Achieving low latency and energy consumption by 5G TDD mode optimization

Eeva Lähetkangas (1), Kari Pajukoski (1), Jaakko Vihriälä (1), Gilberto Berardinelli (2), Mads Lauridsen (2), Esa Tiirola (1), Preben Mogensen (2,3)

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

Abstract— **The target for a new 5G radio access technology is to support multi-Gbps and ms latency connectivity simultaneously at noticeably lower energy consumption and cost compared to the existing 4G technologies, such as LTE-Advanced. Extremely short air interface latency is required to achieve these requirements in a TDD-based local area network. In this paper, we discuss how the required short TDD latency can be achieved and further utilized in 5G physical air interface. First, we investigate the enablers and limits of TDD latency by analyzing the performance of OFDM in different channel environments and discussing on the consequent frame length limits. We then provide a description on how the achieved short TDD latency can further be utilized to enable remarkably low energy consumption. A numerical analysis comparing the battery life time of the suggested 5G TDD air interface and LTE is provided, showing remarkable gains for the 5G air interface concept.**

*Index terms- **5G, latency, RAT, local area, 3GPP, LTE, TDD, OFDM, TTI, cyclic prefix, subcarrier spacing, frame structure, energy consumption, M2M, wake-up time***

I. INTRODUCTION

5G Radio Access Technology (RAT), stretching far beyond year 2020, is expected to enable a highly scalable service experience by allowing people and machines to enjoy a virtually zero latency gigabits data rate on demand. 5G is targeting much higher throughput, spanning towards higher carrier frequencies and wider bandwidths, at the same time reducing energy consumption and costs compared to the existing 4G technologies. [1]

Densification of access points (or base stations), is a classical means for providing increased capacity and improved coverage in wireless networks. This further leads to very dense small cell deployments and focuses the 5G air interface scope especially on local area (LA) networks. Consequently, multi-hop relaying in the form of wireless backhauling links, access links and device-to-device (D2D) links may be required in order to provide sufficient coverage for the Gpbs data flow. Time division duplexing (TDD) is considered a more attractive duplexing method than frequency division duplexing (FDD) in 5G dense deployment environment, as spectrum flexibility is seen more important than coverage. TDD is more flexible than FDD to support bursty traffic demand and new connection types, such as wireless backhaul or D2D. [1]

The air interface latency, concerning both control and data information, becomes critical in order to meet the latency requirements of <~1 ms. [1] This is especially important in a scheduled TDD system where several TDD cycles are required for delivering one round trip transmission related to control signaling or data transmission. Also, the large increase of 5G data throughput leads to the need of transmitting and processing a larger amount of data, consequently imposing demands on baseband processing. A baseband system can cope with the increased throughput demand by decreasing the latency. From the air interface perspective this essentially leads to the transmission of shorter blocks of data in time and wider blocks in frequency.

Besides the achievement of high data rates, latency reductions at the air interface level also become vital to enable energy savings and long battery life time. Fast transitions between sleep and active modes, short active time with high data rate together with low sleep mode power consumption are required to guarantee multiple years of lifetime for a small low cost battery. On air interface level, the need to transmit control and user data quickly in time domain leads again to the demand of fast link direction switching and to short transmission time interval (TTI) length.

The air interface latency of TDD Long Term Evolution Advanced (LTE-A) [2][3] is limited by its physical frame structure. It is possible to include at maximum two uplink (UL)/downlink (DL) switching points inside one 10 ms radio frame, which sets the hard limit for the air interface latency. This is clearly not achieving the 5G physical layer latency target. Evolutions of LTE-A will not be able to support major latency reductions due to the restrictions of incremental evolution. For example, changes in the numerology and frame structure design for enabling latency reductions can hardly be introduced for backwards compatibility reasons. Consequently, there is a need of a new 5G air interface enabling the required physical layer latencies.

In this paper, we discuss the enablers and limits of low air interface latency. We establish our study on the recently proposed TDD- and OFDM-based 5G physical air interface frame structure. [4] Based on channel simulations in both indoor and outdoor environment, we analyze the limitations of the frame length and latency enabled by the proposed frame structure and discuss the further design approaches to overcome these limitations. We also provide an analysis on how the achieved short TDD latency can further be utilized to achieve extremely low energy consumption and compare the obtained 5G battery life time estimations to the corresponding limits of LTE. The paper is structured as follows. In Section II, we describe TDD specific latency requirements and previously proposed solutions. In Section III, we discuss the selection of the most suitable waveform from latency perspective. In Section IV, the limits of the frame length are analyzed and alternatives how to overcome these limit are discussed. Section V provides an analysis of wake up latency and its relationships to power consumption. Finally, Section V concludes the paper.

II. SCHEDULIG LATENCY IN A TDD SYSTEM

Most traffic can be assumed to be user or device-initiated, independently on whether the data is generated or consumed by the user equipment (UE). In a scheduled system, several TDD cycles are required for delivering even one scheduled round trip transmission of control or data signaling, such as request-response or data-acknowledgement pair. Figure 1 illustrates UE initiated data reception / transmission procedure in a scheduled system, requiring 4 TTD cycles from signaling perspective: one TDD cycle for the request signal in UL, possibly one TDD cycle for resource assignment signaling in DL and at least one TDD cycle for the actual data transmission, either to UL or DL, followed by the corresponding acknowledgement. Consequently, the total air interface latency for this procedure is hard limited by the minimum enabled UL/DL switching time, often restricted by a certain UL/DL switching period. This leads to the requirement of flexible and fast link direction switching and increases the importance of short TDD switching guard times between the link directions.

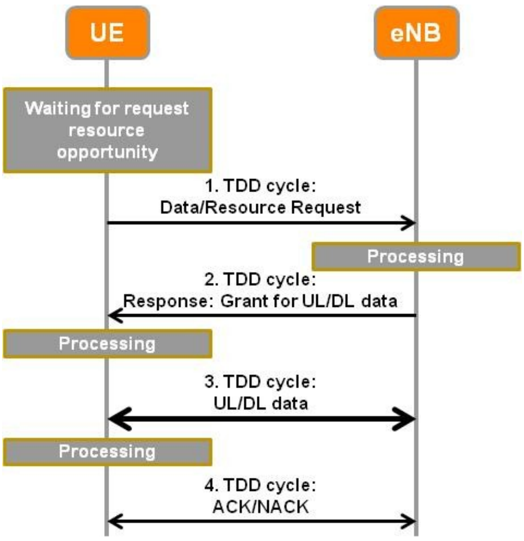

Figure 1. Required TDD cycles for UE initiated data traffic.

Deployment densification with smaller cell sizes and utilization of higher carrier frequencies with larger bandwidth provide remarkable enablers for reducing air interface latency. Smaller cell sizes lead to decreased propagation losses, while delay spread may also be expected to decrease when moving towards utilization of higher carrier frequencies with larger bandwidth. The characteristics of LA environment together with certain expected improvements in component technology within the 5G timeframe, such as shorter TDD link direction hardware switching time, will eventually provide possibility for a TDD numerology optimized for dense deployment, e.g. shorter cyclic prefix (CP) and guard period (GP) times compared to the existing systems. [4] Since the overhead from the guard times becomes significantly smaller, this new 5G optimized numerology further enables shorter frames and more frequent link direction switching. Further, short OFDM symbols correspond to large subcarrier spacing, which is robust to the increased oscillator phase noise at higher carrier frequencies. These properties can further be utilized to design a 5G dense deployment optimized physical frame structure. In [4], we proposed a bi-directional control structure for each TDD frame, enabling opportunity to transmit/receive scheduling information (request/grant) together with synchronization signaling in every frame. The principle is shown in Figure 2. In [4], it was also demonstrated that very short frame lengths, such as 0.25 ms, are feasible with the proposed frame structure from the guard and control overhead point of view, using relatively low 60 kHz subcarrier spacing.

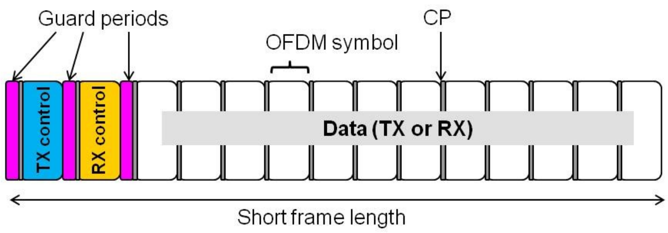

Figure 2. Proposed 5G physical frame structure.

III. RELATIONSHIP BETWEEN WAVEFORMS AND LATENCY IN A TDD SYSTEM

The demand for frequent link direction switching and the consequent need for short TDD guard times sets demands on the modulation methods suitable for a TDD-based air interface for 5G dense deployment. From latency perspective, it would be beneficial to adopt waveforms having good time localization, such that the aforementioned time overheads can be efficiently reduced. In that sense, orthogonal frequency division multiplexing (OFDM) modulation is well known to fulfill such time localization requirement due to the usage of the CP. This is an important benefit compared to filter-bank multi-carrier (FBMC) based waveforms, that are well-localized in frequency but correspondingly dispersed in time dimension. Consequently, transmission of a frame consisting of FBMC symbols is subjected to pre- and post-tails, leading to the need of longer TDD guard times between the link directions. These tails could be shortened but only with a cost of spectral re-growth [5], ultimately leading to the conclusion that from the latency perspective, it is justified to assume the OFDM waveform (and its enhancements) to be the most suitable basis for a 5G air interface designed for a dense deployment environment.

IV. LIMITS DUE TO THE OFDM FRAME LENGTH

The air interface latency of a TDD system is limited by its physical frame structure and the frame length. In this section the limits of frame length with OFDM are analyzed. We first analyze the spectral efficiency performance of OFDM in different channels and then discuss how the obtained numerology affects the frame length of the 5G frame structure illustrated in Figure 2.

The spectral efficiency performance of OFDM is computed taking into account CP overhead and loss due to too short CP. The CP length should match the channel delay spread if no signal to noise ratio (SNR) degradation is accepted. In addition, if timing advance (TA) is not used, also propagation delay needs to be compensated by the CP. The spectral efficiency (SE), measured in bps/Hz, of an OFDM link for one spatial layer is

$$SE = \frac{T_{OFDM}}{T_{CP} + T_{OFDM}} \min\left\{SE_{max}, \log_2\left(1 + \frac{\gamma}{\alpha}\right)\right\}, \qquad (1)$$

where T_{CP} and T_{OFDM} are the cyclic prefix length and OFDM symbol length (inverse of subcarrier spacing), respectively, SE_{max} is the maximum achievable spectral efficiency (limited by largest modulation and coding rate), γ is SNR and α is SNR degradation due to non-ideal modulation and coding.

SNR degradation due to CP length is derived in [6]. The SNR is a function of channel signal-to-noise ratio, CP length, OFDM symbol length, and channel power delay profile. We estimate the SE for two channel power delay profiles (PDPs), namely ITU indoor hotspot (InH) NLOS and ITU urban micro (UMi) NLOS, and four subcarrier spacing (SCS) values. We use 40 dB channel SNR (including EVM). Largest modulation is 256-QAM with rate 0.9, which gives SE_{max} = 7.2 bps/Hz. Implementation loss due to turbo coding and incremental redundancy, α, is estimated to be 2 dB. Results are shown in Figure 3 and Figure 4. Note that overhead from control channels is not taken into account.

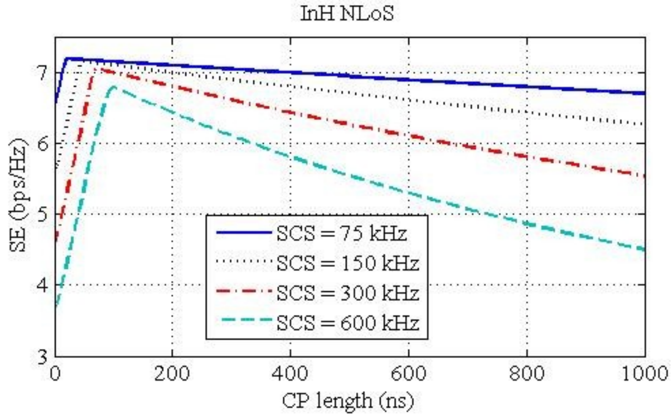

Figure 3. Spectral efficiency as a function of CP length; InH.

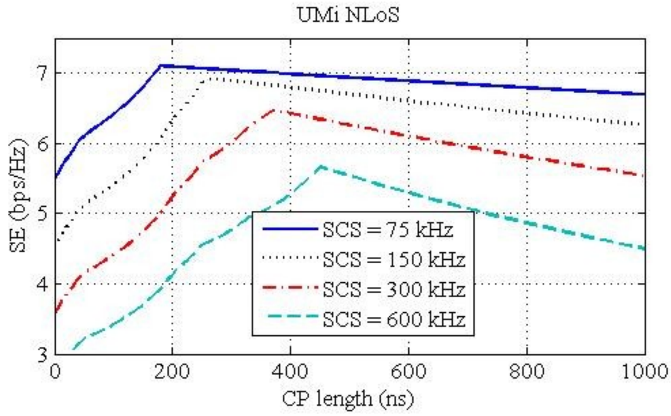

Figure 4. Spectral efficiency as a function of CP length; UMi.

We note that the performance in urban micro is considerably worse than in indoor hot spot. The reason is that UMi has much longer delay spread and requires longer CP which increases overhead. We also note that for both cases the narrowest used subcarrier spacing (75 kHz, corresponding to T_{OFDM} = 13.33 µs) gives the best performance. The reason is that the long OFDM symbol is most resistant to the SNR degradation due to too short CP. We can see that a CP length of ~400 ns is feasible for both indoor and outdoor environments.

Unfortunately, with short frame lengths, large T_{OFDM} also means large control channel overhead, as shown in Figure 2. In order to maintain bi-directional control in each frame, enabling opportunity to both receive and send scheduling information, we need two control symbols per frame. The amount of such control overhead is plotted in Figure 5 as a function of frame length and with different SCS values. We can notice that in order to keep the overhead due to the control symbols in reasonable <15% limits (similar control overhead as in [4]) with small frame lengths, such as on the range of ~100-200 µs, utilization of larger subcarrier spacing values of at least 150 kHz would be required.

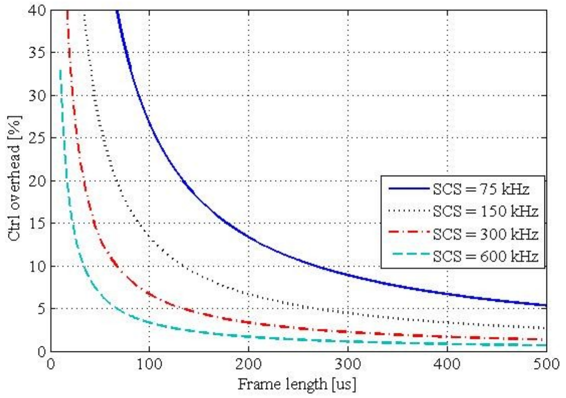

Figure 5. Amount of control overhead vs. frame length.

Besides the control symbols, an obvious overhead of OFDM waveform is represented by the usage of CP. Increasing the subcarrier spacing further increases the relative CP overhead, thus decreasing the achieved spectral efficiency. For example the overhead due to CP and GP increases ~7% when increasing the SCS from 75 kHz to 150 kHz and when using 400 ns CP length and 100 µs frame length. In order for the minimized frame length to be feasible from capacity point of view, this increased guard overhead should be compensated by some means. A few examples of certain design principles that could be used for this compensation are presented in the following. More detailed investigation of these techniques is left for further study.

A. Utilization of different SCS for control and data

The subcarrier spacing used for the frame's control part could be set larger than for the data part, leading to shorter control symbols and to reduced and possibly configurable amount of control overhead, together with longer data symbols with feasible CP overhead. In this approach, control and data planes can be splitted to separate receivers. For example, if assuming 150 kHz subcarrier spacing for the two control symbols and 75 kHz for data symbols, we end up with six data symbols and total overhead of ~20% for 100 µs frame length. This is a 7% improvement compared to the case with constant 150 kHz subcarrier spacing for the whole frame and already a feasible value from SE point of view.

B. Utilization of timing advance

Another alternative for the increased guard overhead compensations is the usage of TA. As mentioned above, the CP duration should cope with the delay spread of the channel as well as the propagation delay in case a TA process is not taking place. The usage of TA can allow for shorter CP durations, but it also increases the system complexity since a hand-shaking procedure between UE and eNB needs to take place. Also, utilization of TA causes time misalignment between DL and UL symbols, which entails that stabilization of the interference covariance matrix for interference rejection purposes [7] is not possible.

C. Utilization of flexible cyclic prefix length

In scheduled systems, the CP duration is hard-coded in the system design in order to fit a predefined frame duration. This may lead to system inefficiencies since such CP length may exceed the instantaneous perceived delay spread, leading to throughput losses. Conversely, in case the delay spread is larger than the CP duration, some Block Error Rate (BLER) increase at the receiver is expected. We have recently proposed the usage of a novel waveform: zero-tail Discrete Fourier Transform – spread – OFDM (ZT DFT-s-OFDM), aiming at dynamically setting the overhead which is needed to cope with the delay spread of the channel rather than relying on hard-coded CP [8]. The main principle of ZT DFT-s-OFDM is that the CP is replaced with a set of very low power samples (nearly zero-power) which are obtained as a part of the inverse fast Fourier transform (IFFT) output. The length of the zero-part can be tuned dynamically according to the delay spread without modifying the system numerology. Surprisingly, such waveform has also the advantage of significantly better frequency localization with respect to baseline OFDM. For further details, refer to [8].

V. Energy Consumption and Wake Up Time

Reduced physical layer latency can be utilized to reduce energy consumption of the network devices. Recent measurements on LTE smartphones, [9], have shown that the power consumption ratio between the UE being ON and sleeping is at least 1:35. Maximizing the UE's sleep time is therefore essential to improve the battery life. When the UE wakes up from the low-power sleep mode to receive or transmit data it follows the procedure illustrated in Figure 1. From the power consumption perspective the procedure can be divided into five different states: sleeping, waking up (wup), synchronizing, transferring data, and powering down (pd). The states are illustrated in Figure 6.

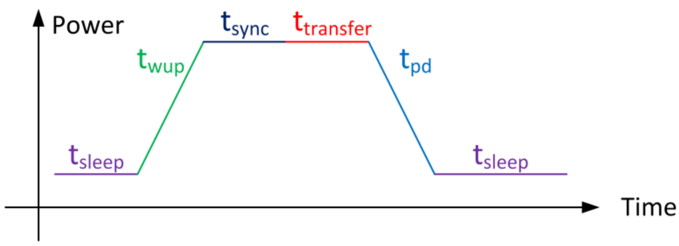

Figure 6. UE power states.

The power up/down states when changing from sleep to active mode and vice versa are device dependent, and not significantly affected by the wireless standard. On the contrary the time duration of the synchronization and transfer states, and therefore the total energy consumption, are directly related to the latency and frame length of the wireless standard.

When an LTE UE exits the deep sleep mode it has to receive and decode the primary and secondary synchronization signals (PSS and SSS) and the Physical Broadcast Channel, which contains the System Frame Number. The PSS and SSS are only sent by the eNodeB every 5 ms, leading to significantly long synchronization time. In [9] it was noted that the t_{sync} is 16 ms, when the UE has been in deep sleep mode. Due to the shorter frames and the enabled allocation of synchronization signals in every frame, the 5G UE will be able to synchronize much faster. This will save energy because it will reduce the total ON time.

The time to transfer data in 5G will also be shorter compared to LTE due to the increased data rates, which reduces the actual transfer time. Furthermore, the shorter 5G frame lengths enable the total time required for the transmission of scheduling request (SR) and scheduling grant (SG) to be reduced. In TDD LTE, the time between when UE sends a SR and receives the SG is not specified. However, the minimum required time is limited by the LTE physical frame structure and depends on the used UL/DL configuration. The number of frames between the grant and the actual transmission is specified to 4-7 frames [3]. Furthermore it has been specified that the base station has at least 4 ms to process the data before it transmits an ACK/NACK, and therefore the total time for a single data transmission is estimated to be at least 10 ms. In practice, the actual wake-up from sleep to RRC connected state requires also radio RRC connection setup procedure with establishment of radio bearers.

In the 5G concept, due to the bidirectional control plane embedded to each frame, the UE initiated data transmission/reception can be achieved within 5 frames + 1 symbol as illustrated in Figure 7. Due to the short frame length, the total time required for this procedure is estimated to be ~1.27 ms (with 0.25 ms frame length) or even shorter with shorter frame length.

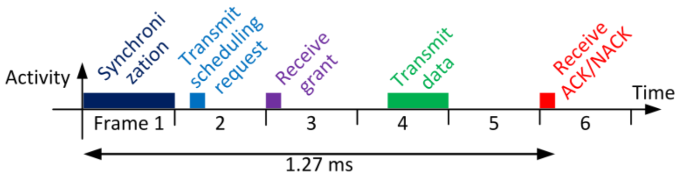

Figure 7. Procedure for UE initiated transmission in 5G.

The short physical layer latency enabling fast transitions between sleep and active modes brings significant gains in terms of power consumption. Figure 8 below shows results of theoretical numerical calculations of the battery life as a function of synchronization and reception opportunity times and the number of transmissions per second. Results are presented for both machine-to-machine (M2M)-optimized LTE [10] and 5G systems with different frame structures but with same power consumption values and with 32400 Ws low cost battery. In the transmission phase, the 5G UE is assumed to follow the procedure according to Figure 7. In the reception phase, the UE is assumed to do only synchronization and thereafter receive a short paging message with length of one OFDM symbol. Based on Figure 8 we can conclude that it is clearly more efficient to use the proposed 5G system over LTE due to the aforementioned reasons.

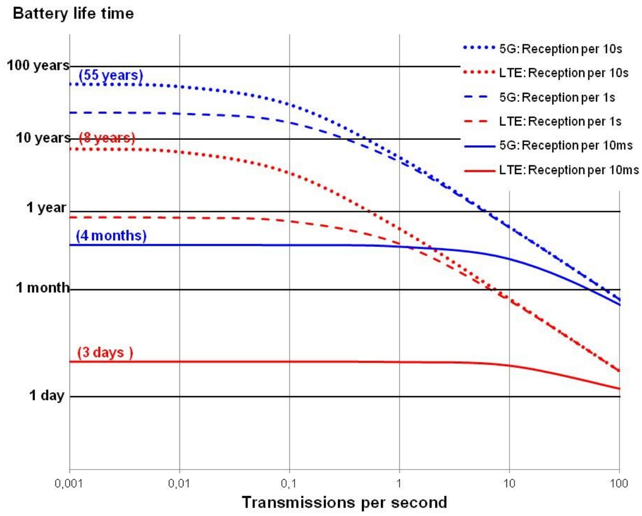

Figure 8. Battery life time comparison between LTE and 5G.

VI. CONCLUSIONS

Latency reductions on air interface level become vital in order to obtain multi-Gbps data rates and to enable energy savings in the 5G TDD-based local area networks. In this paper, we analyzed the performance of OFDM in different channel environments and discussed how the obtained numerology affects the frame length of the previously proposed 5G TDD frame structure. In order to achieve extremely short frame lengths with feasible control overhead, relatively large subcarrier spacing needs to be used. There are several alternatives for decreasing the consequently increasing CP & GP overhead, such as utilization of different subcarrier spacing and receivers for control and data parts of the frame, utilization of TA or usage of adaptive CP length. Further investigation of these techniques is left for further study.

The short TDD latency achieved with the proposed 5G frame structure with very short frame length enables really short UE initiated data transmission/reception times, such as <1.5 ms, including synchronization, scheduling signaling and actual data transmission with acknowledgement. This is a significant improvement compared to LTE and it enables devices to be in the energy-efficient sleep mode for majority of time, consequently leading to very long battery life. A numerical analysis shows that with the same power and battery assumptions the proposed 5G concept has approximately 7-40 times lower energy consumption compared to M2M optimized LTE design.

REFERENCES

[1] P. Mogensen, K. Pajukoski, B. Raaf, E. Tiirola, E. Lähetkangas, I.Z. Kovács, G. Berardinelli, L.G.U. Garcia, L. Hu, A.F. Cattoni, "B4G local area: high level requirements and system design", IEEE Globecom, 2012.

[2] 3GPP TS 36.211, Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation, Available: www.3gpp.org.

[3] 3GPP TS 36.213 v.10.8.0, Physical layer procedures for Evolved Universal Terrestrial Radio Access (E-UTRA). Available: www.3gpp.org.

[4] E. Lähetkangas, K. Parjukoski, E. Tiirola, G. Berardinelli, I. Harjula, J. Vihriälä, ” On the TDD Subframe Structure for Beyond 4G Radio Access Network”, Future Network & Mobile Summit, July 2013.

[5] M. Bellanger, “Efficiency of filter bank multicarrier techniques in bursty radio transmission”, IEEE Globecom 2010, pp. 1-4, December 2010.

[6] M. Batariere, K. Baum, and T. Krauss, "Cyclic Prefix Length Analysis for 4G OFDM Systems," *Proc. of IEEE Vehicular Technology Conference (VTCFall)*, CA, USA, pp. 543-547, 2004.

[7] F. M. L. Tavares, G. Berardinelli, N H. Mahmood, T.B. Sørensen, P. Mogensen, ”On the potential of Interference Rejection Combining in B4G Networks”, accepted in VTC Fall 2013.

[8] G. Berardinelli, F. M. L. Tavares, T. B. Sørensen, P. Mogensen, K. Pajukoski, ”Zero-tail DFT-s-OFDM signals”; accepted to the Broadband wireless access workshop in Globecom 2013.

[9] M. Lauridsen, L. Noël, T.B. Sørensen, P. Mogensen, “An Empirical LTE Smartphone Power Model with a View to Energy Efficiency Evolution”, Intel Technology Journal vol. 18 issue 1, March 2014

[10] T. Tirronen, A. Larmo, J. Sachs, B. Lindoff, N. Wiberg, “Reducing Energy Consumption of LTE Devices for Machine-to-Machine Communication”, Globecom Workshops (GC Wkshps), December 2012.