Part I

Getting Started

1 Introduction

As old as I have become, many developments my eyes have seen. When I was a young man, changes used to take place from time to time, every now and then. Nowadays, they occur so often, life itself seems to change every day.

—A 97-year-old man commented, making reference to the increasing pace of the developments in the communications field through the last century

These comments, although exaggerated at first sight, may not be so, in light of the advancements seen in the telecommunication industry since Graham Bell carried out the first successful bidirectional telephone transmission in 1876 [1]. Since then, society has witnessed

- most long-distance communications having at least a wireless component, freed from wires and operated through air at the speed of light,
- wireless and mobile communications made available to over 8.6 billion connections across the globe [2],
- new types of communications and social interactions emerging through both the Internet [3] and social networking [4] and
- many other breakthroughs, which have certainly changed our everyday lives.

These developments, although of great importance to the 97-year-old man, are probably just small steps towards a new era – the era of digital and pervasive communications – which will continue to change the world we are inhabiting in unpredictable and fascinating manners.

As a matter of fact, today, we are on the brink of another significant societal change. While the network has mainly served humans up to now, this capability will increasingly be extended to machines in the near future too. By 2022, it is expected that there will be not only 8.4 billion handheld or personal mobile-ready devices, but also 3.9 billion machine-to-machine connections [2]. The emergence of this machine-originated data traffic will drive further the demand for network capacity, but also impose additional requirements on network performance, mainly in the areas of end-to-end latency and reliability. These are currently the major challenges for many new applications.

Nowadays, most of the data services reside on the Internet, far away from the user equipment (UE), where the speed of light becomes one of the main factors limiting end-to-end latency. To address this problem, processing will have to move closer to

the UE, e.g. into a cloud computing infrastructure, which will extend – and act as a ramification of – the network. In addition, an intelligent and adaptive network management and a well-designed congestion control can also help to significantly enhance reliability, thus enabling new real-time applications, such as augmented reality or efficient machine communication.

With these new requirements and changes, communication networks are evolving to become our main interface with the virtual world, and increasingly also with the physical one. This future network will simplify and automate many aspects of life, allowing one to effectively "create time," by improving the efficiency in everything we do [5].

Making this vision of the future network a reality will require from a technical perspective both

- ultra-broadband wireless access, providing orders of magnitude improved throughput, delay and reliability as well as quality of service (QoS) control and
- a highly adaptable and remotely programmable cloud computing infrastructure located close to the edge of the network.

Throughout this book, we argue that small cells, and more specifically, ultra-dense deployments are one of the answers to the technological challenges of creating an ultra-broadband wireless access that connects mobile UEs, machines and objects to a processing cloud engine. In more detail, this book serves as a tool to shed new light on the fundamental understanding of ultra-dense networks.

As an introduction to the content of the book, the remainder of this chapter first depicts the current industry capacity challenge and follows with an overview of the small cell technology and its history, from both an industry and an academic perspective. Then, the individual parts and chapters of the book are introduced, as well as their relationship to various aspects of deploying and operating small cell networks. To conclude, some of the key nomenclature used in this book is presented together with a list of the most relevant publications in the field of ultra-dense networks by the authors of this book.

As an important disclaimer, let us note that this book is going to focus on the study of single antenna UEs and base stations (BSs), thus using single-input single-output transmission modes. As a result, this book does not consider either multiple-input multiple-output (MIMO) or multi-cell coordinated transmissions/receptions, and all the statements within are done accordingly.

1.1 The Capacity Challenge

Voice-based services, such as voice over internet protocol (VoIP), were the killer applications at the beginning of this century, demanding an average of tens of kilobits per second per UE for this type of connection [6], while the streaming of high-definition (HD) video is probably the most popular service today, requiring

tens of megabits per second per video feed [7]. Future services, however, such as three-dimensional visualization, augmented and virtual reality, online gaming using multiple displays and the robot-to-robot exchange of HD laser imaging detection and ranging (LIDAR) maps will require much more capacity, with expected average throughputs per UE exceeding 1 Gbps [8] – and who knows what else tomorrow will bring?

With this enormous challenge of improving the average throughput per UE by orders of magnitude, before making any decision on any technology investment, which is likely to be costly, it is advised that network operators and service providers understand well the different dimensions that they have to improve wireless capacity.

In a simplified form, the Shannon–Hartley theorem [9],

$$C = B \cdot \log_2 \left(1 + \frac{S}{N}\right), \tag{1.1}$$

provides an insight into which are the variables that influence the amount of information – capacity, C, in *bps* – that a transmitter can send to a receiver

- over a communication channel of a specified bandwidth, B, in Hertz
- with a received signal power, S, in Watts
- in the presence of an additive white Gaussian noise (AWGN) power, N, in Watts.

From this theorem, it can be inferred that the capacity, C, of a UE can be scaled up by increasing

- the bandwidth, B, per UE and/or
- the signal-to-noise ratio (SNR), $\frac{S}{N}$, of such UE, or more accurately, the signal-to-interference-plus-noise ratio (SINR), $\frac{S}{I+N}$, of the UE in a multi-cell multi-UE network, like the ones that will be studied in this book, where I stands for Gaussian interference in Watts.

Importantly, equation (1.1) also shows that to scale the capacity, C, of a UE, increasing the bandwidth, B, per UE is generally a more promising technique than increasing the SINR, $\frac{S}{I+N}$, of such a UE, since the former yields a linear scaling, while the latter only a logarithmic one.[1]

[1] Even though multi-antenna technology is not considered in this book, for completeness, it should be noted that multiple antennas can be used to either

- leverage spatial multiplexing through MIMO techniques or
- increase the SINR, $\frac{S}{I+N}$, of a UE via beamforming.

In particular, MIMO transmissions/receptions, can take advantage of the spatial resources, and linearly increase the capacity, C, of a UE with the number of spatial streams multiplexed. This can be treated as a "virtual" increase of the bandwidth, B, per UE. When taking a cell or a network perspective, it should also be noted that multi-user MIMO, coordinated beamforming and multi-cell coordinated transmissions/ receptions can be used to increase the spatial multiplexing in the cell or the network and/or the SINR, $\frac{S}{I+N}$, of a UE. Readers interested in related topics are referred to [10] and references therein.

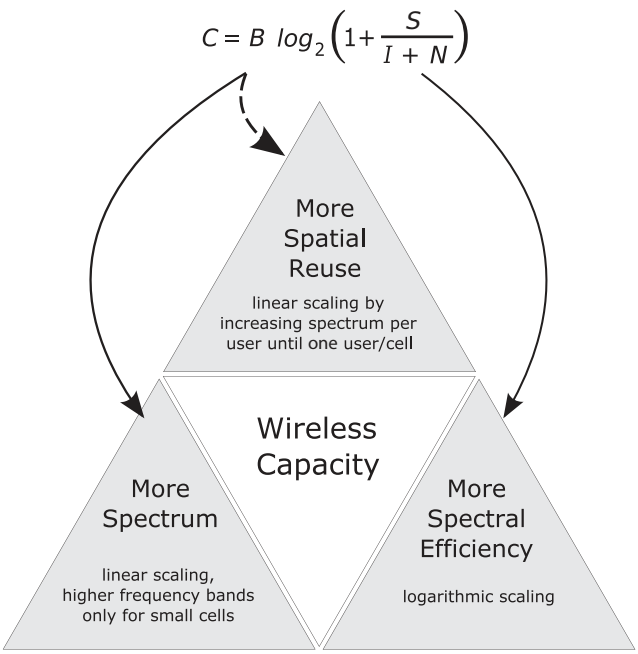

Figure 1.1 Dimensions for scaling the capacity of a UE in a wireless network [8].

With this in mind, a fundamental question arises:

How can we increase the bandwidth, B, per UE?

In a network with multiple UEs, the bandwidth, B, per UE can be scaled up by either increasing

- the amount of frequency resources invested into the network and/or
- the network densification, and in turn, its associated spatial frequency reuse.

For the sake of clarity, it should be noted that the reduced cell size in a denser network results in an improved spatial reuse of the frequency resources, since there are less UEs in each cell to share the available cell bandwidth. As a result, each UE has access to more of such frequency resources.

Overall, as depicted in Figure 1.1, this leaves one with three main approaches to enhance the capacity, C, of the UE, i.e. using a wider bandwidth, deploying a denser network and improving the signal quality – the SINR of the UE, where the two first ones may be more appealing due to their intrinsic linear scaling of capacity.

To put things in context, and show how each of these three degrees of freedom have historically contributed to the increase of the capacity of practical networks, Webb [11] put together an interesting analysis, indicating that, between 1950 and 2000, such network capacity has increased around

- 2700× from densifying the network with smaller cells,

- 15× by using more bandwidth in the sub-6 GHz bands (from 150 MHz to 3 GHz) and
- 10× by improving the spectral efficiency (waveforms and multiple access techniques, modulation, coding and medium access control (MAC) methods such as scheduling, hybrid automatic repeat request (HARQ), etc.).

From this study, it is clear that – by far – the majority of the capacity gains in the past were achieved by increasing the spatial frequency reuse through densifying the network with smaller cells. This leaves one with the following question:

How much further can we increase the spatial reuse by reducing the cell size?

Answering the above question from a theoretical perspective is one of the primary goals of this book.

For completeness, and before proceeding any further, it should be noted here at this point that the operation at higher carrier frequencies, e.g. millimetre wave bands, also offers the possibility of accessing large amounts of spectrum and the associated very wide bandwidths, thus enabling extreme data rates.

Higher carrier frequencies, however, are also associated with higher radio frequency attenuations, which limit their network coverage. Although this can be compensated to some extent by means of multi-antenna technologies – and more in detail through beamforming – a substantial coverage disadvantage will always remain for a network operating at higher carrier frequencies [12].

Another challenge with the operation at higher frequency bands is the regulatory aspects. For non-technical reasons, the rules defining the allowed radiation may change in these higher frequency bands, from a specific absorption rate (SAR)-based limitation to a more effective isotropic radiated power (EIRP)-like limitation. These restrictions may also impose further coverage constraints [13].

More importantly, the millimetre wave technology in general – and the beamforming one in particular – are not mature as of today, or at least, are not cost-effective for ultra-dense deployments. The features required to deal with

- the beam alignment and tracking, possibly needed at both communication ends and
- the related issues arising from unexpected device rotation, blockage and mobility

are still of high complexity, and lack of robust QoS provisioning. This together with the large energy consumption of current millimetre wave access points make this solution still too expensive.

As a result, the more mature and developed sub-6 GHz technology still remains a front runner for ultra-dense deployments, especially due to its ability to satisfy communication requirements in non-line-of-sight (NLoS) and outdoor-to-indoor propagation conditions. For these reasons, we focus on low carrier frequencies in this book, and analyze in detail the impact of network densification and the increase of spatial reuse on network performance, as posed earlier. However, due to its potential, we leave the door open to the analysis of millimetre wave deployments for next editions

of this book, and in particularly to the inter-working of sub-6 GHz and millimetre wave technology.

1.2 Network Densification

In a multi-cell multi-UE network, UEs being served within a cell share the available bandwidth. Thus, reducing the cell size – while deploying more cells to maintain the same level of coverage – also reduces the number of UEs per cell, and in turn, increases the bandwidth per UE. Through this approach, the bandwidth per UE can be increased, until each cell only serves a single UE. When densifying further, beyond this point, only the signal power – and potentially the SINR – of the UE can be improved by reducing the distance between the UE and its serving small cell BS.

Overall, by increasing the bandwidth, B, per UE, the capacity, C, scales up linearly, until the one UE per cell limit is reached, after which the scaling becomes only logarithmic through improvements on the SINRs of the UEs, as indicated by equation (1.1).

Figure 1.2 illustrates this capacity scaling behaviour, showing that with increasing cell densities, the capacity

- initially increases quickly due to the spatial frequency reuse, but then
- slows down, when the one UE per cell limit is reached, and the gains are mainly dominated by improvements on the SINRs of the UEs, through proximity gains [8].

From Figure 1.2, it is also important to note that the results were obtained assuming an active UE density of 300 BSs/km^2, typical in some dense urban scenarios, and that the one UE per cell limit is reached for an inter-site distance (ISD) of around 30–40 m. This indicates that there is still plenty of room for network densification in major cities like Manhattan and London, where the average ISD is around 200 m.

A second aspect of densification is that the required transmit power may reduce to an extent where its contribution to the total energy consumption becomes insignificant, and the processing power of the small cell BS becomes the dominant factor. Moreover, with reduced cell sizes, the required number of small cells to provide coverage increases, and as a result, many of them may not serve any UE for most of the time. However, they still consume energy and transmit unnecessary pilot signals, which may cause inter-cell interference. This issue can be addressed by introducing idle mode capabilities, where small cell BSs are only woken up to actively serve UEs. With efficiently controlled idle modes, the network energy consumption reduces and the SINR of the UE significantly improves.

The main challenge of network densification, however, is the issue of increasing costs for equipment, deployment and operational expenses. In this light, it is important to note that the cost of a small cell BS, estimated in 2015, only accounts for approximately 20% of the total deployment costs associated with outdoor small cells.

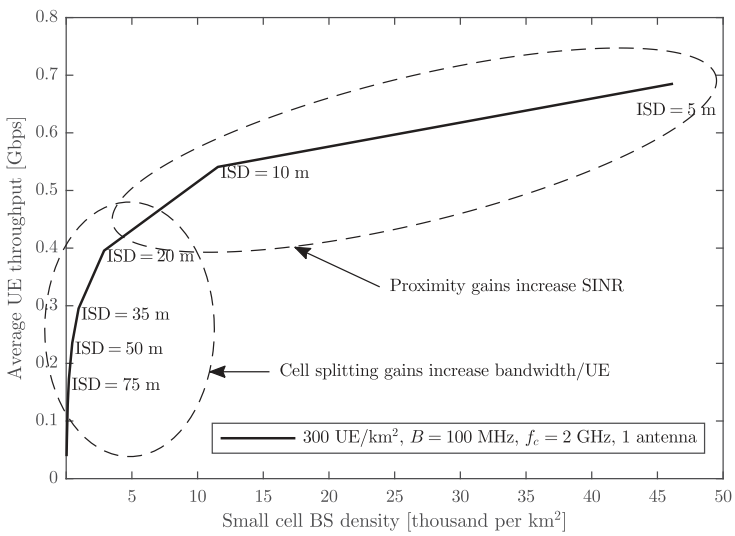

Figure 1.2 Capacity scaling with densification for different inter-site distances in a dense urban scenario considering a hexagonal BS deployment and a semi-clustered UE distribution. For more details on the scenario, models and results, the reader is referred to [8].

The majority of the costs are site leasing (26%), backhaul (26%), planning (12%) and installation (8%) [14]. The good piece of news is that this challenge can be addressed by changing the deployment model from an operator deployment to a "drop and forget" end-user deployment, and reusing the existing power and backhaul infrastructure. In this model, the end-user simply connects the small cell BS to the power and the backhaul, which then triggers a fully automatic configuration, and a continuous self-optimization process during operation. This end-user deployment model is feasible for both the residential and the enterprise markets.

Due to cost and performance reasons, it is also important to highlight that it becomes increasingly important to deploy the small cell BSs wherever the UEs are, since small cells cannot compensate for misplacement as well as larger cells do. If the small cell BSs are not deployed in an intelligent manner following the distribution of UEs, a larger number of small cell BSs will be required to achieve the one UE per cell limit. However, accurate UE demand distributions are hard to derive today, because of the limited accuracy of conventional localization techniques in cellular networks, such as triangulation. One may think of using more accurate techniques such as the global positioning system (GPS) for this purpose, but its performance is poor indoors, where 80% of the traffic demand is located [15]. Advanced planning tools for small cell deployment are still an open challenge.

In summary, densification continues to have a high potential to increase capacity until reaching the one UE per cell limit. To maintain high performance and energy efficiency, idle mode capabilities that switch off small cells when they are not serving

UEs are necessary. Transitioning to a "drop and forget" deployment by the end-user has a high potential to reduce the deployment and operational costs.

1.3 A Brief History of Small Cells

In this section, we provide an overview of the small cell technology and its history, first from an industry perspective, and subsequently from a theoretical one. This last part of the chapter serves as an introduction to the rest of the book, which will be formalized by the outline presented in the next section.

1.3.1 From an Idea to a Market Product

The idea of the small cell has been around for over three decades [16]. Initially, "small cell" was the term used to describe the cell size in a metropolitan area, where a macrocell – with a cell diameter on the order of kilometres – would be split into a larger number of smaller cells with reduced transmit power, known today as metropolitan macrocells or microcells. These small cells had a cell diameter of a few hundreds of metres.

In the 1990s, picocells appeared with even a smaller cell size, between a hundred metres to around a few tens of metres [17]. These "more traditional" small cells were used for coverage and capacity infill, that is, where macrocell penetration was insufficient to give a good connection or where the macrocell was at its capacity limit. These types of small cell BSs were essentially a smaller version of the macrocell BSs, which also had to be planned, managed and interfaced with the network. This last point is probably the most important reason why small cells – other than metropolitan macrocells or microcells – have not gained much popularity for quite some time. Essentially, the costs associated with deploying and running a large number of small cells outweighed the performance advantage that this kind of cellular topology provided.

In the 2000s, new thinking on the deployment and configuration of cellular systems began to address the cost and the operational aspects of small cell deployments, which enabled the cost-effective deployment of even smaller cells [18]. Such thinking crystallized in the home BS concept first [19] and the femtocell one later [20]. A femtocell is a low-cost cellular BS with advanced auto-configuration and self-optimization capabilities, which allows the end-user – without any operator involvement – to deploy this small form factor BS in a plug-and-play manner within the home. Femtocells use a broadband internet connection as backhaul, and connect to the cellular network through dedicated gateways, which enables a better scaling to millions of femtocell BSs. Early results on the performance of 3G universal mobile telecommunication system (UMTS) femtocells were presented in [21–23], which were shortly afterward extended with a bulk of studies on self-optimization and offloading strategies, multiple antenna techniques and energy management methods [20, 24–30]. Soon after, results