5G small cell optimized radio design

Preben Mogensen (1, 2), Kari Pajukoski (3), Esa Tiirila (3), Eeva Lähetkangas (3), Jaakko Vihriälä (3), Seppo Vesterinen (3), Matti Laitila (3), Gilberto Berardinelli (2), Gustavo W. O. Da Costa (2), Luis G. U. Garcia (2), Fernando M. L. Tavares (2), Andrea F. Cattoni (2)

(1) Nokia Siemens Networks, Aalborg, Denmark
(2) Department of Electronic Systems, Aalborg University, Denmark
(3) Nokia Siemens Networks, Oulu, Finland

Abstract — The 5th generation (5G) of mobile radio access technologies is expected to become available for commercial launch around 2020. In this paper, we present our envisioned 5G system design optimized for small cell deployment taking a clean slate approach, i.e. removing most compatibility constraints with the previous generations of mobile radio access technologies. This paper mainly covers the physical layer aspects of the 5G concept design.

I. INTRODUCTION

Historically, new generations of radio access technologies (RATs) have been introduced with an interval of approximately ten years to cope with the exponential increase of the mobile data traffic and to take full advantage of the evolution of the technology components without any legacy burden [1].

The specifications of the Long Term Evolution - Advanced (LTE-A) standard, which is agreed to be the 4th generation (4G) mobile communication technology, were finalized back in 2010 [2]. If history is any indication, a new 5th Generation (5G) radio standard is expected to reach the mass market around 2020 and to last until ~2030, where we may potentially experience a 6th Generation (6G).

In [3] we have predicted the mobile traffic growth to be in the range of ~x150-500 by 2020 (with reference to 2010), and to increase to ~x3000-30,000 by 2030. We also predict the user data rate demand to grow by a factor of ~x10 by 2020 and a factor ~x100 by 2030. Such growth values in both traffic volume and user data rate set demands to both capacity and coverage of 5G.

Within the industry, it is generally anticipated that the 5G network evolution will be heterogeneous in cell types - ranging from macro to pico - and it will also integrate multiple radio access technologies. Our ~x1000 HetNet evolution studies for 2020 indicate that both indoor and outdoor hot spot small cells will start to play a very dominant role to meet the above mentioned capacity and user data rate demands [3]. The studies also clarify the need to operate such ultra dense deployment of small cells in a dedicated spectrum to avoid coexistence issues with high power micro/macro cells. The frequency band from 3.4-4.9 GHz has drawn attention for increasing capacity of International Mobile Telecommunications (IMT) systems in the short term (i.e., WRC-2015 5G agenda point) and potentially even higher frequency bands to be allocated at the WRC-2018/19. The usage of millimeter waves has also drawn attention given the large amount of available bandwidth [4].

Bearing in mind the dedicated spectrum, our 5G studies have focused on a clean slate approach for the design of a novel RAT optimized for small cells. Given the historical evolution of the technology components [5], the classical key performance indicators of 5G are estimated to be significantly better than 4G, i.e.

- peak data rates should be in the order of 10 Gbps;
- Round Trip Time (RTT) should be in the order of 1ms;
- spectral efficiency to be at least ~x2 better than 4G.

However, we also see new significant drivers for the 5G RAT design, such as:

- Very low power consumption of both access points and terminals
- Efficient support of Machine Type Communication (MTC)
- Flexibility in spectrum usage
- Self-optimized ultra-dense deployment of access points
- Support of multi-hop (e.g., self-backhauling)
- Simple and low cost design.

While our previous contribution [3] was focused on the motivation for initiating Beyond 4G research and on a high-level description of its technology enablers, this paper presents our latest updates in the 5G small cell concept design. The paper is structured as follows. The main criteria for the physical layer design are described in Section II along with the proposed frame structure and the envisioned numerology. Section III focuses on the Radio Resource Management (RRM) issues for interference mitigation, while Section IV discusses the feasibility of network synchronization which represents an underlying assumption for our design. Section V presents a general view on the 5G networking aspects. Finally, Section VI resumes the conclusions and states the future work.

II. PHYSICAL LAYER DESIGN

Our previous contribution [3] presented and motivated the main physical layer technology components of our novel Beyond 4G /5G RAT. It is agreed that a set of advanced features such as Multiple-Input-Multiple-Output (MIMO) antenna technology [6], fast link adaptation, Hybrid-Automatic Repeat Request (HARQ) [7] and interference mitigation techniques are to be included in the design. Orthogonal Frequency Division Multiplexing (OFDM) is recognized as the preferred modulation for both uplink and downlink given its multipath mitigation capability and the straightforward extension to MIMO [7]. Moreover, Time Division Duplex (TDD) mode has been preferred to Frequency Division Duplex...
a number of Physical Resource Blocks (PRBs) providing native power consumption. A proper design of the frame structure is then necessary for aligning the radio chains of both Access Point (AP) and User Equipment (UE).

The underlying assumption for our design is a fully synchronized system, i.e. all APs/UEs in the network share the same knowledge of the frame timing. The aforementioned technology components are already well established in the engineering community, and a more detailed description of their usage has been presented in [3]. Here, we focus on the design of the 5G frame structure and numerology.

II.A 5G Frame structure

The 5G physical layer aims at providing high performance in terms of data rate and latency with reduced cost and power consumption. A proper design of the frame structure is then fundamental for achieving our ambitious targets. More specifically, the frame should be designed bearing in mind the following requirements:

- Low latency, below 1 ms. Such requirement establishes a reduction of a factor of ~\(x10\) with respect to the LTE target.
- Support for frequency coordination and reuse. This is meant for dealing with the co-channel interference which may be a strong limiting factor in the 5G performance due to the uncoordinated cell deployment.
- Support for advanced receivers. The usage of such receivers represents a further promising option for counteracting the co-channel interference and then boosting the data rate.
- Support for pipeline-processing at the receiver. This reasonably leads to low computational complexity and reduced latency.
- Low power consumption, allowing longer UE battery life.
- Support for communication links beyond the traditional AP-UE access, e.g., MTC and self-backhauling.

The envisioned frame structure is shown in Figure 1. Note that a short guard period (GP) is inserted at each potential switch of the communication link, in order to accommodate the on-off power transient. In the frequency domain, the system bandwidth is divided into a number of Physical Resource Blocks (PRBs) providing native support for frequency coordination/reuse techniques. The first part of the frame represents the control part while the data part is located next.

In LTE, both Physical Downlink Control Channel (PDCCH) and Physical Uplink Control Channel (PUCCH) are mapped over a set of subcarriers in multiple OFDM symbols which also carry data information [7]; here, the time separation between control part and data part in the frame allows instead separating control and data planes. This enables the cost-effective pipeline processing at the receiver, since the UE can process its dedicated control information while transmitting/receiving in the data part, thus reducing the latency. In case of traditional AP-UE link, the control part can be composed for example of one symbol for the DL and one symbol for the UL, as shown in Figure 1. Further, the control symbols can be selected in a link centric manner to enable further communication directions. For instance, the two symbols can be acceded by devices communicating directly with each other in case of MTC, or by the AP and a relay node for self-backhauling.

The time separation between control and data part is also meant to reduce the power consumption at the UE; the device can turn-off its receiver chain for the rest of the frame in case it does not receive any command or information in the control part, thus reducing its battery consumption.

Examples of control information to be mapped in the control symbols are the scheduling request in the UL, the scheduling grant in the DL, the Modulation and Coding Scheme (MCS) indicator for the Adaptive Modulation and Coding (AMC), the Rank Indicator (RI) which sets the number of streams to be used, etc. Such information can be represented by a small number of bits which can be encoded with a fixed robust modulation (e.g. QPSK 1/6).

Since the system is fully synchronized, multiple APs/UEs may transmit control signaling employing the same PRBs simultaneously, thus leading to harmful collisions; this is precisely where Radio Resource Management (RRM) techniques discussed in Section III come into the picture to ensure a sensible selection of different PRBs. The subsequent data part can be dedicated entirely to either UL transmission or DL transmission. Its first symbol is dedicated to the DeModulation Reference Symbols (DMRS), which are used for channel estimation purposes. The well-known Zadoff-chu sequences [7] can be used as DMRS due to their favorable cross-correlation properties. This design allows the simultaneous estimation of the channel responses of multiple interfering APs/UEs in the data part. Moreover, it also stabilizes the interference pattern within a radio frame, thus being a pivotal enable of the effective usage of interference rejection combining (IRC) receivers [8], as will be further discussed in Section III.

The ambitious latency target can be achieved by assuming a frame duration of 0.25 ms, and an optimized scheduling and HARQ design. Both scheduling and HARQ processes are shown in Figure 2(a) and Figure 2(b), respectively. The UE initiated data transmission requires 3 TDD cycles (scheduling request in the UL, scheduling grant in the DL, data transmission in the UL), for a total of 0.75 ms. Similarly, the round trip time of the HARQ process (AP grant, AP transmission and UE ACK/NACK transmission) requires...
0.75 ms including processing times. Differently from LTE-TDD [7], the HARQ round trip time is here fixed and does not depend on the UL/DL ratio; the control part in each radio frame offers at least one OFDM symbol in each direction for the transmission of acknowledgments. The number of parallel frame offers at least one OFDM symbol in each direction for HARQ processes is 4, while in LTE-TDD it is up to 15; this allows a considerable reduction of the memory circuitry (buffers), which leads to significant cost savings in the baseband chip.

II.B 5G numerology

The envisioned 5G frame numerology is shown in Table 1. LTE, LTE-A and IEEE 802.11 [9] numerologies are also included for the sake of comparison. We assume as a baseline for our initial studies a carrier bandwidth of 200 MHz, though multiple carriers will be required for achieving the maximum throughput target when combined with 4x4 MIMO and 256QAM modulation.

5G subcarrier spacing is set to be 4 times larger than the LTE/LTE-A one, leading to a 4 times shorter time symbol duration. Such large subcarrier spacing allows higher robustness than LTE to the phase noise; while primarily targeting frequency band from 3.4-4.9 GHz, 5G systems can be then set to operate at significantly high carrier frequencies (e.g. 15 GHz) even with relatively cheap devices. On the other side, the 5G symbol duration is still much larger than the IEEE 802.11 one. This allows maintaining a relatively low number of symbols per frame, with the advantage of saving the cumulative overhead given by the Cyclic Prefix (CP) at the beginning of each symbol.

The CP duration is considerably lower than the LTE one, given the shorter expected delay spread/propagation delay in local area scenarios. In LTE the minimum GP duration (in case no timing advance procedure is taking place) is set to the OFDM symbol duration in order to maintain the same numerology on both the data frame and the special subframe where the switching is operated, as well as for compatibility with existing standards operating in the same bandwidth. Our clean slate design approach for 5G removes any backwards compatibility constraint; moreover, the transmit power of the AP in local area is significantly lower than the micro and macro LTE base stations. This leads to the possibility of setting an extremely short GP for the on/off power transient.

Note that in 5G the resultant cumulative overhead represented by CP and GP is approximately the same of LTE and LTE-A, with the advantage of higher robustness to phase noise.

III. RRM AND INTERFERENCE MANAGEMENT

Radio Resource Management (RRM) refers to a variety of procedures with the specific aim of using the radio resources in the most efficient way in terms of target metrics such as throughput, reliability and latency. Broadly, a RRM framework is taking care of performing functions such as:

- Selection of transmission modes and corresponding parameters, e.g. power or Modulation and Coding Scheme (MCS).
- Decision of which nodes will be transmitting, receiving or idle in a particular radio resource (frequency, time, spatial stream or code).
- Channel assignment for particular links and traffic flows.

The possibility of separating control and data planes enabled by the frame structure described in Section II also allows the usage of separate RRM strategies for data and control, which have very different requirements.

A correct reception of the control information is indeed critical in any RAT for enabling data transmission. Control channel reliability is always a design goal, even at the cost of a significantly larger overhead, e.g. the usage of a fixed robust modulation and coding scheme. Frequency domain Inter-cell Interference Coordination (ICIC) techniques are yet another safe-guard to ensure high Signal-to-Interference-plus-Noise Ratio (SINR) for control channels. The selection of a subset of PRBs for control channel operation can be done by efficient distributed techniques at the time of bootstrapping or when more control channel capacity is needed, using for example the mechanism in [10].

A new and different RRM challenge lies in the foreseen usage of cooperative link adaptation techniques or inter-cell interference canceling [11]. In order to work properly, such
schemes require that devices are able to decode the control information from neighboring cells. The synchronized transmission of control information in different PRBs is a first step towards this goal. It enables the receiving APs/UEs to know precisely when to attempt decoding such control information (e.g., scheduling assignment) from neighboring cells. Nevertheless, selecting the right set of PRBs that ensures a reasonable SINR for both intended and neighboring receivers is a topic for further studies.

On the data plane, it is typically worth to be more lax about the interference effects. HARQ works indeed as an insurance against exceptionally congested or faded frames. Moreover, the fast varying nature of traffic will make very persistent interference less likely. In this way, on the data plane one can go after higher risks of losing a single data transmission with higher payoffs in terms of throughput. Compared to the control plane, the data plane has then an almost diametrically opposed RRM strategy: heavier reliance on advanced receivers, link and rank adaptation with a reactive application of resource-limiting ICIC techniques as a self-healing mechanism.

As mentioned in Section II, the interference is stable on a frame basis even though the preferred direction of transmission (UL or DL) is expected to vary quickly. Simply put, the APs/UEs that are sending their reference sequences in the DMRS symbol will also be the ones transmitting data in the rest of the radio frame. This is particularly suited for the usage of advanced receivers, e.g. IRC. Such receivers suppress the interference by exploiting a periodical estimate of the interference-plus-noise covariance matrix [8]. Given our frame structure, such estimate can be performed from the DMRS symbol at every frame, and then used for the computation of the combining matrix that will maximize the SINR in the data part. This design works regardless of the interference source, i.e. the receiver is able to estimate the cross-link interference (e.g., UE-to-UE) as well, solving one of the main concerns related to uncoordinated TDD systems. In a previous work [12], we have studied the potential benefits of the use of IRC receivers in 3GPP-inspired local area scenarios with 40 cells [13]. We compared the performance of IRC receivers to the baseline Maximum Ratio Combining (MRC) receiver [14], which does not exploit any interference estimate and represents the baseline detector, for instance, in LTE. Figure 3 shows the Cumulative Distribution Function (CDF) of the downlink data rates in a full buffer traffic scenario assuming ideal link and rank adaptation with up to 4 MIMO spatial streams. Both cases of frequency reuse 1 (R1) and planned frequency reuse 2 (R2) are displayed. The results show that IRC can significantly improve the data rates (especially for the cells in the worst interference conditions) without bandwidth sacrifices even in very demanding scenarios. Further details are in [12].

Besides the advanced physical layer capabilities, our envisioned 5G system has some additional peculiarities that need to be factored into the RRM design for the 5G data plane:

- multi-Gbps capabilities, which indirectly can lead to wide traffic fluctuations as explained in more detail below;
- fast variability of interference sources, due to independent switching points per cell and topological variations (MTC, self-backhauling).

From the traffic perspective, keeping a flow steadily transmitting at multi-Gbps data rate is a very significant challenge [15]. Networking buffers as well as flow and congestion control mechanisms will tend to make the instantaneous data rate vary quickly from very high to nearly zero. The level of flow aggregation is rather small in a 5G cell due to the presence of few users in local area, leading to an unprecedented burstiness level. The ability to change the transmission direction of the data part every 0.25 ms subframe allows fast adaptation to the traffic demand. The system may more freely schedule the oldest packet in uplink or downlink queues. In LTE-A this is not possible, since the UL/DL configuration determines the direction of each subframe. Furthermore, in LTE-A at most 60% of resources can be allocated to UL [2]. This presents a serious disadvantage in UL-heavy traffic.

However, this flexibility comes at a price: rapid and independent variations of transmission direction pose significant challenges in terms of link/rank adaptation, since the so-called flashlight effect is worsened. Traditional cellular TDD systems have nearby cells with aligned DL/UL switching points to avoid this matter. We are currently investigating novel solutions aiming at preserving the UL/DL flexibility while achieving some degree of predictability of the interference patterns.

Finally, while IRC provides a strong barrier against interference and HARQ represents a second tier of protection, there are still cases in which further interference management such as ICIC is needed. In the case of the data plane, it is preferable to have a reactive mechanism which will improve SINR only when needed and which can respond with certain agility to the traffic variations. One example is described in [16], where essentially a dynamic frequency reuse is achieved in a distributed way. The only requirement for the method in [16] is that each receiver needs to feedback the post-processing
SINR for the transmitter and some interference dependent measurement (e.g., Channel Quality Indicator) for all PRBs.

IV. FEASIBILITY OF NETWORK SYNCHRONIZATION

While link synchronization between AP and UE is needed for coherent data demodulation, network synchronization among multiple APs is the underlying assumption of our design since it allows coordinated operations between APs, e.g. for frequency coordination and interference suppression. In OFDM-based systems, network nodes are considered synchronized in case their time misalignment is within a fraction of the CP duration \( T_{CP} \). In particular, the time misalignment \( \tau_M \) should fulfill the following requirement:

\[
\tau_M < T_{CP} - \tau_D - \tau_{HW} - 2\tau_F
\]

where \( \tau_D \) is the delay spread of the channel, \( \tau_H \) the propagation delay and \( \tau_{HW} \) is the delay response of the hardware filters. Note that in LTE/LTE-A the propagation delay is compensated by using a timing advance technique at the UE, which requires a further hand-shaking procedure before UL transmission can take place [7]. However, given the lower expected propagation delay in local area, we believe it is worth to embed it in the CP at the expense of an extra overhead rather than accepting such additional latency in the UL.

In a local area scenario, root mean square delay spread and propagation delay (assuming a 100 m cell radius) are in the order of \(-100\) ns and \(-170\) ns, respectively [17], while the hardware filter response delay is in the order of \(-50\) ns. As a consequence, by assuming the numerology envisioned in Table I we obtain the following requirement for the time misalignment:

\[
\tau_M < 510 \text{ ns}
\]

Such accuracy level can hardly be achieved with network solutions such as IEEE Precision Time Protocol (PTP) [18]. In LTE-TDD, multiple base stations (BSs) can synchronize to the common reference given by the Global Positioning System (GPS) satellites. However, penetration losses in indoor make not possible to rely on GPS signals for network synchronization. Distributed solutions, where the APs agree on a common timeline without any centralized coordination, have then to be pursued for 5G.

Synchronization at network level can be achieved by ensuring beacon messages exchange among multiple APs, e.g. through the usage of an opportunistic channel. Such channel can be mapped, for instance, on the last OFDM symbol in the frame and used on a contention based manner; the design of the inter-AP communication channel is for further studies. Each AP corrects its local timing upon reception of one or more beacons sent by neighboring APs. In a previous work [19], we have studied different clock update mechanisms for achieving tight OFDM synchronization in a distributed manner in a large network of cells. We evaluated their performance in a local area 3GPP-inspired scenario with 40 cells of apartment [13], by assuming different deployment ratios (DRs), i.e. different probabilities of having an AP at each apartment. We further assumed clocks having nominal precision of 1 part-per-million (PPM) and a 10 ms periodicity of the inter-AP communication channels. Figure 4 shows the CDF of the residual time misalignment between different APs; in 90% of the cases it is possible to achieve a residual error below 200 ns even for DR=100%, and then significantly lower than our requirement. Further details are included in [19]. This justifies our intention of pursuing distributed synchronization for 5G networks.

V. NETWORKING ASPECTS

In the previous sections we have introduced our vision about the lower layers of the 5G RAT. Nevertheless, a more global picture about the architecture that stands behind that, is needed in order to understand how this new system is going to be plugged into the Internet and to the existing mobile networks.

New service paradigms are appearing fast, expanding the current Content Delivery Networks. These paradigms are based on multiple access gateways (GWs) and IP addresses for hosts, such as “multi-homing” service provision. Besides to the traditional networks, Internet of Things (IoT) is boosting the number of connected devices, requiring then the introduction of novel and dynamic network topologies. All the motivations will make IPv6 the standard network protocol in 2020.

Furthermore, local area networks like home- and enterprise networks often utilize shared link IP model for sharing the same subnet prefix with multiple host and to enable shared on-link services like, e.g. printers, file servers, home entertainment services. A 5G system shall naturally support all of it in a smooth way, integrating itself in different types of network topologies, possibly based on heterogeneous systems. For all these reasons, in our vision, a 5G system should be based on an “Ethernet-over-Radio” (EoR) Link Layer (LL).

Figure 5 illustrates the envisioned 5G network based on the switched connectivity, integrated in a heterogeneous, multi-homing IPv6 network.

Such a design has several advantages in supporting the future services. Firstly, an Ethernet-type of LL can be easily translated into an Ethernet IEEE 802.3 LL, allowing the 5G network to be plugged directly into any existing network,
either local area network like home network or the transport network of a Mobile Network Operator (MNO).

Secondly, it would enable mobile Operating Systems (OSs) developers not to care about the used access technology when developing apps or drivers for mobile devices, since many OSs wish to abstract network interfaces as an IEEE 802 type interface by default. Finally, EoR will allow cheaper implementation compared to full IP routing, due to the lower processing power required for operation. But the savings will come also by the smooth integration with the Software Defined Network concept, enabling quick, remote, and eventually automated, network maintenance procedures, e.g. like those allowed by OpenFlow [20] technology.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented our clean slate approach for a concept design of a novel 5G small cell optimized radio access technology, which is envisioned as a TDD-based system. We have proposed a new frame structure which enables low latency, reduced complexity and power consumption, as well as native support for interference coordination schemes and advanced transceivers. A tentative numerology has been presented. The flexibility in the UL/DL allocation offered by the frame structure allows coping with an unprecedented burstiness level. The usage of inter-cell interference coordination techniques is foreseen to be beneficial for the control plane, while the data plane can mostly rely on the native support for multiantenna IRC receivers enabled by our design as well as on HARQ. The potential of IRC in boosting the data rate performance in severe interference limited scenarios has been discussed, as well as the feasibility of tight distributed network synchronization which is the main underlying assumption for our optimized frame based TDD design. Finally, the usage of an Ethernet-over-Radio link layer is proposed for an optimal integration with heterogeneous IPv6 networks. Future work will address the detailed design of novel RRM solutions (e.g., for scheduling/rank/link adaptation) aiming at preserving large flexibility in accommodating the diversity of traffic applications.

REFERENCES