A Flexible Frame Structure for 5G Wide Area

Pedersen, Klaus I.; Frederiksen, Frank; Berardinelli, Gilberto; Mogensen, Preben Elgaard

Published in:
Proceedings of IEEE VTC Fall-2015

DOI (link to publication from Publisher):
10.1109/VTCFall.2015.7390791

Publication date:
2015

Document Version
Accepted author manuscript, peer reviewed version

A Flexible Frame Structure for 5G Wide Area

Klaus Pedersen[(1, 2)], Frank Frederiksen[(1)], Gilberto Berardinelli[(2)], Preben Mogensen[(1,2)]

(1) Nokia Networks, Aalborg, Denmark
(2) Department of Electronic Systems, Aalborg University, Denmark

Abstract—**In this paper we present a 5G frame structure designed for efficient support of users with highly diverse service requirements, including mobile broadband (MBB) data, mission critical communication (MCC), and massive machine communication (MMC). The proposed solution encompasses flexible multiplexing of users on a shared channel with dynamic adjustment of the transmission time interval (TTI) in coherence with the service requirements per link. This allows optimization of the fundamental trade-offs between spectral efficiency, latency, and reliability for each link. The frame structure is based on in-resource physical layer control signaling that follows the corresponding data transmission for each individual user. The principle of in-resource control signaling has numerous advantages; it presents a highly flexible and scalable solution, it allows joint beamforming for control and data transmission, as well as efficient time-frequency inter-cell interference coordination. Numerical results are presented, including simple comparison against LTE.**

Index terms: ***5G, air interface, user multiplexing, service awareness, control channel, latency, overhead.***

I. Introduction

Research towards future standardization of a new 5G air interface is currently in the exploration phase, where academia and industry are presenting their view on possible requirements and candidate techniques to be included in a future system design. Among others, the METIS project has outlined its 5G vision in [1], the 5GNOW project presented their proposal in [2], while the use of more advanced centralized network architectures for 5G has been suggested in [3]. Furthermore, small cell optimized design has been identified as being of particular importance for a new 5G air interface to be able to meet the future mobile broad band traffic requirements [4]-[5].

In this paper, our focus is on presenting a flexible frame structure, which is capable of fulfilling the challenging 5G requirements for efficient support of a mixture of diverse services. Our focus is on macro type of scenarios with wide area coverage, operating on licensed bands for cellular usage below 6 GHz, although striving towards having an agnostic solution that is carrier frequency independent. Efficient use of the spectrum below 6 GHz calls for a flexible air interface design, including harmonized solutions for both frequency division duplexing (FDD) and time division duplexing (TDD), as well as different carrier bandwidths and spectrum aggregation methods. However, in this study the scope is limited to the frame design of a single carrier on continuous spectrum resources. We start by first identifying the main requirements that influence the most on the frame structure, with special emphasis on latency constraints. Following such requirements, a flexible solution is proposed for efficient multiplexing of users having different requirements, assuming an orthogonal frequency division multiple access (OFDMA) air interface structure, where users are scheduled on a time-frequency grid of resources [6]. However, the proposed frame structure is also applicable for other candidate waveforms that offer a time-frequency symbol space for a commonly shared channel per cell. The corresponding relation between physical (PHY) layer control and data channels is outlined and numerical results are presented. Throughout the paper the LTE 4G standard [7]-[8] is used as our reference for motivating and quantifying the benefits of the new 5G frame structure.

The rest of the paper is organized as follows. In Section II we further outline the considered service requirements and introduce the related latency definitions. Section III presents the proposed flexible frame structure, while numerical results appear in Section IV. The paper is closed with concluding remarks in Section V.

II. Service Requirements and Latency Definitions

Efficient support for mobile broadband (MBB) will continue to be important also for 5G. As an example, the International Telecommunications Union (ITU) has recently defined challenging requirements for International Mobile Telecommunications (IMT) at 2020 and beyond [9]. Among others, peak data rates of 20 Gbps and uniform availability of end-user experienced data rates of 100 Mbps to 1 Gbps are listed. Support for MBB requires relative large bandwidth and frequent transmissions. In addition to offering connectivity for humans, 5G should also be designed for efficient machine type of communication (MTC). MTC use cases include massive machine communication (MMC) with large a number of connected low cost devices (e.g. sensors). In this respect, ITU has set a target of being able to support up to 10^6 MTC devices per km^2 [9]. MMC is characterized by infrequent access, typically transmitting only moderate size payloads with relaxed latency requirements. The second class of MTC use cases is mission critical communication (MCC). MCC requires stricter end-to-end latency and high degree of reliability to e.g. support vehicular use cases and factory automation processes. In this context, ITU has set a target to have 1 ms over-the-air communication for a single transmission. Depending on the application, reliability constraints of up to six-sigma (99.99964%) are mentioned [1]. For more information on 5G requirements, see also [10].

Designing a system that supports all of the mentioned MBB, MMC, and MCC targets is rather challenging. Especially since there are fundamental trade-offs in wireless systems between offering high spectral efficiency, low latency, and high reliability [11]. As an example, the performance of MBB can approach the Shannon capacity limit, while there is a cost of reduced spectral efficiency if operating under strict latency and reliability constraints. This essentially calls for a flexible air interface design that allows optimizing each link according to its service requirements. This suggests having a dynamic frame structure that offers the possibility to perform trade-offs between spectral efficiency, latency, and reliability in coherence with requirements per link.

The definition of round trip time (RTT) over the air interface is illustrated in Figs. 1 and 2 for the downlink and the uplink, respectively. Here we have assumed that all scheduling decisions with respect to radio resource allocation will be made on the infrastructure side by the base station. These examples assume a scheduled system design with a common shared channel in the downlink and uplink for each cell. For the downlink, the RTT includes the time for transmitting the data payload (and scheduling grant), denoted the transmission time interval (TTI) – T_{tti}. Subsequently, it takes T_1 seconds for the payload to reach the User Equipment (UE), and for the UE to decode the transmission before it is ready to start sending an Acknowledgement (or Negative acknowledgement). Sending the A/N takes $T_{A/N}$ seconds, followed by T_2 seconds until the base station (BTS) has decoded the A/N and is ready to send a new transmission on the same hybrid automatic repeat request (HARQ) stop-and-wait (SAW) channel, or a corresponding retransmission. Typically, T_1□T_{tti} and T_2□T_{tti}, while $T_{A/N}$□T_{tti}, and hence giving an idea of the required TTI size to meet certain RTT requirements. As an example, for LTE T_{tti}=1ms and RTT=8ms for the downlink [6]-[7], and therefore fails in meeting the 5G latency requirement for MCC.

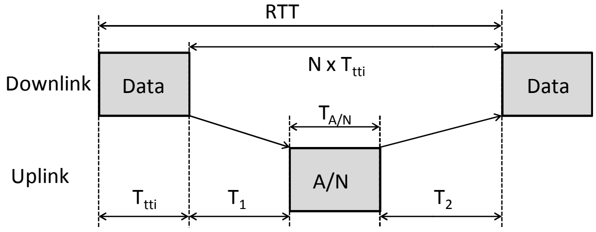

Fig. 1: Downlink round trip time latency budget.

As illustrated in Fig. 2 for the uplink, the BTS first has to schedule the UE by sending a scheduling grant, taking time T_{grant}. Following the scheduling grant, it takes T_3 seconds until the UE is ready to respond and start sending the payload in the uplink. Transmission of the payload takes T_{tti}, followed by T_4 seconds until the BTS has processed the reception and is ready to respond by sending an A/N. As illustrated in Fig. 2, if the payload is erroneously received by the BTS, the negative acknowledgement may (implicitly or explicitly) include a scheduling grant for the corresponding HARQ retransmission. Typically, $T_3 \geqq T_{tti}$ and $T_4 \geqq T_{tti}$, while $T_{grant} \leqq T_{tti}$.

Note that cases with non-scheduled contention based access are outside the scope of this study, although it could be relevant for especially MMC in the uplink.

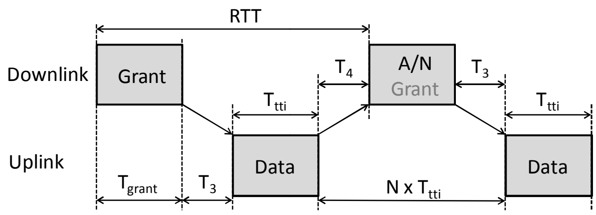

Fig. 2: Uplink round trip time latency budget.

III. Flexible Frame Structure

A. Dynamic multiplexing of users

In order to allow efficient adaptation and optimization for each user in coherence with its service requirements and radio conditions, we propose a flexible frame structure that allows multiplexing of different TTI sizes. The basic concept is illustrated with the time-frequency grid depicted in Fig. 3, where a number of users are flexibly multiplexed over the available resources with different TTI durations. Each tile refers to the smallest allocation unit of time-duration Δt and frequency size Δf. In practice, Δt would equal an integer number of OFDM symbols, while Δf would corresponds to an integer number of subcarriers. In practice those values could equal just a few symbols and/or subcarriers. The value of Δt determines the minimum TTI size for scheduling a user, as well as the resolution for other TTI scheduling options. Given the most stringent latency requirement of 1ms, combined with the practical reasoning in having support for narrow bandwidth and forward error correction coding, there are at least two options for fulfilling the latency budget if doing the reverse calculations based on the delay budgets outlined in Section II (see Figs. 1 and 2). Namely Δt=0.2ms or Δt=0.25ms. Setting Δt=0.2ms leads to the processing time requirement $T_1+T_2 \leqq 0.6$ms for fulfilling the RTT latency budget for the downlink, and $T_3+T_4 \leqq 0.6$ms for the uplink budget. Selecting Δt=0.25ms results in $T1+T2 \leqq 0.5$ms and $T_3+T_4 \leqq 0.5$ms. As will be discussed in greater details in Section IV, the selection of Δt=0.2ms also influences on the relative control overhead. Notice that the 5G small cell concept presented in [4] proposes Δt=0.25ms.

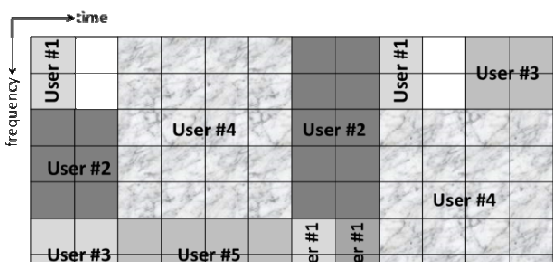

Fig. 3: Flexible multiplexing of users with different requirements.

The proposed frame structure allows to dynamically adjust the TTI size for each scheduling instant of the users. Referring to Fig. 3, user #1 is scheduled with short TTI size of Δt to fulfill the RTT requirement for MCC. However, scheduling all users with this short TTI is not optimal. As mentioned in the previous section, the usage of long TTIs allows benefiting from larger coding gains to approach the Shannon capacity limit, as well as imposes lower control overhead. This comes, however, at the expense of a latency increase; in that respect, the usage of longer TTIs is more beneficial for MBB users. With reference to Fig. 3, users #4 and #5 carry MBB data, while users #2 and #3 correspond to an intermediate case, where scheduling is conducted for a medium size TTI. The fact that the frame structure allows setting the TTI size per scheduling grant furthermore offers the possibility to optimize MBB services using the Transmission Control Protocol (TCP) as follows: During the initial data transmission session, the end user experienced performance is primarily determined by the RTT due to the TCP slow start procedure (i.e. TCP flow control). It would therefore be advantageous to perform first scheduling of MBB TCP users with short TTIs, followed longer TTI sizes when reaching steady state operation.

It should furthermore be noted that although the example pictured in Fig. 3 assumes availability of downlink transmission resources all the time as in FDD, the same principle is applicable also for a TDD system, but naturally having to comply with the constraints on whether the carrier is currently configured for uplink or downlink usage. Secondly, although not shown in Fig. 3, the proposed frame structure also allows users to be scheduled on non-consecutive frequency blocks to benefit from frequency domain scheduling diversity as also known from LTE [8].

B. *In-resource control signaling*

In order to efficiently support the flexible user allocation, we propose the usage of *in-resource PHY signaling* for the downlink control plane. The main idea is to use embedded "on-the-fly" information to the users on its allocated time-frequency resources, as well as the additional information which is needed to decode the data. This is referred to as the users scheduling grant sent on a dedicated physical layer control channel (CCH). The scheduling grant contains information such as e.g. the allocated time-frequency resources for the users (number of consecutive time symbols per TTI, subcarrier allocation), the modulation and coding scheme (MCS), HARQ information and multi-antenna transmission information (e.g. number of spatial streams). The in-resource CCH is mapped at the start of the resource allocation for the user in the first time symbol(s) and over a limited part of the frequency resources, as shown in Fig. 4. In principle, the CCH can be sent with variable settings for the MCS to match the channel quality of the user, thus saving part of the resources for users experiencing favorable channel conditions.

Note that the flexible allocation of in-resource CCH differs significantly from the solutions adopted in the current LTE standard. LTE features a strict periodic time-division separation of the physical layer control and data, by sending the control information in the first set of OFDM symbols [7], [13]. For example, the Physical Downlink Control Channel (PDCCH) is transmitted over the full system bandwidth in the first OFDM symbols of the subframe, having a fixed duration of 1ms. Time separation of the control and data plane has also been suggested for the recently proposed 5G small cells concept optimized for TDD mode [4].

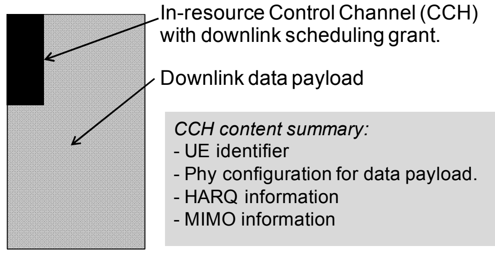

Fig. 4: Position of the in-resource control channel with respect to the data payload.

The same flexibility as illustrated in Fig. 3 should also be supported for uplink data transmissions, so users can be scheduled with different TTI size, as well as with different transmission bandwidths. As an example, a coverage limited user would benefit from being scheduled on a narrow bandwidth with relative long TTI size. Each uplink data transmission needs an uplink scheduling grant that is sent in the downlink. In that respect, we opt for an uplink grant solution as illustrated in Fig. 5. In Fig. 5(a), downlink and uplink grants are multiplexed on the same control resources dedicated to a specific user, with the fundamental difference that the downlink grant provides information for decoding the associated data block, while the uplink grant is pointing to a successive uplink data transmission allocation. In case downlink data transmission does not occur for the user, the uplink grant can be transmitted independently as shown in Fig. 5(b), where multiple uplink scheduling grants are stacked in one downlink resource unit; i.e. the scheduling of users #3 and #4 in the uplink, while scheduling users #1 and #2 in the downlink.

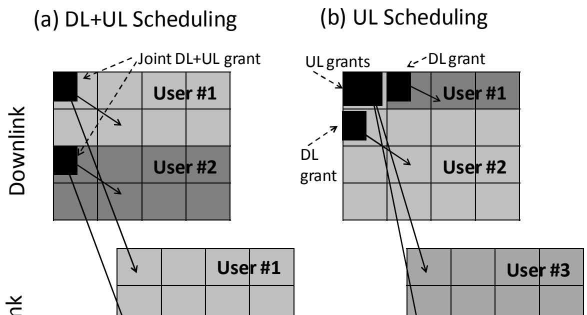

Fig. 5: Downlink in-resource signaling for both downlink and uplink user scheduling.

C. *UE monitoring of scheduling grants*

In the interest of UE power consumption and CCH processing burden, it is advantageous to allow configuration of restrictions for the resource pattern where the UE should listen for potential scheduling grants. It should therefore be possible to instruct users via higher layer signaling to monitor only a subset of the transmission resource grid to alleviate the burden of scheduling grant monitoring. For example, with reference to Fig. 3:

- User #1 could be instructed to monitor at every minimum TTI unit, with limited options in the frequency domain;
- User #2, User #3 and User #4 could be instructed to monitor every second resource unit;
- User #5 could be instructed to monitor every fourth resource unit in time.

As the usage of the in-resource signaling decouples the CCH bandwidth from the system bandwidth, it allows e.g. low cost MTC devices to be instructed to only monitor and operate on part of the bandwidth, which will also be favorable from a terminal energy consumption point of view. This is contrary to LTE, where the UEs need to monitor the full carrier bandwidth as the PDCCH is transmitted on full bandwidth [8].

D. Beamforming and interference coordination

Beamforming and massive multi-antenna techniques is envisioned to be an important technique for 5G. The frame structure must therefore be designed to support such schemes from the very start. The proposed frame structure supports the possibility of using beamforming techniques for strengthening the reliability of both the CCH and data transmission for each specific UE, thus improving the coverage and capacity. This is possible due to the in-resource position of the CCH, which allows using beamforming for both the CCH and the corresponding downlink data transmission in case of single stream transmission, as the same set of reference symbols are applicable for channel estimation (and coherent demodulation) for those PHY channels. In LTE, the PDCCH is transmitted with open loop transmit diversity mode, due to the time-wise disjoint position of the CCH (PDCCH) and data (PDSCH), where common reference symbols (CRS) are used for both the PDCCH and PDSCH transmissions [8]. In LTE-A there is support for dedicated reference symbols for the PDSCH demodulation when using Transmission Mode 9, while the PDCCH is still relying on common reference symbols. Additionally, with LTE-A, there is partial support for beamforming on the CCH through the E-PDCCH, but initial configuration would still need to be addressed through PDCCH.

Furthermore, inter-cell interference is also expected to be a challenge for the 5G-era, calling for both the possibility to use efficient network-based inter-cell interference coordination (ICIC) techniques, as well as receiver-based interference cancellation/suppression schemes. Since the in-resource CCH signaling for the proposed frame structure follows the data allocations, it allows efficient time-frequency domain ICIC for both the CCH and data transmission in case of synchronized base stations. As an example, if cell A mutes a certain set of its time-frequency domain resources, then the users that the neighboring cell B schedules on that set of time-frequency domain resources will experience improved SINR for both the CCH and the data reception. The same flexibility for ICIC is not possible for LTE due to strict time-division of PDCCH and data in each subframe, where the PDCCH transmission is distributed over the full cell bandwidth [8], [12].

IV. Numerical Evaluation

In this section we present a numerical evaluation of the CCH overhead for the proposed frame structure, as well as discuss the comparison against LTE in greater details. We consider cases with scheduling of users with different TTI sizes, on different fractions of the available carrier bandwidth, as well as users in challenging and favorable signal-to-interference-noise ratio (SINR) conditions. As mentioned in the introduction, we focus on wide area macro cases with operation on carrier frequencies below 6 GHz, i.e. similar as used for LTE. As the physical layer numerology for 5G have not yet been fixed, we adopt LTE as our baseline. This allows a fair comparison of the CCH overhead for the proposed 5G frame structure and LTE. We thus assume that within 20 MHz carrier bandwidth and 1ms time-interval, there are a total of 16800 Resource Elements (REs). One RE is equivalent to one subcarrier symbol. Dividing this resource space into segments of Δt=0.2 ms and Δf=2.5 MHz gives 420 REs per allocation unit. In this calculation, we have taken the freedom to assume that the PHY numerology allows such segmentation; a fine tuning of the PHY numerology for 5G wide area is out of scope here and left for future work.

As the scheduling grant on the CCH for the proposed 5G frame structure carry the same information as the LTE scheduling grants on PDCCH, we assume the same structure and air interface decoding performance for 5G as in LTE. In LTE, the PDCCH for a user in good SINR conditions can be sent with QPSK and coding rate 7/10 on a total of 36 REs. This would result in a reception block error rate (BLER) of less than 1% for the CCH decoding if the post-detection SINR is 2 dB, or higher. On the other extreme, the PDCCH could be sent with QPSK rate 1/11 to obtain the same 1% BLER for users in challenging SINR conditions of – 6dB. The latter requires 288 REs. Notice that for the 3GPP defined macro scenarios, less than 1% of the users are having a post-detection SINR of -6 dB (assuming standard 2x2 single-user open loop transmission diversity with subsequent receive diversity operation). More details on the LTE PDCCH performance can be found in [13]. In addition to the CCH overhead, it is assumed that 10% of the REs are used for reference symbols to facilitate channel estimation and coherent demodulation.

Given these assumptions for the required number of REs for the CCH, we can calculate the relative CCH overhead for the proposed 5G frame structure. The relative CCH overhead is defined as the ratio of used REs for the CCH versus the total number of used REs. For the LTE frame structure, the control overhead equals 7%, 14%, or 21% depending on whether 1, 2, or 3 OFDM symbols are configured for CCH usage per TTI.

Figs. 6 and 7 show the CCH overhead as a function of the TTI duration and the relative scheduled bandwidth (compared to 20 MHz). The results in Fig. 6 are for the very challenging case where the user is experiencing bad channel conditions with SINR=-6dB. For the users in bad channel conditions, the CCH overhead is very high (up to 75%) if scheduled with the shortest TTI (0.2ms) on a small fraction of bandwidth. This is simply due to the relative high number of required REs for the CCH. However, if the scheduling resource allocation is increased, the CCH overhead decreases considerably. A minimum of 1.2% overhead is obtained for scheduling on the full bandwidth with a TTI size of 2 ms. In the case where the user experience good channel conditions (2 dB SINR, or higher), the trends are the same but the CCH overhead is kept lower than 10% for all the depicted allocation options (Fig. 7). Note that adopting a larger CCH overhead for short TTIs is tolerable since short TTIs are expected to be set for MCC users with tight latency and lower throughput requirements. MBB users with longer TTIs benefit instead from a low CCH overhead.

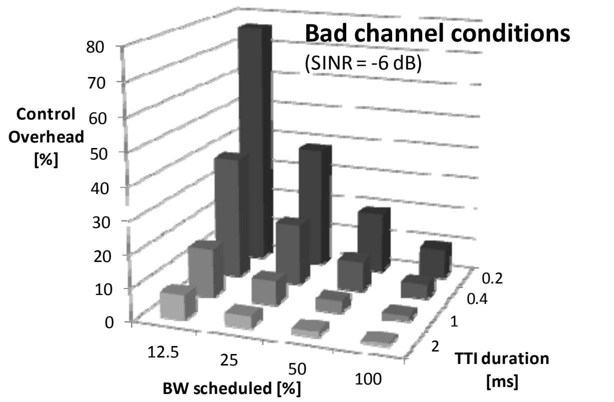

Fig. 6: Control overhead for UEs in bad channel conditions.

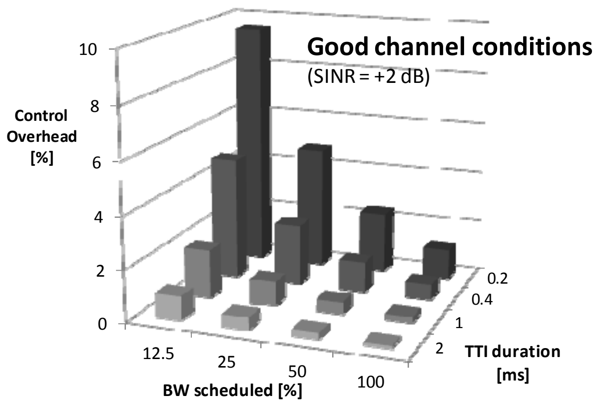

Fig. 7: Control overhead for UEs in good channel condition.

The results in Figs. 6 and 7 essentially show that the CCH overhead scales nicely with the scheduling of users due to the in-resource CCH signaling. It allows trade-offs between CCH overhead and TTI size, or equivalently RTT. The fact that the CCH overhead is not hard limited to values of 7%, 14%, and 21% as in LTE, present a more flexible solution, where CCH blocking is further reduced; see results on LTE PDCCH blocking in [13] with realistic QoS-aware scheduling.

V. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a flexible 5G frame structure. It is designed to serve users with highly diverse service requirements and radio conditions, allowing resource allocation optimization on a per link basis. The concept is based on in-resource physical layer control signaling that follows the corresponding data transmission for each individual user. This principle offers numerous advantages; it presents a highly flexible and scalable solution, it allows joint beamforming for control and data transmission, as well as efficient time-frequency domain ICIC. The numerical results show attractive values for the control overhead, as well as the flexibility to trade such overhead versus use of short TTIs for achieving low RTT.

Given these observations, it is suggested to continue the work on such a frame structure. Among others, it remains to be studied how to arrange downlink common channels like system broadcast information, as well as how to most efficiently facilitate multiplexing of uplink control information such as HARQ feedback, channel state information (CSI), etc. Thus, further research is needed before drawing final conclusions on the 5G frame structure.

REFERENCES

[1] A. Osseiran et al., “Scenarios for 5G Mobile and Wireless Communications: The Vision of the METIS project,” IEEE Communications Magazine, May 2014.

[2] G. Wunder, et.al., “5GNOW: Non-Orthogonal Asynchronous Waveforms for Future Mobile Applications”, IEEE Communications Magazine, February 2014.

[3] C. Lin et al., “Towards Green and Soft: A 5G Perspective”, IEEE Communications Magazine, February 2014.

[4] P. Mogensen et al. “5G small cells optimized radio design”, accepted in Globecom 2013, International workshop on emerging techniques for LTE-Advanced and Beyond 4G.

[5] N. Bhushan, et.al., “Network densification: the dominant theme for wireless evolution into 5G,” IEEE *Communications Magazine,* February 2014.

[6] J. Vihriälä, et.al., “On the waveforms for 5G mobile broadband communications”, in IEEE Proc. VTC-spring, May 2015.

[7] A. Larmo, et.al., ”The LTE Link-Layer Design”, in IEEE Communications Magazine, April 2009.

[8] K.I. Pedersen, et.al., ”An Overview of Downlink Radio Resource Management for UTRAN Long-Term Evolution,” IEEE Communications Magazine, July 2009.

[9] IMT Vision – “Framework and overall objectives of the future development of IMT for 2020 and beyond”, International Telecommunication Union (ITU), Document, Radiocommunication Study Groups, February 2015.

[10] E. Dahlman, et.al., “5G Wireless Access: Requirements and Realization”, IEEE Communications Magazine - Communications Standards Supplement, December 2014.

[11] B. Soret, et.al., ”Fundamental Tradeoffs among Reliability, Latency and Throughput in Cellular Networks”, IEEE Proc. GLOBECOM, Dec. 2014.

[12] G. Boudreau et al., “Interference Coordination and Cancellation for 4G Networks,” IEEE Commun. Mag., vol. 47, no. 4, Apr. 2009.

[13] D. Laselva, et.al., “On the impact of realistic control channel constraints on QoS provisioning in UTRAN LTE”, in IEEE Proc. VTC 2009 fall, September 2009.