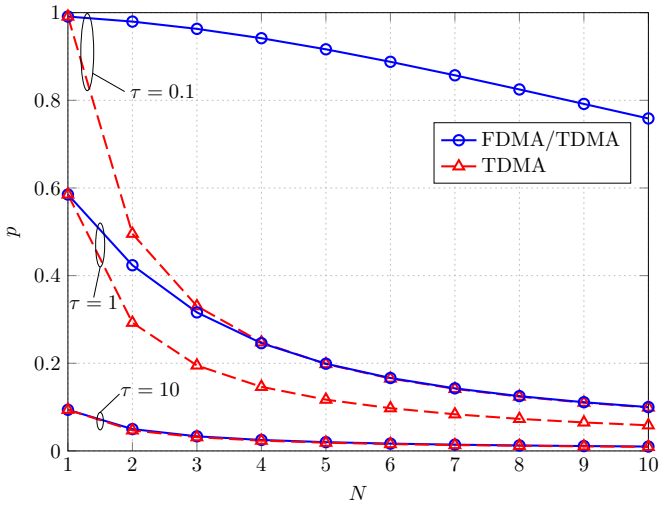


Figure 2.2.: Activity probability p for FDMA/TDMA and TDMA.

2.4.4. Computation of Delay

Computation of delay requires knowledge of the distribution of the total number K_0 of UEs associated with AP 0, *in addition to* the typical UE. The PMF of (2.2) does not hold for AP 0 as conditioning on its area covering the position of the typical UE makes it larger than the cell area of a random AP [18]. Taking this fact into account, the PMF of K_0 can be shown to be given as in the following lemma.

Lemma 2.5. *The PMF of the number K_0 of UEs associated with AP 0, in addition to the typical UE, is [92]*

$$\mathbb{P}(K_0) = \frac{3.5^{4.5}\Gamma(K_0+4.5)\tau^{4.5}}{\Gamma(4.5)K!(1+3.5\tau)^{K_0+4.5}}, K_0 \geq 0. \tag{2.7}$$

For the case of TDMA, it is clear that $L = K_0$, whereas L for FDMA/TDMA is given in the following lemma.

Lemma 2.6. *Define the event $\mathcal{A}_l \triangleq \{l$ UEs assigned on SC 1 in addition to the typical UE$\}$, $l \geq 0$. The PMF of L for FDMA/TDMA equals*

$$\mathbb{P}(L) = \sum_{K_0 \geq 0} \mathbb{P}(K_0)\mathbb{P}(\mathcal{A}_L|K_0), \tag{2.8}$$

with $\Pr\{K_0\}$ as given in Lemma 4,

$$\mathbb{P}(\mathcal{A}_0|K_0) = \begin{cases} 1 & ,0 \leq K_0 \leq N-1, \\ \frac{2N-K_0-1}{N} & ,N \leq K_0 \leq 2N-2, \\ 0 & ,K_0 \geq 2N-1, \end{cases}$$

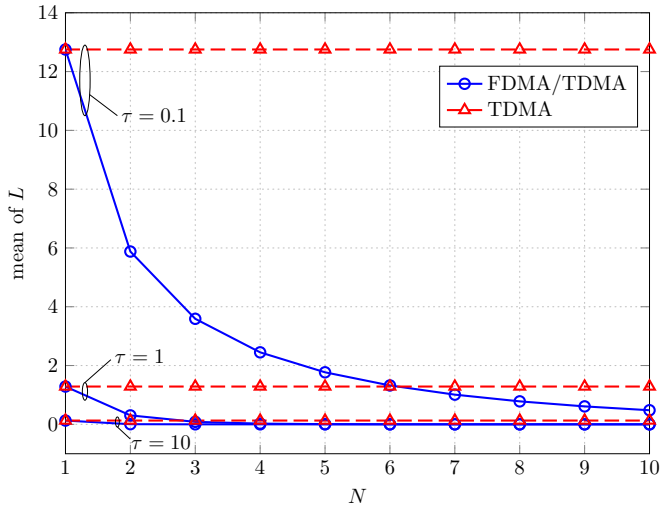


Figure 2.3.: Mean value of delay L for FDMA/TDMA and TDMA.

and

$$\mathbb{P}(\mathcal{A}_l|K_0) = \begin{cases} 0 & ,0 \leq K_0 \leq lN-1, \\ \frac{K_0 - lN + 1}{N} & ,lN \leq K_0 \leq (l+1)N-1, \\ \frac{(l+2)N - K_0 - 1}{N} & ,(l+1)N \leq K_0 \leq (l+2)N-2, \\ 0 & ,K_0 \geq (l+2)N-1, \end{cases}$$

for $l \geq 1$.

Proof. Follows by the same arguments as in the proof of Lemma 2.4. □

Fig. 2.3 shows the mean value of L as a function of N for FDMA/TDMA and TDMA and various values of τ. Both schemes provide reduced delay by increasing τ, since the number of UEs associated with the AP is reduced. For the case of FDMA/TDMA, average L decreases also with N as there are more SCs available to UEs and the probability of time sharing one of them by many UEs is reduced. On the other hand, for TDMA, L is independent of N since the availability of SCs is not exploited for parallel transmissions.

2.4.5. Computation of Rate

Obviously, FDMA/TDMA is advantageous when delay is considered, whereas, TDMA is more robust to interference. However, it is the achieved UE rate that is of more interest and, at this point, there is no clear indication which of the two schemes is

preferable under this performance metric, i.e., what is of more importance, robustness to interference or efficient multiple access. Having specified the statistics of p and L, the CDF of R can now be obtained for both multiple access schemes as follows.

Proposition 2.7. *The CDF of R equals*

$$F_R(r) = \sum_{L \geq 0} \mathbb{P}(L) F_R(r|L), \tag{2.9}$$

with $\mathbb{P}(L)$ as given in Sect. III. D and $F_R(r|L)$ the CDF of R conditioned on the value of L, given by

$$F_R(r|L) = \begin{cases} F_{SIR}(\theta_0) & , r \leq \frac{\bar{R}_{\theta_0}}{N(L+1)} \\ F_{SIR}\left(2^{rN(L+1)} - 1\right) & , r \geq \frac{\bar{R}_{\theta_0}}{N(L+1)} \end{cases} \tag{2.10}$$

where $\bar{R}_{\theta_0} \triangleq \log_2(1+\theta_0)$ is the minimum achievable rate for $N = 1$ and $K_0 = 0$ (no contending UEs), conditioned on UE operation above SIR threshold θ_0.

Proof. See Appendix sec. A.2. □

Note that the upper term of (2.10) indicates that for small values of r, rate outage probability coincides with the SIR outage probability, irrespective of the actual value of r, as for this rate region it is the SIR outage event (strong interference) that prevents UEs from achieving these rates. For larger rates, L appears in the lower term of (2.10), i.e., multiple access also affects performance in addition to interference.

Fig. 2.4 shows $F_R(r)$ for $\theta_0 = 0$ dB, $\alpha = 3$, $N = 1, 5, 10$, and various values of τ, for FDMA/TDMA and TDMA. For the case of $\tau = 1$ each analytical CDF is accompanied by the corresponding empirical CDF (dotted lines) obtained by simulations (simulation results for other τ values are omitted for clarity). The good match between analysis and simulation validates the use of the derived formulas for system analysis and design.

As it can be seen, increasing AP density, i.e., increasing τ, results in improved performance as the distance between UE and serving AP, as well as the number of UEs sharing the resources of a single AP, are reduced, which overbalance the effect of reduced distance from interfering APs. Concerning the dependence of rate on N, it can be seen that setting $N = 1$ (conventional TDMA) is optimal when large data rates are considered, irrespective of τ. However, the shape of the CDF for $N = 1$ indicates a highly unfair system. Increasing N results in a progressively more fair system, favoring the small-rate operational region. In particular, for $\tau = 1$, and assuming a rate outage when the typical UE rate is below 0.1 b/s/Hz (corresponding

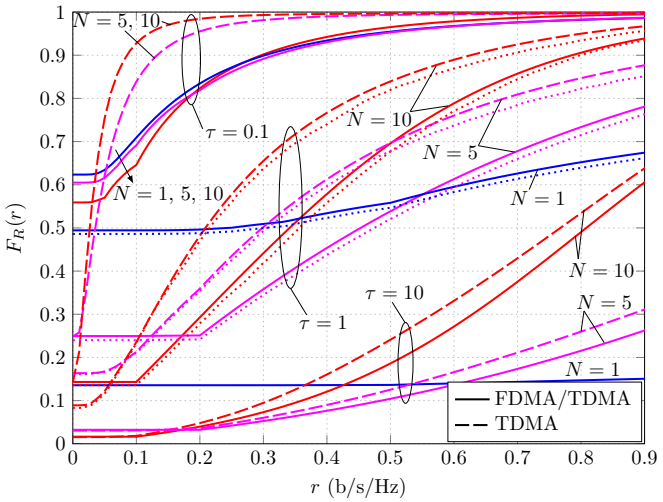


Figure 2.4.: $F_R(r)$ of FDMA/TDMA and TDMA for various values of N and τ ($\theta_0 = 0$ dB, $\alpha = 3$). Dotted lines depict simulation results.

to 2 Mbps in a 20 MHz system bandwidth), the outage probability is about 0.49, 0.25, and 0.15, for $N = 1, 5$, and 10, respectively with FDMA/TDMA.

Comparing FDMA/TDMA and TDMA for the same $N > 1$, it can be seen that TDMA is a better choice at low data rates. As stated above, at this value range it is the SIR outage probability that defines performance and TDMA is preferable as it is more robust to SIR outage events. When higher data rates are considered, TDMA is penalized by the inability of concurrent UE transmissions and FDMA/TDMA becomes a better choice. In Fig. 2.4 this difference in performance is more clearly seen for $\tau = 10$, which results in average K and K_0 equal to 10 and 12.8, respectively. For this load and the values of N considered, FDMA/TDMA utilizes all SCs with high probability, resulting in significantly larger interference compared to that achieved by TDMA which only allows for transmission on a single SC. However, for the same reason, performance of FDMA/TDMA is significantly better for higher rates as it provides much smaller delay than TDMA. Results for $\tau = 1, 10$ show that the performance advantage of FDMA/TDMA in higher rates and of TDMA in lower rates still holds but is less pronounced.

2.5. Optimal Number of Subchannels

By simple examination of Fig. 2.4, it is understood that for a given minimum rate $r_0 > 0$, there is an r_0-dependent, optimal number N^* of SCs that minimizes rate outage probability which can be obtained by numerical search using the analytical expression of Proposition 1. Unfortunately, the highly non-linear dependence of $F_R(r_0)$ on N does not allow for a closed form expression of N^* that holds in the

general case. However, the following proposition, valid under certain operational scenarios to be identified right after, provides some guidelines.

Proposition 2.8. *Under the assumption* $\mathbb{P}(L = 0) = 1$ *(no time-sharing of SCs) and for any* $\theta_0 \geq 0$, $r_0 > 0$, *the optimal number* N^* *of SCs that minimizes* $F_R(r_0)$ *is lower bounded by*

$$N^* \geq N^*_{\text{lb}} \triangleq \max\left\{1, \left\lfloor \overline{R}_{\theta_0}/r_0 \right\rfloor\right\}, \tag{2.11}$$

Proof. Setting $r = r_0$ and keeping only the term $L = 0$ in (2.9), $F_R(\cdot)$ can be written as a function of N as

$$F_R(N) = \begin{cases} F_{\text{SIR}}(\theta_0) & , N \leq \frac{\bar{R}_{\theta_0}}{r_0}, \\ F_{\text{SIR}}(2^{r_0 N} - 1) & , N \geq \frac{\bar{R}_{\theta_0}}{r_0}. \end{cases}$$

As shown in Sect. II. C, $F_{\text{SIR}}(\theta_0)$ is a decreasing function of N for both TDMA and FDMA/TDMA, therefore, so is $F_R(r_0)$ for $N \leq \frac{\overline{R}_{\theta_0}}{r_0}$. □

As discussed in Sect. II. D, L can be made arbitrarily small for both TDMA and FDMA/TDMA by increasing τ, therefore, Proposition 2 holds asymptotically for $\tau \gg 1$, i.e., in an ultra dense AP deployment where $\Pr\{K_0 = 0\} \to 1$. However, FDMA/TDMA can also reduce delay by increasing N. It is easy to see that if, for a given τ, N^*_{lb} is larger than the minimum value of N required for $\Pr\{K_0 \leq N-1\} = 1$, i.e., no UEs sharing a SC, N^* for FDMA/TDMA cannot be smaller than N^*_{lb}. These observations are summarized in the following corollary.

Corollary 2.9. *The bound of (2.11) holds for TDMA when* τ *is sufficiently large so that* $\mathbb{P}(K_0 = 0) \to 1$, *and for FDMA/TDMA when* τ *is sufficiently large so that* $\mathbb{P}(K_0 \leq N^*_{lb} - 1) \to 1$.

Note that the lower bound of (2.11) is inversely proportional to r_0 corresponding to the fact that, when lower user rates are considered, there is no need for large bandwidth utilization and the system can reduce interference by increasing N. For rates $r_0 > \overline{R}_{\theta_0}/2$ the bound becomes trivial, i.e., equal to one, however, these rates may be of small interest in a practical setting as they lead to large rate outage probability, even with optimized N and moderate load (see Fig. 2.4 and Sect. VI).

The following lemma guarantees that an arbitrarily large N cannot provide any non-zero r_0, even though the SIR grows unbounded with N.

Lemma 2.10. *For* $N \to \infty$, $F_R(r_0) \to 1$, *for any* $r_0 > 0$.

Proof. It was shown in [29] that the mean of $\overline{R}$ is a strictly decreasing function of N. Since $0 \leq R \leq \overline{R}$, it follows that the mean of R tends to 0 with increasing N, and applying Markov's inequality completes the proof. □

2.6. Minimum Required Access Point Density for a Given Rate Outage Probability Constraint

A common requirement in practical systems is to provide a minimum rate r_0 to their subscribers with a specified, small outage probability $\epsilon > 0$. As is clear from Fig. 2.4, these system requirements may be such that they cannot be satisfied for a certain τ, even under optimized N. It is therefore necessary to operate in a greater τ, i.e., increase AP density, and it is of interest to know the minimum value, $\tau_{\min}$, that can provide the given requirements. As in the case of N^*, a closed form expression for $\tau_{\min}$ can not be found in closed form for the general case and a two-dimensional numerical search (over τ and N) is necessary. However, under asymptotically small ϵ, an upper bound of $\tau_{\min}$ can be obtained for the case of FDMA/TDMA, as stated in the following proposition.

For FDMA/TDMA and asymptotically small ϵ, the minimum value of τ, $\tau_{\min}$, that can support a UE rate r_0 with $F_R(r_0) \leq \epsilon$ under an SIR threshold θ_0, is upper bounded as

$$\tau_{\min} \leq \begin{cases} \frac{(1-\epsilon)\rho(\theta_0)}{\epsilon\lfloor \bar{R}_{\theta_0}/r_0 \rfloor} & , r_0 \leq \bar{R}_{\theta_0}, \\ \frac{(1-\epsilon)\rho(2^{r_0}-1)}{\epsilon} & , r_0 \geq \bar{R}_{\theta_0}. \end{cases} \tag{2.12}$$

Proof. An upper bound on $\tau_{\min}$ can be obtained by seeking the value of τ that provides the requested outage probability constraint with equality and under $N = N_{\text{lb}}^*$, as given in (2.11), which is not guaranteed to be the optimal choice for N. In addition, only values of τ for which N_{lb}^* results in $\mathbb{P}(L = 0) = 1$ are considered, which effectively places a lower bound on the search space of τ that may be greater than $\tau_{\min}$. Under these restrictions,

$$\begin{aligned} F_R(r_0) &\overset{(a)}{=} \begin{cases} F_{\text{SIR}}(\theta_0) & , r_0 \leq \bar{R}_{\theta_0}, \\ F_{\text{SIR}}(2^{r_0}-1) & , r_0 \geq \bar{R}_{\theta_0}, \end{cases} \\ &= \begin{cases} \frac{1}{1+\tau N_{\text{lb}}^*/\rho(\theta_0)} & , r_0 \leq \bar{R}_{\theta_0}, \\ \frac{1}{1+\tau/\rho(2^{r_0}-1)} & , r_0 \geq \bar{R}_{\theta_0}, \end{cases} \end{aligned} \tag{2.13}$$

where (a) follows from (2.9), (2.10) with $\Pr\{L = 0\} = 1$ and (b) from (2.3) and (2.6) with $\min\{K, N_{\text{lb}}^*\} = K, \forall K$. Setting (2.13) equal to ϵ results in (2.12). Note that the asymptotically small ϵ guarantees that the bound of (2.12) is large enough such that $\mathbb{P}(L = 0) \to 1$, i.e., it is within the restricted search space employed for its derivation. □

A simpler form of the bound can be obtained when asymptotic values of r_0 are considered as shown in the following proposition.

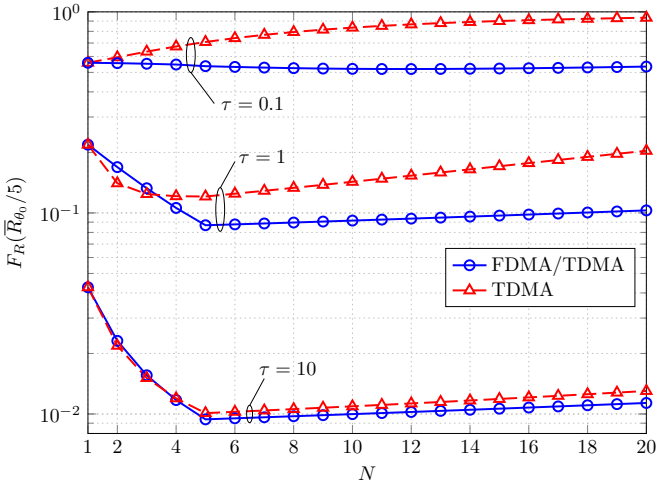


Figure 2.5.: Dependence of $F_R(r_0)$ on N ($r_0 = \overline{R}_{\theta_0}, \theta_0 = -6$ dB, $\alpha = 3$).

Proposition 2.11. *For asymptotically small or large values of r_0, the bound of (2.12) can be approximated by*

$$\tau_{min} \leq \begin{cases} \frac{(1-\epsilon)\rho(\theta_0)}{\epsilon \bar{R}_{\theta_0}} r_0 & , r_0 \ll \bar{R}_{\theta_0}, \\ \frac{(1-\epsilon)2\pi}{\epsilon\alpha \sin(2\pi/\alpha)} 2^{2r_0/\alpha} & , r_0 \gg \bar{R}_{\theta_0}. \end{cases} \tag{2.14}$$

Proof. The upper part of (2.14) can be obtained by noting that $1/\lfloor \overline{R}_{\theta_0}/r_0 \rfloor \approx r_0/\overline{R}_{\theta_0}$, for $r_0 \to 0$, whereas the lower part can be obtained by noting that $\rho(2^{r_0} - 1) \approx \rho(2^{r_0}) = 2^{2r_0/\alpha} \int_{2^{-2r_0/\alpha}}^{\infty} 1/(1+u^{\alpha/2}) du \approx 2^{2r_0/\alpha} \int_0^{\infty} 1/(1+u^{\alpha/2}) du = 2^{2r_0/\alpha} 2\pi \sin(2\pi/\alpha)/\alpha$, for $r_0 \to \infty$ □

Equation (2.14) clearly shows that the bound of $\tau_{\min}$ grows linearly and exponentially with r_0, for asymptotically small and large r_0, respectively.

2.7. Numerical Results and Discussion

This section employs the analytical results obtained previously to examine various aspects of system design. In all cases the path loss exponent is set to $\alpha = 3$ and, unless stated otherwise, the SIR threshold is set to $\theta_0 = -6$ dB ($\overline{R}_{\theta_0} \approx 0.3233$), roughly corresponding to the operational SIR required by the minimum coding rate scheme of a real cellular system [32].

1) Optimal number of SCs: Fig. 2.5 shows $F_R(\overline{R}_{\theta_0}/5)$ as a function of N for FDMA / TDMA and TDMA, and $\tau = 0.1, 1$, and 10. Note that by Proposition 2.8, N^* is lower bounded by $N^*_{\text{lb}} = 5$ when $\mathbb{P}(L = 0) = 1$. Consider first TDMA. It can

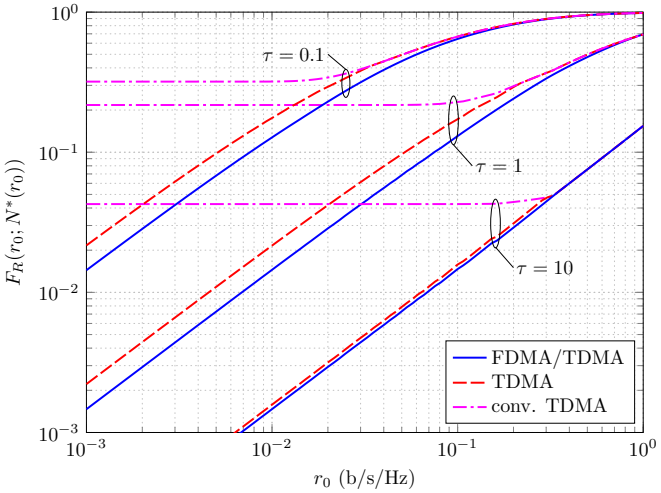


Figure 2.6.: $F_R(r_0)$ with $N^*(r_0)$ ($\theta_0 = -6$ dB, $\alpha = 3$).

be directly calculated that $\mathbb{P}(L = 0) = \mathbb{P}(K_0 = 0) \approx 0.0023, 0.3227, 0.8809$, for $\tau = 0.1, 1, 10$, respectively. Therefore, the operational conditions of Corollary 2.9 hold (approximately) only for the $\tau = 10$ case. It can be seen, that the bound is actually tight for that case, as $N^* = 5$, whereas N^* tends to one as smaller τ values are considered, i.e., N^*_{lb} is a tight bound when $\tau \gg 1$ but is irrelevant for small τ. Turning to the FDMA/TDMA case, $\mathbb{P}(L = 0) = \mathbb{P}(K_0 < N^*_{\text{lb}} - 1) \approx 0.1864, 0.9931, 1$, for $\tau = 0.1, 1, 10$, respectively, i.e., the operational conditions of Corollary 2.9 correspond to the cases of $\tau = 1$ and 10, with N^* actually equal to N^*_{lb}. For $\tau = 0.1$, $N^* = 12$, i.e., N^*_{lb} also serves as a lower bound in this case, albeit a loose one. However, note that performance gain with N^* is only marginal compared to N^*_{lb}. These observations, along with extensive numerical experiments, suggest that setting $N = N^*_{\text{lb}}$ as per (2.11) is a good practice for FDMA/TDMA as it either corresponds to the optimal value or provides performance close to optimal. For TDMA, setting $N = N^*_{\text{lb}}$ for small τ may lead to considerable performance degradation.

2) Comparison of multiple access schemes with optimal N: Fig. 2.6 depicts the minimum rate outage probability provided by FDMA/TDMA and TDMA when the corresponding optimal N for each rate r_0 is employed (found by numerical search). Note that these curves should not be confused as CDF curves since a different N is employed for each rate. Performance of conventional TDMA is also shown. As can be seen, for small to moderate rates ($r_0 < \overline{R}_{\theta_0}$), optimal bandwidth partitioning provides significant benefits compared to conventional TDMA. FDMA/TDMA is shown to outperform TDMA in this regime as it exploits bandwidth more efficiently. For large rates ($r_0 \geq \overline{R}_{\theta_0}$) all schemes have the same performance as N^* becomes one. It is safe to say that, under optimal N, FDMA/TDMA is preferable to TDMA as it provides at least as good performance with the added benefit of reduced delay that is of importance under time-sensitive applications.

3) Effect of SIR threshold: Fig. 2.7 shows $F_R(r_0)$ as a function of SIR threshold θ_0, for

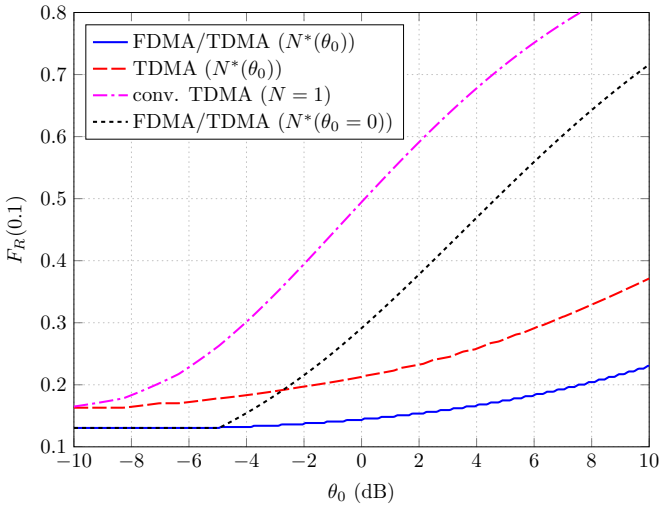


Figure 2.7.: Dependence of $F_R(r_0)$ on θ_0 ($r_0 = 0.1, \tau = 1, \alpha = 3$).

$r_0 = 0.1$, $\tau = 1$, and with N optimized for each θ_0 by numerical search. It can be seen that larger θ_0 values result in degradation of performance for both FDMA/TDMA and TDMA, albeit much less severe than conventional TDMA. Also shown is the performance of FDMA/TDMA when N is optimized assuming $\theta_0 = 0$, i.e., neglecting SIR outage. As expected, performance (significantly) degrades when the actual SIR threshold exceeds a certain value (about -5 dB in this case). Performance of TDMA assuming $\theta_0 = 0$ is not shown as it matches that of conventional TDMA. Similar behavior is observed for other values of r_0, τ and α. These results clearly illustrate the necessity of employing the SIR threshold in system analysis and design.

4) Minimum AP density: Fig. 2.8 shows $\tau_{\min}$ as a function of rate r_0, obtained by a two-dimensional numerical search over τ and N, for FDMA/TDMA, TDMA and conventional TDMA, under an outage constraint $F_R(r_0) \leq 0.1$. In addition, the asymptotic bounds of (2.14) are also shown. Note that, even though (2.14) is derived assuming asymptotically small ϵ, it still provides a very good approximation of $\tau_{\min}$ for this case. Specifically, $\tau_{\min}$ of FDMA/TDMA exhibits the behavior predicted by (2.14), i.e., increases linearly and exponentially with r_0 for asymptotically small and large r_0 respectively. Performance of TDMA follows the same trend with FDMA/TDMA but results in about 1.5 times larger values of $\tau_{\min}$ for small r_0. It is interesting to note that the very good correspondence of the numerical and analytical results for FDMA/TDMA implies that the optimal system parameters (τ and N) for FDMA/TDMA are such that $\Pr\{L = 0\} \approx 1$, i.e., there is small probability of sharing a SC. In contrast, TDMA achieves performance close to FDMA/TDMA with $\Pr\{L = 0\} \approx 0$ for small r_0 ($\tau_{\min}$). Conventional TDMA is clearly out of consideration for the small rate region as it significantly suffers from interference and the only mechanism to reduce it is by employing a large AP density. For rate values equal or greater than $\overline{R}_{\theta_0}$ all schemes coincide as the optimal value of N turns out to be equal to one.

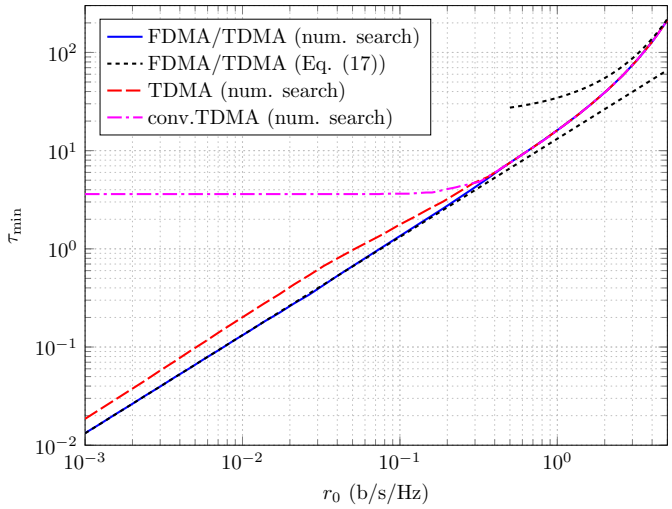


Figure 2.8.: Minimum required τ for rate outage probability $\epsilon = 0.1$ ($\theta_0 = -6$ dB, $\alpha = 3$).

2.8. Conclusion

In this chapter, system parameter selection, namely, number of bandwidth partitions and AP density was investigated for randomly deployed ultra dense wireless networks. The stochastic geometry framework from previous works was incorporated and the user rate distribution was derived analytically, taking into account the UE distribution, multiple access scheme and SIR outage. It was shown that performance depends critically on the number of bandwidth partitions and the way they are utilized by the multiple access scheme. The optimal number of partitions was tightly lower bounded under large AP density, showing that smaller bandwidth utilization is beneficial for interference reduction when small rates are considered. In addition, an upper bound on the minimum AP density required to provide an asymptotically small rate outage probability was obtained, that was shown to provide a very good estimate of the minimum density under moderate probability constraints. When the considered rates are small enough to allow for bandwidth partitioning, the minimum required density is smaller by orders of magnitude compared to the one provided by conventional TDMA.

3. Operational Region of D2D Communications for Enhancing Cellular Network Performance

3.1. Overview

An important enabler towards the successful deployment of any new element/feature to the cellular network is the investigation and characterization of the operational conditions where its introduction will enhance performance. Even though there has been significant research activity on the potential of device-to-device (D2D) communications, there are currently no clear indications of whether D2D communications are actually able to provide benefits for a *wide range* of operational conditions, thus justifying their introduction to the system. This chapter attempts to fill this gap by taking a stochastic geometry approach on characterizing the set (region) of operational conditions for which D2D communications enhance performance in terms of average user rate. For the practically interesting case of a heavy loaded network, the operational region is provided in closed form as a function of a variety of parameters such as maximum D2D link distances and user densities, reflecting a wide range of operational conditions (points). It is shown that under the appropriate deployment scheme, D2D communications can indeed be beneficial not only for the usually considered regime of "proximal communications" but to a wide range of operational conditions that include D2D link distances comparable to the distance to the cellular access point and considerably large user densities.

The results of this chapter were published in

1. S. Stefanatos, A. G. Gotsis, and A. Alexiou, "Operational region of D2D communications for enhancing cellular network performance," *IEEE Transactions on Wireless Communications*, vol. 14 , no. 11, pp. 5984–997, Nov. 2015.
2. S. Stefanatos, A. G. Gotsis, and A. Alexiou, "Operational Region of Overlay D2D Communications," *IEEE International Conference on Communications (ICC)*, London, UK, Jun. 2015.
3. S. Stefanatos, A. G. Gotsis, and A. Alexiou, "Analytical Assessment of Coordinated Overlay D2D Communications," *European Wireless (EW) conference 2014*, Barcelona, Spain, 14-16 May 2014.

3.2. Introduction

The exponential growth of traffic that the cellular networks are currently experiencing has driven academia and industry to consider new channel access methods in addition to the conventional, infrastructure-based access via cellular nodes (access points). One of these methods that have recently attracted a lot of attention is device-to-device (D2D) communications as a means for offloading cellular traffic as well as opportunistically exploiting link proximity [33]. Even though the potential of D2D communications cannot be denied, even more so with the ever increasing number and computational/storage capabilities of devices, their introduction as part of the cellular network introduces new challenges [34] with respect to (w.r.t.) resource sharing between cellular and D2D access modes.

Specification of (optimal) resource sharing is of great importance, nevertheless, it is not alone sufficient to justify the introduction/adoption of a D2D mode as an "add-on" to the (conventional) cellular network, even if the corresponding analysis shows great potential under *certain* operational conditions. For example, it is obvious that exploitation of D2D links with "small" distance compared to the distance from the closest access point is beneficial, which actually provides the incentive to investigate D2D communications in the first place. However, the vague notion of "closeness" (proximity) must be converted to actual figures in order to assess the merits of D2D communications, since it is reasonable to wonder whether the proximity requirements of D2D communications are actually so strict that render the benefits of D2D communications non-existent for realistic/practical conditions. Similar concerns can be posed w.r.t. other relevant parameters such as (relative) densities of concurrent D2D and cellular links.

It becomes clear that a detailed investigation and characterization of the range of operational conditions where introduction of a D2D mode enhances cellular performance is a procedure of critical importance. These operational conditions are typically described by quantities that the system designer/operator has no control over, such as density of users and channel conditions, as well as quantities chosen at a pre-deployment system design stage such as maximum link distance for which D2D transmission can (potentially) be established. Characterization of the D2D operational region will determine whether the introduction of D2D communications is actually beneficial for a (hopefully) large dynamic range of use/business cases and, therefore, plays a key role in the adoption of a D2D mode as part of the cellular network.

3.2.1. Related Work

There has been considerable work on resource sharing for D2D communications under an optimization modeling framework (see [35] for a recent literature review).

Even though providing optimal solutions, this framework is bound to consider finite-area networks, and, therefore, provides limited insight for large-scale network deployments. Toward this end, a second line of works has recently emerged, attempting to describe the merits of D2D communications based on tools from stochastic geometry [18], which have been successfully applied for the analysis of ad-hoc and cellular networks [36, 29].

Out of the relatively small number of notable publications considering analysis and design of D2D communications under a stochastic geometry framework, the most relevant with this work are [37, 38, 39, 40, 41, 42, 43]. An overlay D2D network deployment, i.e., where cellular and D2D transmissions are performed in orthogonal (non-overlapping) bands of the available spectrum, is considered in [37]. The distribution of transmit powers is investigated under a power control scheme, however, without any considerations regarding spectrum sharing between cellular and D2D networks. An underlay D2D network deployment, i.e., where cellular and D2D transmissions are performed in the same bandwidth, is considered in [38]. A simple on-off power control scheme for D2D transmissions is proposed in a setup considering only a single cell serving one uplink user. More comprehensive studies on the co-existence of large-scale deployed cellular and D2D networks are pursued in [39, 40, 41, 42, 43]. In [39, 40, 41], underlay or overlay D2D communications are considered along with an uplink or downlink cellular network. System operation aspects such as resource sharing among networks and mode selection according to various criteria are investigated. In [42], system performance in terms of spectral efficiency (average rate per channel use) is considered under a scheme providing exclusion regions to protect cellular users from D2D-generated interference, whereas in [43], the transmission capacity, i.e., density of cellular and D2D links with signal-to-interference ratio above a certain threshold, is considered. All the above works, even though they provide valuable insights on the potential of introducing a D2D mode to the cellular system, do so by considering only a limited number of (artificially selected) operational points (scenarios) without any clear indication of the *region* of operational points where D2D communications are beneficial to system performance.

3.2.2. Contributions

This chapter considers D2D communications deployed as overlay or underlay to the downlink-dedicated cellular bandwidth. The main goal is to provide a mathematically rigorous specification of the operational region where introduction of a D2D mode enhances cellular network performance in terms of average user rate. The operational region is described in terms of the following parameters (stated in no particular order) whose possible combinations of values can be mapped to a wide variety of use/business cases of potential interest:

- density of users,

- density of cellular access points,
- propagation conditions,
- signal-to-interference-ratio (SIR) threshold for successful communications, and
- maximum D2D link distance.

A number of D2D deployment schemes that trade-off complexity for performance are investigated, characterized by whether (non-trivial) mode selection and/or D2D channel access procedures are employed. For the mode selection procedure, two alternatives are considered depending on the availability of D2D link distance information. Based on accurate analytical expressions for the average rates achieved by the cellular and D2D links, a tractable expression for the average user rate of a D2D-enabled cellular network is obtained. For the practically interesting case of a heavy loaded system (large user density), the operational regions for each D2D deployment scheme are obtained in closed form, allowing for investigating the effect on system performance of the two most important parameters related to D2D use/business cases, namely, maximum D2D link distance and density of users (devices) capable for D2D communications.

Important outcomes of this analysis include the following:

1. D2D communications can provide benefits for operational conditions of practical interest including D2D link distances of the order of the distance to the closest access point.
2. Employing both a mode selection and a D2D channel access procedure does not necessarily provide significant gains compared to employing one of them only.
3. Exploitation of D2D link distance information alone for mode selection purposes is sufficient for D2D communications to be beneficial for any operational point.

The key notation of this chapter is summarized in Tab. 3.1 (next page).

3.3. System Model

3.3.1. System Nodes Description and Channel Access Mechanisms

A D2D-enabled cellular system is considered that allows for overlay or underlay downlink-inband D2D transmissions, i.e., transmissions are performed either in cellular or D2D mode. In particular, the system serves two types of user equipments (UEs) [34]. The first type, referred to in the following as C-UEs, request data from sources that do not have the ability to establish a direct communication link with the

Notation	Description
$\mathbb{R}, \mathbb{R}^+, \emptyset$	set of reals, set positive reals, empty set
$\mathbb{P}(\cdot), \mathbb{E}(\cdot), \mathbb{I}(\cdot)$	probability measure, expectation operator, indicator function
$\lvert x \rvert$	Euclidean norm of $x \in \mathbb{R}^2$
$\lambda_a, \lambda_d, \lambda_c$	densities of APs, D-UEs, and C-UEs, respectively
r_d	distance of a random (potential) D2D link ($r_d \in (0, r_{d,\max}]$)
$r_{d,\text{th}}$	threshold distance used by the distance-based mode selection procedure ($r_{d,\text{th}} \in (0, r_{d,\max}]$)
P_a	(common) AP transmit power
p	probability of D-UE selecting D2D mode ($p \in (0, 1]$)
q	probability of transmission of a D2D link ($q \in (0, 1]$)
α	path loss exponent ($\alpha > 2$)
K, K_0	number of cellular UEs within a random cell and the cell containing the origin, respectively
η_c	portion of bandwidth dedicated to cellular transmissions ($\eta_c \in (0, 1]$)
κ	$\triangleq (2\pi/\alpha)/\sin(2\pi/\alpha)$
$\rho(\theta)$	$\triangleq \theta^{2/\alpha} \int_{\theta^{-2/\alpha}}^{\infty} \frac{1}{1+u^{\alpha/2}} du$
γ	$\triangleq 1, 2$, for probabilistic and distance-based mode selection, respectively
θ_0	SIR threshold below which communication is considered unsuccessful ($\theta_0 > 0$)
$\mathcal{P}_{\text{op}}, \mathcal{P}_{\text{d}}$	system operational point, system design parameters
R, R_c, R_d	rate of typical link, typical cellular link, and typical D2D link, respectively
R_{noD2D}	rate of typical cellular link when the network does not support D2D transmissions
$\mathcal{R}^i_{\text{D2D}}$	Operational region of D2D communications using Scheme $i \in \{1, 2, 3, 4\}$

Table 3.1.: Summary of Key Notation

requesting C-UE, due to being located far away and/or not having transmit capabilities. C-UEs can only utilize downlink cellular transmissions in order to obtain their data. The second type of UEs, referred to in the following as D-UEs, request data from sources that are physically proximal and have transmit capabilities, providing D-UEs the flexibility to establish either a D2D or a dual-hop (uplink and downlink) cellular link for communication. Full buffers are assumed for all communication links.

The positions of C-UEs and D-UEs are modeled as independent homogeneous Poisson point processes (HPPPs) Φ_c, $\Phi_d \subset \mathbb{R}^2$, of densities λ_c, $\lambda_d > 0$, respectively, whereas the position of the source for a D-UE located at $x \in \Phi_d$ is modeled as an independent random variable uniformly distributed within the closed ball $\{y \in \mathbb{R}^2 : |y - x| \leq r_{d,\max}\}$. By application of the displacement theorem for HPPPs [18], it can be shown that the sources of D-UEs also constitute an HPPP of density λ_d which, with a slight abuse of notation, will be also denoted as Φ_d. The maximum

possible D2D link distance $r_{d,\max} > 0$ provides an exact description of the notion of "proximity" and is determined by the relevant D2D use cases and/or imposed by considerations regarding, e.g., power consumption or system performance. Note that parameters λ_c, λ_d and $r_{d,\max}$ are correlated, e.g., increasing $r_{d,\max}$ can potentially "transform" some of the C-UEs into D-UEs. However, there are currently no well-established models available in the literature for modeling this correlation. Therefore, in this chapter, these parameters are treated as independent variables, with the corresponding analysis easily extended under any model describing their interrelation.

The cellular system infrastructure consists of access points (APs) randomly deployed according to an independent HPPP $\Phi_a \subset \mathbb{R}^2$ of density $\lambda_a > 0$ and each UE employing cellular transmissions is served by its closest AP. In order to avoid intracell interference, time is (universally) divided into discrete slots of equal duration and UEs within a cell employing cellular transmissions are served by a time-division-multiple-access (TDMA) scheme with no priorities among them (round-robin scheduling). All active APs, i.e., APs with at least one served UE, transmit with a fixed power P_a, without any coordination/cooperation among them.

In contrast to cellular communications where intracell interference is eliminated by centralized scheduling, a *probabilistic* channel access scheme [36, 85] is employed for D2D communications. Specifically, D2D communications utilize the same time slots as the cellular system (perfect synchronization is assumed between D2D and cellular slots in the underlay case), with the source of any established D2D link attempting to transmit in each slot with a probability $q \in (0, 1]$. Active D2D links employ open loop power control so that the effect of path loss, averaged over small-scale fading, is eliminated (large-scale path loss inversion) [45, 41].

3.3.2. Mode Selection and D2D Network Deployment Schemes

Since D-UEs have the flexibility to choose between two access methods, a mode selection procedure should be employed. In this chapter, the following two simple procedures are considered:

Definition 3.1. *Probabilistic mode selection*: Each D-UE selects D2D mode by independently tossing a biased coin with bias $p_{\text{prob}} \in (0, 1]$.

Definition 3.2. *Distance-based mode selection*: Each D-UE selects D2D mode if and only if the distance $r_d \in (0, r_{d,\max}]$ from its source is smaller than a pre-defined threshold $r_{d,\text{th}} \in (0, r_{d,\max}]$.

Note that both mode selection procedures can be viewed as thinning operations [18] on Φ_d with retention probability equal to p_{prob} and $\mathbb{P}(r_d \leq r_{d,\text{th}}) = (r_{d,\text{th}}/r_{d,\max})^2$ for the probabilistic and distance-based case, respectively. This allows for a unified treatment of both mode selection schemes under a common notation $p \in (0, 1]$ for

their corresponding D2D mode selection probabilities. Parameters p and q constitute the so called *D2D mode parameters* and their (optimal) values will be shown to require knowledge of parameters such as AP and UE densities. Therefore, p_{prob}, $r_{d,\text{th}}$, and q are provided to the D-UEs via a broadcast mechanism.

Since incorporation of mode selection and/or channel access procedures comes with implementation cost, e.g., for time slot synchronization or acquisition of D2D link distance information, it is of interest to examine and compare the merits of each of the following D2D network deployment schemes, trading off flexibility on design of D2D mode parameters to complexity:

- Scheme 1 (Baseline): $p = 1$, $q = 1$,
- Scheme 2 (D-UEs employ D2D mode by default): $p = 1$, $q \in (0, 1]$,
- Scheme 3 (D2D links always active): $p \in (0, 1]$, $q = 1$,
- Scheme 4 (Most general case): $(p, q) \in (0, 1]^2$.

Scheme 3 has two versions, according to whether probabilistic or distance-based mode selection is employed, that will be denoted, when necessary, as 3-p and 3-d, respectively, with similar notation also used for Scheme 4. It is noted that Schemes 3-d and 4-p were first examined in [39] and [40], respectively, under different system model and/or performance metrics than this chapter. Clearly, Scheme 4 is optimal under any performance criterion as it incorporates the other schemes as special cases. However, the implementation cost of operating under Scheme 4 with both $p < 1$ and $q < 1$ must be justified by a significant performance gain compared to the simpler schemes, an issue that will be examined in the following sections.

Remark: Unless stated otherwise, results and definitions in the following sections correspond to operation under Scheme 4. Translation to the other schemes can be made by trivially setting p and/or q equal to 1.

3.3.3. Signal-to-Interference Ratio

The standard approach of conditioning on the existence of a (typical) UE located without loss of generality (w.l.o.g.) at the origin will be employed for the following analysis. Note that, by the properties of HPPP, this conditioning does not have any effect on the distribution of the other UEs [18]. Treating interference as noise and neglecting the effect of thermal noise due to the system operating in the interference limited region, the SIR experienced by the typical UE when served by a transmitter (either an AP or a proximal source) located at $y_0 \in \mathbb{R}^2$ equals

$$\text{SIR} = \frac{\left(P_a \mathbb{I}(y_0 \in \tilde{\Phi}_a) + |y_0|^{\alpha} \mathbb{I}(y_0 \in \tilde{\Phi}_d)\right) g_{y_0} |y_0|^{-\alpha}}{\sum\limits_{y \in \tilde{\Phi}_a \setminus y_0} P_a g_y |y|^{-\alpha} + \sum\limits_{y \in \tilde{\Phi}_d \setminus y_0} |x_y - y|^{\alpha} g_y |y|^{-\alpha}}, \tag{3.1}$$

where $\tilde{\Phi}_a \subseteq \Phi_a$, $\tilde{\Phi}_d \subseteq \Phi_d$ denote the sets of positions of APs and D-UE sources, respectively, that access the channel at the considered time slot and bandwidth, g_y

is the channel gain corresponding to a transmitter located at $y \in \tilde{\Phi}_a \cup \tilde{\Phi}_d$, x_y is the position of a D-UE receiving data from its source located at $y \in \tilde{\Phi}_d$, and $\alpha > 2$ is the path loss exponent. The channel gains are assumed independent and exponentially distributed (Rayleigh fading) with mean one.

Note that w.l.o.g. the useful received power of a D-UE in D2D mode, averaged over channel fading, is normalized to one, which allows to treat P_a as the only variable that controls the relative interference levels of cellular and underlay D2D transmissions. Under this power control model, the transmit power of the source of a D2D link may become unrealistically large when D2D links of excessively large distances are established. In order to study the fundamental gains/limitations of employing a D2D mode as part of a cellular network w.r.t. to average user rate, no constraint on D2D maximum transmit power is set in this chapter. However, as it will be shown later for operational scenarios of practical interest, significant performance gains are only provided by establishing D2D links of distances up to about the order of the distance of a D-UE from its closest AP, corresponding to power requirements similar to uplink cellular transmissions.

3.4. Average Cellular and D2D Rates

The SIR is a critical measure of performance as increased SIR translates to improved spectral efficiency (in bits/Hz/channel use). In particular, the commonly employed mapping $\text{SIR} \mapsto \log_2(1+\theta_0)\mathbb{I}(\text{SIR} \geq \theta_0)$ will be used for determining the spectral efficiency of a link [24, 36, 50], where $\theta_0 > 0$ is the SIR threshold below which communications are considered unsuccessful. However, the actual rate achieved by a UE is also affected by the multiple access procedures, namely, TDMA for cellular links and probabilistic channel access for D2D links.[1] This section provides closed form expressions for the average rate of cellular and D2D links of the typical UE, which, in turn, will determine the more important metric of average UE rate that will be considered in the next section.

3.4.1. SIR Distribution of Cellular and D2D Links

The SIR distribution of cellular and D2D links, required to characterize the statistical properties of the corresponding spectral efficiencies, are provided in Lemmas 1–2.

Lemma 3.3. *The SIR distribution of the typical UE when employing a cellular link in a system with underlay D2D communications is*

$$\mathbb{P}(SIR \geq \theta) \approx \frac{1}{1 + \mathbb{P}(K>0)\rho(\theta) + \frac{qp^{\gamma}\kappa\lambda_d r_{d,\max}^2}{2\lambda_a}\left(\frac{\theta}{P_a}\right)^{2/\alpha}}, \tag{3.2}$$

[1] The effect of issues such as signaling, e.g., for establishment of transmission links, is ignored.

for $\theta > 0$, where K is the number of cellular receivers (RXs), composed from C-UEs and D-UEs selecting cellular mode, within a random cell of the network, $\gamma = 1, 2$ for probabilistic and distance-based mode selection, respectively, and κ, $\rho(\theta)$ as defined in Tab. 3.1. In case of overlay D2D communications, Eq. (3.2) holds with the last term of the denominator removed.

Proof. See Appendix sec. B.1. □

Lemma 3.4. *The SIR distribution of the typical UE when employing a D2D link in a system with underlay D2D communications is*

$$\mathbb{P}(SIR \geq \theta) \approx \exp[-\kappa\pi\theta^{2/\alpha}((1/2)qp^{\gamma}\lambda_d r_{d,\max}^2 + \lambda_a \mathbb{P}(K > 0)P_a^{2/\alpha})], \quad (3.3)$$

for $\theta > 0$, where $\gamma = 1, 2$ for probabilistic and distance-based mode selection, respectively. In case of overlay D2D communications, (3.3) holds with the last term inside parenthesis of the exponential removed.

Proof. Proof follows the lines of [40] and is omitted. □

Remark: The expressions of Lemmas 1 and 2 are approximate since they are based on the assumption that $\tilde{\Phi}_a$ is generated by a thinning of Φ_a with a retention probability $\mathbb{P}(K > 0)$ for each $x \in \Phi_a$, independent of everything else, which was shown in [24] to be a very good approximation. For the case when $\mathbb{P}(K > 0) = 1$, i.e., when all APs serve at least one UE each, and, therefore, $\tilde{\Phi}_a = \Phi_a$ almost surely, (3.2) and (3.3) are exact [40]. Note that an approximate closed form expression for $\mathbb{P}(K > 0)$ with very good accuracy is available [92], namely,

$$\mathbb{P}(K > 0) \approx 1 - \left(1 + \frac{\lambda_c + (1-p)\lambda_d}{3.5\lambda_a}\right)^{-3.5}. \quad (3.4)$$

3.4.2. Average Rate of Cellular Link

Exact computation of the rate achieved by the typical UE when employing cellular communications is difficult since there exist cases when the end-to-end link will consist not only of a downlink, but also of an uplink hop, e.g., when the typical UE is a D-UE selecting cellular mode. In order to simplify the problem, the following assumption is made.

Assumption 3.5. The cellular system is downlink limited in terms of user rate.

Assumption 1 is reasonable when the number of cellular RXs requesting data from the network core (e.g., an internet server) is (much) larger than the number of cellular RXs requesting data from another UE, with the end-to-end cellular rate essentially limited by the downlink TDMA process. Towards describing the effect of TDMA on the average rate of the typical UE employing cellular transmissions, the following lemma will be of use.

Lemma 3.6. *Let K_0 and K denote the number of cellular RXs positioned within the cell containing the origin and any other randomly selected cell, respectively. It holds*

$$\mathbb{E}\left(\frac{1}{K_0+1}\right) = \frac{\lambda_a \mathbb{P}(K>0)}{\lambda_c + (1-p)\lambda_d}. \tag{3.5}$$

Proof. See Appendix sec. B.2. □

Note that the term $1/(K_0+1) \leq 1$ of the above Lemma represents the rate reduction of the typical cellular UE rate due to resource sharing via TDMA.

Remark: Previous works considering the effect of TDMA either relied on numerical evaluations based on complicated, approximate expressions for the distribution of K_0 [47, 48], or employed heuristic approximations of $\mathbb{E}(1/(K_0+1))$ with the corresponding expressions equal to Eq. (3.5) with $\mathbb{P}(K>0)$ replaced by 1 [49] or 7/9 [40, 47]. The *exact* expression of Lemma 3 shows that these approximations do not hold in the general case.

The average rate of the typical UE when employing cellular transmissions, normalized by the total bandwidth dedicated for downlink cellular and D2D transmissions can now be obtained as follows.

Proposition 3.7. *The average normalized rate achieved by the typical UE when employing cellular mode equals*

$$R_c \triangleq \eta_c \mathbb{E}\left(\frac{1}{K_0+1}\mathbb{I}(SIR \geq \theta_0)\right)\log_2(1+\theta_0) \tag{3.6}$$

$$\approx \frac{\eta_c \lambda_a \mathbb{P}(K>0)}{\lambda_c + (1-p)\lambda_d}\mathbb{P}(SIR \geq \theta_0)\log_2(1+\theta_0), \tag{3.7}$$

in bits/s/Hz, where η_c is the portion of the total bandwidth where downlink cellular transmissions take place ($0 < \eta_c < 1$ and $\eta_c = 1$ for overlay and underlay D2D communications, respectively) and $\mathbb{P}(SIR \geq \theta_0)$, $\mathbb{P}(K>0)$ are given by Eqs. (3.2) and (3.4), respectively.

Proof. By ignoring the correlation between K_0 and SIR due to their common dependence on Φ_a, (3.6) can be approximated as

$$R_c \approx \eta_c \mathbb{E}\left(\frac{1}{K_0+1}\right)\mathbb{P}(\text{SIR} \geq \theta_0)\log_2(1+\theta_0). \tag{3.8}$$

The result then follows by application of Lemmas 1 and 3. □

Remark: The expression on the right hand side of (3.8) was used as the actual definition of cellular rate in previous works, e.g., [40, 49], implicitly suggesting that K_0 and SIR are uncorrelated (or even independent) which does not hold. As will be shown in Sec. VII, this approximation is very accurate, especially for the overlay D2D case.

3.4.3. Average Rate of D2D Link

Considering the same SIR threshold θ_0 as for cellular communications, the average normalized D2D link rate, R_d, is as follows (compare with the corresponding expression of R_c).

Lemma 3.8. *The average normalized rate achieved by the typical UE when employing overlay D2D mode equals*

$$\begin{aligned} R_d &\triangleq (1-\eta_c)\mathbb{E}\left[\mathbb{I}(\textit{D2D link active})\mathbb{I}(\textit{SIR} \geq \theta_0)\right]\log_2(1+\theta_0) \\ &= (1-\eta_c)q\mathbb{P}(\textit{SIR} \geq \theta_0)\log_2(1+\theta_0) \end{aligned} \tag{3.9}$$

in bits/s/Hz, with η_c as in Proposition 1 and $\mathbb{P}(\textit{SIR} \geq \theta_0)$ as given in Eq. (3.3). The rate expression for underlay D2D transmissions is the same as overlay with the term $1-\eta_c$ replaced by 1.

3.5. Average UE Rate and System Design Considerations

In this chapter, incorporation of D2D communications is considered as a means for enhancing the performance of the conventional cellular system. Towards this end, the average normalized rate achieved by the typical UE without *a-priori* information of its type (C-UE or D-UE) is considered as the system metric of interest, equal to [39, 40]

$$R \triangleq \frac{\lambda_c}{\lambda_c+\lambda_d}R_c + \frac{\lambda_d}{\lambda_c+\lambda_d}\left(pR_d + (1-p)R_c\right) \tag{3.10}$$

in bits/s/Hz, with R_c, R_d, the average rate of the cellular and D2D links, respectively, as described in Sec. III. Note that the rate achieved by the typical UE when a D2D mode option is not available, denoted as R_{noD2D}, is a degenerate case of (3.10) for $p=0$ and $\eta_c=1$, and equals

$$R_{\text{noD2D}} \approx \frac{\lambda_a\mathbb{P}(K>0)\log_2(1+\theta_0)}{(\lambda_c+\lambda_d)(1+\mathbb{P}(K>0)\rho(\theta_0))}, \tag{3.11}$$

as can be easily verified using the results of Sec. III.

Average rate R depends on a large number of parameters (through R_c and R_d) which are not shown explicitly for ease of notation. These parameters can be partitioned into two categories. The first category describes the *system operational point* and is represented by the tuple

$$\mathcal{P}_{\text{op}} \triangleq (\lambda_c, \lambda_d, \lambda_a, r_{d,\max}, \theta_0, \alpha),$$

which consists of parameters that the system operator has no control over, namely, λ_c, λ_d, and α, and parameters which are chosen at system design stage and remain fixed during operation, namely, λ_a, $r_{d,\max}$, and θ_0. The second category of parameters is represented by the tuple

$$\mathcal{P}_\mathrm{d} \triangleq (p, q, \eta_c, P_a),$$

which contains the *system design parameters* that can be selected by the operator in accordance to $\mathcal{P}_\mathrm{op}$. Note that specification of P_a is irrelevant for system design with overlay D2D, whereas $\eta_c = 1$ by default with underlay D2D. Therefore, $\mathcal{P}_\mathrm{d}$ essentially consists of the two D2D mode parameters and one parameter reflecting the resource sharing between cellular and overlay/underlay D2D networks.

Since the D2D mode is considered as an enhancement option to the conventional cellular network, it is natural to impose that its introduction should not degrade the performance experienced by UEs employing cellular transmissions [34, 38]. This can be reflected by the following restriction on the values of $\mathcal{P}_\mathrm{d}$.

Constraint 1. $\mathcal{P}_\mathrm{d}$ is allowed to take values for which the average normalized rate of the cellular links is unaffected by the introduction of a D2D mode, i.e., $R_c = R_\mathrm{noD2D}$.

Constraint 1 is convenient as it allows for eliminating one element of $\mathcal{P}_\mathrm{d}$, thus reducing the design space and simplifying the problem. In particular, by employing Eqs. (3.7) and (3.11), the resource sharing design parameters P_a and η_c can be determined in closed form as functions of p and q, namely,

$$P_a = \theta_0 \left(\frac{(\lambda_c + (1-p)\lambda_d)\, \kappa r_{d,\max}^2 q p^{\gamma-1}}{2\lambda_a \left(1 + \mathbb{P}(K>0)\rho(\theta_0)\right)} \right)^{\alpha/2}, \tag{3.12}$$

for the underlay D2D case, with $\gamma = 1, 2$ for probabilistic and distance-based mode selection, respectively, and

$$\eta_c = 1 - \frac{p\lambda_d}{\lambda_c + \lambda_d}, \tag{3.13}$$

for the overlay D2D case, irrespective of the mode selection scheme. Notice how the complicated interference environment that a UE on cellular mode experiences in the underlay D2D case is reflected on the expression for P_a, which increases as $\mathcal{O}(r_{d,\max}^{\alpha})$ in order to compensate for the correspondingly increasing transmit power of D2D links. In contrast, the expression for η_c in the overlay case is much simpler, having the intuitive interpretation that bandwidth is proportionally partitioned according to the relative densities of cellular and D2D links.

Specification of the resource sharing parameters by Eqs. (3.12) and (3.13) leaves the D2D mode parameters as the only variable elements of $\mathcal{P}_\mathrm{d}$. Their optimal values

$$(p^*, q^*) \triangleq \arg \max_{(p,q)\in(0,1]^2} R \tag{3.14}$$

can then in principle be found by a numerical search using the analytical formulas of the previous section for any operational point $\mathcal{P}_{\text{op}}$ of interest. Another quantity of significant importance is the *D2D operational region* defined as follows.

Definition 3.9. The D2D operational region, $\mathcal{R}_{\text{D2D}}$, is the set of system operational points for which incorporation of a D2D mode can increase the average UE rate provided by the (conventional) cellular network, i.e.,

$$\mathcal{R}_{\text{D2D}} \triangleq \{\mathcal{P}_{\text{op}} : \exists\, (p,q) \in (0,1]^2 \text{ for which } R(\mathcal{P}_{\text{op}}) > R_{\text{noD2D}}\}. \tag{3.15}$$

Note that $\mathcal{R}_{\text{D2D}}$ includes operational points where an arbitrarily small rate increase is achieved. These points, even though of small interest in that sense, may still be considered as beneficial due to implicit gains achieved w.r.t. other metrics, e.g., signaling levels, latency.

3.6. Average UE Rate and Ordering of D2D Deployment Schemes for Heavy Loaded System

3.6.1. Average UE Rate for Heavy Loaded System

The closed form expressions of the previous sections allow for efficiently computing (p^*, q^*) for any $\mathcal{P}_{\text{op}}$ of interest. On the other hand, obtaining $\mathcal{R}_{\text{D2D}}$ numerically is a rather complicated task as the domains of (p, q) and $\mathcal{P}_{\text{op}}$ are uncountable. In order to obtain closed form expressions for (p^*, q^*) and $\mathcal{R}_{\text{D2D}}$ the following assumption will be employed.

Assumption 3.10. The density λ_c of C-UEs is sufficiently large so that $\mathbb{P}(K > 0) \approx 1$ (irrespective of the density λ_d of D-UEs), where K is the number of cellular RXs in a random cell of the system.

The above assumption is mathematically convenient as it eliminates the use of the complicated expression of Eq. (3.4) in computing R. The cost of doing so is that attention is restricted only to operational points where each AP is almost surely active due to the presence of C-UEs alone. However, this heavy loaded system scenario is actually of the most interest for introducing a D2D mode to the conventional cellular network as a traffic offloading method. In addition, it follows from Eq. (3.4) that $\mathbb{P}(K > 0) > 0.955$ with $\lambda_c/\lambda_a \geq 5$, i.e., a very good approximation of the requirement of Assumption 2 is achieved by a moderate/reasonable value of C-UEs traffic load.

Remark: In general, the requirement $\mathbb{P}(K > 0) \approx 1$ of Assumption 2 can be achieved by an appropriately large $\lambda_c + (1-p)\lambda_d$, i.e., the constraint on λ_c can be relaxed by taking into account the cellular load due to D-UEs. However, this approach

imposes a constraint on the mode selection probability p that depends on λ_c and λ_d, complicating the closed form solution of the design problem.

The following proposition provides R in a simple closed form expression.

Proposition 3.11. *The average normalized UE rate under Constraint 1 and Assumption 2 equals*

$$R \approx R_{noD2D}\left(1+\frac{\lambda_d}{\lambda_c+\lambda_d}f(p,q)\right), \tag{3.16}$$

with R_{noD2D} as in Eq. (3.11) with $\mathbb{P}(K>0)=1$, and

$$f(p,q) \triangleq \begin{cases} c_1 p^2 q e^{-c_2 q p^{\gamma}} - p & \text{for overlay D2D,} \\ \bar{c}_1 p q e^{-(\bar{c}_2 q p^{\gamma} + \bar{c}_3 q p^{\gamma-1})} - p & \text{for underlay D2D,} \end{cases} \tag{3.17}$$

with $\gamma = 1$, 2 for probabilistic and distance-based mode selection, respectively, $c_1 \triangleq \frac{\lambda_d}{\lambda_a}(1+\rho(\theta_0))$, $c_2 \triangleq (1/2)\lambda_d \pi r_{d,\max}^2 \kappa \theta_0^{2/\alpha}$, $\bar{c}_1 \triangleq c_1(1+\lambda_c/\lambda_d)$, $\bar{c}_2 \triangleq c_2(1-\kappa\theta_0^{2/\alpha}/(1+\rho(\theta_0)))$, and $\bar{c}_3 \triangleq c_2\kappa\theta_0^{2/\alpha}(1+\lambda_c/\lambda_d)/(1+\rho(\theta_0))$.

Proof. Result follows after some lengthy but tedious algebra by substituting Eqs. (3.7), (3.9), (3.12), and (3.13) with $\mathbb{P}(K>0)=1$ into Eq. (3.10). □

Proposition 2 introduces the function f which plays an important role in system design as it is only via f that the D2D mode parameters affect R, i.e., the optimization of R w.r.t. (p,q) is equivalent to the optimization of f w.r.t. (p,q). In addition, f alone indicates whether the incorporation of a D2D mode results in $R > R_{\text{noD2D}}$ or $R \leq R_{\text{noD2D}}$ for a given $\mathcal{P}_{\text{op}}$, depending on whether it is positive or non-positive, respectively. Interestingly, the effect of $\mathcal{P}_{\text{op}}$ on f is compactly described by the set of the coupled parameters c_1, $c_2 \in \mathbb{R}^+$, for the overlay, and $\bar{c}_1$, $\bar{c}_2$, $\bar{c}_3 \in \mathbb{R}^+$, for the underlay D2D case. Note that f is independent of λ_c for the overlay D2D case.

The above observations significantly simplify the representation (and investigation) of $\mathcal{R}_{\text{D2D}}$ which can be equivalently written for the overlay case as

$$\begin{aligned} \mathcal{R}_{\text{D2D}} &= \{(c_1,c_2) \in \mathbb{R}^{+2} : \exists\ (p,q) \in (0,1]^2 \text{ for which } f(p,q) > 0\} \\ &= \{(c_1,c_2) \in \mathbb{R}^{+2} : \max_{(p,q)\in(0,1]^2} f(p,q) > 0\}, \end{aligned} \tag{3.18}$$

where the second equality follows by noting that $f(p,q)$ is uniformly continuous on $(0,1]^2$. The operational region for the underlay D2D case can be similarly described in terms of $(\bar{c}_1,\bar{c}_2,\bar{c}_3) \in \mathbb{R}^{+3}$.

3.6.2. Ordering of D2D Deployment Schemes for Heavy Loaded System

Before proceeding with a detailed investigation of the operational regions and optimal mode parameters for each of the four D2D deployment schemes considered in Sec. II.B, significant insights on their merits can already be made at this stage of the analysis. The first is provided by the following proposition which shows that the flexibility offered by employing both a mode selection and a channel access procedure provides much less gains than what would probably be expected.

Proposition 3.12. *There do not exist operational points in $\mathcal{R}_{D2D}$ for which the maximization of R is achieved with $(p^*, q^*) \in (0,1)^2$ when overlay D2D (either with probabilistic or distance-based mode selection) or underlay D2D with distance-based mode selection are employed.*

Proof. A necessary condition for $(p^*, q^*) \in (0,1)^2$ is $\frac{\partial f(p^*,q)}{\partial q}|_{q=q^*} = \frac{\partial f(p,q^*)}{\partial p}|_{p=p^*} = 0$. For the Schemes referred to in the proposition, this system of equations is either inconsistent or gives a solution for which $f(p^*, q^*) = 0$. □

The above result essentially states that Schemes 4-d and 4-p for the overlay case and Scheme 4-d for the underlay case need not be considered explicitly as they achieve their optimal performance when operating as their simplified versions which do not allow for $(p, q) \in (0,1)^2$ by definition. For the Schemes remaining into consideration, an ordering can be made based on the following binary relation.

Definition 3.13. Relation X $\succeq$ Y ("X greater than Y"), where X and Y belong to the set of D2D deployment schemes described in Sec. II.B, implies that the maximum rate provided by Scheme X is equal to or greater than the maximum rate provided by Scheme Y, for any operational point.

Note from Eq. (3.18) that X $\succeq$ Y also implies $\mathcal{R}^{\mathrm{X}}_{\mathrm{D2D}} \supseteq \mathcal{R}^{\mathrm{Y}}_{\mathrm{D2D}}$, where $\mathcal{R}^{\mathrm{X}}_{\mathrm{D2D}}$, $\mathcal{R}^{\mathrm{Y}}_{\mathrm{D2D}}$ are the D2D operational regions of Schemes X and Y, respectively.

Proposition 3.14. *The set of D2D deployment schemes forms a totally ordered set with*

Scheme 3-d $\succeq$ Scheme 2 $\succeq$ Scheme 3-p $\succeq$ Scheme 1,

for the overlay case, and a partially ordered set with

Scheme 3-d $\succeq$ Scheme 4-p $\succeq$ Scheme 2 $\succeq$ Scheme 1,

and

Scheme 3-d $\succeq$ Scheme 4-p $\succeq$ Scheme 3-p $\succeq$ Scheme 1,

for the underlay case. Scheme 3-d of the underlay D2D case is greater than Scheme 3-d of the overlay D2D case.

Proof. See Appendix sec. B.3. □

Scheme 3-d is therefore the optimal choice for both overlay and underlay D2D deployments, showing the significance of mode selection based on instantaneous per link information instead of employing probabilistic approaches for mode selection and/or channel access procedures. In addition, Scheme 3-d achieves its best performance with an underlay D2D deployment suggesting that the latter exploits the system resources more efficiently than an overlay deployment.

3.7. Optimal D2D Mode Parameters and Operational Region for Heavy Loaded System

This section explicitly investigates the optimal D2D mode parameters and operational region of each scheme which, in addition to being of interest in their own right for the system designer/operator, will allow to obtain insights on the dependance of system performance on λ_d and $r_{d,\max}$, the most critical operational parameters related to D2D use/business cases.

3.7.1. Scheme 1 (Baseline)

The baseline scheme is the simplest approach for incorporating D2D communications and has been routinely employed in D2D studies, which provides the incentive to investigate it even though it is a special case of the other schemes. The next lemma follows directly from Eq. (3.17) by setting $p = q = 1$.

Lemma 3.15. *The operational region of Scheme 1 equals*[2]

$$\mathcal{R}^1_{D2D} = \{c_1 > e^{c_2}\},$$

for overlay D2D, and

$$\mathcal{R}^1_{D2D} = \{\bar{c}_1 > e^{\bar{c}_2+\bar{c}_3}\}.$$

for underlay D2D.

Lemma 5 provides a simple description of the operational region of Scheme 1, allowing the system operator to check whether the introduction of a baseline D2D mode

[2]The notation of operational region will discriminate among schemes by use of a superscript, with no differentiation between overlay and underlay D2D cases as this will be clear from context. For simplicity, the sets describing the operational region for overlay D2D will be compactly denoted only in terms of constraints on (c_1, c_2) (as defined in Proposition 2) with the understanding that $(c_1, c_2) \in \mathbb{R}^{+2}$, and similarly for the underlay case.

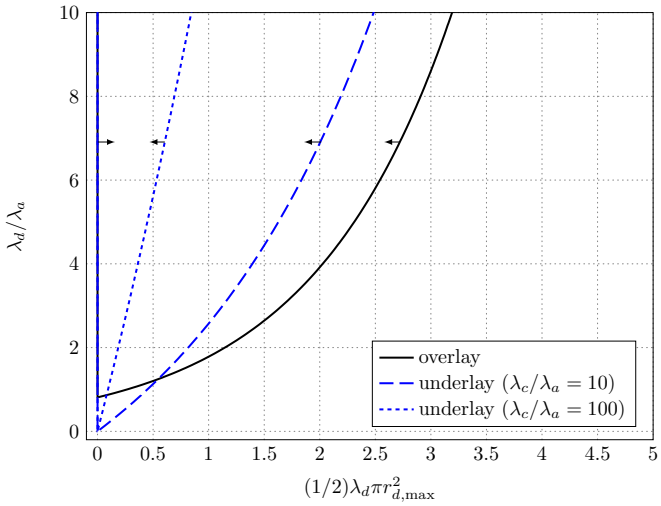


Figure 3.1.: $\mathcal{R}^1_{\text{D2D}}$ for overlay and underlay D2D ($\alpha = 4$, $\theta_0 = -6$ dB). Lines depict the boundary and arrows point to the interior of the region, respectively.

can provide benefits for a system operational point of interest (the latter may be determined by potential D2D-related business/use cases). To aid the visual comparison between overlay and underlay D2D options and examine the effects of λ_d and $r_{d,\max}$ on performance, Fig. 3.1 depicts $\mathcal{R}^1_{\text{D2D}}$ in terms of quantities λ_d/λ_a and $\lambda_d \mathbb{E}(\pi r_d^2) = (1/2)\lambda_d \pi r_{d,\max}^2$ for the test case of $\alpha = 4$ and $\theta_0 = -6$ dB, the latter value corresponding to the minimum SIR supported by current wireless standards [32]. Note that λ_d/λ_a equals the average number of D-UEs within a random cell [18], whereas $\lambda_d \mathbb{E}(\pi r_d^2)$ can be interpreted as the average number of D-UE sources that are closer to a randomly selected D-UE than its own source. For the underlay case, the region is parameterized by the average number of C-UEs within a random cell, λ_c/λ_a (overlay performance is independent of λ_c).

The expressions of Lemma 5 verify the intuition that baseline D2D communications cannot provide benefits with excessively large $r_{d,\max}$, although the maximum supported $r_{d,\max}$ increases with λ_d/λ_a. Interestingly, the operational region for overlay D2D case requires a minimum λ_d/λ_a, even when $r_{d,\max} \to 0$ ($c_2 \to 0$), as it must always hold $c_1 > 1 \Rightarrow \lambda_d/\lambda_a > 1/(1+\rho(\theta_0))$. This might be surprising as with an arbitrarily small D2D link distance all D2D transmissions are successful and one might expect that D2D communications would therefore be beneficial irrespective of λ_d in that case. However, for sufficiently small λ_d, the portion of bandwidth dedicated to overlay D2D communications according to Eq. (3.13) is so small that the average rate achieved by the reliable D2D links is in fact smaller than the one provided by the less reliable and time shared, but of much higher bandwidth, cellular links.

In contrast, noting that $\bar{c}_1 > \lambda_c/\lambda_a$ by definition and $\lambda_c/\lambda_a > 1$ under Assumption 2, underlay D2D is able to accommodate any λ_d/λ_a for arbitrarily small $r_{d,\max}$ ($\bar{c}_2 + \bar{c}_3 \to 0$). However, underlay operation imposes stricter constraints on the max-

imum supported $r_{d,\max}$ than overlay operation when moderate to large λ_d/λ_a are considered. This difference becomes more pronounced for increasing λ_c/λ_a with underlay $\mathcal{R}^1_{\mathrm{D2D}} \to \varnothing$ asymptotically. Considering D2D use/business cases that require operation with large D-UE density and D2D link distances, it follows that the overlay option is more appropriate when Scheme 1 is considered for D2D deployment.

3.7.2. Scheme 2 (D-UEs Employ D2D Mode by Default)

It is straightforward to show the following lemma by setting $p = 1$ in Eq. (3.17).

Lemma 3.16. *The operational region of Scheme 2 equals* $\mathcal{R}^2_{D2D} = \mathcal{R}^1_{D2D} \cup \hat{\mathcal{R}}^2_{D2D}$, *with* $\mathcal{R}^1_{D2D}$ *as in Lemma 5,*

$$\hat{\mathcal{R}}^2_{D2D} = \{c_1 > ec_2, c_2 > 1\}$$

for overlay D2D, and

$$\hat{\mathcal{R}}^2_{D2D} = \{\bar{c}_1 > e(\bar{c}_2 + \bar{c}_3), \bar{c}_2 + \bar{c}_3 > 1\}$$

for underlay D2D. The rate maximizing access probability equals

$$q^* = \begin{cases} \frac{1}{c_2} < 1 & \text{for overlay D2D,} \\ \frac{1}{\bar{c}_2 + \bar{c}_3} < 1 & \text{for underlay D2D,} \end{cases}$$

when the system operates in $\hat{\mathcal{R}}^2_{D2D}$, *and* $q^* = 1$ *otherwise.*

As expected, $\mathcal{R}^2_{\mathrm{D2D}} \supset \mathcal{R}^1_{\mathrm{D2D}}$, i.e., introduction of a channel access mechanism to the baseline scheme can only result in an increase of the operational region. This region augmentation is due to the inclusion of $\hat{\mathcal{R}}^2_{\mathrm{D2D}}$ which is depicted in Fig. 3.2. As can be seen, $\hat{\mathcal{R}}^2_{\mathrm{D2D}}$ does not include operational points with small $r_{d,\max}$ (for finite λ_c/λ_a), as in that case the SIR of a D2D link is large enough and use of a channel access mechanism with $q < 1$ incurs performance loss. It can also be shown that $\hat{\mathcal{R}}^2_{\mathrm{D2D}} \cap \mathcal{R}^1_{\mathrm{D2D}} \neq \varnothing$ (compare Fig. 3.1 and Fig. 3.2), i.e., Scheme 2 not only expands $\mathcal{R}^1_{\mathrm{D2D}}$ but also improves performance in the subset of $\mathcal{R}^1_{\mathrm{D2D}}$ corresponding to larger $r_{d,\max}$ values.

Regarding the overlay/underlay comparison for Scheme 2, it can be easily verified that the maximum R provided by the underlay version is greater than the corresponding one provided by the overlay version for any operational point, which also implies a greater operational region for underlay Scheme 2 as can also be seen in Fig. 3.2.

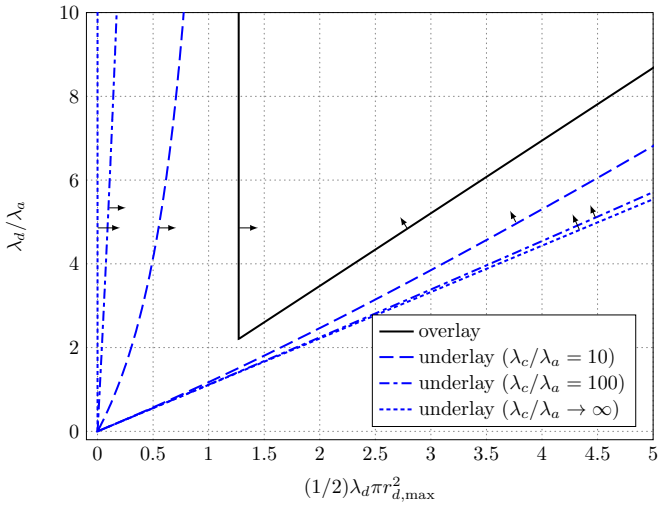


Figure 3.2.: $\hat{\mathcal{R}}^2_{\mathrm{D2D}}$ for overlay and underlay D2D ($\alpha = 4$, $\theta_0 = -6$ dB). Lines depict the boundary and arrows point to the interior of the region, respectively.

3.7.3. Scheme 3-p (Probabilistic Mode Selection with D2D Links Always Active)

Proposition 3.17. *The operational region of Scheme 3-p equals* $\mathcal{R}^{3\text{-}p}_{D2D} = \mathcal{R}^1_{D2D} \cup \hat{\mathcal{R}}^{3\text{-}p}_{D2D}$, *with* $\mathcal{R}^1_{D2D}$ *as in Lemma 5,*

$$\hat{\mathcal{R}}^{3\text{-}p}_{D2D} = \left\{ c_2 e < c_1 < \frac{e^{c_2}}{2 - c_2}, 1 < c_2 < 2 \right\} \cup \{ c_2 e < c_1, c_2 \geq 2 \}$$

for overlay D2D, and

$$\hat{\mathcal{R}}^{3\text{-}p}_{D2D} = \left\{ e^{\bar{c}_3} < \bar{c}_1 < \frac{e^{\bar{c}_2 + \bar{c}_3}}{1 - \bar{c}_2}, \bar{c}_2 < 1 \right\} \cup \{ e^{\bar{c}_3} < \bar{c}_1, \bar{c}_2 \geq 1 \}$$

for underlay D2D. The rate maximizing D2D mode selection probability equals

$$p^* = \begin{cases} \frac{x^*}{c_2} < 1 & \text{for overlay D2D,} \\ \frac{x^*}{\bar{c}_2} < 1 & \text{for underlay D2D,} \end{cases} \tag{3.19}$$

when the system operates in $\hat{\mathcal{R}}^{3\text{-}p}_{D2D}$, *where* x^* *is the unique solution of* $xe^{-x}(2-x) = c_2/c_1$, $1 < x < \min\{2, c_2\}$, *and* $e^{-x}(1-x) = e^{\bar{c}_3}/\bar{c}_1$, $0 < x < \min\{1, \bar{c}_2\}$, *for overlay and underlay D2D, respectively. For all other operational points of* $\mathcal{R}^{3\text{-}p}_{D2D}$, $p^* = 1$.

Proof. See Appendix sec. B.4. □

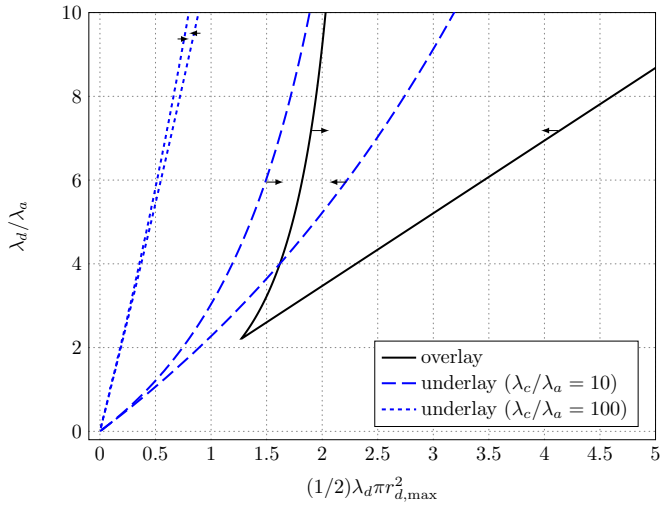


Figure 3.3.: $\hat{\mathcal{R}}_{\text{D2D}}^{\text{3-p}}$ for overlay and underlay D2D ($\alpha = 4$, $\theta_0 = -6$ dB). Lines depict the boundary and arrows point to the interior of the region, respectively.

Fig. 3.3 depicts $\hat{\mathcal{R}}_{\text{D2D}}^{\text{3-p}}$ for overlay and underlay D2D. Similar observations as for Scheme 1 also hold here, i.e., overlay and underlay regions of Scheme 3-p are partially overlapping, with the overlay version supporting larger $r_{d,\max}$ for moderate to large λ_d/λ_a, whereas underlay version is severely limited with increasing λ_c/λ_a.

An interesting observation that directly follows from Lemma 6 and Proposition 5 is that, for the overlay case only, $\mathcal{R}_{\text{D2D}}^{\text{3-p}} = \mathcal{R}_{\text{D2D}}^{2}$, i.e., *overlay versions of Schemes 2 and 3-p provide the same operational region even though under a completely different approach.* However, as shown in Proposition 4, Scheme 2 is a better choice in terms of maximum R. For the underlay case, a clear-cut relation between Schemes 2 and 3-p does not exist, as there are operational points included in $\mathcal{R}_{\text{D2D}}^{\text{3-p}}$ and not in $\mathcal{R}_{\text{D2D}}^{2}$ and vice versa. However, it can be shown that for the operational points belonging to the operational regions of both underlay Scheme 2 and Scheme 3-p, Scheme 2 provides the largest maximum R.

3.7.4. Scheme 3-d (Distance-Based Mode Selection with D2D Links Always Active)

The results for this scheme are presented separately for its overlay and underlay versions as the expression of f for the latter does not allow for exact closed form expressions (as was the case for the schemes examined so far) and necessitates the use of bounding techniques.

Proposition 3.18. *The operational region of Scheme 3-d for overlay D2D equals* $\mathcal{R}_{D2D}^{3\text{-}d} = \mathcal{R}_{D2D}^{1} \cup \hat{\mathcal{R}}_{D2D}^{3\text{-}d}$, *with* $\mathcal{R}_{D2D}^{1}$ *as in Lemma 5 and*

$$\hat{\mathcal{R}}_{D2D}^{3\text{-}d} = \left\{ \sqrt{2c_2 e} < c_1 < \frac{e^{c_2}}{2(1-c_2)}, \frac{1}{2} < c_2 < 1 \right\} \cup \{c_1 > \sqrt{2c_2 e}, c_2 \geq 1\}. \tag{3.20}$$

The optimal D2D mode selection probability equals

$$p^* = \sqrt{\frac{x^*}{c_2}} < 1 \tag{3.21}$$

when operating in $\hat{\mathcal{R}}_{D2D}^{3\text{-}d}$, where x^* is the unique solution of $\sqrt{x}e^{-x}(1-x) = \sqrt{c_2}/(2c_1)$, $1/2 < x < \min\{1, c_2\}$, and $p^* = 1$ otherwise.

Proof. Proof is similar to the case of Scheme 3-p and is omitted. □

Proposition 3.19. *The operational region of Scheme 3-d for underlay D2D equals*

$$\mathcal{R}_{D2D}^{3\text{-}d} = \{\bar{c}_1 > 1\}.$$

The region $\hat{\mathcal{R}}_{D2D}^{3\text{-}d}$ for which $p^* < 1$ is bounded as $\hat{\mathcal{R}}_{D2D}^{3\text{-}d,(1)} \subseteq \hat{\mathcal{R}}_{D2D}^{3\text{-}d} \subseteq \hat{\mathcal{R}}_{D2D}^{3\text{-}d,(2)}$, where

$$\hat{\mathcal{R}}_{D2D}^{3\text{-}d,(\beta)} = \left\{ 1 < \bar{c}_1 < \frac{e^{\bar{c}_2+\bar{c}_3}}{1-\beta(\bar{c}_2+\bar{c}_3)}, \bar{c}_2+\bar{c}_3 < \frac{1}{\beta} \right\} \cup \left\{ \bar{c}_1 > 1, \bar{c}_2+\bar{c}_3 \geq \frac{1}{\beta} \right\} \tag{3.22}$$

for $\beta = 1, 2$.

Proof. See Appendix sec. B.5. □

Proposition 7 provides $\mathcal{R}_{\text{D2D}}^{\text{3-d}}$ in terms of a simple constraint, namely, $\bar{c}_1 > 1$, which, as was stated in the analysis of Scheme 1, holds for any operational point under Assumption 2, leading to the following conclusion.

Corollary 3.20. *Underlay Scheme 3-d enhances cellular system performance in terms of R for any operational point.*

The above result is of significant importance as it shows that *exploitation of knowledge of link distances between D-UEs and their sources with an underlay D2D deployment is sufficient to achieve optimal performance w.r.t. D2D operational region.* In this respect, incorporation of other/additional information, e.g., distance of D-UEs from their closest AP [41], is not necessary (although it may lead to a greater maximum R).

Fig. 3.4 depicts $\hat{\mathcal{R}}_{\text{D2D}}^{\text{3-d}}$ for overlay D2D and the corresponding region bounds for the underlay D2D case. Note that the bounds are tight, with the operational points not included in $\hat{\mathcal{R}}_{\text{D2D}}^{\text{3-d}}$ corresponding to small $r_{d,\max}$ values. For the overlay case, exploitation of D2D link distance information provides significant region enlargement (compare with previous figures) although there exist operational points corresponding to small λ_d/λ_a and large $r_{d,\max}$ that are not included.

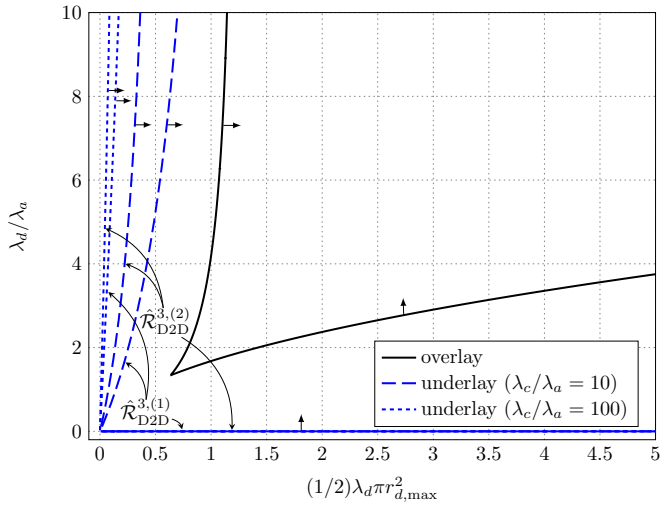


Figure 3.4.: $\hat{\mathcal{R}}_{\text{D2D}}^{\text{3-d}}$ with probabilistic mode selection for overlay and underlay D2D ($\alpha = 4$, $\theta_0 = -6$ dB). Lines depict the boundary and arrows point to the interior of the region, respectively.

3.7.5. Scheme 4-p (Probabilistic Mode Selection and Channel Access for D2D Links)

According to Proposition 3, Scheme 4-p for underlay D2D, is the only candidate scheme for which there exist points within the D2D operational region achieving maximum R with $(p, q) \in (0, 1)^2$. This is indeed the case, as described by the following proposition whose proof is omitted.

Proposition 3.21. *The operational region of Scheme 4-p for underlay D2D equals* $\mathcal{R}_{D2D}^{4\text{-}p} = \mathcal{R}_{D2D}^{2} \cup \mathcal{R}_{D2D}^{3\text{-}p} \cup \hat{\mathcal{R}}_{D2D}^{4\text{-}p}$*, with* $\mathcal{R}_{D2D}^{2}$*,* $\mathcal{R}_{D2D}^{3\text{-}p}$ *as in Lemma 6 and Proposition 5, respectively, and*

$$\hat{\mathcal{R}}_{D2D}^{4\text{-}p} = \left\{ \max\left(\frac{e}{\bar{c}_3}, e\bar{c}_3\right) < \bar{c}_1 < \frac{e(\bar{c}_2 + \bar{c}_3)^2}{\bar{c}_3}, \bar{c}_2 + \bar{c}_3 > 1 \right\}. \tag{3.23}$$

The optimal D2D mode parameters equal

$$p^* = \frac{1}{\bar{c}_2}\left(\sqrt{\frac{\bar{c}_2\bar{c}_3}{e}} - \bar{c}_3\right) < 1, \tag{3.24}$$

$$q^* = \sqrt{\frac{e}{\bar{c}_2\bar{c}_3}} < 1, \tag{3.25}$$

when the system operates in $\hat{\mathcal{R}}_{D2D}^{4\text{-}p}$*,* $p^* = 1$ *and* q^* *as in Lemma 6 when the system operates in* $\mathcal{R}_{D2D}^{2} \setminus \hat{\mathcal{R}}_{D2D}^{4\text{-}p}$*, and* $q^* = 1$ *and* p^* *as in Proposition 5, otherwise.*

The region $\hat{\mathcal{R}}_{\text{D2D}}^{\text{4-p}}$ corresponding to operational points for which $(p^*, q^*) \in (0, 1)^2$ is shown in Fig. 3.5. As can be seen, the region expansion offered by Scheme 4-p

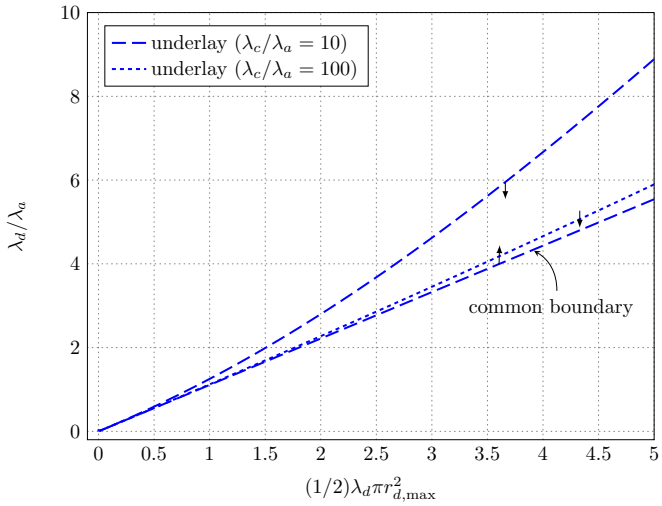


Figure 3.5.: $\hat{\mathcal{R}}_{\text{D2D}}^{\text{4-p}}$ for underlay D2D ($\alpha = 4$, $\theta_0 = -6$ dB). Lines depict the boundary and arrows point to the interior of the region, respectively.

is rather limited, with $\hat{\mathcal{R}}_{\text{D2D}}^{\text{4-p}} \to \varnothing$ for increasing λ_c/λ_a. Note that, even though underlay Scheme 4-p allows for both non-trivial mode selection and D2D channel access procedures, its region does not include all possible operational points which can be attributed to the purely probabilistic nature of both schemes that do not exploit per link information as in the case of underlay Scheme 3-d.

Having examined all schemes, Fig. 3.6 compares the operational regions of overlay and underlay D2D deployments achieved with either probabilistic or distance-based mode selection. In addition to the already established fact that underlay D2D with Scheme 3-d is the best option that can support any operational point, it can be seen that an overlay deployment with probabilistic mode selection provides the smaller operational region, whereas the regions of overlay D2D with distance-based mode selection and underlay D2D with probabilistic mode selection partially overlap. With small operational $r_{d,\max}$, probabilistic mode selection may be preferred as it provides benefits without the cost associated with obtaining distance information, however, distance-based mode selection always provides larger rates (Proposition 5) and is mandatory for large operational $r_{d,\max}$.

Also shown as thin dashed lines in Fig. 3.6 for the case of overlay D2D deployment with probabilistic mode selection, are the maximum-gain level sets, i.e., operational points where the ratio R/R_{noD2D} under optimal (p, q) is constant (each level set is labeled by the value of achieved gain). The level curves of the other schemes show similar trends and are omitted for clarity. It can be seen that incorporation of D2D communications in the system provides largest gains for operational points corresponding to small D2D links and large D-UE density, as expected. Note that small perturbations (e.g., due to changes in D-UE density) of operational points corresponding to large D2D link distances result in small changes of gain, whereas

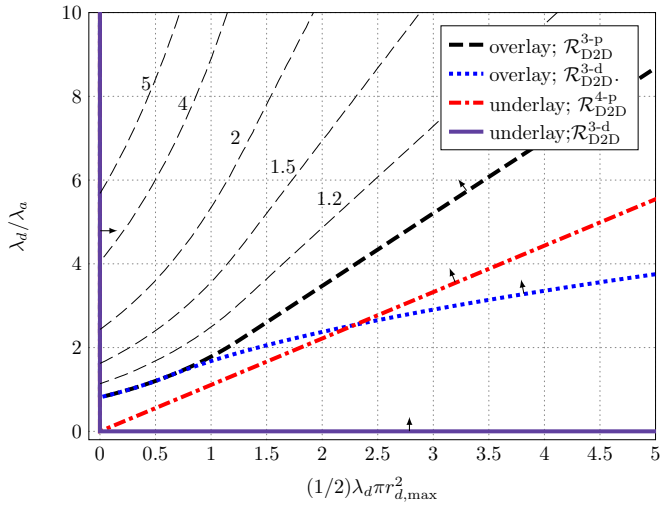


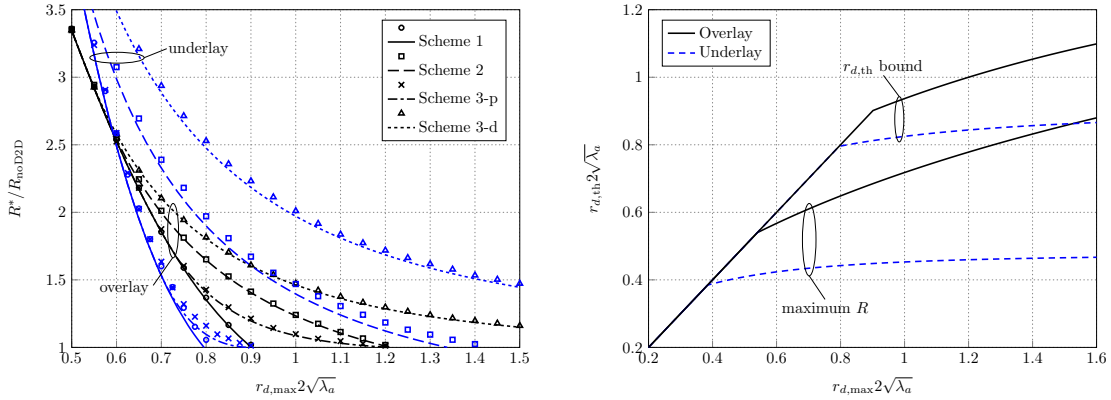
Figure 3.6.: Operational regions of overlay and underlay D2D ($\alpha = 4$, $\theta_0 = -6$ dB, $\lambda_c/\lambda_a = 10$). Thick Lines depict the boundary and arrows point to the interior of the region, respectively. Thin lines represent the maximum-gain level sets of overlay D2D deployment with probabilistic mode selection.

gain is highly sensitive with perturbations of operational points corresponding to small D2D link distances.

3.8. Performance Dependence on Maximum D2D Link Distance ans D-UEs Density

This section further investigates the effect on system performance of $r_{d,\max}$ and λ_d, this time w.r.t. the performance gain provided by the introduction of a D2D mode in terms of maximum R, denoted as R^*. In addition, the presented test cases will also serve to validate the accuracy of the analytical expressions of the previous sections by comparison with Monte Carlo evaluation of Eq. (3.10). The latter was obtained under the system model of Sec. II under Assumption 1, i.e., the cellular uplink was not considered. In addition, the optimal D2D mode parameters of each scheme were employed for each operational point as given in Sec. VI (p^* was obtained by numerical search for the case of underlay Scheme 3-d). In all cases, test values of $\alpha = 4$ and $\theta_0 = -6$ dB were considered, with Monte Carlo results obtained by averaging over 10^5 independent system realizations, each with 30 APs on average. Performance of underlay Scheme 4-p is not depicted as it provides insignificant gains compared to Scheme 2 for the test cases considered.

1) Dependence on $r_{d,\max}$: 3.7a shows the gain ratio R^*/R_{noD2D} provided by the incorporation of a D2D mode to the conventional cellular network as a function of the normalized maximum D2D link distance $r_{d,\max}2\sqrt{\lambda_a}$, where $1/(2\sqrt{\lambda_a})$ is the



(a) **(b)**

Figure 3.7.: Dependence on normalized $r_{d,\max}$ of (a) rate improvement of D2D-enabled system and (b) normalized threshold distance employed by Scheme 3-d. Lines depict analytical results and markers depict Monte Carlo evaluations ($a = 4, \theta_0 = -6$ dB, $\lambda_c/\lambda_a = \lambda_d/\lambda_a = 10$).

average distance of a UE from its closest AP [45]. This normalization is convenient, not only because it relates the distances involved in setting up cellular and D2D links, but also since the achieved rate R depends only on the ratios of densities, λ_c/λ_a and λ_d/λ_a, and not on their absolute values, as can be directly verified from Eqs. (3.16) and (3.17). For the results of 3.7a, a test case of $\lambda_c/\lambda_a = \lambda_d/\lambda_a = 10$ was considered, i.e., there are on average 10 C-UEs and 10 D-UEs within each cell, corresponding to a (future) operational scenario where D2D use cases constitute a significant part of the system load.

It can be seen that the analytical results for the overlay case match almost exactly the Monte Carlo evaluations, whereas for the underlay case, they slightly underestimate performance. Similar correspondence of analytical and Monte Carlo evaluation results was observed for all operational points that were tested (not shown).

As expected, the performance gain of all schemes (either overlay or underlay) diminishes with increasing $r_{d,\max}$. For $r_{d,\max} \to 0$, all schemes boil down to Scheme 1, with $R^*/R_{\text{noD2D}} \approx 12.82, 6.67$, for underlay and overlay D2D, respectively (not shown in the figure). This superiority of underlay D2D holds for increasing $r_{d,\max}$ values up to $r_{d,\max} \approx 0.6 \times 1/(2\sqrt{\lambda_a})$. For larger $r_{d,\max}$, the ordering of the overlay/underlay schemes becomes involved and dependent on $r_{d,\max}$.

In accordance with the analysis of the previous section, all schemes with the exception of Scheme 3-d provide gains up to a certain maximum $r_{d,\max}$. An important observation is that these maximum values are of the order of the distance from the closest AP, i.e., *introduction of D2D communications can enhance cellular network performance even for distances that are not restricted to the common notion of proximal communications assuming distance significantly smaller that the distance from*

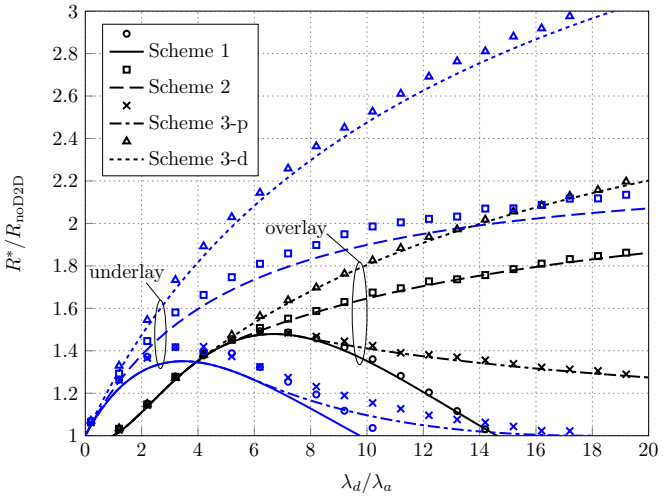


Figure 3.8.: Performance gain of the various D2D deployment schemes as a function of λ_d/λ_a. Lines depict analytical results and markers depict Monte Carlo evaluations ($\alpha = 4$, $\theta_0 = -6$ dB, $\lambda_c/\lambda_a = 10$).

the cellular APs.

Regarding Scheme 3-d, it should be emphasized that even though it is able to provide gains for large values of $r_{d,\max}$, it does so by only allowing the establishment of D2D links of distance $r_d \leq r_{d,\text{th}} = \sqrt{p} r_{d,\max}$, which is strictly smaller than $r_{d,\max}$ for $p < 1$. 3.7b shows the (normalized) threshold value $r_{d,\text{th}}$ when $p = p^*$ as a function of the (normalized) maximum D2D link distance $r_{d,\max}$ (curves labeled "maximum R"). In particular, for $r_{d,\max}$ equal to 1.5 times the average distance from the closest AP, underlay Scheme 3-d only permits establishment of D2D links of distances up to about 0.45 times the average distance from the closest AP, with the overlay version being less conservative. Note that if larger values of $r_{d,\text{th}}$ are needed, e.g., due to application requirements, these can be obtained by allowing some performance degradation w.r.t. R. In particular, $r_{d,\text{th}}$ may be set to any positive value not exceeding $\sqrt{\sup(p)} r_{d,\max}$ with $\sup(p)$ the supremum of p for which $R > R_{\text{noD2D}}$ holds. The latter can be found numerically and the corresponding bound on $r_{d,\text{th}}$ is shown in 3.7b (curves labeled "$r_{d,\text{th}}$ bound"). It can be seen that relaxing the performance gain provided by D2D transmissions can significantly enlarge $r_{d,\text{th}}$, however, still remaining strictly smaller than $r_{d,\max}$ when the latter exceeds a certain threshold.

2) Dependance on λ_d/λ_a: Fig. 3.8 shows R^*/R_{noD2D} as a function of λ_d/λ_a for $\lambda_c/\lambda_a = 10$, and $r_{d,\max} = 0.8 \times 1/(2\sqrt{\lambda_a})$. This study corresponds to a scenario of interest for network operators, where the network is called to accommodate for an increasing number of D-UEs due to the increased penetration of D2D use cases and introduction of D2D enabled devices.

Note that for this scenario, performance of both conventional and D2D-enabled cellular network decreases with increasing λ_d/λ_a as there are more UEs competing for

the same set of resources. However, as can be seen from Fig. 3.8, R^*/R_{noD2D} is an increasing function of λ_d/λ_a when the latter is small, i.e., D2D-enabled cellular network performance degrades slower than the conventional network since it exploits the direct communication possibilities. For the overlay case, this advantage of D2D-enabled network is observed only above a threshold on λ_d/λ_a (see corresponding discussion following Lemma 5). Interestingly, Schemes 1 and 3-p provide increasing performance up to a certain value of λ_d/λ_a above which the performance degrades due to the increasingly high interference from the D2D transmissions that cannot be handled by these two schemes. In contrast, both schemes 2 and 3-d show a monotonically increasing performance gain as λ_d/λ_a increases. This trend can be verified analytically by direct substitution of the expression for q^* as provided in Lemma 5 into in Eq. (3.16) and the fact that Scheme 3-d $\succeq$ Scheme 2. Intuitively, Scheme 3-d provides the largest gains since the increased number of D-UEs in conjunction with the distance-based mode selection procedure results in establishing many D2D links of small link distances, essentially exploiting a form of multiuser diversity.

3.9. Conclusion

This chapter investigated and characterized the operational region of D2D communications for enhancing the performance of the conventional cellular network in terms of average user rate. Various D2D deployment schemes were examined for both overlay and underlay options. For the important case of a heavy loaded network, the optimal D2D mode parameters as well as the operational region of every scheme were obtained analytically. It was shown that exploitation of D2D link distance for mode selection is alone sufficient to render D2D communications beneficial for any operational point of interest. Under the appropriate D2D deployment scheme, significant performance gains can be achieved even with significant number of D2D-enabled devices and D2D link distances of the order of the distance of UEs to their closest AP. These observations suggest that the introduction of a D2D mode as an add-on to the conventional cellular network has significant potential for success as the operational points where performance enhancement is observed are not restricted to the typically-considered regime corresponding to small distances compared to the distance from the closest AP and user densities.

4. Cross-Mode Interference Characterization in Cellular Networks with Voronoi Guard Regions

4.1. Overview

Advances in cellular networks such as device-to-device communications and full-duplex radios, as well as the inherent elimination of intra-cell interference achieved by network-controlled multiple access schemes, motivates the investigation of the cross-mode interference under a guard region corresponding to the Voronoi cell of an access point. By modeling the positions of interfering access points (APs) and user equipments (UEs) as Poisson distributed, analytical expressions for the statistics of the interference generated by either APs or UEs are obtained based on appropriately defined density functions. The considered system model and analysis are general enough to capture many operational scenarios of practical interest, including conventional downlink/uplink transmissions with nearest AP association, as well as transmissions where not both communicating nodes lie within the cell. Analysis provides insights on the level of protection offered by a Voronoi guard region and its dependence on type of interference and receiver position. Various examples demonstrate the validity/accuracy of the analysis in obtaining the system coverage probability for operational scenarios of practical interest.

The results of this chapter were published in

1. S. Stefanatos, A. G. Gotsis, and A. Alexiou, "Cross-Mode Interference Characterization in Cellular Networks with Voronoi Guard Regions," *IEEE Transactions on Wireless Communications*, Feb. 2017, submitted.

4.2. Introduction

Characterization of the interference experienced by the receivers of a wireless network is of critical importance for system analysis and design [45]. This is a task that

has been performed by engineers throughout the last decades resulting in the successful deployment of wireless networks with the most prominent example being the cellular network. Cellular networks are currently experiencing fundamental changes in their architecture, technology, and operation [52], which have significant impact on the interference footprint. Interference characterization under these new system features is of the utmost importance in order to understand their potential merits as well as their ability to co-exist.

Towards increasing the spatial frequency reuse, two of the most prominent techniques/features considered for the future cellular network are device-to-device (D2D) communications [53] and full-duplex (FD) radios [54]. Although promising, application of these techniques introduces additional, cross-mode interference. For example, an uplink transmission is no longer affected only by interfering uplink transmissions but also by interfering D2D and/or downlink transmissions. Although it is reasonable to expect that the current practice of coordinated transmissions within the cell of each access point (AP) as a means to eliminate intra-cell interference will also hold in the future [55], the continually increasing density of APs and user equipment (UEs) suggest that inter-cell interference will be the major limiting factor of D2D- and/or FD-enabled cellular networks, rendering its statistical characterization critical.

4.2.1. Previous Work

Stochastic geometry is by now a well accepted framework for analytically modeling interference in large-scale wireless networks [56]. Under this framework, most of the numerous works on D2D-enabled cellular networks (without FD radios) consider the interfering D2D nodes as uniformly distributed on the plane, i.e., there is no spatial coordination of D2D transmissions (see, e.g., [40, 39, 38, 91]). Building on the approach of [58], various works consider the benefits of employing spatially coordinated D2D transmissions where for each link in the system a circular guard region (zone) is established, centered at either the receiver or transmitter, within which no interfering transmissions are performed [42, 59]. This type of guard zone can be viewed as a model for a CSMA-type transmission protocol [45, Sec. 3.7], which implies that D2D links operate as an independent-from-APs tier. However, when the D2D links are network controlled [55], a more natural and easier to establish guard region is the (Voronoi) cell of a coordinating AP. Under a non-regular AP deployment [60], this approach results in a random polygon guard region, which makes the interference characterization a much more challenging task.

Interference characterization for this type of guard region has only been partially investigated in [61, 62, 63] where conventional uplink transmissions with nearest AP association and one active UE per cell are considered. As the interfering point process is distributed as a Voronoi perturbed lattice process (VPLP) [64, 65], which is analytically intractable, it is approximated as a Poisson point process (PPP) of an

appropriately defined density. This approach is shown to have reasonable accuracy for uplink interference characterization, which suggests the same approach also for the case when the considered receiver is located at an arbitrary position in the plane (not that of an AP). However, the interferer density functions considered in the works on uplink analysis are heuristically proposed and, as shown in [65], they may change drastically when a receiver position other than that of the coordinating AP is considered. In addition, the uplink analysis does not hold when the interference is generated by APs.

Investigation of interference statistics under a cell guard region is also missing in the (much smaller) literature on FD-enabled cellular networks, which typically assumes no spatial coordination for the UE transmissions (see e.g., [66, 67, 68]).

4.2.2. Contributions

This chapter considers a stochastic geometry framework for modeling the interference power experienced at an arbitrary receiver position due to transmissions by APs or UEs and under the spatial protection of a Voronoi guard region. The system model is general enough to capture operational scenarios such as conventional downlink and uplink communications with nearest AP association, as well as communications where the receiver location is not guaranteed to lie within the guard region. The latter case has not been previously investigated in the literature and is of particular interest in, e.g., D2D communications where cross-cell links may be established [69].

For this system model, the interference power at the receiver is statistically characterized in terms of its Laplace transform. The exact Laplace transform is obtained for the case of AP-generated interference, whereas a tight lower bound is provided for the Laplace transform of the UE-generated interference power. Both Laplace transforms are uniquely determined by appropriately defined *equivalent* densities of interferers, which also serve as an intuitive quantity for understanding the interference properties. It is shown that the equivalent interferer density and, hence, interference properties, significantly change when the type of interference (AP- or UE-generated) and/or receiver position change. Emphasis is given on the more challenging for analysis UE-generated interference scenario, where various special instances of the system model are investigated in detail, including the standard Planck communication scenario.

Analysis provides insights, among others, on the validity of the heuristic density functions considered previously for the analysis of the conventional uplink scenario, the effectiveness of a Voronoi guard region formed around the receiver or transmitter, and the position within the Voronoi cell where the interference in its close vicinity is minimal. As an application of the Laplace transform expressions, the coverage probability for certain operational scenarios of practical interest, including that of potentially cross-cell D2D communications, is presented.

4.2.3. Notation

The origin of the two-dimensional plane $\mathbb{R}^2$ will be denoted as o. The Euclidean norm of $x \in \mathbb{R}^2$ is denoted as $|x|$ with operator $|\cdot|$ also used to denote the absolute value of a scalar or the Lebesgue measure (area) of a bounded subset of $\mathbb{R}^2$. The polar form representation of $x \in \mathbb{R}^2$ will be denoted as $(|x|, \angle x)$ or $|x|\angle x$, where $\angle x$ is the principal branch of the angular coordinate of x taking values in $[-\pi, \pi)$. The open ball in $\mathbb{R}^2$, centered at $x \in \mathbb{R}^2$ and of radius $R > 0$, is denoted as $\mathcal{B}(x, R) \triangleq \{y \in \mathbb{R}^2 : |y - x| < R\}$, whereas its boundary is denoted as $\mathcal{C}(x, R) \triangleq \{y \in \mathbb{R}^2 : |y - x| = R\}$. $\mathbb{I}(\cdot)$ is the indicator $(0-1)$ operator, $\mathbb{P}(\cdot)$ is the probability measure, and $\mathbb{E}(\cdot)$ is the expectation operator. Functions $\arccos(\cdot) : [-1, 1] \to [0, \pi]$ and $\arcsin(\cdot) : [-1, 1] \to [-\pi/2, \pi/2]$ are the principal branches of the inverse cosine and sine, respectively. The Laplace transform of a random variable z equals $\mathcal{L}_z(s) \triangleq \mathbb{E}(e^{-sz})$, for all $s \in \mathbb{R}$ for which the expectation exists. A function f defined over some set $\mathcal{S} \subset \mathbb{R}^2$ will be referred to as circularly-symmetric with respect to (w.r.t.) $z \in \mathcal{S}$ if it holds $f(z + x) = f_z(|x|)$, $x \in \mathcal{S}$, where f_z is a function defined over $[0, \infty)$.

4.3. System Model

A large-scale model of a cellular network with APs and UEs positioned over $\mathbb{R}^2$ is considered. The positions of APs are modeled as a realization of a homogeneous PPP (HPPP) $\Phi_a \subset \mathbb{R}^2$ of density $\lambda_a > 0$ (average number of APs per unit area) and the positions of the UEs are modeled as a realization of another, independent HPPP $\Phi_u \subset \mathbb{R}^2$ of density $\lambda_u > 0$ (average number of UEs per unit area). Within the cellular network there exists a node of interest (either AP or UE), referred to in the following as typical node. Without loss of generality (w.l.o.g.), the typical node will be assumed to be positioned at the origin o. The typical node establishes a communication link with another node (AP or UE) arbitrarily positioned on the plane. The receiver position of the link will be denoted as $x_R \in \mathbb{R}^2$ ($x_R = o$ in the case when the typical node is in receive mode).

The receiver experiences interference generated by the other transmitting nodes (APs and UEs) in the network. For simplicity, it will be assumed that all APs and UEs in the system are in transmit mode with a fixed power $P_a > 0$ and $P_u > 0$, respectively. Generalization of the analysis in the case when the transmit powers of the nodes of the same type (APs or UEs) are modeled as independent and identically distributed (i.i.d.) nonnegative random variables is straightforward based on the theory of independently marked PPPs [18].

Let $x^* \triangleq \arg\min_{x \in \Phi_a} |x|$ denote the position of the closest AP to the the typical node ($x^* = o$ in case the typical node is an AP) and let $\mathcal{V}^*$ denote its Voronoi cell generated by the Poisson-Voronoi tessellation of the plane from APs, i.e.,

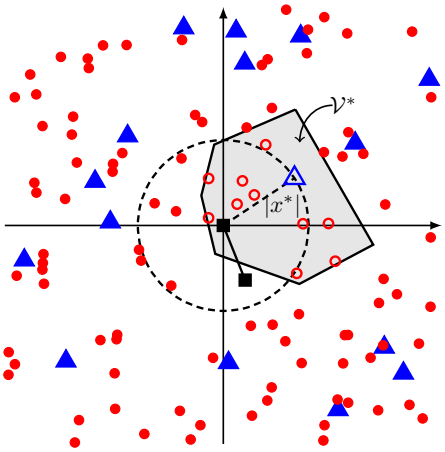


Figure 4.1.: Random realization of the system model. APs and UEs are shown as triangle and circle markers, respectively. The typical node and its associated receiver are UEs shown as square markers. The shaded polygon area indicates the Voronoi guard region $\mathcal{V}^*$ within which no node (indicated by open marker) transmits. All nodes outside $\mathcal{V}^*$ (filled markers) produce interference.

$$\mathcal{V}^* \triangleq \{y \in \mathbb{R}^2 : |y - x^*| < |y - x|, \text{ for all } x \in \Phi_a \setminus \{x^*\}\}.$$

In order to reduce the interference experienced at the receiver position x_R, the AP at x^* coordinates transmissions such that $\mathcal{V}^*$ effectively forms a spatial guard region within which no interference is generated. An example of a Voronoi guard region is shown in Fig. 4.1. Note that $\mathcal{V}^*$ is a *random* guard region with no guarantee that x_R lies within $\mathcal{V}^*$, i.e., it holds $\mathbb{P}(x_R \in \mathcal{V}^*) < 1$, unless x_R is a point on the line segment joining the origin with x^*.

Denoting $I_{x_R,a}$ and $I_{x_R,u}$ the interference power experienced at x_R due to transmissions by APs and UEs that lie outside $\mathcal{V}^*$, respectively, the standard interference model is adopted in this chapter, namely [45]

$$I_{x_R,k} \triangleq \sum_{x \in \Phi_k \setminus \mathcal{V}^*} P_k g_x |x - x_R|^{-\alpha_k}, k \in \{a, u\}, \tag{4.1}$$

where $g_x \geq 0$ the channel gain of a transmission generated by a node at $x \in \mathbb{R}^2$, assumed to be an independent of x, exponentially distributed random variable of mean 1 (Rayleigh fading), and $\alpha_k > 2$ is the path loss exponent. Note that the model is general enough so that, by appropriately choosing x_R and/or x^*, $I_{x_R,a}$ and $I_{x_R,u}$ correspond to many instances of (cross-mode) interference experienced in D2D/FD-enabled as well as conventional cellular networks. For example, with

$x_R = o$ and $x^* \neq o$, $I_{x_R,a}$ is the interference power experienced by a standard downlink transmission with nearest AP association [29], whereas $I_{x_R,u}$ is the cross-mode interference power generated by UEs in, e.g., D2D mode. Various other cases of practical interest will be discussed in the following sections.

For characterization of the performance of the communication link as well as obtaining insights on the effectiveness of the guard region $\mathcal{V}^*$ for reducing interference, it is of interest to characterize the statistical properties of the random variables $I_{x_R,a}$ and $I_{x_R,u}$. This is the topic of the following section, where the marginal statistics of $I_{x_R,a}$ and $I_{x_R,u}$ *conditioned on* x^* *and treating* x_R *as a given, but otherwise free, parameter*, are investigated in detail. The respective unconditional interference statistics can be simply obtained by averaging over the distribution of x^*, which corresponds to a uniformly distributed $\angle x^*$ and a Rayleigh distributed $|x^*|$ of mean $1/(2\sqrt{\lambda_a})$ [18]. Characterization of the joint distribution of $I_{x_R,a}$ and $I_{x_R,u}$ is left for future work.

4.4. Interference Characterization

Towards obtaining a tractable characterization of the distribution of the (AP or UE) interference power, or, equivalently, its Laplace transform, the following standard result in PPP theory is first recalled [18].

Lemma 4.1. *Let* $I \triangleq \sum_{x\in\tilde{\Phi}} Ph_x|x-z|^{-\alpha}$, $z \in \mathbb{R}^2$, $P > 0$, *with* $\tilde{\Phi}$ *an inhomogeneous PPP of density* $\lambda : \mathbb{R}^2 \to [0,\infty)$, *and* $\{h_x\}_{x\in\Phi}$ *i.i.d. exponential random variables of mean* 1. *The Laplace transform of* I *equals*

$$\mathcal{L}_I(s) = \exp\left\{-\int_{\mathbb{R}^2} \lambda(x) g(sP|x-z|^{-\alpha})dx\right\} \tag{4.2}$$

$$= \exp\left\{-2\pi \int_0^\infty \lambda_z(r) r g(sPr^{-\alpha})dr\right\}, \tag{4.3}$$

where $g(t) \triangleq 1 - \frac{1}{1+t}$, *and the second equality holds only in the case of a circularly-symmetric density function w.r.t. point* z, *i.e.,* $\lambda(z+x) = \lambda_z(|x|)$, *for an appropriately defined function* $\lambda_z : [0,\infty) \to [0,\infty)$.

The above result provides a complete statistical characterization in integral form of the interference experienced at a position $z \in \mathbb{R}^2$ due to Poisson distributed interferers. Of particular interest is the case of a circularly-symmetric density, which only requires a single integration over the radial coordinate and has previously led to tractable analysis for various wireless network models of interest such as mobile ad-hoc [45] and downlink cellular [29]. In the following it will be shown that even though the interference model of Sec. II suggests a non circularly-symmetric interference density due to the random shape of the guard region, the Laplace transform formula of (4.3) holds exact for $I_{x_R,a}$ and is a (tight) lower bound for $I_{x_R,u}$ with appropriately defined circularly-symmetric equivalent density functions.

4.4.1. AP-Generated Interference

Since, by definition, no other AP lies within $\mathcal{V}^*$ except the one at x^*, the actual shape of the guard region $\mathcal{V}^*$ has no effect on AP-generated interference apart from restricting the AP at x^* to transmit. However, since x^* is the closest AP position to the origin, the interfering APs, given x^*, are effectively distributed as an inhomogeneous PPP with density [29]

$$\tilde{\lambda}_a(x) = \lambda_a \mathbb{I}(x \notin \mathcal{B}(o, |x^*|)), x \in \mathbb{R}^2, \tag{4.4}$$

i.e., an implicit circular guard zone is introduced around the typical node (see Fig. 4.1). This observation may be used to obtain the Laplace transform of the AP-generated interference experienced at x_R directly from the formula of (4.2). However, the following result shows that the two-dimensional integration can be avoided as the formula of (4.3) is also valid in this case with an appropriately defined equivalent density function.

Proposition 4.2. *The Laplace transform, $\mathcal{L}_{I_{x_R,a}}(s \mid x^*) \triangleq \mathbb{E}(e^{-sI_{x_R,a}} \mid x^*)$, of the AP-generated interference power $I_{x_R,a}$, conditioned on x^*, equals the right-hand side of (4.3) with $P = P_a$, $\alpha = \alpha_a$ and $\lambda_z(r) = \lambda_{x_R,a}(r)$, where*

$$\lambda_{x_R,a}(r) \triangleq \begin{cases} \lambda_a & , r > |x^*| + |x_R| \\ \begin{cases} 0 & , |x^*| > |x_R|, \\ \lambda_a & , |x^*| < |x_R|, \\ \lambda_a/2 & , |x^*| = |x_R| \end{cases} & , r \leq ||x^*| - |x_R|| \\ \lambda_a \left(1 - \arccos\left(d\right)\right) & , \text{otherwise}, \end{cases} \tag{4.5}$$

with $d \triangleq \frac{r^2 - (|x^*|^2 - |x_R|^2)}{2r|x_R|}$, for $x_R \neq o$, and

$$\lambda_{x_R,a}(r) = \lambda_a \mathbb{I}(r \geq |x^*|), \tag{4.6}$$

for $x_R = o$.

Proof. See Appendix sec. C.1. □

The following remarks can be made:

1. The interference experienced at an arbitrary position $x_R \neq o$ under the considered guard region scheme is equal in distribution to the interference experienced at the origin due to interferers distributed as an inhomogeneous PPP of density given by (4.5).

2. Even though derivation of the statistics of $I_{x_R,a}$ was conditioned on x_R and x^*, the result depends only on their norms $|x_R|$ and $|x^*|$.

3. Function $\lambda_{x_R,a}(r)$ appearing in Prop. 4.2 can be considered as an *equivalent* interferering AP density experienced at x_R. However, $\lambda_{x_R,a}(r)$ is equal to the *actual* interferering AP density experienced at x_R when $\angle x_R$ is uniformly distributed in $[-\pi, \pi)$.

4. $\lambda_{x_R,a}(r)$ is a decreasing function of $|x^*|$, corresponding to a (statistically) smaller AP interference power due to an increased guard zone area.

5. For $x_R = o$, the interference corresponds to the one experienced at the receiver of a downlink cellular transmission with nearest AP association and Prop. 4.2 coincides with the analysis of [29]. The case of a downlink transmission without nearest AP association, i.e., where the receiver is not guaranteed to lie in the cell of the serving AP is also captured by the analysis by setting $x^* = o$.

6. The case $x_R = x^*$ corresponds to the nearest-neighbor transmission scenario considered in [70] where the validity of $\lambda_{x_R,a}(r)$ in (4.5) as an equivalent density function for computation of the Laplace transform of the interference power was not observed.

7. In [71], the Laplace transform of the interference power experienced at $x_R \in \mathbb{R}^2$ due to a Poisson hole process, i.e., with interferers distributed as an HPPP over $\mathbb{R}^2$ except in the area covered by randomly positioned disks (holes), was considered. The holes were assumed to not include x_R and a lower bound for the Laplace transform was obtained by considering only a single hole [71, Lemma 5], which coincides with the result of Prop. 4.2. This is not surprising as the positions of the APs, conditioned on x^*, are essentially distributed a single-hole Poisson process. Note that Prop. 4.2 generalizes [71, Lemma 5] by allowing x_R to be covered by the hole and considers a different proof methodology.

Fig. 4.2 shows the normalized density function $\lambda_{x_R,a}(r)/\lambda_a$ for various values of $|x_R|$ and assuming that $|x^*| = 1/(2\sqrt{\lambda_a})$. Recalling that in the absence of guard region, the interfering AP density would be constant over all $\mathbb{R}^2$ and equal to λ_a, it can be seen that the presence of a guard region results in reducing the interferer density in certain intervals of the radial coordinate r, depending on the value of $|x_R|$. In particular, when $|x_R| < |x^*|$, it is guaranteed that no APs exist within a radius $|x^*| - |x_R| > 0$ from x_R. In contrast, when $|x_R| > |x|$ there is no protection from APs in the close vicinity of x_R.

The case $|x_R| = |x^*|$ is particularly interesting since it corresponds to the case when the receiver of interest is the AP at x^*. For $x^* \neq o$, corresponding to an uplink transmission by the typical node to its nearest AP, it can be easily shown that it

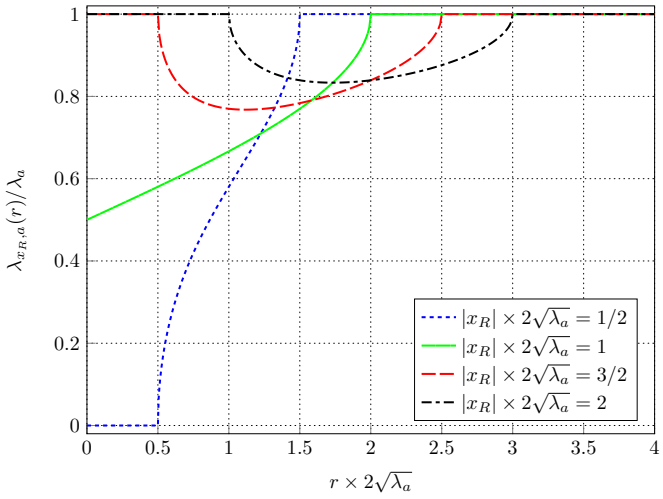


Figure 4.2.: Equivalent radial density of interfering APs experienced at various $x_R \in \mathbb{R}^2$, conditioned on $|x^*| = 1/(2\sqrt{\lambda_a})$.

holds

$$\lambda_{x^*,a}(r) = \lambda_a \left(\frac{1}{2} + \frac{r}{2\pi|x^*|} \right) + \mathcal{O}(r^3), r \to 0, \tag{4.7}$$

i.e., the guard region results in the serving AP experiencing about half of the total interfering APs density in its close vicinity. Fig. 4.3 shows the normalized $\lambda_{x^*,a}(r)$ for various values of $|x^*|$, where its linear asymptotic behavior as well as the advantage of a larger $|x^*|$ are clearly visible.

4.4.2. UE-Generated Interference

The positions of interfering UEs, conditioned on the realization of Φ_a, are distributed as an inhomogeneous PPP of density

$$\tilde{\lambda}_u(x|\Phi_a) = \lambda_u \mathbb{I}(x \notin \mathcal{V}^*), x \in \mathbb{R}^2. \tag{4.8}$$

Although the expression for $\tilde{\lambda}_u$ is very similar to the that of the density $\tilde{\lambda}_a$ of interfering APs given in (4.4), the random Voronoi guard region $\mathcal{V}^*$ appearing in (4.8) is significantly more complicated than the deterministic circular guard zone $\mathcal{B}(o, |x^*|)$, which renders the analysis more challenging. To simplify the following exposition, it will be assumed that a rotation of the Cartesian coordinate axes is performed such that $\angle x^* = 0$. Note that this rotation has no effect in the analysis

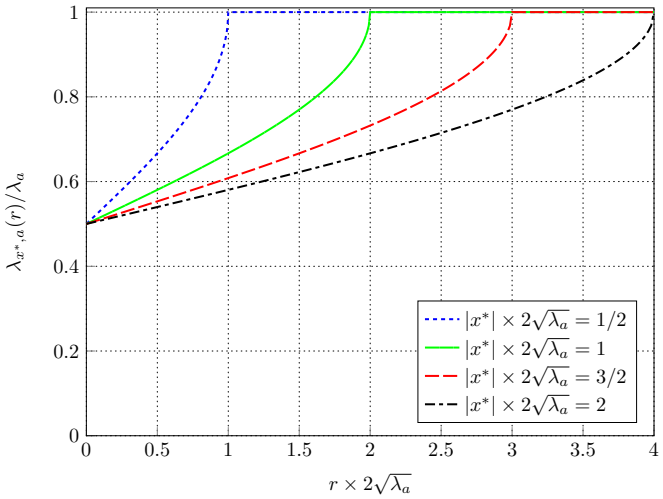


Figure 4.3.: Equivalent radial density of interfering APs experienced at x^*.

due to the isotropic property of the HPPP [45] and immediately renders the following results independent of the value of $\angle x^*$ in the original coordinate system.

Since it is of interest to examine the UE interference statistics conditioned only on x^*, a natural quantity to consider, that will be also of importance in the interference statistics analysis, is the *probabilistic cell area* (PCA) function $p_c(x \mid x^*), x \in \mathbb{R}^2$. This function gives the probability that a point $x \in \mathbb{R}^2$ lies within $\mathcal{V}^*$ conditioned only on x^*, i.e.,

$$p_c(x \mid x^*) \triangleq \mathbb{P}(x \in \mathcal{V}^* \mid x^*).$$

Lemma 4.3. *For all $x \in \mathbb{R}^2$, the PCA function equals*

$$p_c(x \mid x^*) = \begin{cases} 1 & , |x| \leq |x^*|, \angle x = 0, \\ e^{-\lambda_a |\mathcal{A}|} < 1 & , \text{otherwise}, \end{cases}$$

where $\mathcal{A} \triangleq \mathcal{B}(x, |x - x^*|) \setminus \mathcal{B}(o, |x^*|)$. For all $x \neq x^*$, it holds

$$|\mathcal{A}| = \begin{cases} r_*^2(|\angle x| + \theta_*) - |x^*|^2 |\angle x| + |x||x^*| \sin(|\angle x|) & , x^* \neq o, \\ \pi |x|^2 & , x^* = o, \end{cases}$$

with $r_* \triangleq |x - x^*| = \sqrt{|x|^2 + |x^*|^2 - 2|x||x^*| \cos(\angle x)}$ and

$$\theta_* \triangleq \begin{cases} \pi - \arcsin\left(\frac{|x| \sin(|\angle x|)}{r_*}\right) & , |x| \cos(\angle x) > |x^*|, \\ \arcsin\left(\frac{|x| \sin(|\angle x|)}{r_*}\right) & , |x| \cos(\angle x) \leq |x^*|. \end{cases}$$

Proof. The probability that the point $x \in \mathbb{R}^2$ belongs to $\mathcal{V}^*$ is equal to the probability that there does not exist a point of $\Phi_a \setminus \{x^*\}$ within the set $\mathcal{A}$, which equals $e^{-\lambda_a \pi |\mathcal{A}|}$ [18]. For $|x| \leq |x^*|$ and $\angle x = 0$, it is a simple geometrical observation that $\mathcal{A} = \emptyset$, therefore, $|\mathcal{A}| = 0$. When $x^* = o$, $\mathcal{A} = \mathcal{B}(x, |x|) \setminus \mathcal{B}(o, 0) = \mathcal{B}(x, |x|)$, and, therefore, $|\mathcal{A}| = \pi |x|^2$. For all other cases, $|\mathcal{A}|$ can be computed by the same approach as in the proof of [72, Theorem 1]. The procedure is straightforward but tedious and is omitted. □

A simple lower bound for $p_c(x \mid x^*)$ directly follows by noting that $\mathcal{A} \subseteq \mathcal{B}(x, |x - x^*|)$ for all $x \in \mathbb{R}^2$.

Corollary 4.4. *The PCA function is lower bounded as*

$$p_c(x \mid x^*) \geq e^{-\lambda_a \pi |x - x^*|^2}, x \in \mathbb{R}^2, \tag{4.9}$$

with equality if and only if $x^* = o$.

Remark: The right-hand side of (4.9) equals the probability that x belongs to the Voronoi cell of the AP positioned at x^* when the latter is *not* conditioned on being the closest AP to the origin or any other point in $\mathbb{R}^2$.

Fig. 4.4 depicts $p_c(x \mid x^*)$ for the case where $|x^*| = 1/\sqrt{\lambda_a}$ (behavior is similar for other values of $|x^*| > 0$). Note that, unless $x^* = o$, $p_c(x \mid x^*)$ is not circularly symmetric w.r.t. any point in $\mathbb{R}^2$, with its form suggesting that points isotropically distributed in the vicinity of x^* are more probable to lie within $\mathcal{V}^*$ than points isotropically distributed in the vicinity of o.

The lower bound $e^{-\lambda_a \pi |x - x^*|^2}, x \in \mathbb{R}^2$, is also shown in Fig. 4.4 , clearly indicating the *probabilistic expansion effect* of the Voronoi cell of an AP, when the latter is conditioned on being the closest to the origin. This cell expansion effect is also demonstrated in Fig. 4.5 where the conditional average cell area, equal to

$$\begin{aligned} \mathbb{E}(|\mathcal{V}^*| \mid x^*) &= \mathbb{E}\left(\int_{\mathbb{R}^2} \mathbb{I}(x \in \mathcal{V}^*) dx \middle| x^* \right) \\ &= \int_{\mathbb{R}^2} p_c(x | x^*) dx, \end{aligned} \tag{4.10}$$

is plotted as a function of the distance $|x^*|$, with the integral of (4.10) evaluated numerically using Lemma 4.3. Simulation results are also depicted serving as a verification of the validity of Lemma 4.3. It can be seen that the average area of the guard region increases with $|x^*|$, which implies a corresponding increase of the average number of UEs that lie within the region. However, as will be discussed in the following, even though resulting in more UEs silenced on average, an increasing

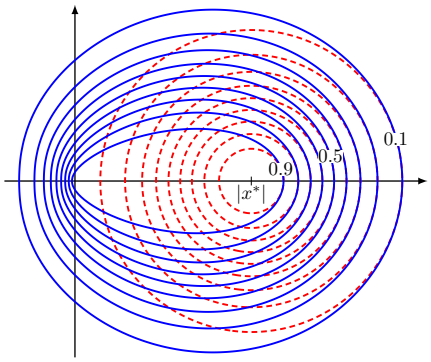


Figure 4.4.: Contours of $p_c(x \mid x^*)$ (solid lines), corresponding to probabilities 0.1 to 0.9 in steps of 0.1, for $|x^*| = 1/\sqrt{\lambda_a}$. The points of the line segment joining o to x^* are the only ones with $p_c(x|x^*) = 1$. The corresponding contours of $e^{-\lambda_a \pi |x-x^*|^2}$ are also shown (dashed lines).

$|x^*|$ *does not necessarily* imply improved protection from UE interference, depending on the receiver position.

The interference statistics of the UE-generated interference power are given in the following result.

Proposition 4.5. *The Laplace transform* $\mathcal{L}_{I_{x_R,u}}(s \mid x^*) \triangleq \mathbb{E}(e^{-sI_{x_R,u}} \mid x^*)$ *of the UE-generated interference power* $I_{x_R,u}$*, conditioned on* x^**, is lower bounded as the right-hand side of (4.3) with* $P = P_u$*,* $\alpha = \alpha_u$ *and* $\lambda_z(r) = \lambda_{x_R,u}(r)$*, where*

$$\lambda_{x_R,u}(r) \triangleq \lambda_u \left(1 - \frac{1}{2\pi} \int_0^{2\pi} p_c\left((r,\theta) + x_R \mid x^*\right) d\theta\right), \tag{4.11}$$

with $p_c(x \mid x^*)$ *as given in Lemma 4.3.*

Proof. See Appendix sec. C.2. □

The following remarks can be made:

1. The statistics of the interfering UE point process when averaged over $\mathcal{V}^*$ (equivalently, over $\Phi_a \setminus \{x^*\}$) *do not* correspond to a PPP, even though this is the case when a fixed realization of $\mathcal{V}^*$ is considered. Therefore, it cannot be expected that Lemma 4.1 applies for $\mathcal{L}_{I_{x_R,u}}(s \mid x^*)$. However, as Prop. 4.5 shows, when the interfering point process is *treated* in the analysis as an PPP of an appropriately defined equivalent density function $\lambda_{x_R,u}(r)$, a tractable lower bound for $\mathcal{L}_{I_{x_R,u}}(s \mid x^*)$ is obtained, which will be shown to be tight in the numerical results section.

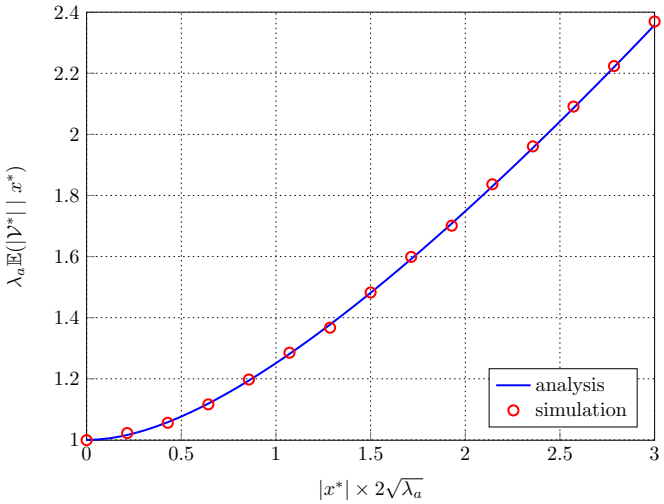


Figure 4.5.: Average area of $\mathcal{V}^*$ conditioned on x^* as a function of the (normalized) distance of x^* from the origin.

2. For $x^* \neq o$, the non circularly-symmetric form of the PCA function results in an integral-form expression for the equivalent density $\lambda_{x_R,u}(r)$ that depends in general on $\angle x_R$, in contrast to the case for $\lambda_{x_R,a}(r)$ (see Sec. III.A, Remark 2). A closed form expression for the equivalent density is only available for $x^* = o$ discussed below.

3. When the receiver position is specified as $|x_R|\angle x_R$ with $\angle x_R$ uniformly distributed over $[-\pi, \pi)$, a straightforward extension of the proof of Prop. 4.5 results in the same lower bound for the conditioned Laplace transform of $I_{x_R,u}$ as in Prop. 4.5, with $\lambda_{x_R,u}(r)$ replaced by

$$\lambda_{|x_R|,u}(r) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} \lambda_{x_R,u}(r) d\angle x_R, \tag{4.12}$$

which is independent of $\angle x_R$.

Although $\lambda_{x_R,u}(r)$ is not available in closed form when $x^* \neq o$, the numerical integration over a circular contour required in (4.11) is very simple. With $\lambda_{x_R,u}(r), r \geq 0$, pre-computed, evaluation of the lower bound for $\mathcal{L}_{I_{x_R,u}}(s \mid x^*)$ is of the same (small) numerical complexity as the evaluation of $\mathcal{L}_{I_{x_R,a}}(s \mid x^*)$. Moreover, a closed-form upper bound for $\lambda_{x_R,u}(r)$ is available, which can in turn be used to obtain a looser lower bound for $\mathcal{L}_{I_{x_R,u}}(s \mid x^*)$ that is independent of $\angle x_R$. The bound is tight in the sense that it corresponds to the exact equivalent density when $x^* = o$.

Proposition 4.6. *The equivalent density $\lambda_{x_R,u}(r)$ of Prop. 4.5 is upper bounded as*

$$\lambda_{x_R,u}(r) \leq \lambda_u \left[1 - e^{-\lambda_a \pi (|x^*|^2 + |x_R|^2 + r^2)} \mathcal{I}_0(\lambda_a 2\pi \bar{x} r)\right], \tag{4.13}$$

where $\bar{x} \triangleq \max(|x^*|, |x_R|)$ *and* $\mathcal{I}_0(\cdot)$ *denotes the zero-order modified Bessel function of the first kind. Equality holds only when* $x^* = o$.

Proof. See Appendix sec. C.3. □

An immediate observation of the above result is that, given x_R, the equivalent interfering UE density decreases for any $x^* \neq o$ compared to the case $x^* = o$. This behavior is expected due to the cell expansion effect described before. However, as the bound of (4.13) is independent on $\angle x_R$, it does not offer clear information on how the interferering UE density changes when different values of $x^* \neq o$ are considered for a given x_R. In order to obtain futher insights on the UE-generated interference properties, $\lambda_{x_R,u}(r)$ is investigated in detail in the following section for certain special instances of x_R and/or x^*, which are of particular interest in cellular networks.

4.5. Analysis of Special Cases of UE-Generated Interference

4.5.1. $x_R = x^*$

This case corresponds to an uplink cellular transmission with nearest AP association, no intra-cell interference, and interference generated by UEs outside $\mathcal{V}^*$ operating in uplink and/or D2D mode. For this case, a tighter upper bound than the one in Prop. 4.6 is available for $\lambda_{x^*,u}(r)$, which, interestingly, is independent of $|x^*|$, in the practical case when $x^* \neq o$.

Lemma 4.7. *For* $x_R = x^*$*, the equivalent density* $\lambda_{x_R,u}(r) = \lambda_{x^*,u}(r)$ *of Prop. 4.5 is upper bounded as*

$$\lambda_{x^*,u}(r) \leq \lambda_u \left(1 - e^{-\lambda_a \pi r^2}\right), \tag{4.14}$$

with equality if and only if $x^* = o$.

Proof. Follows by replacing the term $p_c\left((r,\theta) + x_R \mid x^*\right) = p_c\left((r,\theta) + x^* \mid x^*\right)$ appearing in (4.11) with its bound given in Cor. 4.4, which evaluates to $e^{-\lambda_a \pi r^2}$. □

The bound of (4.14) indicates that $\lambda_{x^*,u}(r)$ tends to zero at least as fast as $\mathcal{O}(r^2)$ for $r \to 0$, *irrespective of* x^*. The following exact statement on the asymptotic behavior of $\lambda_{x^*,u}(r)$ shows that $\lambda_{x^*,u}(r) \sim cr^2, r \to 0$, with the value of c independent of $|x^*|$ when $x^* \neq o$.

Proposition 4.8. *For $x_R = x^*$, the equivalent density $\lambda_{x_R,u}(r) = \lambda_{x^*,u}(r)$ of Prop. 4.5 equals*

$$\lambda_{x^*,u}(r) = \lambda_u \lambda_a b \pi r^2 + \mathcal{O}(r^3), r \to 0, \tag{4.15}$$

with $b = 1$ for $x^* = o$ and $b = 1/2$ for $x^* \neq o$.

Proof. For $x^* = o$, the result follows directly from Lemma 4.7. For $x^* \neq o$, it can be shown by algebraic manipulation based on Lemma 4.3, that the term $p_c\left((r,\theta) + x_R \mid x^*\right) = p_c\left((r,\theta) + x^* \mid x^*\right)$ appearing in (4.11) equals

$$\lambda_u \lambda_a \left(q\pi + (-1)^q \arccos(\sin(\theta)) + \cos(\theta)\sin(\theta)\right) r^2 + \mathcal{O}(r^3),$$

for $r \to 0$, where $q = 0$ for $\theta \in [-\pi, -\pi/2] \cup [\pi/2, \pi)$ and $q = 1$ for $\theta \in (-\pi/2, \pi/2)$. Substituting this expression in (4.11) and performing the integration leads to the result. □

Equation (4.15) indicates that imposing a guard region $\mathcal{V}^*$ when $x^* \neq o$ reduces the interfering UE density by half in the close vicinity of x^*, compared to the case $x^* = o$. This effect is similar to the behavior of $\lambda_{x^*,a}(r)$ discussed in Sec. III. A. However, $\lambda_{x^*,a}(0) = \lambda_a/2$, whereas $\lambda_{x^*,u}(0) = 0$, clearly demonstrating the effectiveness of the guard region for reducing UE-generated interference in the vicinity of x^*.

For non-asymptotic values of r, the behavior of $\lambda_{x^*,u}(r)$ can only be examined by numerical evaluation of (4.11). Fig. 4.6 depicts the normalized equivalent density $\lambda_{x^*,u}(r)/\lambda_u$ for various values of $|x^*|$ in the range of practical interest $[0, 3/(2\sqrt{\lambda_a})]$ (note that $\mathbb{P}(|x^*| > 3/(2\sqrt{\lambda_a})) = e^{-\lambda_a \pi \left(3/(2\sqrt{\lambda_a})\right)^2} \approx 8.5 \times 10^{-4}$). It can be seen that for non-asymptotic values of r, $\lambda_{x^*,u}(r)$ is a decreasing function $|x^*|$ for all $r > 0$, corresponding to a reduced UE-generated interference. This is expected since, as evident from Fig. 4.4, the cell expansion effect with increasing $|x^*|$ can only improve the interference protection at x^*.

Interestingly, the dependence of $\lambda_{x^*,u}(r)$ on $|x^*|$, although existing, can be seen to be rather small. This observation along with Prop. 4.8 strongly motivates the consideration of a single curve as an approximation to the family of curves in Fig. 4.6. A natural approach is to consider the upper bound of (4.14), which has the benefit of being available in closed form. This is essentially the approach that was heuristically employed in [63] without observing that it actually corresponds to a (tight) bound for the equivalent interferer density.

Another approach is to consider a piecewise-linear curve, which, in addition to having a simple closed form, results in a closed form approximate expression for the bound of $\mathcal{L}_{I_{x^*,u}}(s \mid x^*)$ as given below (details are straightforward and omitted).

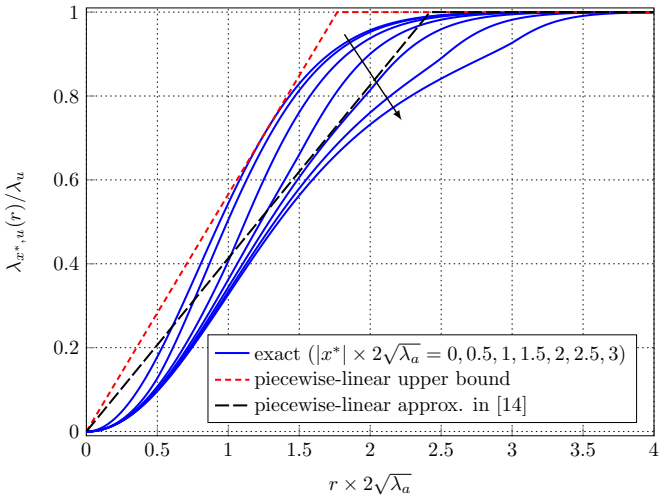


Figure 4.6.: Equivalent density function $\lambda_{x^*,u}(r)$ for various $|x^*|$ (solid curves), with the direction of the arrow corresponding to increasing $|x^*|$. A piecewise-linear approximation and bound are also shown.

Lemma 4.9. *With the piecewise-linear approximation*

$$\lambda_{x^*,u}(r) \approx \lambda_u \min\left(\delta r, 1\right), r \geq 0, \tag{4.16}$$

for some $\delta > 0$, the lower bound for $\mathcal{L}_{I_{x^*,u}}(s \mid x^*)$ equals

$$\exp\left\{-\lambda_u \pi \left[C(P_u s)^{2/\alpha} - \frac{1}{\delta^2}\left(\frac{2}{3}\tilde{F}\left(3\right) - \tilde{F}\left(2\right)\right)\right]\right\}$$

where $C \triangleq (2\pi/\alpha)/\sin(2\pi/\alpha)$ and $\tilde{F}(x) \triangleq {}_2F_1\left(1, \frac{x}{\alpha}, 1+\frac{x}{\alpha}, \frac{-1}{sP_u\delta^\alpha}\right)$ with ${}_2F_1\left(\cdot\right)$ denoting the hypergeometric function.

A piecewise-linear approximation as in (4.16) was heuristically proposed in [62] for a slightly more general system model than the one considered here. A closed form formula for δ was provided, only justified as obtained by means of curve fitting without providing any details on the procedure. For the system model considered in this chapter, this formula results in $\delta \approx 0.82687\sqrt{\lambda_a}$ and the corresponding piecewise-linear approximation of the equivalent density is also depicted in Fig. 4.6. It can be seen that this approximation is essentially an attempt to capture the average behavior of $\lambda_{x^*,u}(r)$ over $|x^*|$, thus explaining its good performance reported in [62]. However, it is not clear why this particular value of δ is more preferable than any other (slightly) different value. A more rigorous approach for the selection of δ is to consider the tightest piecewise-linear upper bound for $\lambda_{x^*,u}(r)$, which leads to another (looser) lower bound for $\mathcal{L}_{I_{x_R,u}}(s \mid x^*)$. This bound corresponds to a value of $\delta \approx 1.13118\sqrt{\lambda_a}$, found by numerical optimization, and is also shown in Fig. 4.6.

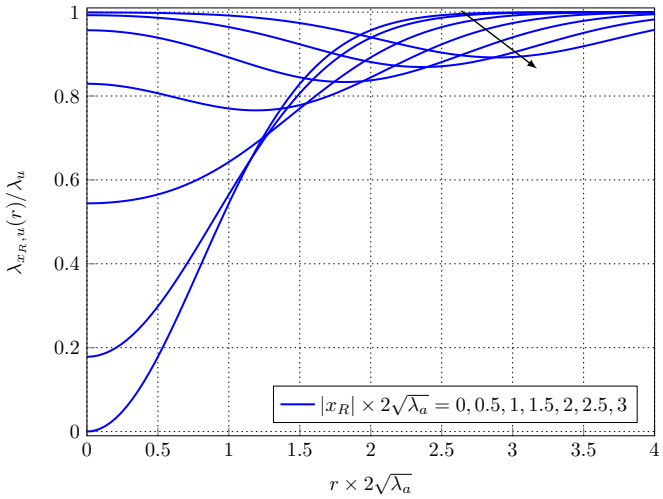


Figure 4.7.: Equivalent density function $\lambda_{x_R,u}(r)$ when $x^* = o$ for various $|x_R|$.

4.5.2. $x^* = o$

This case corresponds to the typical node being an AP, i.e., a typical AP. Note that, in contrast to the cases when $x^* \neq o$, the Voronoi cell of the typical AP is *not* conditioned to cover any point in $\mathbb{R}^2$, therefore, no cell expansion effect is observed and, for $x_R \neq o$, it is possible that the receiver lies outside the guard region. This is a different operational scenario from the common downlink with nearest AP association [29] where the intended receiver by default lies within the Voronoi cell of the serving AP. It may arise, for example, in case of D2D communications, where both communicating UEs receive data (e.g., control information) from the same AP.

The equivalent interfering UE density is given in Prop. 4.6 and is depicted in Fig. 4.7 for various values of $|x_R|$. Note that the case $x_R = o$ corresponds to the typical AP in receive mode, whereas for $x_R \neq o$ the typical AP transmits. It can be seen that increasing $|x_R|$ effectively results in reduced protection in the close vicinity of the receiver since the latter is more probable to lie near the edge or even outside the Voronoi guard region.

4.5.3. $x_R = o$

This case corresponds to the typical node being at receive mode. One interesting scenario which corresponds to this case is downlink cellular transmission with nearest AP association, where interference is due to transmitting UEs (e.g., in uplink or D2D mode) that lie outside the serving cell.

When $x^* = o$, the setup matches the case $x_R = x^*$ with $x^* = o$, discussed above, and it holds $\lambda_{o,u}(r|x^*) = \lambda_u \left(1 - e^{-\lambda_a \pi r^2}\right)$. For $x^* \neq o$, the asymptotic behavior of $\lambda_{o,u}(r|x^*)$ can be obtained by the same approach as in the proof of Prop. 4.8.

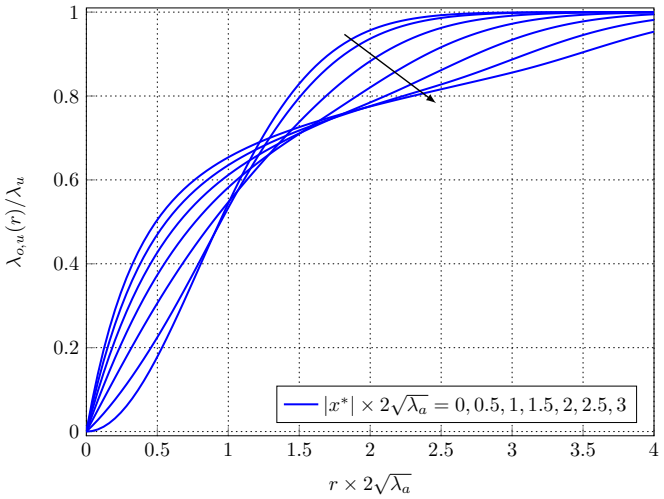


Figure 4.8.: Equivalent density function $\lambda_{o,u}(r)$ for various $|x^*|$, with the direction of the arrow corresponding to increasing $|x^*|$.

Proposition 4.10. *For $x^* \neq o$ the equivalent density function $\lambda_{x_R,u}(r) = \lambda_{o,u}(r|x^*)$ of Prop. 4.5 equals*

$$\lambda_{o,u}(r) = \lambda_u \lambda_a \frac{8}{\pi} |x^*| r + \mathcal{O}(r^2), r \to 0.$$

In stark contrast to the behavior of $\lambda_{x^*,u}(r)$, $\lambda_{o,u}(r)$ tends to 0 only linearly (instead of quadradically) as $r \to 0$ when $x^* \neq o$, and with a rate that is also proportional to (instead of independent of) $|x^*|$. This implies that the typical node is not as well protected from interfering UEs as its nearest AP, experiencing an increased equivalent interferer density in its close vicinity as $|x^*|$ increases. This can be understood by noting that with increasing $|x^*|$, the origin, although contained in a guard region of (statistically) increasing area, is more probable to lie near the cell edge (Fig. 4.1 demonstrates such a case.)

The strong dependence of $\lambda_{o,u}(r)$ on $|x^*|$ observed for $r \to 0$ also holds for all values of r. This is shown in Fig. 4.8 where the numerically evaluated $\lambda_{o,u}(r)$ is depicted for the same values of $|x^*|$ as those considered in Fig. 4.6. It can be seen that $\lambda_{o,u}(r)$ increases with $|x^*|$ for all r up to approximately $1 - 1.5$ times the average value of $|x^*|$ (equal to $1/(2\sqrt{\lambda_a})$), whereas it decreases with $|x^*|$ for greater r .

This behavior of $\lambda_{o,u}(r)$, especially for $r \to 0$, does not permit the use of single function to be used a reasonable approximation or bound for $\lambda_{o,u}(r)$, irrespective of $|x^*|$, as was done for the case for $\lambda_{x^*,u}(r)$. Noting that the upper bound of (4.13) has the same value whenever $|x^*|$ or $|x_R|$ are zero, it follows that the curves shown in Fig. 4.7 with $|x_R|$ replaced by $|x^*|$ are upper bounds for the corresponding curves of Fig. 4.8. Unfortunately, it can be seen that the bound is very loose for small r and/or large $|x^*|$.

4.5.4. $|x_R| \in (0, |x^*|), \angle x_R = 0$

The previous cases demonstrate how different the properties of the UE-generated interference become when different receiver positions are considered. This observation motivates the question of which receiver position is best protected from UE interfernce given the positions o and $x^* \neq o$ of the typical node and nearest AP, respectively. This question is relevant in, e.g., relay-aided cellular networks where the communication between a UE and its nearest AP is aided by another node (e.g., an inactive UE) [73]. Although the performance of a relay-aided communication depends on multiple factors including the distances among the nodes involved and considered transmission scheme [74], a reasonable choice for the realy position is the one experiencing less interference.

By symmetry of the system geometry, this position should lie on the line segment joining the origin with x^* (inclusive). However, as can be seen from Fig. 4.6 and Fig. 4.8, the shape of the equivalent interferer density does not allow for a natural ordering of the different values of x_R. For example, $\lambda_{x^*,u}(r)$ is smaller than $\lambda_{o,u}(r)$ for small r, whereas the converse holds for large r. Noting that the receiver performance is mostly affected by nearby interference [45], a natural criterion to order the interfering UE densities for varying x_R is their asymptotic behavior as $r \to 0$. For $|x_R| = |x^*|$ and $|x_R| = o$, this is given in Props. 4.8 and 4.10, respectively. For all other positions the asymptotic density is given by the following result.

Proposition 4.11. *For $|x_R| \in (0, |x^*|), \angle x_R = 0$, and $x^* \neq o$, the equivalent density function $\lambda_{x_R,u}(r)$ of Prop. 4.5 equals*

$$\lambda_{x_R,u}(r) = \lambda_u \lambda_a \frac{8}{9\pi} \frac{(|x^*|/|x_R|)^2}{|x^*| - |x_R|} r^3 + \mathcal{O}(r^4), r \to 0. \tag{4.17}$$

It follows that the optimal receiver position must have $|x_R| \in (0, |x^*|)$ since, in this case, the density scales as $\mathcal{O}(r^3)$ instead of $\mathcal{O}(r)$ and $\mathcal{O}(r^2)$ when $|x_R| = 0$ and $|x_R| = |x^*|$, respectively. Its value can be easily obtained by minimization of the expression in (4.17) w.r.t. $|x_R|$.

Corollary 4.12. *The receiver position that experiences the smallest equivalent UE interferer density in its close vicinity lies on the line segment joining the origin to x^* and is at distance $|x^*|/\sqrt{2}$ from the origin. The corresponding equivalent density equals $\lambda_{x_R,u}(r) = \lambda_u \lambda_a \frac{32}{9\pi} r^3 + \mathcal{O}(r^4), r \to 0$.*

4.6. Numerical Examples

In this section numerical examples will be given demonstrating the performance of various operational scenarios in cellular networks that can be modeled as special

cases of the analysis presented in the previous sections. The system performance metric that will be considered is $\mathcal{L}_{I_{x_R,k}}(\rho^{\alpha_k}\theta|x^*)$, with $k \in \{a,u\}$ depending on whether AP or UE interference is considered, and $\rho > 0$, $\theta > 0$ given variables. Note that this metric corresponds to the *coverage probability* that the signal-to-interference ratio (SIR) at x_R is greater than θ, when the distance between transmitter and receiver is ρ, the direct link experiences a path loss a_k and Rayleigh fading, and the transmit power is equal to P_k [45]. For simplicity and w.l.o.g., all the following examples correspond to $\lambda_a = 1$ and $\alpha_a = \alpha_u = 4$.

4.6.1. D2D Link Not Necessarily Contained in a Single Cell

In this example, UE-generated inteference is considered at a receiver position $x_R \neq o$ with $\angle x_R$ uniformly distributed in $[-\pi, \pi)$ and $\rho = |x_R|$. This case models a D2D link where the typical node is a UE that directly transmits to a receiver isotropically distributed at a distance $|x_R| > 0$. A guard region $\mathcal{V}^*$ is established by the closest AP to the typical node, however, it is not guaranteed to include x_R, i.e., the D2D nodes may not be contained within the same cell, which is a common scenario that arises in practice, referred to as cross-cell D2D communication [69].

With interference generated from other UEs in D2D and/or uplink mode, a lower bound for $\mathcal{L}_{I_{x_R,u}}(\rho^{\alpha_u}\theta|x^*)$ can be obtained using Prop. 4.5 with an equivalent density function $\lambda_{|x_R|,u}(r)$ as given in (4.12). Fig. 4.9 shows this lower bound for $\lambda_u = \lambda_a$ (4.9a) and $\lambda_u = 10\lambda_a$ (4.9b). The position x^* is assumed to be isotropically distributed with $|x^*| = 1/(2\sqrt{\lambda_a})$ and various values of the ratio $|x_R|/|x^*|$ are considered. For each case, the exact value of $\mathcal{L}_{I_{x_R,u}}(\rho^{\alpha_u}\theta|x^*)$, obtained via Monte Carlo simulation, is also shown. It can be seen that the quality of the analytical lower bound depends on λ_u. For $\lambda_u = \lambda_a$, it is very close to the exact coverage probability, whereas for $\lambda_u = 10\lambda_a$, it is reasonably tight and is able to capture the behavior of the exact coverage probability over varying $|x_R|$. In both cases, a performance degradation is observed with increasing $|x_R|$, due to both increasing path loss of the direct link as well as reduced interference protection by $\mathcal{V}^*$.

4.6.2. AP-to-D2D-Receiver Link

In this example, AP-generated inteference is considered at a receiver position $x_R \neq o$ with $\angle x_R$ uniformly distributed in $[-\pi, \pi)$ and $\rho = |x^* - x_R|$. This case is similar to the previous, however, considering AP interference and the receiver at x_R receiving data from the AP node at x^*. This case models the scenario where the typical node establishes a D2D connection with a node at x_R, and the AP at x^* sends data to x_R via a dedicated AP link (e.g., for control purposes or for implementing a cooperative transmission scheme). Note that, in contrast to the previous case, the link distance ρ is random and equal to $\rho = |x^*| + |x_R| - 2|x^*||x_R|\cos\psi$, with ψ uniformly distributed

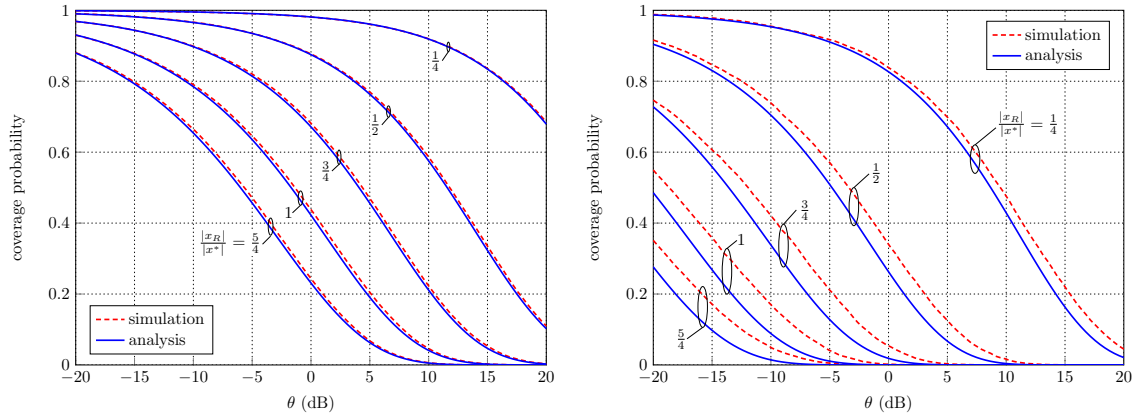


Figure 4.9.: Coverage probability of the link between the typical node and an isotropically positioned receiver under UE interference for (a) $\lambda_u = \lambda_a$, (b) $\lambda_u = 10\lambda_a$.

in $[-\pi, \pi)$. The exact coverage probability of this link can be obtained using Prop. 4.2 followed by a numerically computed expectation over ψ.

Fig. 4.10 shows $\mathbb{E}(\mathcal{L}_{I_{x_R,a}}(\rho^{\alpha_a}\theta|x^*))$ for an isotropically distributed x^* with $|x^*| = 1/(2\sqrt{\lambda_a})$, and for various values of the ratio $|x_R|/|x^*|$. Note that the case $|x_R| = 0$ corresponds to the standard downlink transmission model with nearest AP association [29]. Monte Carlo evaluation of the coverage probability (not shown here) perfectly matches the analytical curves. It can be seen that increasing $|x_R|$ reduces the coverage probability for small SIR but increases it for high SIR. This is in direct proportion to the behavior of the equivalent interferer density $\lambda_{x_R,a}(r)$ with increasing $|x_R|$ shown in Fig. 4.2.

4.6.3. Uplink with Nearest AP Association

In this example, AP-generated interference is considered at a receiver position $x_R = x^*$ with $\rho = |x^*|$ and $\lambda_u = \lambda_a$. This case can be viewed as an approximation of the conventional uplink transmission scenario with nearest AP association and one active UE per cell. Note that the actual intereferer process in the uplink scenario is distributed as a VPLP [64, 65], which is difficult to analytically characterize. Fig. 4.11 shows the lower bound of $\mathcal{L}_{I_{x_R,u}}(\rho^{\alpha_u}\theta|x^*)$ obtained by Prop. 4.5 as well as the looser, but more easily computable, lower bound obtained using the closed form expression given in Lemma 4.7 with $\delta = 1.13118\sqrt{\lambda_a}$ (resulting in the tightest bound possible with a piecewise-linear equivalent density function). The coverage probability is computed for an AP at a distance $|x^*|{=}c/(2\sqrt{\lambda_a})$ from the typical node, with $c = 1/2, 1, 2$, roughly corresponding to a small, average, and large uplink distance (performance is independent of $\angle x^*$). In addition, the coverage probability

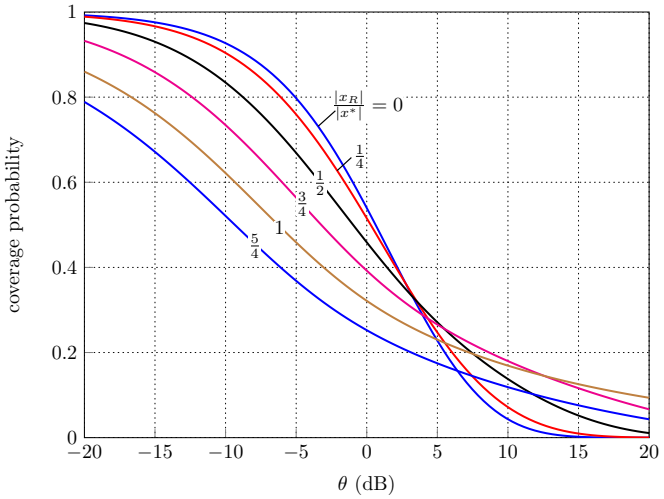


Figure 4.10.: Coverage probability for the link between the AP positioned at x^* ($|x^*| = 1/(2\sqrt{\lambda_a})$) and an isotropicaly positioned receiver.

obtained by Monte Carlo simulation under both the PPP and VPLP interferer positions models are also shown.

As was the case in 4.9a, Prop. 4.5 provides a very tight lower bound for the actual coverage probability (under the PPP model for the interfering UE positions). The bound of Lemma 4.7, although looser, is nevertheless a reasonable approximation of the performance, especially for smaller values of $|x^*|$. Compared to the VPLP model, the PPP model provides an optimistic performance prediction that is, however, reasonable tight, especially for large $|x^*|$. This observation motivates its usage as a tractable approximation of the actual inteference statistics as was also reported in [62, 63]. Interestingly, the bound of Lemma 4.7 happens to provide an even better approximation for the VPLP performance for $|x^*|$ close to or smaller than $1/(2\sqrt{\lambda_a})$.

4.6.4. Effect of Guard Region on UE-Generated Interference Protection

In order to see the effectiveness of a Voronoi guard region for enhancing link quality under UE-generated interference, the performance under the following operational cases is examined.

- Case A (no guard region is imposed): This results in the standard transmission model under an HPPP of interferers positions of density λ_u [45]. The exact coverage probability is well known (see, e.g., [45, Eq. 3.29]).
- Case B (transmitter imposes a Voronoi guard region): This case corresponds to $x^* = o$, $x_R \neq o$.

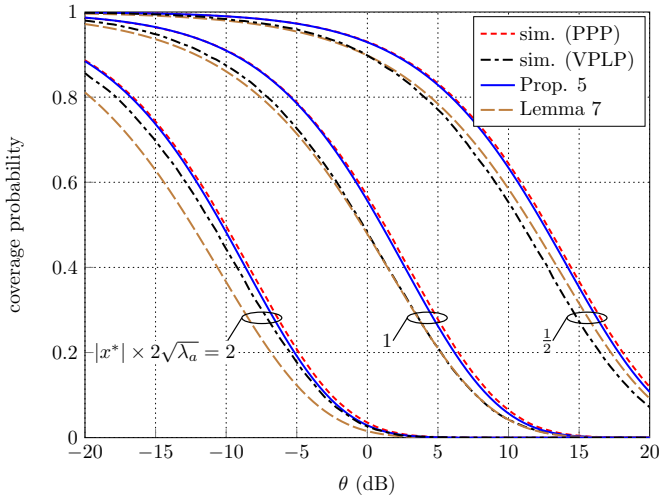


Figure 4.11.: Coverage probability of the uplink between the typical node its nearest AP ($\lambda_u = \lambda_a$).

- Case C (receiver imposes a Voronoi guard region): This case corresponds to $x_R = x^* = o$.

Cases B and C correspond to the analysis considered in Sec. V. B. Assuming the same link distance ρ for all cases, Fig. 4.12 shows the analytically obtained coverage probability (exact for Case A, lower bound for Cases B and C), assuming $\lambda_u = \lambda_a$. It can be seen that imposing a Voronoi guard region (Cases B and C) is always beneficial, as expected. However, for large link distances, Case B provides only marginal gain as the receiver is very likely to be located at the edge or even outside the guard region. In contrast, the receiver is always guaranteed to be protected under case C, resulting in the best performance and significant gains for large link distances.

4.7. Conclusion

This chapter considered the analytical characterization of cross-mode inter-cell interference experienced in future cellular networks. By employing a stochastic geometry framework, tractable expressions for the interference statistics were obtained that are exact in the case of AP-generated interference and serve as a tight lower bound in the case of UE-generated interference. These expressions are based on appropriately defined equivalent interferer densities, which provide an intuitive quantity for obtaining insights on how the interference properties change according to the type of interference and receiver position. The considered system model and analysis are general enough to capture many operational scenarios of cellular networks, including conventional downlink/uplink transmissions with nearest AP association as well as D2D transmissions between UEs that do not necessarily lie in the same cell. The

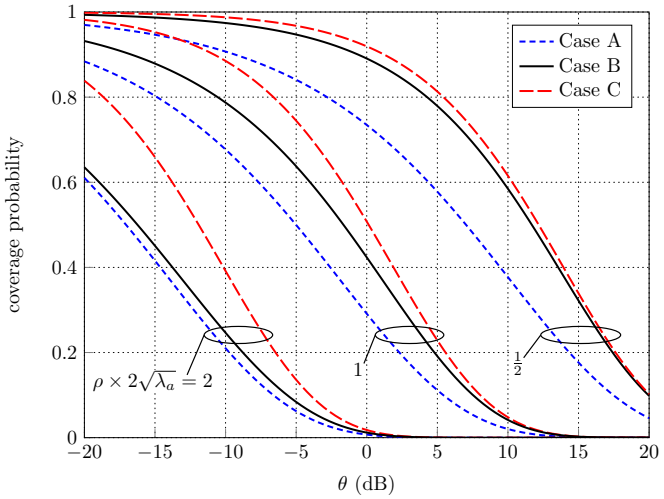


Figure 4.12.: Coverage probability under UE interference for the links corresponding to the cases described in Sec. V. D.

analytical expressions of this chapter can be used for sophisticated design of critical system aspects such as mode selection and resource allocation towards getting the most out D2D- and/or FD-enabled cellular communications.

5. On the Competition Among Small Cell Wireless Operators with Large Scale Deployments

5.1. Overview

Recent trends in wireless networking such as small cells and network function virtualization, strongly suggest that the future cellular network will consist of many sub-networks controlled by self-interest operators competing for a share of the revenues associated with the high demand for service. This chapter attempts to analyze the future network performance under these new market conditions, taking into account not only engineering properties of the sub-networks, e.g., spectral efficiency, but economic properties as well, i.e., service pricing. Going beyond previous related works, large scale infrastructure deployments over wide geographical areas, targeting a large number of users, are explicitly considered. The focus of the analysis is on the effects of different deployment densities (in units of average number of small cells per unit area) among competing operators as well as random placement of small cells. A game-theoretic system analysis is employed with market equilibrium characterization obtained analytically for the practical cases of two competing operators with arbitrary network characteristics and multiple competing operators with similar network characteristics. Closed form expressions for user and operator utilities at equilibrium are provided, revealing, among others, that extreme infrastructure densification, even though desirable from a purely engineering perspective, may not lead to stable market conditions under the considered competition model.

The results of this chapter were published in

1. S. Stefanatos and A. Alexiou, "On the Competition Among Small Cell Wireless Operators with Large Scale Deployments," *IEEE Transactions on Mobile Computing*, 2017, to appear.

2. S. Stefanatos and A. Alexiou, "Duopoly Competition between Small Cell Operators with Large Scale Deployments," *IEEE Signal Proces. Adv. In Wireless Commun. (SPAWC)*, UK, Jul. 2016.

5.2. Introduction

The exponential increase of wireless traffic observed in recent years, as well as market predictions suggesting that this trend will be sustained in the future, challenge the traditional way of advancing cellular networks based primarily on improvements in transmission technology. It is now widely accepted that the success of the future cellular network will rely heavily (if not mostly) on the use of massive infrastructure deployment (small cell networks) [51, 75] and the co-existence of multiple heterogeneous networks [22]. Research on these two new system attributes has been extensive, focusing on the implications their introduction to the system has with respect to (w.r.t.) engineering metrics, such as capacity, coverage, and energy efficiency, which are critical indicators of the network capabilities.

However, apart from technological implications, dominance of small cells will also introduce changes in the way the future cellular network is operated, rendering the legacy operational model, where infrastructure and spectrum are in the control of a few major operators (OPs), irrelevant [76]. In particular, the low deployment/operational costs of small cells, the recent emergence of the concept of virtualization in wireless networking [77], and the increase of wireless devices/applications strongly suggest that the future network will be a market place where many players will attempt to grab a share of the revenues associated with the high demand for service. Clearly, in this highly competitive market environment, the level of service experienced by the users depends not only on the engineering capabilities of the competing networks but also on their economy-driven attributes such as service pricing. It is therefore important to investigate the characteristics of the resulting wireless market and how these are reflected upon metrics of interest such as user level of service and revenues of OPs.

5.2.1. Related Work

The study of economic aspects of communication networks has reached a high level of maturity [78], however, with the focus on wireline networks. Since wireless networks have important special characteristics such as the broadcast nature of the wireless medium and user mobility, an economics oriented study of wireless networks requires approaches that explicitly take them into account [79]. This has resulted in a number of recent works that apply game theory – the natural tool to model interactions among competitive entities (players) [80] – to a wireless networks setting. In [81], the scenario where two OPs compete (duopoly) for attracting a finite set of users via bandwidth and pricing decisions is investigated. It is shown that the existence and uniqueness of a market equilibrium depends on parameters such as difference between the bandwidth leasing prices the two OPs pay to acquire bandwidth. A similar setup is considered in [82] where the undesirable *price war* phenomenon between two OPs is investigated and appropriate regulation measures are suggested

to eliminate it. Pricing competition among multiple (more than two) OPs with the presence of a regulator-controlled *neutral* OP is considered in [83]. Regulator actions for steering the competing OP strategies towards providing improved user services with smaller costs are proposed. The co-existence of a primary and a secondary OP, where the latter can lease part of the spectrum of the former, is considered in [84]. In contrast to the previous references, an infinite number of users is assumed, which requires appropriate equilibrium notions for system analysis. An infinite number of users is also considered in [85, 86] for a finite, but otherwise arbitrary, number of competing OPs.

A notable common aspect of the above mentioned studies is that they, either explicitly or implicitly, consider a small geographical area that each competing OP can cover with a single access point (AP). Therefore, generalization of results for the case of large scale deployments over wide geographical areas essentially requires the assumption that all OPs deploy identical networks w.r.t. number and position of APs. Clearly, this is not a realistic assumption, even more so in the case of small cell networks where APs of a single OP are typically introduced in the system in an unplanned and opportunistic manner, resulting in an essentially random deployment [60]. It is therefore necessary to explicitly take into account these topological properties of the networks in order to obtain insights on their effects on system performance. Interestingly, the random AP placement allows for tractable, analytic evaluation of the engineering (information theoretic) capabilities of a (single OP) network [29] using tools from stochastic geometry [18]. However, even though the stochastic geometry framework has been widely employed in recent works on cellular networks performance [21], very few of them investigate the economic impact of a random infrastructure deployment, and they only consider the case of a single OP (monopoly) [87, 23].

5.2.2. Contributions

By combining tools from stochastic geometry and game theory, this work proposes a simplified yet rigorous mathematical model for the interactions among self-interest OPs and users, in an effort to obtain insights on how these affect system performance in terms of appropriate utilities for all the entities involved. The main contributions of this work are the following.

- *Game formulation for OP deployments over a wide geographical area.* A large (mathematically modeled as infinite) geographical area is considered where OPs deploy their infrastructure. An appropriate system model and game formulation is proposed that explicitly considers the effects of random placement of (infinite) APs and users, parameterized by the competing networks attributes, namely, deployment densities, bandwidth leasing prices, and transmission technologies (spectral efficiencies).

- *Evaluation of average user rate.* As the commonly employed, single cell approach for obtaining the (average) user rate, namely, dividing the sum rate provided by an AP by the finite number of co-served users, is not appropriate in the considered setting, a new closed-form approximate expression for the average user rate is derived using tools from stochastic geometry. This expression allows for defining tractable user utility functions that can be used for system analysis.
- *Analytic characterization of market equilibria.* For certain special cases of practical interest, characterization of market equilibrium is obtained by analytical means. This allows to determine how the competing network parameters affect the existence and uniqueness of a market equilibrium as well as the corresponding user and OP utilities. Special emphasis is given on the investigation of the role of infrastructure deployment densities, which is an issue that has been neglected in the literature.
- *Network parameters optimization towards improved services.* The analytic expressions of users' utilities allow for determining the optimal competing network parameters w.r.t. the quality of the user perceived service. This analysis is of importance to a regulator attempting to steer market parameters towards maximizing the social (users') welfare by providing appropriate incentives and/or imposing appropriate constraints to the competing OPs.

5.3. System Model

5.3.1. Network Characteristics

A finite number of $N \geq 2$ wireless system OPs, arbitrarily ordered by an index $n \in \{1, 2, \ldots, N\}$, is considered. With the goal of achieving maximum profit, each OP attempts to attract (part of) the user equipments (UEs) over a large geographical area of interest by deploying an exclusively owned infrastructure of APs. The positions of the APs of OP n are modeled as a realization of an independent from everything else, homogeneous Poisson point process (HPPP) $\Phi_n \subset \mathbb{R}^2$ of density $\lambda_n > 0$ (in units of average number of APs per unit area). This stochastic model has been shown to provide an accurate representation of real AP deployments [88, 29] and has been widely adopted in the literature [21]. Note that an infinite deployment area is considered for mathematical convenience with the corresponding analysis holding approximately true for the case of a finite deployment area that is sufficiently large so that edge effects are negligible.

It is also natural to model the positions of the UEs as another independent of everything else HPPP $\Phi_u \subset \mathbb{R}^2$ of density $\lambda_u > 0$ (in units of average number of UEs per unit area) [23]. For simplicity and without loss of generality (w.l.o.g.), a normalized unit area such that $\lambda_u = 1$ will be considered in the following, which

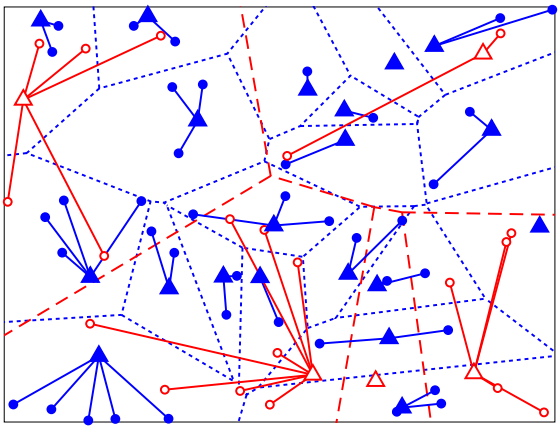


Figure 5.1.: Distribution of APs and subscribers association for the case of $N = 2$ OPs in a random subset of $\mathbb{R}^2$. Filled (open) triangles and filled (open) circles denote positions of APs and subscribers of OP 1 (OP 2), respectively. Solid lines connect pairs of UEs and serving APs, and dotted (dashed) lines represent the cell boundaries for the APs of OP 1 (OP 2).

results in λ_n having the interpretation of ratio of average number of APs of OP n to average number of UEs in any bounded subset of $\mathbb{R}^2$.

Any UE subscribed to OP n is associated with the AP of OP n that is closest to its position (see Fig. 5.1). It is assumed that the AP density of each OP is sufficiently large so that there are no coverage issues, i.e., all UEs are able to attach to their subscribed networks irrespective of their actual positions. This assumption is appropriate for (large scale) small cell network deployment, rendering the achievable user rate as the fundamental engineering metric of system performance [22]. The achievable UE rate is clearly affected by the multiple-access scheme employed at each AP. A simple time-division multiple-access (TDMA) scheme without priorities (round-robin) is assumed in this work for all APs.[1]

Apart from the exclusively owned infrastructure, each OP utilizes a dedicated portion of the spectrum for serving subscribers (no spectrum sharing is considered among OPs). This dedicated bandwidth is made available by spectrum provider(s), e.g., a regulatory agency, who own spectrum and provide it by leasing agreements with the OPs involving a fee (price) per requested unit bandwidth per unit area.

The final important characteristic of the network of OP n is the employed transmission technology, which is reflected by the spectral efficiency (in bits/second/Hz) of the network transmissions. Even though the spectral efficiency of a transmission link depends on factors such as link distance and channel realization, in this study, a constant value, $c_n > 0$, will be considered for the spectral efficiency of the network of OP n, i.e., with the effects of distance and channel averaged. This is also

[1] Any other orthogonal multiple-access scheme, e.g., in frequency or code domain, could be used instead, with no effect on the analysis and results.

the approach considered, either explicitly or implicitly, in most related studies, e.g., [82, 83, 84, 86]. Note that c_n is also considered to be *independent* of λ_n, which was actually shown in [29] to hold for the system model considered in this study, under the common assumptions of intereference-limited operation and Rayleigh fading.

5.3.2. Pricing Schemes

OPs achieve revenues by charging their subscribers with a fee for employing their network. A pricing scheme based on the *peak* rate that a subscriber (potentially) experiences will be employed in this work. The peak rate of a UE subscribed to OP n corresponds to ideal operating conditions where the UE is the only one served by its associated AP, and is equal to $c_n w$ (in bits/second) [83, 84], where $w > 0$ (in Hz) is the utilized transmission bandwidth (possibly different among subscribers of the same OP).

Even though, under small to moderate infrastructure deployment densities, a UE will typically have to share the serving AP resources, thus experiencing reduced rates, use of "flat-rate" measures such as peak rate are commonly employed in practice for pricing, as they allow for straightforward and easily implementable pricing schemes [89] (compared to, say, a scheme based on the exact achieved rate of a UE). In particular, each OP is considered to employ a differentiated services pricing scheme [78], offering subscribers a multitude of possible peak rates to be served with, each priced proportionally to its value. For analytical simplicity, infinite granularity of available peak rates will be assumed, i.e, UEs may choose any real-valued (positive) peak rate that optimally balances cost with performance according to their needs. Note that a peak rate request can be translated to an equivalent bandwidth request given the OP's spectral efficiency. As it will turn out, the subscribers of a certain OP are only interested in the same peak rate value, rendering the infinite granularity assumption mild, with the corresponding peak rate value of interest being an approximation of the closest peak rate value when a finite number of choices are available.

5.4. Game Formulation

5.4.1. Players and Strategies

Under the assumption that all entities (players) involved in the above setting, namely, OPs and UEs, are non-cooperating, self-interested, and rational, it is natural to employ a game theoretic approach in order to investigate the attributes of the system. Note that, in contrast to the standard game-theoretic modeling assumption of a finite number of players [80], an infinite number of players is implied by the

system model of the previous section, which is more appropriate for the scenario of large scale network deployments targeting a very large number of UEs.

Similar to previous works [81, 84, 86], the game is modeled as taking place in two consecutive stages:

1. Initially, the OPs select their strategy (action) from the set

$$\mathcal{S}_{\mathrm{OP}} \triangleq \left\{(w_{\mathrm{OP}}, p_{\mathrm{UE}}) \in \mathbb{R}^+ \times \mathbb{R}^+\right\},$$

   where w_{OP} represents the amount of bandwidth requested by an OP to the spectrum provider in order to serve all (expected) peak rate requests from subscribers,[2] and p_{UE} is the price per unit of requested peak rate that the OP will charge the subscribers with.

2. After the OPs announce their prices, each UE selects a strategy (action) from the set $\mathcal{S}_{\mathrm{UE}} \triangleq \{(n, w_{\mathrm{UE}}) \in \{1, 2, \ldots, N\} \times \mathbb{R}^+\}$, where n is the index of the chosen OP and w_{UE} is the bandwidth corresponding to a peak rate request given the spectral efficiency of the selected OP.

This two-step procedure can model situations of practical interest where, due to some event, such as when (some of) the established OPs in the market simultaneously adopt new transmission technologies or a new OP enters the market, all OPs and UEs re-evaluate their strategies accordingly in order to maximize their utility (revenue) in the new market conditions.

Note that, according to the definition of $\mathcal{S}_{\mathrm{UE}}$, *every UE will subscribe to some OP*, which implicitly suggests a market environment where the services provided by the competing OPs exceed a certain quality threshold. Even though this model is not flexible enough to account for UEs differing from subscribing to any OP due to poor services, the corresponding analysis is still insightful: it will determine the service quality experienced by UEs (in terms of appropriate utility functions) which, in turn, will reveal how system parameters such as OP deployment densities and bandwidth leasing cost(s) should be modified in order to allow for *all* UEs to experience high quality services. This is an issue of critical importance to entities such as regulators who are interested in maximizing social welfare.

5.4.2. UE Utility function

In order to complete the game formulation, utility functions must be specified for the players, which, of course, should depend on the strategies of all players involved. Although there do not exist well-established mathematical expressions (metrics) for the utility of the service that a UE experiences, it is clear that any reasonable

[2] It is assumed that the central spectrum frequencies of each OP have been assigned beforehand and they are sufficiently apart so that partitioning of the spectrum is possible for any (practical) profile of OP and UE strategies.

UE utility function should involve the *actual* rate experienced by the UE. The rate experienced by a UE positioned at $x_0 \in \Phi_u$ and adopting the strategy $(n, w_{\text{UE}}) \in \mathcal{S}_{\text{UE}}$ equals

$$r_n(x_0) \triangleq \frac{c_n w_{\text{UE}}}{1 + K_{0,n}}, \tag{5.1}$$

in bits/sec, where $K_{0,n} \geq 0$ denotes the number of additional UEs subscribed to OP n and served by the same AP as the considered UE. Clearly, $K_{0,n} = K_{0,n}(x_0, \Phi_n, \Phi_{u,n})$, where $\Phi_{u,n} \subseteq \Phi_u$ denotes the positions of UEs selecting OP n.

Unfortunately, due to the infinite number of UEs and the randomness of Φ_n and Φ_u, direct utilization of the expression of (5.1) in the utility function of a UE leads to an intractable problem (no matter what utility function is used). In order to alleviate these difficulties, an approach similar to a mean field game formulation [90] is employed in this work. Specifically, the utility of a UE positioned at $x_0 \in \Phi_{u,n}$ under strategy $(n, w_{\text{UE}}) \in \mathcal{S}_{\text{UE}}$ is considered dependent on $\mathbb{E}(r_n(x_0))$ (instead of $r_n(x_0)$) with the expectation taken over the distributions of Φ_n and $\Phi_{u,n}$ (the latter conditioned on the existence of a point at x_0). In order to obtain $\mathbb{E}(r_n(x_0))$, the following modeling assumption will be employed.

Assumption 5.1. The point process $\Phi_{u,n} \subseteq \Phi_u$, $n \in \{1, 2 \ldots, N\}$, is distributed as an independent from everything else HPPP of density $q_n \in [0, 1]$, with $\sum_{n=1}^{N} q_n = 1$.

Note that the value of density q_n equals the percentage of the UE population that is subscribed to OP n. Assumption 1 is mathematically convenient, as it allows to analytically obtain the average of $r_n(x_0)$ as per Lemma 5.2 below. In addition, by the stationarity and ergodicity properties of the HPPP [18], the (enseble) average rate, $\mathbb{E}(r_n(x_0))$, is actually independent of x_0 and equals the spatial average of r_n for any realization of Φ_n and $\Phi_{n,u}$. Note that the spatial average of the UE rate is a natural metric to incorporate in the UE utility for the scenario where the UEs have non-fixed positions and expect good quality services on average over the whole OP deployment area.

Lemma 5.2. *The average rate experienced by a UE employing strategy* $(n, w_{UE}) \in \mathcal{S}_{\text{UE}}$ *is approximately equal to*

$$\mathbb{E}(r_n) = c_n w_{\text{UE}} \left(1 - e^{-\lambda_n / q_n}\right). \tag{5.2}$$

Proof. See Appendix sec. D.1. □

The term q_n/λ_n appearing in inversion in (5.2) equals the average number of subscribers that a random AP of OP n is associated with [93] and can be considered

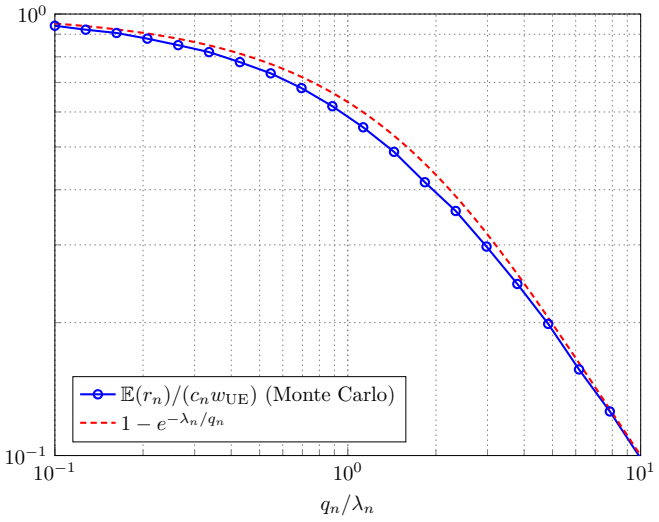


Figure 5.2.: Dependence of average UE rate on OP load (exact and approximate)

as an indicator of the *traffic load* of OP n. Equation (5.2) shows that OPs with larger investments w.r.t. deployment density and/or transmission technology have increased potential to provide large average rates to their subscribers as a long as their traffic load is not overwhelming. The accuracy of the approximate expression of Lemma 5.2 is verified in Fig. 5.2 where it is compared with Monte Carlo evaluations of $\mathbb{E}(r_n)$ obtained by averaging over 10^5 independent system realizations conditioned on the existence of a subscriber at the origin (any other point considered instead would give the same results). It is seen that the relatively simple expression of (5.2) provides a good approximation of the actual average rate.

With the average UE rate analytically evaluated, the utility function of a UE employing strategy $(n, w_{\mathrm{UE}}) \in \mathcal{S}_{\mathrm{UE}}$ can now be specified. In this work, the considered UE utility equals

$$\begin{aligned} u_{\mathrm{UE},n} &\triangleq \log(\mathbb{E}(r_n)) + \beta \log \lambda_n - p_{\mathrm{UE},n} c_n w_{\mathrm{UE}} \\ &= \log\left(c_n w_{\mathrm{UE}} \lambda_n^{\beta} \left(1 - e^{-\lambda_n/q_n}\right)\right) - p_{\mathrm{UE},n} c_n w_{\mathrm{UE}}, \end{aligned} \tag{5.3}$$

where $p_{\mathrm{UE},n} > 0$ is the price that OP n sets per unit of requested peak rate and $\beta \geq 0$ is a parameter representing the competing advantage (if any) of OPs with large infrastructure densities. This advantage reflects potential bias of UEs towards major brands and also provides a simple generalization of results to a scenario where coverage issues do exist and larger infrastructure density provides benefits in this respect. Similar UE utility functions have been widely used in the related literature (see, e.g., [83, 84, 87]), with the $\log(\cdot)$ operator capturing the typical behavior of a UE being reluctant to pay for more resources when the perceived service quality is already at a high level.

5.4.3. OP Utility function

The utility function that will be considered in the following for OP $n \in \{1, 2, \ldots, N\}$ employing strategy $(p_{\mathrm{UE},n}, w_{\mathrm{OP},n}) \in \mathcal{S}_{\mathrm{OP}}$ is (similar to, e.g., [84, 87])

$$u_{\mathrm{OP},n} \triangleq q_n p_{\mathrm{UE},n} c_n \langle \{w_{\mathrm{UE},n}\} \rangle - p_{\mathrm{OP},n} w_{\mathrm{OP},n}, \tag{5.4}$$

where $\langle \{w_{\mathrm{UE},n}\} \rangle$ is the mean of the infinite sequence of bandwith requests from the subscribers of OP n,[3] and $p_{\mathrm{OP},n} > 0$ is the price that OP n pays to the corresponding spectrum provider per unit bandwidth per unit area. Note that the first term of (5.4) represents the average revenue from subscriber fees per unit area and the second term the cost of bandwidth leasing per unit area. In order to be able to accomodate for all bandwidth (peak rate) requests from subscribers, it is assumed that $w_{\mathrm{OP},n} \geq \max\{w_{\mathrm{UE},n}\}$ for all $n \in \{1, 2, \ldots, N\}$.

Even though the utility function of (5.4) can be equal to any real value in principle, an OP will in practice require that this value is strictly greater than a minimum threshold reflecting, e.g., operational costs. Otherwise, no profit can be obtained and the OP will withdraw from the competition. For simplicity and w.l.o.g., a zero minimum utility threshold will be assumed for all OPs in the following, unless stated otherwise.

5.5. Solution Concepts (General Case)

As in every game-theoretic study, it is of interest to characterize the equilibrium (eq.) of the game in terms of player strategies and utilities. This characterization will allow to obtain insights on how the market is affected by parameters such as number of OPs, infrastructure deployment densities, and spectral efficiencies.

In order to obtain the eq. (solution) of the two-stage game described in Sec. 3. 1, the common approach of backward induction will be employed: Initially, the *UE game* will be analyzed, obtaining the optimal strategies and utilities of UEs for any strategy profile of the OPs. Given the information about how UEs respond to OP strategies, the *OP game* can then be examined, in order to obtain the optimal OP strategies that will determine the utilities of OPs and UEs when market eq. is established.

5.5.1. UE game

In the UE game, UEs are informed about the prices $\{p_{\mathrm{UE},n}\}_{n=1}^{N}$ offered by the OPs and have to decide on which OP to subscribe to as well as the requested bandwidth.

[3] This mean exists since $\{w_{\mathrm{UE},n}\}$ is bounded for all practical cases.

Regarding the latter parameter, it is trivial to show that requesting bandwidth equal to $w_{\mathrm{UE},n} = 1/(p_{\mathrm{UE},n}c_n)$ when selecting OP n maximizes the UE utility function of (5.3). This observation simplifies the UE strategy as one of OP selection only, and the utility of a UE selecting OP n can be expressed as

$$\begin{aligned} u_{\mathrm{UE},n} &= \log\left(\frac{1}{p_{\mathrm{UE},n}}\lambda_n^{\beta}\left(1-e^{-\lambda_n/q_n}\right)\right) - 1 \\ &\doteq \tilde{p}_{\mathrm{UE},n}\lambda_n^{\beta}\left(1-e^{-\lambda_n/q_n}\right), \end{aligned} \tag{5.5}$$

where $\doteq$ denotes equality after application of invertible operations that have no effect in the analysis, and $\tilde{p}_{\mathrm{UE},n} \triangleq 1/p_{\mathrm{UE},n} \in [0,\infty)$ is the inverse of the price for employing the services of OP n, which will be employed for convenience in place of $p_{\mathrm{UE},n}$ in the following. Comparing (5.5) with (5.2) shows that the (equivalent) utility function of a UE selecting OP n can be interpreted as the rate actually experienced by the UE for a peak rate service equal to $\tilde{p}_{\mathrm{UE},n}\lambda_n^{\beta}$.

In order to specify the OP selection strategy of the UEs, the type of the UE game eq. must be defined. Consideration of an infinite number of UEs and utilities that depend only on the distribution $\{q_n\}$ of UEs among OPs and not on the individual strategy of each UE implies that the UEs are infinitesimal, i.e., the impact of a *single* UE switching from OP n to OP $m \neq n$ on the utility of the other UEs subscribed to these OPs is negligible. (However, the utility of the transitioning UE will change in general.) Under this setting, a natural model for the steady state of the UE game, based on Wardrop's first principle [94], is the following.

Definition 5.3 (UE game equilibrium)**.** When the UE eq. is established, the UE distribution $\{q_n^*\}_{n=1}^N$ among OPs will be such that all UEs achieve the same utility (the *UE game value*), $u_{\mathrm{UE}}^* > 0$, irrespective of the OP they are associated with.

This eq. definition only specifies the distribution of the strategies of UEs and does not provide the means to explicitly determine the strategy of each individual UE. However, the latter is of no practical interest since the utility of each UE is the same irrespective of the actual OP it is associated with when the eq. has been established.[4]

By this approach, the infinite set of UEs is essentially translated into a single player (the mean field) whose strategy belongs to the set of N-dimensional probability distributions. Note that it is possible to achieve a UE game eq. by more than one distributions, each resulting in possibly different UE utility values. However, it is reasonable to assume that the UEs will eventually reach the eq. of maximum

[4]This eq. will eventually be reached by some evolutionary process whose model and investigation is out of the scope of this work.

utility (Wardrop's second principle), implying that characterization of the UE eq. is equivalent to solving the following optimization problem.

$$\underset{\left\{u_{\mathrm{UE}},\mathcal{N},\{q_n\}_{n=1}^N\right\}}{\text{maximize}} \quad u_{\mathrm{UE}} \tag{5.6a}$$

$$\begin{aligned}
\text{subject to } & u_{\mathrm{UE}} > 0, \\
& \mathcal{N} \subseteq \{1,2,\ldots,N\}, \\
& q_n \in (0,1], n \in \mathcal{N}, \\
& q_n = 0, n \notin \mathcal{N}, \\
& \sum_{n\in\mathcal{N}} q_n = 1, && \text{(5.6b)} \\
& \tilde{p}_{\mathrm{UE},n}\lambda_n^{\beta}\left(1-e^{-\lambda_n/q_n}\right) = u_{\mathrm{UE}}, n \in \mathcal{N}, && \text{(5.6c)}
\end{aligned}$$

where $\mathcal{N}$ has the interpretation of the set of OPs attracting a non-zero percentage of UEs at the UE equilibrium. Ties appearing when the maximum UE utility can be achieved by more than one UE distributions over OPs, are resolved arbitrarily.

The above multi-variable constrained optimization problem does not have a closed form solution. However, the following proposition shows that it can be efficiently solved numerically by performing $2^N - 1$ one-dimensional searches.

Proposition 5.4. *The solution* $\left\{u_{\mathrm{UE}}^*, \mathcal{N}^*, \{q_n^*\}_{n=1}^N\right\}$ *of the optimization problem (8) is*

$$\mathcal{N}^* = \arg \max_{\mathcal{N}\subseteq\{1,2,\ldots,N\}} u_{\mathrm{UE}}(\mathcal{N}), \tag{5.7}$$

$$u_{\mathrm{UE}}^* = u_{\mathrm{UE}}(\mathcal{N}^*), \tag{5.8}$$

$$q_n^* = \begin{cases} \frac{-\lambda_n}{\log(1-u_{\mathrm{UE}}^*\lambda_n^{-\beta}/\tilde{p}_{\mathrm{UE},n})} & , n \in \mathcal{N}^*, \\ 0 & , n \notin \mathcal{N}^*, \end{cases} \tag{5.9}$$

where $u_{\mathrm{UE}}(\mathcal{N}), \mathcal{N} \subseteq \{1,2,\ldots,N\}$*, is the unique solution of* $\sum_{n\in\mathcal{N}} \frac{-\lambda_n}{\log(1-x\lambda_n^{-\beta}/\tilde{p}_{\mathrm{UE},n})} = 1, 0 < x < \min\{\tilde{p}_{\mathrm{UE},n}\lambda_n^{\beta}\}_{n\in\mathcal{N}}$*, when it exists, or equals* 0*, otherwise.*

Proof. One approach for solving the mixed integer optimization problem is to consider the $2^N - 1$ subproblems resulting from (8) by setting $\mathcal{N}$ to each of its possible non-empty configurations, finding the optimal solutions, $\{u_{\mathrm{UE}}(\mathcal{N})\}_{\mathcal{N}\subseteq\{1,2,\ldots,N\}}$, and obtaining $\mathcal{N}^*$ and u_{UE}^* as in (5.7) and (5.8), respectively, with (5.9) following directly from (5.6c).

Determination of $u_{\mathrm{UE}}(\mathcal{N})$ can be obtained as follows. Combining (5.6b) and (5.6c), it follows that the optimal solution for each subproblem (if it exists) must satisfy the condition

$$\sum_{n\in\mathcal{N}} \frac{-\lambda_n}{\log(1-u_{\mathrm{UE}}(\mathcal{N})\lambda_n^{-\beta}/\tilde{p}_{\mathrm{UE},n})} = 1, \tag{5.10}$$

with $u_{\mathrm{UE}}(\mathcal{N}) < \min\{\tilde{p}_{\mathrm{UE},n}\lambda_n^{\beta}\}_{n\in\mathcal{N}}$, in order to guarantee that each of the summands on the left hand side of (5.10), corresponding to $q_n, n \in \mathcal{N}$, is real and positive. Noting that $\sum_{n\in\mathcal{N}} \frac{-\lambda_n}{\log(1-x\lambda_n^{-\beta}/\tilde{p}_{\mathrm{UE},n})}$ is a monotonically decreasing function of $x > 0$, if follows that if the subproblem corresponding to some $\mathcal{N}$ is feasible, then its optimal value is the unique solution of (5.10), whereas if the subproblem is not feasible such a solution cannot be found and this configuration of $\mathcal{N}$ can be neglected or, equivalently, its optimal value is set to 0. This procedure specifies $u_{\mathrm{UE}}(\mathcal{N}), \mathcal{N} \subseteq \{1, 2, \ldots, N\}$ and completes the proof. □

Proposition 5.4 shows that when maximization of u_{UE} is achieved by a unique $\mathcal{N}^*$, the UE distribution over OPs is also unique. In addition, it directly follows from Proposition 5.4, that u_{UE}^* is bounded as $u_{\mathrm{UE}}^* < \min\{\tilde{p}_{\mathrm{UE},n}\lambda_n^{\beta}\}_{n\in\mathcal{N}^*}$, providing a simple expression on how the OP strategies (prices) affect the UE game value. In order to obtain additional insight on how strategic choices $\{\tilde{p}_{\mathrm{UE},n}\}$ as well as network densities $\{\lambda_n\}$ of OPs affect the UE game value, the following proposition is of value.

Proposition 5.5. *The UE game value, given* $\mathcal{N}^*$, $\{\tilde{p}_{\mathrm{UE},n}\}_{n\in\mathcal{N}^*}$ *and* $\{\lambda_n\}_{n\in\mathcal{N}^*}$ *equals*

$$u_{\mathrm{UE}}^* = \frac{\sum_{n\in\mathcal{N}^*} \lambda_n^{1+\beta}\tilde{p}_{\mathrm{UE},n}}{1+\kappa\sum_{n\in\mathcal{N}^*}\lambda_n}, \tag{5.11}$$

where $\kappa \in [1/2, 1]$.

Proof. Bounding the expression of (5.9) for q_n^* using the inequality $\frac{1}{x}-1 \leq -1/\log(1-x) \leq \frac{1}{x}-\frac{1}{2}$ that holds for $0 < x < 1$, and since $\sum_{n\in\mathcal{N}^*} q_n^* = 1$, it follows that quantity

$$\sum_{n\in\mathcal{N}^*} \lambda_n \left(\frac{\tilde{p}_{\mathrm{UE},n}\lambda_n^{\beta}}{u_{\mathrm{UE}}^*} - s\right)$$

is less or equal than 1 for $s = 1$ and greater or equal than 1 for $s = 1/2$, from which u_{UE}^* can be bounded as

$$\frac{\sum_{n\in\mathcal{N}^*} \lambda_n^{1+\beta}\tilde{p}_{\mathrm{UE},n}}{1+\sum_{n\in\mathcal{N}^*}\lambda_n} \leq u_{\mathrm{UE}}^* \leq \frac{\sum_{n\in\mathcal{N}^*} \lambda_n^{1+\beta}\tilde{p}_{\mathrm{UE},n}}{1+\frac{1}{2}\sum_{n\in\mathcal{N}^*}\lambda_n}.$$

The form of this bound implies the expression of (5.11). □

The simple expression of (5.11) provides u^*_{UE} with an ambiguity in terms of parameter κ, which depends, in general, on $\{\tilde{p}_{\mathrm{UE},n}\}_{n\in\mathcal{N}^*}$ and $\{\lambda_n\}_{n\in\mathcal{N}^*}$. However, (5.11) is tight for $\sum_{n\in\mathcal{N}^*}\lambda_n \ll 1$ (small OP infrastructure investments) with the ratio of upper and lower bounds, corresponding to $\kappa = 1$ and $\kappa = 1/2$, respectively, not exceeding 3/2 for $\sum_{n\in\mathcal{N}^*}\lambda_n \leq 2$, which is the region of practical interest as it will be shown in the analysis of the OP game. Equation (5.11) provides an intuition on how each OP taking part at the UE eq. contributes to the UE game value, namely, the presence of OPs with large ratio of deployment density to offered service price results in large UE game values.

5.5.2. OP game

The analysis and solution of the UE game provides the best response of the UEs to any price profile $\{\tilde{p}_{\mathrm{UE},n}\}_{n=1}^{N}$ employed by the OPs. This is an information that is available to all OPs and clearly plays an important role in shaping their own strategy. In particular, since *every* subscriber of OP n requests bandwidth $w_{\mathrm{UE},n} = \tilde{p}_{\mathrm{UE},n}/c_n$, it directly follows from (5.4) that it is optimal for OP n to request the exact same bandwidth from the spectrum provider. This, in turn, simplifies the strategy specification problem of OP n to the determination of the (inverse) price $\tilde{p}_{\mathrm{UE},n}$, with the utility function of (5.4) becoming

$$u_{\mathrm{OP},n}(\{\tilde{p}_{\mathrm{UE},m}\}_{m=1}^{N}) = q^*_n - p^c_{\mathrm{OP},n}\tilde{p}_{\mathrm{UE},n}, \tag{5.12}$$

where $q^*_n \in [0,1]$ is the market share of OP n when the UE eq. has been established and $p^c_{\mathrm{OP},n} \triangleq p_{\mathrm{OP},n}/c_n > 0$ is the *normalized bandwidth leasing price* that OP n pays for acquiring bandwidth, reflecting the advantage of employing a highly efficient transmission technology.

Towards determining the OP strategies, the following definition for the OP game eq. will be employed, according to the well-known concept of Nash eq. for a finite number of competitive players [80].

Definition 5.6 (OP game equilibrium)**.** At the eq. of the OP game, no OP has the incentive to change strategy, i.e.,

$$\tilde{p}^*_{\mathrm{UE},n} = \arg\max_{\tilde{p}_{\mathrm{UE},n}\geq 0} u_{\mathrm{OP},n}\left(\tilde{p}_{\mathrm{UE},n}, \{\tilde{p}^*_{\mathrm{UE},m}\}_{m\neq n}\right), \tag{5.13}$$

for $n \in \{1,2,\ldots,N\}$, where $\tilde{p}^*_{\mathrm{UE},n}$ denotes the eq. strategy (best response) of OP n. A value of $\tilde{p}^*_{\mathrm{UE},n} = 0$ implies that OP n withdraws from the market.

Fig. 5.3 depicts $u_{\mathrm{OP},1}$, as a function of $\tilde{p}_{\mathrm{UE},1}$ and parameterized by $\tilde{p}_{\mathrm{UE},2}$, for the example case of $N = 2$, $\lambda_1 = \lambda_2 = 0.4$, $\beta = 0$, and $p^c_{\mathrm{OP},1} = 0.01$ (similar curves are

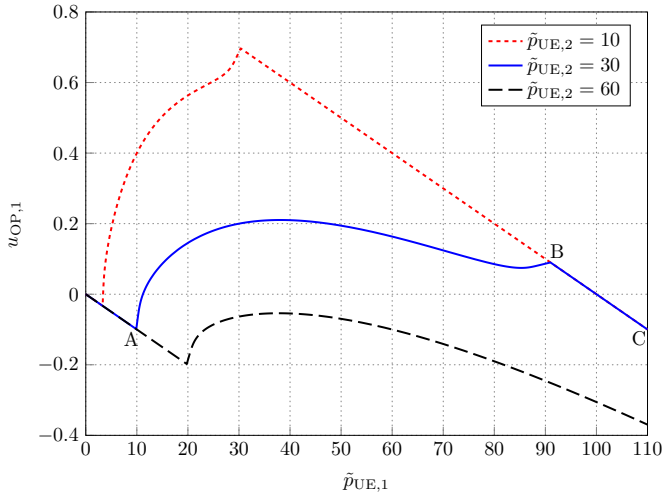


Figure 5.3.: Utility function of OP 1 for $N = 2$, $\lambda_1 = \lambda_2 = 0.4$, $p^{\mathrm{c}}_{\mathrm{OP},1} = 0.01$.

obtained for arbitrary values of these parameters). It can be seen that the interval $[0, \infty)$, where $\tilde{p}_{\mathrm{UE},1}$ takes values in, is partitioned into three regions according to the resulting value of q_1^* (see curve corresponding to $\tilde{p}_{\mathrm{UE},2} = 30$ in Fig. 5.3):

1. $q_1^* = 0$ (segment 0A): The inverse price $\tilde{p}_{\mathrm{UE},1}$ is so small so that the solution of the UE eq. problem results in $\mathcal{N}^* = \{2\}$, i.e., all UEs subscribe to OP 2. It follows from (5.12) that the utility of OP 1 is non-positive in this region.
2. $0 < q_1^* < 1$ (segment AB): The inverse price $\tilde{p}_{\mathrm{UE},1}$ is large enough so that OP 1 attracts part of the UE population ($\mathcal{N}^* = \{1, 2\}$). Note that this does not necessarily mean a positive utility for OP 1 as q_1^* may not be sufficiently large so that revenues from subscriber payments can compensate for the bandwidth leasing cost.
3. $q_1^* = 1$ (segment BC): For sufficiently large $\tilde{p}_{\mathrm{UE},1}$ (point B), $\mathcal{N}^* = \{1\}$ at the UE eq., i.e., OP 1 acquires monopoly status. It follows from (5.12) that further increase of $\tilde{p}_{\mathrm{UE},1}$ reduces utility of OP 1 linearly. Monopoly status of an OP does not guarantee a positive utility for the same reason as in the $0 < q^* < 1$ case.

Note that, depending on $\tilde{p}_{\mathrm{UE},2}$, which reflects the competing strength of OP 2, the best response of OP 1 may result in either (a) withdraw from, (b) share, or (c) monopolize the market. These three cases are shown in Fig. 5.3, for $\tilde{p}_{\mathrm{UE},2} = 60, 30,$ and 10, respectively.

It is clear from the simple case considered above, that the OP utility function does not possess desirable properties such as differentiability and (quasi-)convexity (see shape of curves in Fig. 5.3). In addition, it can be verified that the OP game specification does not correspond to a potential or modular game. This means that "off-the-self" theorems that are commonly employed for determining important properties of the game eq. such as existence and uniqueness [95] cannot be applied in

this case. Therefore, characterization of the OP game eq. must be performed either by an analytic approach tailored to the particular problem, or by using (generic purpose) numerical techniques. An analytic characterization is in general preferable (when available) for determining existence and uniqueness, as well as general properties of the OP game eq., since a numerical approach may fail to determine the eq. under certain system conditions even when it exists. Unfortunately, an analytical investigation of the OP game eq. for the general case appears to be intractable. However, analytical insights can be obtained for some special, but also of practical importance, cases, which are treated in the following sections.

5.6. Two Competing Operators

This section provides an analytical investigation of the OP game eq. in the case of $N = 2$ competing OPs, which models the scenario of practical importance where there are two dominant OPs in the market with their total share of UEs not affected by the presence of other OPs of small market power. For example, these two OPs may explicitly address the need a certain type of application/service not considered by other OPs.

In this scenario, if an eq. of the OP game exists, there can only be two possibilities:

- A monopoly eq. is established, i.e., the strategy of one of the OPs does not allow the other to obtain a positive utility for any choice of strategy, or
- A duopoly eq. [81] is established, i.e., both OPs achieve positive utilities by sharing the market.

The focus in the following will be on characterizing the duopoly eq. as it is typically preferable from a social welfare perspective. However, analysis will also provide insights on monopoly equilibrium.

5.6.1. Characterization of Duopoly equilibria

The following theorem shows that for certain system parameters, establishment of a duopoly equilibrium is possible and is actually unique, i.e., no OP prefers to establish monopoly instead.

Theorem 5.7. *If a duopoly equilibrium of the OP game for $N = 2$ exists, i.e., with both OPs achieving positive utility, it is unique, and it holds*

$$1 - e^{-\lambda_1} < \frac{p^c_{\mathrm{OP},1}\lambda_2^\beta}{p^c_{\mathrm{OP},2}\lambda_1^\beta} < \frac{1}{1 - e^{-\lambda_2}}, \tag{5.14}$$

with OP eq. strategies given as

$$\tilde{p}^*_{\mathrm{UE},1} = \frac{q_1^{*2}\left(e^{\lambda_1/q_1^*}-1\right)}{p^c_{\mathrm{OP},1}\left(\lambda_1 + \frac{p^c_{\mathrm{OP},1}}{p^c_{\mathrm{OP},2}}\left(\frac{q_1^*}{q_2^*}\right)^2 \frac{\lambda_2^{1+2\beta}}{\lambda_1^{2\beta}} e^{\lambda_1/q_1^* - \lambda_2/q_2^*}\right)}, \tag{5.15}$$

$$\tilde{p}^*_{\mathrm{UE},2} = \frac{p^c_{\mathrm{OP},1}\lambda_2^\beta}{p^c_{\mathrm{OP},2}\lambda_1^\beta}\tilde{p}^*_{\mathrm{UE},1}, \tag{5.16}$$

where $q_1^* \in (0,1)$ *and* $q_2^* = 1 - q_1^* \in (0,1)$ *are the market shares of OPs* 1 *and* 2*, respectively, given by the unique solution of*

$$p^c_{\mathrm{OP},2}\lambda_1^\beta\left(1 - e^{-\frac{\lambda_1}{q_1^*}}\right) = p^c_{\mathrm{OP},1}\lambda_2^\beta\left(1 - e^{-\frac{\lambda_2}{1-q_1^*}}\right). \tag{5.17}$$

Proof. See Appendix sec. D.2. □

Theorem 5.7 provides a semi-analytic formula for computing the eq. strategies of OPs, based on the numerical solution of (5.17) that provides the corresponding UE game eq. strategy. These strategy values can then be employed to determine the UE and OP utilities from (5.6c) and (5.12), respectively. An interesting special case that permits a complete, closed form characterization of the OP game eq. is for $\beta = 0$ and $p^c_{\mathrm{OP},1} = p^c_{\mathrm{OP},2}$, corresponding to the case where UEs are indifferent to infrastructure densities and both OPs employ the same transmission technology and acquire spectrum with the same price (e.g., set by a regulator). The eq. description directly follows from Theorem 5.7.

Corollary 5.8. *Under a duopoly equilibrium with* $N = 2$*,* $\beta = 0$*, and* $p^c_{\mathrm{OP},1} = p^c_{\mathrm{OP},2} = p^c_{\mathrm{OP}}$*, the OP and UE strategies are*

$$\tilde{p}^*_{\mathrm{UE},1} = \tilde{p}^*_{\mathrm{UE},2} = \frac{\lambda_1\lambda_2(e^{\lambda_1+\lambda_2}-1)}{p^c_{\mathrm{OP}}(\lambda_1+\lambda_2)^3},$$
$$q_n^* = \frac{\lambda_n}{\lambda_1+\lambda_2}, n \in \{1,2\},$$

respectively, with corresponding utilities

$$u^*_{\mathrm{OP},n} = \frac{\lambda_n}{\lambda_1+\lambda_2} - \frac{\lambda_1\lambda_2(e^{\lambda_1+\lambda_2}-1)}{(\lambda_1+\lambda_2)^3}, n \in \{1,2\},$$
$$u^*_{\mathrm{UE}} = \frac{2\lambda_1\lambda_2\left(\cosh(\lambda_1+\lambda_2)-1\right)}{p^c_{\mathrm{OP}}(\lambda_1+\lambda_2)^3}.$$

It can be seen that for this special case, both OPs will announce *equal* service prices (and request same bandwidth), even when their deployment densities are different. This is because at the eq. the UE distribution between OPs is proportional to their densities, resulting in both OPs providing the same quality of service (average UE rate) due to the balanced traffic load. In addition, the OP utilities at the eq. are independent of the normalized bandwidth leasing price. However, the latter does affect the UE game value, with a smaller price translated into an increased utility for the UEs. This is an important observation from the perspective of a regulator who is interested in the UE welfare as it has implications on a) how the bandwidth leasing price is set, i.e., a smaller price may be preferred leading to an increased UE welfare that counterbalances direct monetary loss, and b) providing other incentives than those reflected in the OP utility considered in this work for investing in advanced transmission technologies .

5.6.2. Duopoly Region

Condition (5.14) implies that a duopoly eq. cannot exist for arbitrary values of λ_1, λ_2, $p^c_{\mathrm{OP},1}/p^c_{\mathrm{OP},2}$, and β. However, (5.14) is introduced in the proof of Theorem 5.7 as the necessary and sufficient condition for the existence of a (unique) solution, $q_1^* \in (0,1)$, of (5.17), which does not guarantee that the corresponding OP strategies of (5.15) and (5.16) will be both positive. In order to precisly determine conditions leading to both OPs achieving positive utility the notion of duopoly region is introduced.

Definition 5.9. The duopoly region $\mathcal{L} \subseteq \mathbb{R}^+ \times \mathbb{R}^+$ is the set of pairs of deployment densities $(\lambda_1, \lambda_2) \in \mathbb{R}^+ \times \mathbb{R}^+$ for which an OP eq. exists with $u^*_{\mathrm{OP},1} > 0$ and $u^*_{\mathrm{OP},2} > 0$.

Of course, the above definition can be generalized to the case where OPs require their utility any positive real number. Note that the uniqueness of a duopoly equilibrim implies the following result

Corollary 5.10. *The set $(\mathbb{R}^+ \times \mathbb{R}^+) \backslash \mathcal{L}$ corresponds to pairs of deployment densities where if an eq. exists it must be a monopoly.*

With the exception of the symmetric case corresponding to Corollary 5.8, $\mathcal{L}$ can only be obtained numerically by examining, for each $(\lambda_1, \lambda_2) \in \mathbb{R}^+ \times \mathbb{R}^+$, the values of $u^*_{\mathrm{OP},1}$ and $u^*_{\mathrm{OP},2}$ under the strategies of Theorem 5.7. Fig. 5.4 shows $\mathcal{L}$ for $p^c_{\mathrm{OP},1}/p^c_{\mathrm{OP},2} = 1, 5, 10$, $\beta = 0, 1$ and minimum utility threshold for both OPs equal to 0 and 0.1. Note that values of $p^c_{\mathrm{OP},1}/p^c_{\mathrm{OP},2}$ greater than 1 correspond to the cases where OP 1 has a disadvantage due to increased bandwidth leasing cost and/or outdated transmission technology. Graphs of $\mathcal{L}$ corresponding to $p^c_{\mathrm{OP},1}/p^c_{\mathrm{OP},2}$ values smaller than 1 are not depicted for clarity as they are the mirror images of the graphs corresponding to larger than 1 values w.r.t. the $\lambda_1 = \lambda_2$ axis.

The following observations are in order:

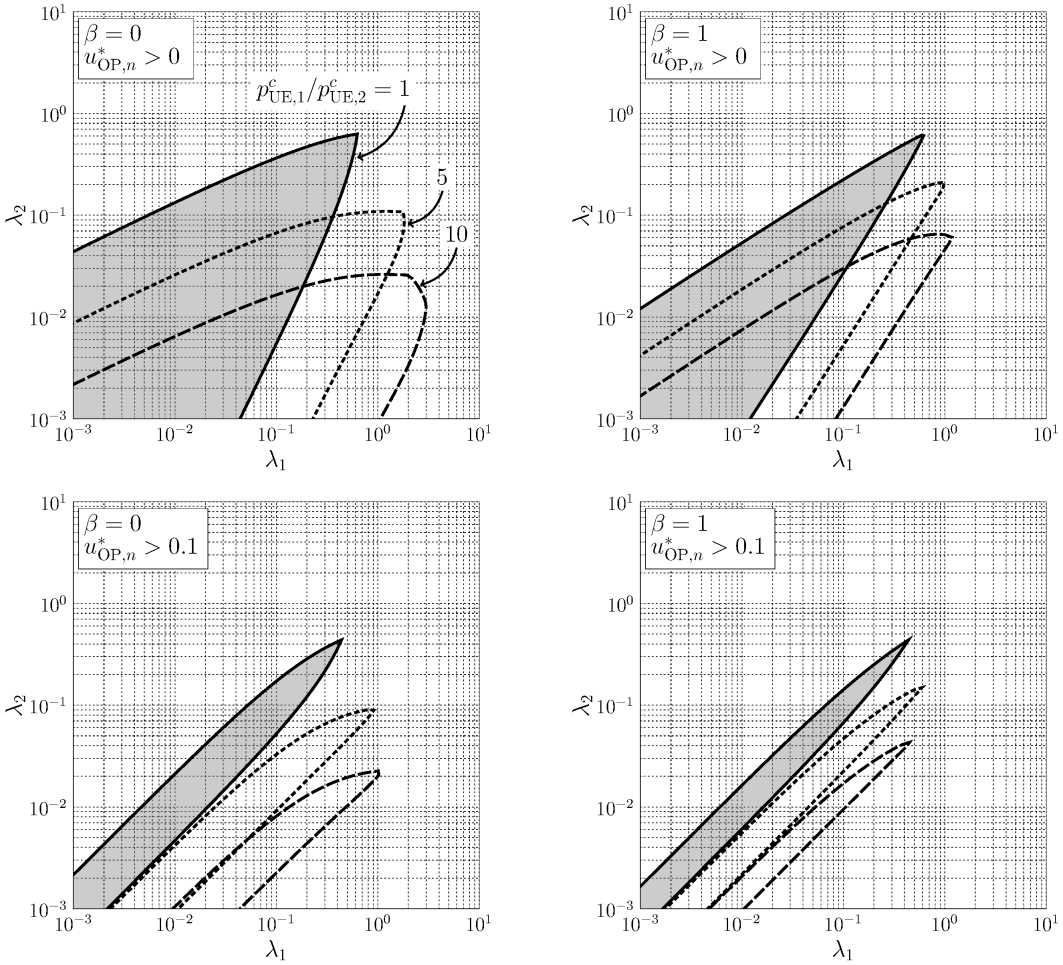


Figure 5.4.: Duopoly region, $\mathcal{L}$ for various values of $p^c_{\mathrm{OP},1}/p^c_{\mathrm{OP},2}$, β, and minimum OP utility thresholds (same for both). Shaded area corresponds to the region interior for the case of $p^c_{\mathrm{OP},1} = p^c_{\mathrm{OP},2}$. Only the region boundary is depicted for the other cases (their interior is obvious).

1. The duopoly region $\mathcal{L}$ is bounded, connected and does not includes points where λ_1 and λ_2 or $|\lambda_1 - \lambda_2|$ are sufficiently large. The maximum *total* deployment density of both OPs for which a duopoly eq. exists is of order 1 in all cases. For example, for the case of $p^c_{\mathrm{OP},1}/p^c_{\mathrm{OP},2} = 1$, $\beta = 0$, and a zero OP utility threshold, $\max\{\lambda_1\} = \max\{\lambda_2\} \approx 0.62826$, corresponding to an average of $2 \times 0.62826 \approx 1.25$ APs per subscriber for each OP.
2. For the regime of dense infrastructure deployments relative to the total UE density, with, say, $\max\{\lambda_1, \lambda_2\} \geq 0.1$, duopoly requires comparable OP densities as well as comparable normalized bandwidth leasing prices.
3. For the regime of sparse infrastructure deployments, with, say, $\max\{\lambda_1, \lambda_2\} < 0.1$, under small values of β and minimum OP utility threashold, duopoly can be established even with significant difference (orders of magnitude) between

λ_1 and λ_2 and/or $p^c_{\mathrm{OP},1}$ and $p^c_{\mathrm{OP},2}$.

4. Increasing $p^c_{\mathrm{OP},1}/p^c_{\mathrm{OP},2}$ results in duopoly regions where points with $\lambda_2 \gg \lambda_1$ are not included, as the only way OP 1 can compensate for the large normalized bandwidth leasing price is with a sufficiently large q^*_1, which, in turn, requires λ_1 to be at least comparable, if not greater, than λ_2.

5. The duopoly region with a positive minimum OP utility threshold is a subset of the region with a zero threshold, as differences in infrastructure densities result in an unbalanced UE distribution with the number of subscribers to the OP with the smaller density not large enough to provide sufficiently large utility.

6. Increasing β results in a "thinning" of the duopoly region for the same reason as in the case of an increasing minimum utility. In addition, since increasing β results in a more unbalanced distribution among OPs in the case when $\lambda_1 \neq \lambda_2$ compared to the $\beta = 0$ case, a corresponding "shift" of the duopoly region towards the $\lambda_1 = \lambda_2$ axis is observed.

The first observation has implications on the level of infrastructure densification of future networks, not only for establishment of duopoly but a stable market as well. In particular, it is possible that for $(\lambda_1, \lambda_2) \notin \mathcal{L}$ even monopoly cannot be established. To see that, consider the symmetric case with $p^c_{\mathrm{OP},1}/p^c_{\mathrm{OP},2} = 1$ and $\lambda_1 = \lambda_2$. Numerical evaluation of the OP strategies under a duopoly in this case shows that increasing density results in OPs increasing their inverse price. This behavior can be explained by noting that, under conditions with large infrastructure densities, an OP can attract a large percentage of the market when offering services with even a slightly more attractive price than its competitor. Accordingly, both OPs decrease their prices (increase their inverse price) with increasing density to eliminate the incentive of their competitor for offering an even smaller price. This ultimately results in an unacceptably small utility for both OPs when the infrasturcrure density becomes sufficiently large implying that duopoly cannot be established. However, the symmetry of this case suggests that a monopoly eq. can not exist either, i.e., the market is unstable. This shows that arbitrarily large densification levels, including cases with orders of magnitude larger AP than UE densities, that are typically considered in engineering studies due to the associated benefits in rate performance [51, 24], may not be a viable option when the market interactions are taken into account. This suggests that if extremely high network densification levels are required, a different market model should be employed, e.g., with (partial) cooperation among OPs.

Focusing on the UE welfare under an duopoly eq., Fig. 5.5 shows the normalized UE game value $u^*_{\mathrm{UE}} p^c_{\mathrm{OP},2}$ as a function of $p^c_{\mathrm{OP},1}/p^c_{\mathrm{OP},2} \geq 1$, for $\beta = 0$, arbitrary $p^c_{\mathrm{OP},2}$, and various example cases of (λ_1, λ_2) (results are similar for other deployment density values). Note that the graph corresponding to a specific value of (λ_1, λ_2) is plotted only for the (open) interval of values of $p^c_{\mathrm{OP},1}/p^c_{\mathrm{OP},2}$ for which a duopoly

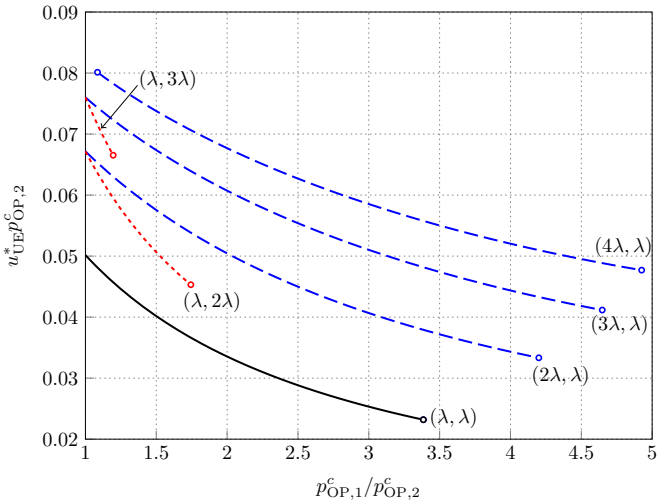


Figure 5.5.: Normalized UE game value as a function of $p^c_{\mathrm{OP},1}/p^c_{\mathrm{OP},2}$ for various values of OP deployment densities (λ_1, λ_2) ($\lambda = 0.1$, $\beta = 0$, $p^c_{\mathrm{OP},2}$ is arbitrary).

exists.[5] Consider first the case of $(\lambda_1, \lambda_2) = (0.1, 0.1)$ as the reference. It is seen that increasing $p^c_{\mathrm{OP},1}/p^c_{\mathrm{OP},2}$, results in a decrease of the UE game value, a result that can be accounted to the unequal market power of the two OPs. For a fixed value of $p^c_{\mathrm{OP},1}/p^c_{\mathrm{OP},2}$ the UE game value increases when either λ_1 or λ_2 increase, as expected. However, for the cases considered, it is the increase of λ_1 that is more beneficial to the UEs as in this way OP 1 counterbalances the disadvantage of having a larger normalized bandwidth leasing price. In contrast, when λ_2 is the one increasing, the benefits to the UEs become smaller as the difference among competitive powers of OPs becomes even larger.

5.7. UE Game Value Maximization Under Multiple OPs Competition

From a purely engineering perspective, dense infrastructure constitutes one of the new and of critical importance attributes of the future network [51, 75]. This observation has resulted in the well-known trend of network infrastructure densification. However, as demonstrated for the two OP case in the previous section, arbitrarily large densification levels, even though preferable for enhancing UE experience in terms of rate may not correspond to a stable market.

This section investigates how the well-known engineering benefits of infrastructure densification are also applicable under a competitive market framework. Focusing

[5] For values of $p^c_{\mathrm{OP},1}/p^c_{\mathrm{OP},2}$ where a monopoly eq. may be established, it is clear that the resulting UE game will not be greater than the one corresponding to the closest value of $p^c_{\mathrm{OP},1}/p^c_{\mathrm{OP},2}$ where duopoly is established.

on maximizing u^*_{UE}, the optimal distribution of deployment densities among $N \geq 2$ competing OPs resulting in a stable market is examined. Apart from potentially providing insights for mechanism design from a regulator perspective, this analysis can also be viewed as an upper performance bound of future competing networks from a UE utility (rather than purely engineering) perspective. For analytical simplicity, a value of $\beta = 0$ will be considered throughout this section with the more general case left open for future investigation.

The next theorem provides a description of the properties of the OP game eq. when the total AP density is optimally (w.r.t. u^*_{UE}) "distributed" among the competing OPs.

Theorem 5.11. *Under the OP density distribution $\{\lambda^*_n\}_{n=1}^N$ that maximizes u^*_{UE}, given $\sum_{n=1}^N \lambda^*_n = \lambda_{\mathrm{tot}}$, where $N \geq 2$ is the number of OPs with positive utility at the OP game equilibrium, it must hold*

$$\frac{\lambda^*_n(\lambda_{\mathrm{tot}} - \lambda^*_n)}{p^c_{\mathrm{OP},n}} = \frac{\lambda^*_m(\lambda_{\mathrm{tot}} - \lambda^*_m)}{p^c_{\mathrm{OP},m}} = \rho \tag{5.18}$$

for all $m, n \in \{1, 2, \ldots, N\}$, with OP strategies and UE distribution among OPs given by

$$\tilde{p}^*_{\mathrm{UE},n} = \frac{\rho\left(e^{\lambda_{\mathrm{tot}}} - 1\right)}{\lambda^3_{\mathrm{tot}}}, \tag{5.19}$$

$$q^*_n = \frac{\lambda^*_n}{\lambda_{\mathrm{tot}}}, \tag{5.20}$$

for $n \in \{1, 2, \ldots, N\}$, respectively.

Proof. See Appendix sec. D.3. □

Equation (5.19) shows that under the u^*_{UE} maximizing deployment density distribution, all OPs will employ the same strategy, implying that all OPs have the same market power. Note that for the case of $N = 2$, condition (5.18) can be satisfied only with $p^c_{\mathrm{OP},1} = p^c_{\mathrm{OP},2}$, which is the scenario considered in Corollary 5.8 from which it is easy to show that maximization of u^*_{UE} is achieved with $\lambda^*_1 = \lambda^*_2 = \lambda_{\mathrm{tot}}/2$. For the case of $N \geq 3$, (5.18) does not imply any simple constraint on the values of $\{p^c_{\mathrm{OP},n}\}_{n=1}^N$, rendering numerical methods necessary to obtain the optimal density distribution. Motivated by the $N = 2$ case, the next Lemma provides closed form characterization of the OP game eq. when all OPs are equal in terms of normalized bandwidth leasing prices.

Lemma 5.12. *When $p^c_{\mathrm{OP},n} = p^c_{\mathrm{OP}}$, for all $n \in \{1, 2, \ldots, N\}$, where N is the number of OPs with non zero UE share at the OP game equilibrium, the density distribution among OPs that maximizes u^*_{UE} is $\lambda^*_n = \lambda$, for all $n \in \{1, 2, \ldots, N\}$, with OP and UE utilities, respectively,*

$$u^*_{\mathrm{OP},n} = \frac{1}{N} - \frac{(N-1)\left(e^{N\lambda} - 1\right)}{N^3\lambda}, n \in \{1, 2, \ldots N\}, \tag{5.21}$$

$$u^*_{\mathrm{UE}} = \frac{2(N-1)(\cosh(N\lambda) - 1)}{p^c_{\mathrm{OP}} N^3 \lambda}. \tag{5.22}$$

Proof. The case for $N = 2$ has already been discussed. Considering an arbitrary value for $N \geq 3$, condition (5.18), under common normalized bandwidth leasing prices, can be written as $\lambda_{\mathrm{tot}}(\lambda_m - \lambda_n) = \lambda_n^2 - \lambda_m^2$, for all m, n. Obviously, this condition is satisfied with $\lambda_m = \lambda_n = \lambda$, for all m, n. Assuming that this condition is also satisfied by a density distribution for which there exists m, n with $\lambda_m \neq \lambda_n$, it follows that it must hold $\lambda_{\mathrm{tot}} = \lambda_m + \lambda_n$, which is impossible. The OP and UE utilities and utilities can be obtained by application of Theorem 5.11 with uniform density distribution and the utility expressions of (5.5) and (5.12). □

Lemma 5.12 shows that UEs benefit from an OP competition involving players of equal market power in terms of deployment densities in addition to normalized bandwidth leasing costs. Note that the equal deployment density scenario can be viewed as a generalization of the setup, considered in previous related works, of a single cell containing an AP for each of the competing OPs, however, explicitly taking into account the randomness of AP positions in wide area system deployments.

Examination of (5.22) shows that the UE game value achieved by OPs having equal deployment densities is an increasing function of N and λ for a fixed common normalized bandwidth price. However, this does not mean that these parameters should jointly be increased to maximize u^*_{UE} since this operation may result in negative OP utilities. Optimization of u^*_{UE} with respect to N and λ, subject to $u^*_{\mathrm{OP},n} > 0$, $n \in \{1, 2, \ldots N\}$, can be performed numerically. Fig. 5.6, shows $\max_N u^*_{\mathrm{UE}} p^c_{\mathrm{OP}}$ as a function of λ. Note that the range of values of λ is partitioned into half-open intervals corresponding to different maximizing value of N. Interestingly, larger values of λ correspond to smaller maximizing N, with the maximum values of $u^*_{\mathrm{UE}} p^c_{\mathrm{OP}}$ achieved for $N = 2$ and sufficiently large λ not exceeding 0.62826 (larger values of λ cannot result in an eq., as discussed in Sec. 5). In addition, Fig. 5.6 depicts the lower bound of $u^*_{\mathrm{UE}} p^c_{\mathrm{OP}}$ obtained from (5.22) by fixing $N = 2$. It can be seen that the presence of multiple competing OPs can provide up to almost two times higher UE utility when OP deployment densities are small.

Another question of importance is how efficient the market is in terms of the UE game value. In particular, how much does u^*_{UE} resulting from the competition among OPs differs from the cooperative setup where the OP strategies are *jointly* configured

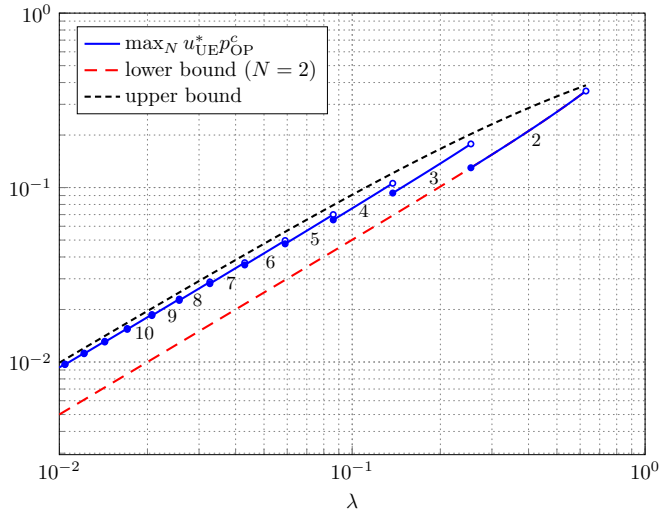


Figure 5.6.: Normalized UE game value under a common density λ and a common normalized bandwith leasing price p^c_{OP} for all OPs. Numbers correspond to the optimal N for each partition of the λ range.

to maximize u^*_{UE} instead of their own utility (under the constraint that the latter is positive). For the case where all involved OPs have the same deployment density and normalized bandwidth leasing price as in the setup considered in Lemma 5.12, the cooperative u^*_{UE} can be upper bounded by a simple expression.

Proposition 5.13. *The UE game value for which $u_{\mathrm{OP},n} > 0$, with $\lambda_n = \lambda$ and $p^c_{\mathrm{OP},n} = p^c_{\mathrm{OP}}$, for all $n \in \{1, 2, \ldots, N\}$, is upper bounded as*

$$u^*_{\mathrm{UE}} < \frac{\lambda}{p^c_{\mathrm{OP}}\,(1+\lambda)}. \tag{5.23}$$

Proof. Clearly, the UE game value is maximized under some OP strategy profile resulting from all OPs cooperating towards this goal without any self interest for maximizing their own utility function as long as it is positive. Denoting this strategy profile as $\{\tilde{p}'_n\}_{n=1}^N$ and the corresponding UE distribution among OPs as $\{q'_n\}_{n=1}^N$ at the UE game eq., the UE game value can be bounded, based on the expression of (5.11), as

$$\begin{aligned} u^*_{\mathrm{UE}} &< \frac{\lambda \sum_{n=1}^N \tilde{p}'_n}{1 + \frac{1}{2} N \lambda} \\ &< \frac{\lambda \sum_{n=1}^N q'_n}{p^c_{\mathrm{OP}} \left(1 + \frac{1}{2} N \lambda\right)} \end{aligned}$$

where the last inequality follows by noting from (5.12) that it must hold $\tilde{p}'_n < q'_n/p^c_{\mathrm{OP}}$ in order to have $u_{\mathrm{OP},n} > 0$, for all $n \in \{1, 2, \ldots, N\}$. Equation (5.23) follows since $\sum_{n=1}^{N} q'_n = 1$ and noting that $N = 2$ maximizes the bound. □

The upper bound of (5.23) is also shown in Fig. 5.6. It can be seen that, with the optimal number of competing OPs, the resulting UE game value is close to the upper bound, i.e., the market efficiency is high, especially for deployment densities where multiple OPs are involved.

5.8. Conclusion

This chapter considered the implications that competition among small cell wireless OPs with large scale deployments has on the quality of services provided by the future network. An appropriate mathematical model was employed in order to accurately model the effects of different infrastructure densities among competing networks and randomness of AP positions on UE rates. Modeling the interactions among OPs and UEs as a two stage game and employing game-theoretic analysis, the system (market) equilibrium properties were investigated, providing insights on how parameters such as number of OPs, deployment densities, spectral efficiencies and bandwidth leasing costs contribute in the utilities experienced by UEs and OPs. For special cases of practical interest, a (semi-) analytic investigation of the market equilibrium was provided, revealing that extreme infrastructure densification levels may not lead to a stable market under the considered competition model.

6. Conclusion

6.1. Contributions

This thesis investigated various important aspects of future cellular networking, in particular, performance of conventional downlink transmissions, operational region of D2D communications, cross-mode inter-cell interference statistics, and multi-operator competition. In all cases, a stochastic geometry framework was employed in order to take into account and obtain insights on the effect of (extreme) system densification, the most important new attribute of the future network.

The results of the studies performed in this thesis provide insights related to the fundamental questions regarding future network operation posed in sec. 1.3. In particular, the main insights obtained are the following:

1. Even though extreme infrastructure densification has severe consequences on the interference properties and intra-cell resource allocation due to the resulting irregular AP deployment, it is still able to provide significant benefits in terms of downlink user rate. In particular, under a well-selected intra-cell scheduling scheme, the user rate (statistically) improves as λ_a/λ_u (ratio of AP-to-users densities) increases. Considering the maximum downlink user rate, r_0, corresponding to a given rate-outage probability threshold, it was shown that r_0 increases linearly with λ_a/λ_u in the regime $\lambda_a/\lambda_u \ll 1$, whereas it increases logarithmically with λ_a/λ_u in the regime $\lambda_a/\lambda_u \geq 1$. This effectively means that in the ultra-dense regime where the number of APs are comparable with the number of UEs, a further increase of the AP density provides only marginal benefits, rendering it an impractical approach for enhancing performance when issues such installation/maintenance costs are taken into account. This suggests that densification alone is not able to provide unbounded performance benefits, i.e., it should be complemented with spectrum increase and/or advanced transmission techniques.

2. D2D communications, even though motivated by the concept of proximal communications between users, can actually be beneficial to the *overall* network performance when D2D links distances of the order of user-to-closest-AP distance are allowed. However, in order to achieve the corresponding performance gains an appropriate mode selection and/or D2D channel access scheme is required. These observations were made clear by the explicit identification of the so-called operational region of D2D communications under various options

of mode selection, channel access, and cellular/D2D spectrum sharing. The optimal scheme, both in terms of maximum operational region as well as average user rate, utilizes a mode selection procedure based on the knowledge of the distance of potential D2D links with underlay D2D transmissions (cellular and D2D links operate on the same bandwidth). It was demonstrated that a well-selected mode selection or D2D channel access scheme can cope with the increase of (potential) D2D users density, despite the corresponding increase of interference, providing improved system performance compared to a non-D2D-enabled network. These results strongly motivate the introduction of D2D communications to the cellular network as a means to achieve technical benefits as well as providing new market opportunities related to the development of D2D-based applications.

3. The randomness of the cellular (Voronoi) cell shape due to the random AP positions has a significant impact on the cross-mode inter-cell interference properties, especially in the case when interference is generated by mobile users. A intuitive characterization of the interference properties can be obtained by means of the so called equivalent interferer density, which is available in closed form for the case of AP-generated interference and can be easily computed numerically for the case of user-generated interference. Analysis shows that a large distance between a mobile receiver and its closest AP implies (a) reduced experienced AP-generated interference due to the increased effective guard region and (b) increased experienced user-generated interference due to the receiver most likely located near the cell edge. The analysis provides the means to efficiently compute metrics of interest such as SIR coverage under various operational scenarios relevant to the future network, ranging from conventional uplink to cross-cell D2D transmissions.

4. Under a simple model for the techno-economic interactions among competing operators and self-interested users, it was shown that there exist conditions, characterized by the density, technology, and service pricing of the competing subnetworks, where multiple operators can achieve profit. In particular, for the case of competition between two operators, the so called duopoly region was identified, consisting of the set of density pairs where operators can choose strategies (service prices) that will result in a profitable operation for both. As expected, this scenario is possible only when the difference between densities is sufficiently small. Interestingly, for the case when both operators have networks of exactly the same characteristics, operation with an arbitrarily large AP density does not lead to stable market conditions. This implies that ultra-dense networking, although desirable from an engineering perspective may require a different market model than the one considered in this study, e.g., based on (partial) cooperation among operators. In terms of user experience, competition among multiple operators with similar subnetwork characteristics is the optimal scenario, resulting in a user performance close to the utopia scenario where all operators cooperate for maximum user satisfaction.

This thesis has contributed towards understanding the impact of (extreme) densification on future cellular networking, both in terms of users as well as infrastructure. The main conclusion is that densification, although resulting in a fundamentally different interference landscape than the one encountered in traditional cellular engineering, can be exploited in order to obtain a better cellular network in terms of the quality of services offered to the end users. However, maximum densification gains can be achieved only via sophisticated network operation with respect to resource allocation, as the excessive interference levels due to the multitude parallel transmissions from both APs and users can easily result in network failure if not properly controlled.

6.2. Future Work

The topic of system level aspects in cellular networking is without a doubt a vast one, with numerous topics worth of investigating, especially when one is interested in future operational scenarios. Regarding the particular topics considered in this thesis, the following comments can be made on future research directions.

- *Infrastructure densification*: During the period of this thesis, the topic of densification was addressed by numerous publications in the stochastic geometry literature to the extent that the topic could be considered as mature and well understood. However, all these works have not considered critical aspects of the system model, such as OFDM(A)-based transmissions and resource allocation. This is due to the inherent limitation of stochastic geometry tools to accommodate for correlated decisions among APs and/or users. An effort is required to overcome this limitation. Another important aspect of cellular networking that has not been sufficiently addressed by stochastic geometry works is cooperation/coordination among APs.
- *D2D communications*: The work in this thesis considered the operation region of D2D communications when the latter are performed in-band with the cellular *downlink*. It would be interesting to investigate the case when the cellular *uplink* is considered instead and perform a comparison. One expects that the optimal solution would be to incorporate D2D communications as in-band to both uplink and downlink. Another important issue that has not been addressed in the literature is that of bidirectional communication between D2D users. Although it is expected that D2D links will in many cases be bidirectional, only single direction D2D links have been considered so far for tractability reasons. Of course, bidirectional communications also appear in conventional user-to-AP links, which is a topic that has also received almost minimal attention in the literature.
- *Cross-mode interference*: The cross-mode interference characterization provided in this thesis can be used as an analytical framework for designing

sophisticated resource allocation algorithms, coordinating concurrent uplink, downlink and D2D transmissions over the same spectrum. In addition, the cross-mode interference characterization can be used to determine the effects of novel transmission/communications concepts that are based on D2D links, beyond the standard proximal communication scenario, such as user-assisted relaying and caching.

- *Techno-economic properties of the network*: The results of this thesis indicate that extreme infrastructure densification, although desirable from an engineering perspective, may not lead to stable market conditions due to operator competition. It is interesting to investigate the techno-economic properties of the network under a cooperative management model among operators using tools from, e.g., cooperative game theory. Also, the system model could be improved by considering a more realistic model for the spectral efficiency (e.g., taking into account distance from serving AP) and more sophisticated utility functions for both UEs and APs.

A. Proofs of Chapter 2

A.1. Proof of Lemma 2.4

Activity probability for TDMA is obtained by simply noting that

$$p = \mathbb{P}(\text{transmission on SC 1}|K > 0)\mathbb{P}(K > 0).$$

For the case of FDMA/TDMA, consider a random AP associated with K indexed UEs and let $p(K)$ denote the probability of assigning at least one UE on SC 1. Clearly, $p(K) = 1$ for $K \geq N$ and $p(K) = 0$ for $K = 0$. For the case $0 < K < N$, define the mutually exclusive events $\mathcal{B}_m \triangleq \{m\text{-th UE is assigned SC } 1\}, m = 1, 2, \ldots, K$. It is easy to see that

$$\mathbb{P}(\mathcal{B}_m) = \frac{1}{N-(m-1)} \prod_{r=1}^{m-1} \left(1 - \frac{1}{N-(r-1)}\right), \tag{A.1}$$

and

$$p(K) = \sum_{m=1}^{K} \mathbb{P}(\mathcal{B}_m) = K/N, 0 < K < N, \tag{A.2}$$

where the last equality follows by simple algebra. Averaging $p(K)$ over K results in the form of (2.6).

A.2. Proof of Proposition 2.7

Denoting the SIR outage event $\{\text{SIR} < \theta_0\}$ and its complement, $\{\text{SIR} \geq \theta_0\}$, as $\mathcal{O}$ and $\overline{\mathcal{O}}$, respectively, $F_R(r|L)$ can be written as

$$F_R(r|L) = F_{\text{SIR}}(\theta_0)F_R(r|L, \mathcal{O}) + (1 - F_{\text{SIR}}(\theta_0))F_R(r|L, \overline{\mathcal{O}}). \tag{A.3}$$

From (2.5), $R = 0$ conditioned on $\mathcal{O}$, therefore,

$$F_R(r|L, \mathcal{O}) = 1, \forall r, L, \tag{A.4}$$

whereas, conditioned on $\overline{\mathcal{O}}$,

$$
\begin{aligned}
F_R(r|L,\overline{\mathcal{O}}) &= \mathbb{P}(\bar{R}/(L+1) < r|\bar{\mathcal{O}}) \\
&= F_{\mathrm{SIR}}(\tilde{\theta}|\mathrm{SIR} \geq \theta_0) \\
&= \begin{cases} \frac{F_{\mathrm{SIR}}(\tilde{\theta})-F_{\mathrm{SIR}}(\theta_0)}{1-F_{\mathrm{SIR}}(\theta_0)} & ,\tilde{\theta} \geq \theta_0, \\ 0 & ,\tilde{\theta} < \theta_0, \end{cases}
\end{aligned} \tag{A.5}
$$

where $\tilde{\theta} \triangleq 2^{rN(L+1)}-1$ and the last equality follows from basic probability theory and the continuity of $F_{\mathrm{SIR}}(\theta)$. Combining (A.3)–(A.5) leads to (2.10) and application of the total probability theorem gives (2.9).

B. Proofs of Chapter 3

B.1. Proof of Lemma 3.3

Consider the underlay D2D case as the overlay case follows without taking into account the effect of D2D-generated interference. Derivation follows the same steps as in [40] and is briefly described here as there are a few details that are different due to system model differences. Let $I_{\text{cell}} \triangleq \sum_{y\in\tilde{\Phi}_a\setminus y_0} P_a g_y |y|^{-a}$, $I_{\text{D2D}} \triangleq \sum_{y\in\tilde{\Phi}_d} |y - x_y|^{\alpha} g_y |y|^{-a}$ with all quantities as defined in Eq. (3.1). It follows from Eq. (3.1) and the independence of I_{cell}, I_{D2D}, that $\mathbb{P}(\text{SIR} \geq \theta) = \mathbb{E}\left[\mathcal{L}_{I_{\text{cell}}}(\theta|y_0|^{\alpha}/P_a)\mathcal{L}_{I_{\text{D2D}}}(\theta|y_0|^{\alpha}/P_a)\right]$, where $\mathcal{L}_x(s)$ denotes the Laplace transform of variable x and the expectation is over $|y_0|$, which is Rayleigh distributed with mean $1/(2\sqrt{\lambda_a})$ [29]. Approximating $\tilde{\Phi}_a$ as generated from a thinning of Φ_a with retention probability $\mathbb{P}(K > 0)$ for each $x \in \Phi_a$ [24], it follows that [29]

$$\mathcal{L}_{I_{\text{cell}}}\left(\theta|y_0|^{\alpha}/P_a\right) \approx \exp\left(-\pi\mathbb{P}(K>0)\lambda_a\rho(\theta)|y_0|^2\right).$$

It is easy to see, based on the system model and fundamental properties of HPPPs [45, 18], that $\tilde{\Phi}_d$ forms a HPPP of density $qp\lambda_d$ with each node transmitting with an independent, identically distributed (i.i.d.) power P_d, equal to r_d^{α} and $r_d^{\alpha}\mathbb{I}(r_d \leq r_{d,\text{th}})$ under probabilistic and distance-based mode selection, respectively. Starting from the Laplace transform of the interference power generated by an HPPP with i.i.d. node powers[45],

$$\begin{aligned}\mathcal{L}_{I_{\text{D2D}}}\left(\frac{\theta|y_0|^{\alpha}}{P_a}\right) &= \exp\left(-qp\lambda_d\kappa\pi\mathbb{E}(P_d^{2/\alpha})\left(\frac{\theta}{P_a}\right)^{\frac{2}{\alpha}}|y_0|^2\right)\\ &= \exp\left(-\frac{1}{2}qp^{\gamma}\lambda_d\kappa\pi r_{d,\max}^2\left(\frac{\theta}{P_a}\right)^{\frac{2}{\alpha}}|y_0|^2\right),\end{aligned}$$

with $\gamma = 1, 2$ for probabilistic and distance-based mode selection, respectively, where the second equality follows by noting that $\mathbb{E}(r_d^2) = (1/2)r_{d,\max}^2$ and $\mathbb{E}(r_d^2\mathbb{I}(r_d \leq r_{d,\text{th}})) = (1/2)r_{d,\text{th}}^2 = (1/2)pr_{d,\max}^2$. The result of (3.2) then follows by evaluating the expectation over $|y_0|$.

B.2. Proof of Lemma 3.6

Normalize w.l.o.g. the unit area so that the density of APs becomes $\lambda_a' = 1$. By the properties of HPPPs [18], the positions of UEs employing cellular transmissions

constitute an HPPP of density $\lambda_c' \triangleq (\lambda_c + (1-p)\lambda_d)/\lambda_a$. Let $f_Y(y)$, $f_X(x)$ denote the probability density functions of the area Y of the cell containing the origin and the area X of any other randomly selected cell, respectively. It can be shown that $f_Y(y) = y f_X(y)$ [92]. Then,

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{K_0+1}\right) &= \sum_{k=0}^{\infty} \frac{1}{k+1}\mathbb{P}(K_0=k) \\
&\overset{(a)}{=} \sum_{k=0}^{\infty} \frac{1}{k+1}\int_0^{\infty} f_Y(y)\frac{(\lambda_c' y)^k}{k!}e^{-\lambda_c' y}dy \\
&\overset{(b)}{=} \frac{1}{\lambda_c'}\int_0^{\infty} f_X(y)e^{-\lambda_c' y}\left(\sum_{k=0}^{\infty}\frac{(\lambda_c' y)^{k+1}}{(k+1)!}\right)dy \\
&= \frac{1}{\lambda_c'}\int_0^{\infty} f_X(y)e^{-\lambda_c' y}\left(e^{\lambda_c' y}-1\right)dy \\
&= \frac{1}{\lambda_c'}\left(1-\int_0^{\infty} f_X(y)e^{-\lambda_c' y}\right)dy,
\end{aligned}$$

where (a) is due to K_0 being Poisson distributed with mean $\lambda_c' y$ given $Y = y$, and (b) follows by interchanging integral and summation (Fubini's theorem). Noting that the last integral equals $\mathbb{P}(K=0)$ completes the proof.

B.3. Proof of Proposition 3.14

Consider the overlay D2D case. Starting from f provided by Scheme 1, the following inequalities hold for any operational point $(c_1, c_2) \in \mathbb{R}^{+2}$ (not necessarily included in the operational region of a scheme).

$$\begin{aligned}
f(1,1) &= c_1 e^{-c_2} - 1 \\
&\leq \max_{p\in(0,1]}\{c_1 p^2 e^{-c_2 p} - p\} &\text{(B.1)}\\
&= \max_{p\in(0,1]}\{p(c_1 p e^{-c_2 p} - 1)\} \\
&\leq \max_{p\in(0,1]}\{c_1 p e^{-c_2 p} - 1\} &\text{(B.2)}\\
&= \max_{p\in(0,1]}\{c_1 p^2 e^{-c_2 p^2} - 1\} \\
&\leq \max_{p\in(0,1]}\{c_1 p^2 e^{-c_2 p^2} - p\}. &\text{(B.3)}
\end{aligned}$$

Noting that the right hand side of Eqs. (B.1), (B.2), (B.3) correspond to the maximum f provided by Schemes 3-p, 2, and 3-d, respectively, completes the proof. The orderings for the underlay case and the superiority of the underlay version of Scheme 3-d over its overlay version are shown similarly.

B.4. Proof of Proposition 3.17

Consider the overlay D2D case. For the operational points where $p^* < 1$, it must hold $\frac{\partial f(p,1)}{\partial p}|_{p=p^*} = 0$, which is equivalent to

$$xe^{-x}(2-x) = \frac{c_2}{c1}, \tag{B.4}$$

where $x \triangleq p^* c_2 < c_2$. Since the right hand side of (B.4) is positive it must hold $x < \min\{2, c_2\}$. Assuming this is the case, $f(p^*, 1) = f(x/c_2, 1) = \frac{x(x-1)}{c_2(2-x)}$ which requires $x > 1$ in order to have a positive value, resulting in $x \in (1, \min\{2, c_2\})$, which in turn requires $c_2 > 1$. Assuming the latter holds, and noting that $xe^{-x}(2-x)$ is a strictly decreasing function of x in $1 < x < 2$, it is easy to see that a unique solution of Eq. (B.4) exists as long as $c_2/c_1 > 1/e$, for $c_2 \geq 2$ and $1/e < c_2/c_1 < c_2 e^{-c_2}(2-c_2)$ for $1 < c_2 < 2$, which provide the description for $\hat{\mathcal{R}}^3_{\text{D2D}}$. Derivation of the result for the underlay case is similar.

B.5. Proof of Proposition 3.19

Clearly, $\mathcal{R}^{\text{3-d}}_{\text{D2D}} = \mathcal{R}^1_{\text{D2D}} \cup \hat{\mathcal{R}}^{\text{3-d}}_{\text{D2D}}$, where $\hat{\mathcal{R}}^{\text{3-d}}_{\text{D2D}}$ is the set of operational points for which $p^* < 1$ and $\mathcal{R}^1_{\text{D2D}}$ as in Lemma 5. Noting that the maximum f for underlay Scheme 3-d can be bounded as $\max_{p\in(0,1)}\{\bar{c}_1 p e^{-(\bar{c}_2+\bar{c}_3)p} - p\} \leq \max_{p\in(0,1)} f(p,1) \leq \max_{p\in(0,1)}\{\bar{c}_1 p e^{-(\bar{c}_2+\bar{c}_3)p^2} - p\}$, it follows that $\hat{\mathcal{R}}^{\text{3-d},(1)}_{\text{D2D}} \subseteq \hat{\mathcal{R}}^{\text{3-d}}_{\text{D2D}} \subseteq \hat{\mathcal{R}}^{\text{3-d},(2)}_{\text{D2D}}$, with $\hat{\mathcal{R}}^{\text{3-d},(\beta)}_{\text{D2D}}$, $\beta = 1, 2$, denoting the sets of operational points where the lower and upper bounds are positive, respectively. The bounding sets can be obtained as in Eq. (3.22) using a similar procedure as previous derivations and it is easy to see that $\mathcal{R}^1_{\text{D2D}} \cup \hat{\mathcal{R}}^{\text{3-d},(\beta)}_{\text{D2D}} = \{\bar{c}_1 > 1\}$ for $\beta = 1, 2$.

C. Proofs of Chapter 4

C.1. Proof of Proposition 4.2

Let $\tilde{\Phi}_a \triangleq \Phi_a \setminus \{x^*\}$ denote the interfering AP point process. Given x^*, $\tilde{\Phi}_a$ is an inhomogeneous PPP of density $\tilde{\lambda}_a(x), x \in \mathbb{R}^2$, defined in (4.4). Density $\tilde{\lambda}_a(x)$ is circularly symmetric w.r.t. $x_R = o$, which, using Lemma 4.1, directly leads to the Laplace transform expression of (4.3) with radial density as in (4.6). Considering the case $x_R \neq o$, it directly follows from Lemma 4.1 and (4.2) that

$$\mathcal{L}_{I_{x_R,a}}(s) = \exp\left\{-\int_{\mathbb{R}^2} \tilde{\lambda}_a(x + x_R) g(sP_a|x|^{-\alpha_a}) dx\right\}$$

by a change of integration variable. By switching to polar coordinates for the integration, the right-hand side of (4.3) results with $P = P_a$, $\alpha = \alpha_a$ and $\lambda_{x_R}(r) = \lambda_{x_R,a}(r)$, where

$$\begin{aligned}
\lambda_{x_R,a}(r) &= \frac{1}{2\pi} \int_0^{2\pi} \tilde{\lambda}_a((r,\theta) + x_R) d\theta \\
&= \frac{\lambda_a}{2\pi} \int_0^{2\pi} \mathbb{I}\left\{(r,\theta) + x_R \notin \mathcal{B}(o, |x^*|)\right\} d\theta \\
&= \frac{\lambda_a}{2\pi} \int_0^{2\pi} \mathbb{I}\left\{(r,\theta) \notin \mathcal{B}(-x_R, |x^*|)\right\} d\theta \\
&= \frac{\lambda_a}{2\pi} \int_0^{2\pi} \mathbb{I}\left\{(r,\theta) \notin \mathcal{B}((|x_R|, 0), |x^*|)\right\} d\theta.
\end{aligned}$$

The last equality follows by noting that the integral does not depend on the value of $\angle x_R$. Evaluation of the final integral is essentially the evaluation of an angle (see Fig. C.1), which can be easily obtained by elementary Euclidean geometry, resulting in the expression of (4.5).

C.2. Proof of Proposition 4.5

It holds

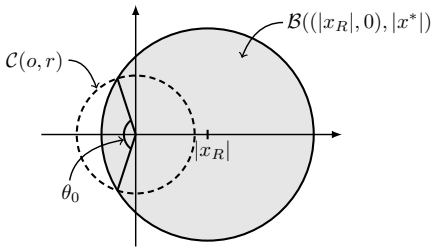


Figure C.1.: Geometrical figure depicting the angle θ_0, which is equal to quantity $\int_0^{2\pi} \mathbb{I}\left((r,\theta) \notin \mathcal{B}((|x_R|,0),|x^*|)\right) d\theta$ appearing in the proof of Prop. 4.2.

$$\begin{aligned}\mathcal{L}_{I_{x_R,u}}(s \mid x^*) &= \mathbb{E}\left(\mathbb{E}\left(e^{-sI_{x_R,u}} \mid \Phi_a\right) \mid x^*\right) \\ &\overset{(a)}{=} \mathbb{E}\left(\exp\left\{-\int_{\mathbb{R}^2} \tilde{\lambda}_u(x|\Phi_a) g(sP_u|x-x_R|^{-\alpha_u}) dx\right\} \middle| x^*\right) \\ &\geq \exp\left\{-\int_{\mathbb{R}^2} \mathbb{E}\left(\tilde{\lambda}_u(x+x_R|\Phi_a) \mid x^*\right) g(sP_u|x|^{-\alpha_u}) dx\right\},\end{aligned}$$

where (a) follows by noting that, conditioned on Φ_a, the interfering UE point process equals $\Phi_u \setminus \mathcal{V}^*$, which is distributed as an inhomogeneous PPP of density $\tilde{\lambda}_u(x|\Phi_1), x \in \mathbb{R}^2$ as given in (4.8), and using Lemma 4.1. The inequality is an application of Jensen's inequality with a change of integration variable. Result follows by noting that $\mathbb{E}\left(\tilde{\lambda}_u(x|\Phi_a) \mid x^*\right) = \lambda_u\left[1 - p_c(x \mid x^*)\right]$ and switching to polar coordinates for the integration.

C.3. Proof of Proposition 4.6

Replacing the PCA function in (4.11) with its lower bound as per Cor. 4.4 results in the equivalent density bound

$$\begin{aligned}&\lambda_{x_R,u}(r) \\ \leq&\lambda_u\left(1 - \frac{1}{2\pi}\int_0^{2\pi} e^{-\lambda_a\pi|(r,\theta)+x_R-x^*|^2} d\theta\right) \\ \overset{(a)}{\leq}&\lambda_u\left(1 - \frac{e^{-\lambda_a\pi|x^*|^2}}{2\pi}\int_0^{2\pi} e^{-\lambda_a\pi|(r,\theta)+x_R|^2} d\theta\right) \\ \overset{(b)}{=}&\lambda_u\left(1 - \frac{e^{-\lambda_a\pi|x^*|^2}}{2\pi}\int_0^{2\pi} e^{-\lambda_a\pi|(r,\theta)+(|x_R|,0)|^2} d\theta\right) \\ \overset{(c)}{=}&\lambda_u\left(1 - \frac{e^{-\lambda_a\pi(|x^*|^2+|x_R|^2+r^2)}}{2\pi}\int_0^{2\pi} e^{\lambda_a\pi 2r|x_R|\cos\theta} d\theta\right),\end{aligned}$$

where (a) follows by application of the triangle inequality, (b) by noting that the integral is independent of $\angle x_R$ and (c) since $|(r,\theta) + (|x_R|, 0)|^2 = r^2 + |x_R|^2 - 2r|x_R|\cos(\pi - \theta)$ (cosine law). The integral in the last equation is equal to $2\pi\mathcal{I}_0(\lambda_a 2\pi|x_R|r)$. By exchanging the roles x^* and x_R in (a) and following the same reasoning, another upper bound of the form of (c) results, with the only difference that the integrand has $|x^*|$ instead of $|x_R|$ and evaluates to $2\pi\mathcal{I}_0(\lambda_a 2\pi|x^*|r)$. Considering the minimum of these two bounds leads to the expression of (4.13).

D. Proofs of Chapter 5

D.1. Proof of Lemma 5.2

Applying the expectation operator to both sides of (5.1), it follows that for any $x_0 \in \mathbb{R}^2$,

$$\begin{aligned}\mathbb{E}(r_n(x_0)) &= c_n w \mathbb{E}\left(\frac{1}{1+K_{0,n}}\middle| x_0 \in \Phi_{u,n}\right) \\ &\overset{(a)}{=} c_n w \frac{\lambda_n}{q_n}\mathbb{P}\left(K_n > 0\right), \end{aligned} \tag{D.1}$$

where (a) is a direct application of [91, Lemma 3] with K_n denoting the number of subscribers associated with any random AP of OP n. A closed form expression of $\mathbb{P}(K_n > 0)$ is not available, whereas the accurate approximate expression derived in [92] does not allow for analytical manipulations. Therefore, a simpler approximate expression is pursued as follows.

$$\begin{aligned}\mathbb{P}(K_n > 0) &= 1 - \mathbb{P}(K_n = 0) \\ &= 1 - \mathbb{E}\left(\mathbb{P}\left(K_n = 0 | A_n\right)\right) \\ &\overset{(b)}{=} 1 - \mathbb{E}\left(e^{-q_n A_n}\right) \\ &\approx 1 - e^{-q_n \mathbb{E}(A_n)} \\ &\overset{(c)}{=} 1 - e^{-q_n/\lambda_n}, \end{aligned} \tag{D.2}$$

where $A_n \in \mathbb{R}^+$ denotes the area of the (Voronoi) cell of a random AP of OP n, (b) follows from the void probability of the HPPP [18, Theorem 1.1.5] and (c) by noting that the average area of a Voronoi cell of an HPPP of density λ equals $1/\lambda$ [18, Corollary 4.2.3]. Equation (5.2) follows by plugging (D.2) into (D.1) and employing the approximation $\frac{\lambda_n}{q_n}\left(1 - e^{-q_n/\lambda_n}\right) \approx 1 - e^{-\lambda_n/q_n}$.

D.2. Proof of Theorem 5.7

Assume that an equilibrium exists such that the UE distribution between OPs is non-degenerate, i.e., $q_1^* = 1 - q_2^* \in (0, 1)$. This distribution can be obtained, according

to (5.6c), as the solution of

$$f \triangleq \tilde{p}^*_{\mathrm{UE},1}\lambda_1^{\beta}\left(1-e^{-\frac{\lambda_1}{q_1^*}}\right)-\tilde{p}^*_{\mathrm{UE},2}\lambda_2^{\beta}\left(1-e^{-\frac{\lambda_2}{1-q_1^*}}\right)=0. \tag{D.3}$$

It is easy to see that (D.3) can only have a unique solution $q_1^* \in (0,1)$ if and only if $(\tilde{p}^*_{\mathrm{UE},1},\tilde{p}^*_{\mathrm{UE},2}) \in \mathcal{S}_d \triangleq \{(\tilde{p}_{\mathrm{UE},1},\tilde{p}_{\mathrm{UE},2}) \in \mathbb{R}^+\times\mathbb{R}^+ : 1-e^{-\lambda_2} < \frac{\tilde{p}_{\mathrm{UE},1}\lambda_1^{\beta}}{\tilde{p}_{\mathrm{UE},2}\lambda_2^{\beta}} < \frac{1}{1-e^{-\lambda_1}}\}$. Direct application of the implicit function theorem [96] shows that q_1^* is a continuously differentiable function on some neighborhood of any $(\tilde{p}^*_{\mathrm{UE},1},\tilde{p}^*_{\mathrm{UE},2}) \in \mathcal{S}_d$ for which (D.3) holds, with continuous partial derivatives equal to

$$\frac{\partial q_1^*}{\partial \tilde{p}^*_{\mathrm{UE},n}} = -\frac{\partial f}{\partial \tilde{p}^*_{\mathrm{UE},n}}\Big/\frac{\partial f}{\partial q_1^*}, n=1,2. \tag{D.4}$$

This, in turn, implies that the OP utility function (5.12) is also continuously differentiable w.r.t. $(\tilde{p}^*_{\mathrm{UE},1},\tilde{p}^*_{\mathrm{UE},2}) \in \mathcal{S}_d$ for which (D.3) holds, which allows for the incorporation of the first order necessary conditions [97] for solving the OP game eq. optimization problem (5.13), namely,

$$\left\{\frac{\partial u_{\mathrm{OP},n}(\tilde{p}^*_{\mathrm{UE},1},\tilde{p}^*_{\mathrm{UE},2})}{\partial \tilde{p}^*_{\mathrm{UE},n}} = 0 \overset{(5.12)}{\Rightarrow} \frac{\partial q_n^*}{\partial \tilde{p}^*_{\mathrm{UE},n}} = p^c_{\mathrm{OP},n}\right\}_{n=1,2}. \tag{D.5}$$

Solving the the first order conditions for $n=1$ w.r.t. $\tilde{p}^*_{\mathrm{UE},1}$ and treating $\tilde{p}^*_{\mathrm{UE},2}$ and q_1^* as independent variables gives

$$\tilde{p}^*_{\mathrm{UE},1} = \frac{\left(e^{\frac{\lambda_1}{q_1^*}}-1\right)\lambda_1^{\beta} - p^c_{\mathrm{OP},1}\tilde{p}^*_{\mathrm{UE},2}\lambda_2^{1+\beta}\frac{e^{\frac{\lambda_1}{q_1^*}-\frac{\lambda_2}{q_2^*}}}{q_2^{*2}}}{\lambda_1^{1+\beta}p^c_{\mathrm{OP},1}/q_1^{*2}} \tag{D.6}$$

In addition, by incorporating (D.4) into (D.5), (D.5) can be written as

$$\left\{(-1)^n\frac{\partial f}{\partial \tilde{p}^*_{\mathrm{UE},n}} = p^c_{\mathrm{OP},n}\frac{\partial f}{\partial q_1^*}\right\}_{n=1,2}. \tag{D.7}$$

Equation (5.17) follows after dividing equations (D.7) by parts, from which it follows that the market share distribution $\{q_n^*\}_{n=1,2}$ at the OP game eq. exists and is unique, if and only if (5.14) holds. Comparing (5.17) with (D.3) results in (5.16) and substituting the latter into (D.6) results in (5.15).

In order to prove the uniqueness of the duopoly equilibrium, first note from (D.3) that since q_1^* is unique, the OP eq. strategy profile $(\tilde{p}^*_{\mathrm{UE},1}, \tilde{p}^*_{\mathrm{UE},2})$ is also unique. This means that $\tilde{p}^*_{\mathrm{UE},1}$ is the unique best response to $\tilde{p}^*_{\mathrm{UE},2}$ out of all possible $\tilde{p}_{\mathrm{UE},1}$ such that $(\tilde{p}_{\mathrm{UE},1}, \tilde{p}^*_{\mathrm{UE},2}) \in \mathcal{S}_d$ but does not exclude the possibility that OP 1 may have the incentive to choose a strategy $\tilde{p}_{\mathrm{UE},1}$ such that $(\tilde{p}_{\mathrm{UE},1}, \tilde{p}^*_{\mathrm{UE},2}) \notin \mathcal{S}_d$, i.e., establish monopoly. This would be the case if $u_{\mathrm{OP},1}(\frac{\tilde{p}^*_{\mathrm{UE},2}(\lambda_2/\lambda_1)^\beta}{1-e^{-\lambda_1}}, \tilde{p}^*_{\mathrm{UE},2}) > u_{\mathrm{OP},1}(\tilde{p}^*_{\mathrm{UE},1}, \tilde{p}^*_{\mathrm{UE},2})$, where $\frac{\tilde{p}^*_{\mathrm{UE},2}(\lambda_2/\lambda_1)^\beta}{1-e^{-\lambda_1}}$ is the minimum inverse price set by OP 1 for which $(\tilde{p}_{\mathrm{UE},1}, \tilde{p}^*_{\mathrm{UE},2}) \notin \mathcal{S}_d$. However, by the continuity of the OP utility function (see Fig. 5.3), this would imply that there exists a $\tilde{p}_{\mathrm{UE},1} \neq \tilde{p}^*_{\mathrm{UE},1}$ in the neighborhood of $\frac{\tilde{p}^*_{\mathrm{UE},2}(\lambda_2/\lambda_1)^\beta}{1-e^{-\lambda_1}}$ such that $(\tilde{p}_{\mathrm{UE},1}, \tilde{p}^*_{\mathrm{UE},2}) \in \mathcal{S}_d$ and $u_{\mathrm{OP},1}(\tilde{p}_{\mathrm{UE},1}, \tilde{p}^*_{\mathrm{UE},2}) > u_{\mathrm{OP},1}(\tilde{p}^*_{\mathrm{UE},1}, \tilde{p}^*_{\mathrm{UE},2})$, which is a contradiction. A similar argument w.r.t. OP 2 completes the proof.

D.3. Proof of Theorem 5.11

Initially, the differentiability of u^*_{UE} w.r.t. $\{\lambda_n\}_{n=1}^N$ will be shown. To this end, let $\{\tilde{p}^*_{\mathrm{UE},n} > 0\}_{n=1}^N$ denote the strategy profile of the OPs at the OP game equilibrium. When the UE game eq. has been established, it follows from (5.6b), (5.6c) that the UE distribution $\{q_n^* > 0\}_{n=1}^N$ and the UE game value u^*_{UE} must be a solution of a non-linear homogeneous system of $N+1$ equations $\{f_n = 0\}_{n=1}^{N+1}$, where $f_n \triangleq \tilde{p}^*_{\mathrm{UE},n}\left(1 - e^{-\lambda_n/q_n^*}\right) - u^*_{\mathrm{UE}}$, for $n = 1, 2, \ldots, N$, and $f_{N+1} \triangleq \sum_{n=1}^N q_n^* - 1$. Denoting by $\frac{\partial\{g_1,\ldots,g_K\}}{\partial\{x_1,\ldots,x_K\}}$ the determinant of the square matrix with (i,j) element $\frac{\partial g_i}{\partial x_j}$, $i, j \in \{1, 2, \ldots, K\}$ (i.e., the Jacobian of $\{g_1, \ldots, g_K\}$) and noting by direct computation that $\frac{\partial\{f_1,\ldots,f_{N+1}\}}{\partial\{q_1^*,\ldots,q_N^*,u^*_{\mathrm{UE}}\}} \neq 0$, it follows from the implicit function theorem that u^*_{UE} is continuously differentiable w.r.t. $\{\lambda_n\}_{n=1}^N$ and it holds

$$\frac{\partial u^*_{\mathrm{UE}}}{\partial \lambda_n} = -\frac{\partial\{f_1, \ldots, f_{N+1}\}}{\partial\{q_1^*, \ldots, q_N^*, \lambda_n\}} \Big/ \frac{\partial\{f_1, \ldots, f_{N+1}\}}{\partial\{q_1^*, \ldots, q_N^*, u^*_{\mathrm{UE}}\}}, \tag{D.8}$$

for $n \in \{1, 2, \ldots, N\}$.

With the differentiability of u^*_{UE} proven, the first order necessary optimallity conditions can now be applied. In particular, let $* \triangleq u^*_{\mathrm{UE}} - \mu\left(\sum_{n=1}^N \lambda_n^* - \lambda_{\mathrm{tot}}\right)$ denote the Lagrangian of the constrained optimization problem, where $\mu \in \mathbb{R}$ is a Lagrange multiplier. Then, a set $\{\lambda_n^*\}_{n=1}^N$ with $\sum_{n=1}^N \lambda_n^* = \lambda_{\mathrm{tot}}$, that maximizes u^*_{UE} must satisfy the set of equations $\{\frac{\partial *}{\partial \lambda_n^*} = 0\}_{n=1}^N$, which is easy to show that it corresponds to

$$\begin{aligned}
\frac{\partial u^*_{\mathrm{UE}}}{\partial \lambda_n^*} &= \frac{\partial u^*_{\mathrm{UE}}}{\partial \lambda_m^*}, \\
\overset{\text{(D.8)}}{\Rightarrow} \frac{\partial\{f_1, \ldots, f_{N+1}\}}{\partial\{q_1^*, \ldots, q_N^*, \lambda_n^*\}} &= \frac{\partial\{f_1, \ldots, f_{N+1}\}}{\partial\{q_1^*, \ldots, q_N^*, \lambda_m^*\}}, \\
\Rightarrow \frac{q_n^*}{q_m^*} &= \frac{\lambda_n^*}{\lambda_m^*},
\end{aligned} \tag{D.9}$$

for all $m, n \in \{1, 2, \ldots, N\}$, where (D.9) follows by direct computation, leading to (5.20). Substituting these values of $\{q_n^*\}_{n=1}^N$ into the system of equations $\{f_n = 0\}_{n=1}^N$ shows that the OP strategies should be the same, i.e., $\tilde{p}_{\mathrm{UE},n}^* = \tilde{p}^*$, for all n. In addition, since these OP strategies are OP game eq. strategies, they must also satisfy the system of equations

$$\left\{ \frac{\partial q_n^*}{\partial \tilde{p}_{\mathrm{UE},n}^*} = p_{\mathrm{OP},n}^c \right\}_{n=1}^N \tag{D.10}$$

All the partial derivatives included in the above system exist and can be obtained by a methodology similar to that for deriving (D.8). Their evaluation results in a new set of equations involving $\{\tilde{p}_{\mathrm{UE},n}^*\}$, $\{\lambda_n^*\}$ and $\{q_n^*\}$. Equation (5.19) and condition (5.18) follows by substituting $\tilde{p}_{\mathrm{UE},n}^* = \tilde{p}^*$ and $q_n^* = \lambda_n^*/\lambda_{\mathrm{tot}}$, for all $n \in \{1, 2, \ldots, N\}$ in the system of (D.10).

Bibliography

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update 2014-2019 white paper," 2015.

[2] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for Next Generation Wireless Systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186-195, Feb. 2014.

[3] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for future radio access," *IEEE VTC spring 2013*, Jun. 2013.

[4] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!" *IEEE Access,* vol. 1, no. 1, pp. 335–349, Aug. 2013.

[5] J. G. Andrews, X. Zhang, G. D. Durgin, and A. K. Gupta, "Are we approaching the fundamental limits of wireless network densification?," *IEEE Commun. Mag.*, vol. 54, no. 10, pp. 184-190, Oct. 2016.

[6] N. Bhushan, L. Junyi, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavai, C. Patel, and S. Geirhofer, "Network densification: the dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82–89, Feb. 2014.

[7] S. F. Yunas, M. Valkama, and J. Niemeläm, "Spectral and energy efficiency of ultra-dense networks under different deployment strategies," *IEEE Commun. Mag.*, vol. 53, pp. 90–100 , Jan. 2015.

[8] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-Dense Networks: A Survey," IEEE Communications & Tutorials, vol. 18, no. 4, Fourth quarter 2016.

[9] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Infor. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[10] M. Ji, G. Caire, and A. Molisch, "Wireless device-to-device caching networks: Basic principles and system performance," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 176–189, 2016.

[11] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Prentice Hall Englewood Clis, 2nd edition, 2001.

[12] H. Viswanathan and S. Mukherjee, "Performance of cellular networks with relays and centralized scheduling," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2318–2328, Sept 2005.

[13] G. P. Fettweis, "The tactile internet: Applications and challenges," IEEE Vehicular Technology Magazine, vol. 9, no. 1, pp. 64–70, Mar. 2014.

[14] S. Mukherjee, *Analytical Modeling of Heterogeneous Cellular Networks*. Cambridge University Press, 2014.

[15] A. I. Hwang, B. Song, and S. S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 20–27, Jun. 2013.

[16] S. Ahmadi, *LTE-Advanced: A Practical Systems Approach to Understanding 3GPP LTE Releases 10 and 11 Radio Access Technologies,* Academic Press, 2014.

[17] S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*, 3rd ed., Wiley, 2013.

[18] F. Baccelli and B. Błaszczyszyn, *Stochastic geometry and wireless networks*. Now Publishers Inc., 2009.

[19] H. ElSawy, A. Sultan-Salem, M.-S. Alouini, and M. Z. Win, "Modeling and analysis of cellular networks using stochastic geometry: A tutorial," CoRR, vol. abs/1604.03689, 2016. [Online]. Available: http://arxiv.org/abs/1604.03689

[20] J. G. Andrews, A. K. Gupta, and H. S. Dhillon, "A primer on cellular network analysis using stochastic geometry," arXiv preprint, 2016, available online: arxiv.org/abs/1604.03183.

[21] H. ElSawy, E. Hossain, and M. Haenggi,"Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: a survey," *IEEE Communications Surveys & Tutorials*, vol. 15, pp. 996–1019, Jul. 2013.

[22] J. G. Andrews, "Seven ways that HetNets are a cellular paradigm shift," *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 136–144, Mar. 2013.

[23] S. Lee and K. Huang, "Coverage and economy in cellular networks with many base stations," *IEEE Commun. Lett.*, vol. 16, no. 7 pp. 1038–1040, Jul. 2012.

[24] C. Li, J. Zhang, and K. B. Letaief, "Throughput and energy efficiency analysis of small cell networks with multi-antenna base stations," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2502–2517, May 2014.

[25] D. Cao, S. Zhou, and Z. Niu, "Optimal base station density for energy-efficient heterogeneous cellular networks," in *Proc. of IEEE Int. Conf. on Commun. (ICC)*, Otawwa, Canada, Jun. 2012.

[26] S. Singh, H. S. Dhillon and J. G. Andrews, "Offloading in heterogeneous networks: modeling, analysis and design insights", *IEEE Trans. on Wireless Commun.*, vol. 12, no. 5, pp. 2484–2497, May 2013.

[27] N. Jindal, J. G. Andrews, and S. P. Weber, "Bandwidth partitioning in decentralized wireless networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5408–5419, Jul. 2008.

[28] K. Huang, V. K. N. Lau, and Y. Chen, "Spectrum sharing between cellular and mobile ad hoc networks: Transmission-capacity trade-off," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1256–1267, Sep. 2009.

[29] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.

[30] A. Lozano, R. Heath, and J. Andrews, "Fundamental limits of cooperation," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5213–5226, Sep. 2013.

[31] S. M. Yu and S.-L. Kim, "Downlink capacity and base station density in cellular networks," in *Proc. of IEEE WiOpt Workshop on Spatial Stochastic Models for Wireless Networks (SpaSWiN)*, 2013.

[32] G. Piro, L. Alfredo Grieco, G. Boggia, F. Capozzi, and P. Camarda, "Simulating LTE cellular systems: an open-source framework," *IEEE Trans. on Veh. Technol.*, vol. 60, no. 2, pp 498–513, Feb. 2011.

[33] S. Andreev, A. Pyattaev, K. Johnsson, O. Galinina, and Y. Koucheryavy, "Cellular traffic offloading onto network-assisted device-to-device connections," *IEEE Commun. Mag.*, vol. 52, no. 4, pp. 20–31, Apr. 2014.

[34] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Enable device-to-device communications underlaying cellular networks: challenges and research aspects," *IEEE Commun. Mag.*, vol. 52, no. 6, pp. 90–96, Jun. 2014.

[35] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1801–1819, Fourth Quarter 2014.

[36] F. Baccelli, B. Błaszczyszyn, and P. Mühlethaler, "An ALOHA protocol for multihop mobile wireless networks," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 421–436, Feb. 2006.

[37] M. C. Ertuk, S. Mukherjee, H. Ishii, and H. Arslan, "Distributions of transmit power and SINR in device-to-device networks," *IEEE Commun. Lett.*, vol. 17, no. 2, pp. 273–276, Feb. 2013.

[38] N. Lee, X. Lin, J. G. Andrews, and R. W. Heath Jr., "Power control for D2D underlaid cellular networks: modeling, algorithms and analysis," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 1, pp. 1–13, Jan. 2015.

[39] X. Lin, J. G. Andrews, and A. Ghosh, "Spectrum sharing for device-to-device communication in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 12, pp. 6727–6740, Dec. 2014.

[40] Q. Ye, M. Al-Shalash, C. Caramanis, J. G. Andrews, "Resource optimization in device-to-device cellular systems using time-frequency hopping," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5467–5480, Oct. 2014.

[41] H. ElSawy, E. Hossain, and M.-S. Alouini, "Analytical modeling of mode selection and power control for underlay D2D communication in cellular networks," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 4147–4161, Nov. 2014.

[42] G. George, R. K. Mungara, and A. Lozano, "An analytical framework for device-to-device communication in cellular networks," Jul. 2014, online: http://arxiv.org/abs/1407.2201.

[43] M. Sheng, J. Liu, X. Wang, Y. Zhang, H. Sun, and J. Li, "On transmission capacity region of D2D integrated cellular networks with interference management," to appear in *IEEE Trans. Commun.*, 2015.

[44] X. Zhang and M. Haenggi, "Random Power Control in Poisson Networks," *IEEE Trans. Commun.*, vol. 60, pp. 2602–2611, Sep. 2012.

[45] M. Haenggi and R. K. Ganti, *Interference in Large Wireless Networks*, NOW (Foundations and Trends in Networking, Vol. 3), 2008.

[46] S. M. Yu and S.-L. Kim, "Downlink capacity and base station density in cellular networks," in *Proc. of IEEE WiOpt Workshop on Spatial Stochastic Models for Wireless Networks (SpaSWiN)*, 2013.

[47] S. Singh, H. S. Dhillon and J. G. Andrews, "Offloading in heterogeneous networks: modeling, analysis and design insights," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2484–2497, May 2013.

[48] D. Cao, S. Zhou, and Z. Niu, "Optimal combination of base station densities for energy-efficient two-tier heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4350–4362, Sep. 2013.

[49] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3484–3495, Oct. 2012.

[50] Y. Zhong, W. Zhang, and M. Haenggi, "Managing interference correlation through random medium access," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 928–941, Feb. 2014.

[51] I. Hwang, B. Song, and S. S. Soliman, "A holistic view on hyper-dense heterogeneous and small cell networks," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 20–27, Jun. 2013.

[52] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, pp. 74–80, Feb. 2014.

[53] K. Doppler, M. Rinne, C. Wijting, C. Ribeiro, and K. Hugl, "Device-to-Device Communication as an Underlay to LTE-Advanced Networks," *IEEE Commun. Mag.*, vol. 50, no. 3, pp. 170–177, Mar. 2012.

[54] A. Sabharwal, P. Schniter, D. Guo, D. Bliss, S. Rangarajan, and R. Wich- man, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Select. Areas Commun.*, vol. 32, pp. 1637–1652, Sept. 2014.

[55] L. Lei and Z. Zhong, "Operator controlled device-to-device communications in LTE-Advanced networks," *IEEE Wireless Commun.*, vol. 19, no. 3, pp. 96–104, Jun. 2012.

[56] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *EEE J. Select. Areas Commun.*, vol. 27, pp. 1029-1046, Sept. 2009.

[57] S. Stefanatos, A. Gotsis, and A. Alexiou, "Operational region of D2D communications for enhancing cellular network performance," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 5984–5997, Nov. 2015.

[58] A. Hasan and J. G. Andrews, "The guard zone in wireless ad hoc networks," *IEEE Trans. Wireless Commun.,* vol. 6, no.3, pp. 897–906, Mar. 2007.

[59] Z. Chen and M. Kountouris, "Decentralized opportunistic access for D2D underlaid cellular networks," Jul. 2016. [Online]. Available: http://arxiv.org/abs/1607.05543.

[60] J. G. Andrews, R. K. Ganti, N. Jindal, M. Haenggi, and S. Weber, "A primer on spatial modeling and analysis in wireless networks," IEEE Commun. Mag., vol. 48, no. 11, pp. 156–163, Nov. 2010.

[61] H. ElSawy and E. Hossain, "On stochastic geometry modeling of cellular uplink transmission with truncated channel inversion power control," *IEEE Trans. Wireless Commun.*, vol. 13, no. 8, pp. 4454–4469, Aug. 2014.

[62] H. Lee, Y. Sang, and K. Kim, "On the uplink SIR distributions in heterogeneous cellular networks," *IEEE Commun. Let.*, vol. 18, no. 12, pp. 2145–2148, Dec. 2014.

[63] S. Singh, X. Zhang, and J. G. Andrews, "Joint rate and SINR coverage analysis for decoupled uplink-downlink biased cell associations in HetNets," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5360–5373, Oct. 2015.

[64] B. Błaszczyszyn and D. Yogeshwaran, "Clustering comparison of point processes with applications to random geometric models," in *Stochastic Geometry, Spatial Statistics and Random Fields*, vol. 2120, Lecture Notes in Mathematics. Cham, Switzerland: Springer-Verlag, 2015, pp. 31–71.

[65] M. Haenggi, "User point processes in cellular networks," Nov. 2016. [Online]. Available: http://arxiv.org/abs/1611.08560.

[66] Z. Tong and M. Haenggi, "Throughput analysis for full-duplex wireless networks with imperfect self-interference cancellation," *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 4490–4500, Nov. 2015.

[67] C. Psomas, M. Mohammadi, I. Krikidis, and H. A. Suraweera, "Impact of directionality on interference mitigation in full-duplex cellular networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 487–502, Jan. 2017.

[68] K. S. Ali, H. ElSawy, and M.-S. Alouini, "Modeling cellular networks with full-duplex D2D communication: a stochastic geometry approach," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4409–4424, Oct. 2016.

[69] H. Zhang, Y. Ji, L. Song, and H. Zhu "Hypergraph based resource allocation for cross-cell device-to-device communications," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[70] M. Haenggi, "The local delay in poisson networks," *IEEE Trans. Inf. Theory*, vol. 59, pp. 1788–1802, Mar. 2013.

[71] Z. Yazdanshenasan, H. S. Dhillon, M. Afshang, and P. H. J. Chong, "Poisson hole process: theory and applications to wireless networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7531–7546, Nov. 2016.

[72] S. Sadr and R. S. Adve, "Handoff rate and coverage analysis in multi-tier heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2626–2638, May 2015.

[73] H. E. Elkotby and M. Vu, "Uplink User-Assisted Relaying Deployment in Cellular Networks," *IEEE Trans. on Wireless Commun.*, vol. 14, no. 10, pp. 5468–5483, Oct. 2015.

[74] A. E. Gamal and Y.-H. Kim, *Network Information Theory*, Cambridge University Press, 2012.

[75] J. G. Andrews *et al.*, "What will 5G be?," *IEEE J. Sel. Areas Commun.*, vol. 32, pp. 1065–1082, Jun. 2014.

[76] L. E. Doyle, J. Kibiłda, T. K. Forde, and L. A. DaSilva, "Spectrum without bounds, networks without borders," *Proc. IEEE*, vol. 102, no. 3, pp. 351–365, Mar. 2014.

[77] C. Liang and F. R. Yu, "Wireless network virtualization: a survey, some research issues and challenges," *IEEE Commun. Surveys & Tutorials*, First Quarter 2015.

[78] C. Courcoubetis and R.Weber, *Pricing Communication Networks*. NewYork: Wiley, 2003.

[79] J. Huang and L. Gao, *Wireless Network Pricing*, Synthesis Lectures on Communication Networks, Morgan & Claypool, 2013.

[80] M. J. Osborne and A. Rubenstein, *A Course in Game Theory*, MIT Press, 1994.

[81] L. Duan, J. Huang, and B. Shou, "Duopoly competition in dynamic spectrum leasing and pricing," *IEEE Trans. Mobile Comput.*, vol. 11, no. 11, pp 1706–1718, Nov. 2012.

[82] S. M. Yu and S.–L. Kim, "Game-theoretic understanding of price dynamics in mobile communications services," *IEEE Trans. Wireless Commun.*, vol., 13, no. 9, pp. 5120–5131, Sep. 2014.

[83] O. Korcak, G. Iosifidis, T. Alpcan, and I. Koutsopoulos, "Competition and regulation in a wireless operator market: an evolutionary game perspective," in *Proc. 6th International Conf. on Network Games, Control and Optimization (NetGCooP)*, 2012.

[84] L. Guijarro, V. Pla, B. Tuffin, P. Maillé, and J. R. Vidal, "Competition and bargaining in wireless networks with spectrum leasing," in *Proc. IEEE Global Communications Conference (GLOBECOM 2009),* Dec. 2009.

[85] F. Zhang and W. Zhang, "Competition between wireless service providers: pricing, equilibrium and efficiency," in *Proc. WiOpt,* 2013.

[86] L. Rose, E. V. Belmega, W. Saad, and M. Debbah, "Pricing in heterogeneous wireless networks: hierarchical games and dynamics," *IEEE Trans. Wireless Commun.*, vol., 13, no. 9, pp. 4985–5001, Sep. 2014.

[87] J. Park, S.–L. Kim, and J. Zander, "Asymptotic Behavior of Ultra-Dense Cellular Networks and Its Economic Impact," in *Proc. IEEE Global Communications Conference (GLOBECOM 2014)*, Austin, Dec. 2014.

[88] C.-H. Lee, C.-Y. Shih, and Y.-S. Chen, "Stochastic geometry based models for modeling cellular networks in urban areas," Wireless Networks, vol. 19, no. 6, pp. 1063–1072, August 2013

[89] S. Li and J. Huang, "Usage-Based Pricing Differentiation for Communication Networks: Incomplete Information and Limited Pricing Choices," in *Smart Data Pricing,* Wiley, 2014, ch. 8, pp. 195–240.

[90] P. Semasinghe and E. Hossain, "Downlink power control in self-organizing dense small cells underlaying macrocells: a mean field game," *IEEE Trans. Mobile Comput.*, vol. 15, no. 2, pp. 350–363, Jan. 2016.

[91] S. Stefanatos, A. G. Gotsis, and A. Alexiou, "Operational region of D2D communications for enhancing cellular network performance," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 5984–5997, Nov. 2015.

[92] S. M. Yu and S.-L. Kim, "Downlink capacity and base station density in cellular networks," in *Proc. of IEEE WiOpt Workshop on Spatial Stochastic Models for Wireless Networks (SpaSWiN)*, 2013.

[93] S. G. Foss and S. A. Zuyev, "On a Voronoi aggregative process related to a bivariate Poisson process", *Advances in Applied Probability*, vol. 28, no. 4, pp. 965–981, Dec. 1996.

[94] S. Lasaulce and H. Tembine, *Game Theory and Learning in Wireless Networks: Fundamentals and Applications*. Elsevier Academic Press, Oct. 2011.

[95] S. Lasaulce, M. Debbah, and E. Altman, "Methodologies for analyzing equilibria in wireless games," *IEEE Signal Processing Mag.*, vol. 26, no. 5, pp. 41–52, Sep. 2009.

[96] J. Callahan, *Advanced Calculus: A Geometric View*. Springer, 2010.

[97] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd ed., Springer, 2006.