Centimeter-wave concept for 5G
ultra-dense small cells

Preben Mogensen (1, 2), Kari Pajukoski (3), Esa Tiirola (3), Jaakko Vihriälä (3), Eeva Lähetkangas (3), Gilberto Berardinelli (2), Fernando M. L. Tavares (2), Nurul H. Mahmoood (2), Mads Lauridsen (2), Davide Catania (2), Andrea F. Cattoni (2)

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

Abstract— Ultra-dense small cells are foreseen to play an essential role in the 5th generation (5G) of mobile radio access technology, which will be operating over different bands with respect to established systems. The natural step for exploring new spectrum is to look into the centimeter-wave bands as well as exploring millimeter-wave bands. This paper presents our vision on the technology components for a 5G centimeter-wave concept for ultra-dense small cells. Fundamental features such as optimized short frame structure, multi-antenna technologies, interference rejection, rank adaptation and dynamic scheduling of uplink/downlink transmission are discussed, along with the design of a novel flexible waveform and energy-saving enablers.

I. INTRODUCTION

The wireless data traffic demand is expected to increase by a factor of approximately ~3000-30,000 by 2030 (with reference to 2010) [1]. Existing radio technologies such as Long Term Evolution – Advanced (LTE-A) and WiFi have inner design limitations which make their potential enhancements unable to cope with the huge capacity requirements. This justifies the design of a new 5th generation (5G) radio access technology (RAT), whose exploration phase has already started. 5G is to be introduced around 2020 and shall fulfill the demand of wireless services for the next decade, i.e. beyond 2030. It is expected to be significantly faster than the previous 4th generation (4G) and to provide higher capacity; the peak data rate requirement is set to 10 Gbps and the latency reduced to the order of 1 ms [2]. Further, the spectral efficiency is set to be at least ~x2 better than 4G.

Following the historical network evolution, it is generally anticipated by the industry that boosting the number of cells will be the larger contributor for increasing the overall network capacity and provide coverage with high data rates [1]. Hence, ultra-dense small cells are expected to play a major role in reaching the above 5G targets. Secondly, significantly larger spectrum is required. A certain amount of unused spectrum is available in the centimeter-wave (cm-wave) frequency range between 3 GHz -30 GHz. Further, the millimeter-wave range (30 GHz- 300 GHz) is also currently being explored [3]. Finally, spectral efficiency can be improved by relying on Multiple-Input-Multiple-Output (MIMO) antenna technology and advanced receivers.

Further, we foresee the following additional requirements for 5G: low power consumption of both access points and terminal, self-optimization of the network, flexible spectrum usage, low cost design and support of new paradigms such as Machine Type Communication (MTC) [2].

In this paper, we focus on a cm-wave concept for 5G, optimized for ultra-dense small cells. In particular, we address the main qualitative differences with the 5G mm-wave design, and describe our vision on the key technology components for enabling in a cost-effective manner multi-Gbps wireless transmission with low latency.

The paper is structured as follows. Section II recalls the basic challenges of both mm-wave and cm-wave concepts. The key technology components enabling multi-Gbps data rate and low latency in the 5G cm-wave concept are described in Section III. Section IV discusses the design of a flexible waveform to be adopted in our concept, while Section V addresses energy consumption issues. Finally, Section VI resumes the conclusions.

II. MM-WAVE VS. CM-WAVE CHALLENGES

Both cm-wave and mm-wave concepts aim at multi-Gbps data rates and reduced latency. Radio-propagation characteristics at mm-waves are still being explored with extensive measurements campaigns (e.g., [4]). To counteract the severe link budget due to paramount propagation losses and heavy attenuation of the potential obstructions, the presence of a robust line-of-sight (LOS) component is proved to be fundamental for establishing the communication link. The link budget can be significantly improved by using large antenna arrays steering highly directional beams at both ends of the communication link. In that respect, operating at extremely high frequency is beneficial for placing such large antenna array in a small physical area. The beamforming will be likely to be performed in the analog radio-frequency domain given the huge cost of analog-to-digital (ADC) and digital-to-analog (DAC) converters at mm-wave frequencies [3]. Clearly, mm-wave concept does not meet requirements like low power consumption and cost efficiency required for MTC. Nevertheless, the possibility of exploiting a large frequency spectrum alleviates the necessity of relying on the spatial multiplexing of data streams for conveying large data rate, thus significantly reducing the computational complexity.
of the baseband processing. Further, in mm-waves the inter-cell interference is not considered a major limiting factor given the possibility of using the aforementioned highly directional beamforming.

Conversely, the cm-wave concept is expected to be applied over bands whose propagation characteristics are quite similar to the ones of already known bands. Establishing the communication link at cm-wave frequencies does not require LOS conditions. Given the lower amount of spectrum at cm-wave frequencies, the usage of MIMO spatial multiplexing becomes fundamental for achieving the targeted data rates. On the other side, the maximum number of spatial data streams is limited in the cm-wave concept given the practical difficulties in placing a large number of antenna ports operating e.g. at below 6 GHz in a hand-held device. Similar limitations may apply also to the small cell access point. We then believe MIMO 4x4 to be a realistic target for 2020 [1]. Further, the presence of large number of significant scattered components besides the LOS component makes inter-cell interference the main limiting factor for the cm-wave concept.

A careful design for both cm-wave and mm-wave concepts should then tackle their respective issues in an agile and cost-effective manner. Further, significant research effort is also being spent on the harmonization of both concepts from a numerology perspective. This would allow using some of the baseband component for both technologies regardless of the different specific radio-frequency front-ends. The rest of the paper will be focused on our envisioned cm-wave concept.

III. KEY-TECHNOLOGIES FOR 5G CM-WAVE CONCEPT

The 5G RAT will be optimized for local area, assuming scenarios characterized by a dense deployment of small cells. Time Division Duplex (TDD) has been recognized as the preferred operational mode given its flexibility and the possibility of exploiting unpaired frequency bands [1]. In this section, we present our envisioned enablers of the cm-wave concept, bearing in mind their cost-effectiveness as a fundamental requirement.

III.A Optimized frame structure

The frame structure is a critical element in the design of a novel RAT since it significantly affects the latency and the required baseband processing for detecting the data. The ms latency target of 5G is justified by envisioned applications such as tactile internet [5], which require extremely short round trip times. Further, the low latency reduces the necessities of storing large blocks of data, e.g. for acknowledgment transmissions. This allows for a significant reduction the size of the buffers, which represent the most expensive component in the baseband chip.

In [1], we have proposed a frame structure which is comprised of a control part located at the beginning and time-separated from the data part, as shown in Figure 1. The control part features in a TDD fashion both downlink (DL) and uplink (UL) control, while the data part can be assigned only to one transmission direction per frame. A short Guard Period (GP) is inserted between every switch of the transmission direction to allow the radio circuitry to power on/off. Multiple access points (APs) or User Equipments (UEs) can transmit simultaneously their control information on the control part by using, for instance, orthogonal frequency blocks. Example of control information 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. Note that the first time symbol in the data part is dedicated to the transmission of DeModulation Reference Sequences (DMRSs), for enabling channel estimation at the receiver for coherent detection. The constraint of allowing an unique transmission direction per frame allows stabilizing the interference pattern within the frame itself. This is beneficial for the usage of interference rejection receivers as it will be further elaborated later.

The time-separation between control and data allows for a straightforward separation of their respective control planes. This enables 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. Note that this is different, from instance, from LTE, where both Physical Uplink Control Channel (PUCCH) and Physical Downlink Control Channels (PDCCH) are multiplexed in the same time symbol and mapped over different frequency resources, forcing UEs and base station to detect both data and control for extracting the needed information [6]. The UE initiated data transmission requires 3 TDD cycles (scheduling request in the UL, scheduling grant in the DL, data transmission in the UL), with a total delay of 0.75 ms. Similarly, the RTT of the Hybrid Automatic Repeat request (HARQ) process (AP grant, AP transmission and UE ACK/NACK transmission) requires 0.75 ms including processing times. Differently from LTE-TDD, the HARQ RTT is here fixed and is independent from the UL/DL ratio; the control part in each radio frame offers indeed the possibility of transmitting acknowledgements at each frame. The number of parallel 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. The benefits in terms of energy consumption of our envisioned frame structure will be addressed in Section V.

![Figure 1. Envisioned 5G cm-wave frame structure.](image-url)
III.B. Interference mitigation

As mentioned in Section II, 4x4 MIMO is foreseen as a realistic target for cm-wave frequencies. It can be shown that the 10 Gbps data rate requirement can be accomplished by assuming 256QAM modulation over at least 2 carriers of 200 MHz each and using the maximum of four simultaneous spatial streams. However, such target can only be achieved in practice in very favorable conditions. In most of the cases, the strong inter-cell interference due to the uncoordinated dense deployment can hinder the possibility of reaching such ambitious data rates. Inter-cell interference represents then a risky show-stopper for the 5G cm-wave concept if efficient interference management techniques are not used.

Frequency reuse techniques, where the available bandwidth is divided in a number of orthogonal channels which are assigned based on network topology, are the typical approach for managing inter-cell interference. Distributed self-organizing techniques where the cells dynamically decide their spectrum allocation without any preliminary planning have also emerged in the last decade, and they have been proved to significantly boost the outage data rate [7].

On the other side, the inter-cell interference can be tackled from a pure receiver perspective, by relying on advanced baseband processing. One approach is the use of Interference Rejection Combining (IRC) receivers, based on a linear filter which rejects the interfering signals by projecting them onto an orthogonal subspace with respect to the desired signals. Such processing requires the receiver to estimate the interference covariance matrix (ICM). The frame structure presented in Section III.A is particularly suited for an efficient application of the IRC detector. By assuming neighbor cells to be time synchronized, the interference is stable within the frame since the data part can be dedicated only to one transmission direction. It is possible then to align the DMRSs of desired and interference signals and then to estimate multiple channel responses within an unique time symbol (provided the used reference sequences are orthogonal in the code domain). This enables an instantaneous calculation of the ICM which can then be used for tuning the IRC filter at each frame, rather than estimating it based on long term statistics. It is worth to mention that the assumption of achieving time alignment among neighbor cells in a distributed manner with fraction of µs precision has been proved to be feasible in our recent contribution [8].

Recent results have shown that the use of IRC receivers can significantly boost the overall network performance in uncoordinated dense local area deployments, improving especially the throughput of cells in harsh interference conditions [9]. IRC receivers have also been recognized to be robust to non-idealities such as receiver-front end limitations and poor channel estimation [10].

The use of IRC also reduces the need for interference mitigation methods based on frequency reuse. Figure 2 depicts the Cumulative Distribution Function (CDF) of the average cell throughput assuming an IRC receiver and different frequency reuse schemes. It is assumed a network with 40 cells, and each cell has 75% probability of transmission, with transmission direction (UL or DL) selected randomly. It is clear that the network may operate in frequency reuse 1 mode without degradation of the outage throughput performance, while it would be necessary to use, for example, planned frequency reuse in the case of interference-unaware MIMO detector such as the Maximum Ratio Combining (MRC) receiver [9].

Another possible approach to deal with inter-cell interference would be the use of Successive Interference Cancellation (SIC) receivers. In SIC receivers, the streams received with strongest power are decoded first and then used for canceling their interference contribution on the weakest streams. Besides the increased computational complexity, such solution requires the receiver to be aware of the transmission format of the interfering signals (specifically, the used MCS as well as the transmission rank). A detailed investigation on benefits and drawbacks of SIC receivers in cm-wave 5G networks is still topic for future research.

III.C Rank adaptation

The IRC receiver has been proved to be effective in rejecting a number of interfering streams when part of the MIMO degrees of freedom are dedicated to that purpose. Ideally, it would be beneficial to transmit the maximum number of streams when no significant interfering transmission occurs, while reducing them in case of non-negligible interference for exploiting the IRC capabilities. Rank adaptation aims at balancing the trade-off between the spatial multiplexing gain of MIMO and the increased interference dimension due to the transmission of multiple streams. The known approaches presented in literature are based on an estimate of the Signal-to-Interference plus Noise (SINR) ratio, which in turn requires an estimate of the ICM. However, the aforementioned possibility of estimating the ICM just-on-time cannot be exploited for rank adaptation due to the classic “chicken and egg” problem: in principle, the rank should be decided prior to transmission, but the ICM can only be efficiently estimated at the end of the current transmission as detailed in Section III.B.

This does not represent a significant issue in current LTE networks since the rank adaptation is performed at much slower rate than the transmission time interval (e.g., in the range of ~100 ms), and the ICM can be based on long term
An efficient rank adaptation algorithm that addresses the aforementioned “chicken and egg” problem is proposed in [11]. Rather than using the received ICM to calculate the SINR directly, the proposed algorithm relies on information about the interferers’powers and their rank to estimate an average SINR. Thereafter, the SINR estimate is used as an input to the proposed rank adaptation algorithm that returns the rank that is most likely to maximize the user’s throughput. The information on the rank of each transmitter can be mapped on the control part of our envisioned frame, which also enables estimation of their power. This algorithm has the further advantage of a lower computational complexity with respect to known solutions based on ICM estimation, e.g [12]. As an example of the performance results for the proposed algorithm, Figure 3 shows the CDF of the achievable network spectral efficiency for an interfering network with 4 cells. The optimum spectral efficiency obtained through a brute-force (BF) search across all possible rank combinations and the performance with an ICM based algorithm similar to that presented in [12] are also presented for comparison. Further, the network spectral efficiency obtained with a fixed rank (i.e. no RA) is also included. The proposed algorithm is found to outperform the fixed rank transmission as well as the known ICM-based algorithm [12], which was not designed to cope with fast interference variability. This allows to significantly reduce the gap with the BF search case, while having the advantage of simplicity and reduced computational complexity. For further details, we refer to [11].

III.D Dynamic UL/DL frame allocation

The level of data flow aggregation is expected to be rather low in local area given the presence of few users per cell, leading to an unprecedented burstiness. Our envisioned frame structure copes with such unpredictable and large traffic fluctuations by enabling the possibility of the changing the link direction every 0.25 ms. This subsumes the presence of a flexible scheduling algorithm which can efficiently decide whether to transit in UL or DL direction at each frame. Such flexibility leads obviously to larger variations in the interference experienced by each receiving node, which is however expected to be efficiently mitigated by the IRC receiver.

An adaptive scheduling algorithm that allows the AP to decide the transmission direction can be designed based on the following principle. Let us assume that the generated data packets are stored in a buffer waiting to be transmitted, and the first packet is denoted as the head-of-line (HOL). A delay threshold for the HOL is set to a desired value, e.g. 1 ms. The AP can be informed of the UL incremental HOL delay and the buffer size with every scheduling request received from the UE. At the time of deciding the transmission direction, the AP inspects the HOL delay of the unutilized direction, and the direction is switched if it exceeds the HOL delay threshold.

A preliminary insight on the performance of this algorithm has been obtained by simulating a 10x2 grid network containing an AP and an UE per room. Light and heavy data rates are defined, the latter being three times higher. Each cell carries a traffic load corresponding to 25% of the capacity, and the traffic imbalance is captured by setting 50% of the cells as downlink heavy and uplink light and vice versa. The packet size is set to 1500 bytes representing the Ethernet Maximum Transmission Unit (MTU). UDP is used as our transport protocol. We further assume to operate in an unacknowledged mode. This setup captures the expected traffic variation at each cell and allows us to observe the benefits of an adaptive scheme. Rank 1 with IRC and planned reuse 2 are used to minimize the SINR variations and keep our focus on the potential advantages of such a flexible UL/DL scheme.

Figure 4 depicts the CDF of the delay in the reception of the packet (note that the processing delays are not taken into account). For the sake of comparison, we include in the analysis a fixed allocation scheme where the frames are allocated alternatively to UL and DL. In most of the cases our adaptive scheme leads to a shorter delay compared to the fixed allocation scheme. This is due to the fact that the fixed allocation might not be able to deal with large traffic bursts; data are then buffered and the transmission completed on subsequent frames, hence increasing the delay. The gain in flexibly adapting to the traffic on such a fast scale becomes immediately apparent. The lower tail of the CDF represents the links which are not heavily loaded and therefore benefit from the fixed allocation scheme, since the buffers are flushed immediately. Future research will address the inclusion of the multi-user aspect in the presented algorithm.
IV. FLEXIBLE WAVEFORM DESIGN

The selection of the radio waveform has a significant impact on the system design since it affects the transceiver complexity and the link performance. Though several novel waveforms with attractive properties have emerged in the literature, in a recent contribution [13] we have motivated our conservative choice of relying on the well-known Orthogonal Frequency Division Multiplexing (OFDM) as the 5G waveform. This is mainly due to its capability of coping with the multipath channel in a cost-effective manner as well as its straightforward extension to MIMO [14]. However, the OFDM waveform has also a number of drawbacks such as system overhead due to the presence of a hard-coded cyclic prefix (CP) appended at the beginning of each time symbol, large peak-to-average power ratio (PAPR), sensitivity to hardware impairments and large out-of-band (OOB) emissions.

Nevertheless, the aforementioned demerits can be partially overcome by straightforward enhancements which do not significantly increase the low complexity. It is well known, for instance, the Discrete Fourier Transform–spread–OFDM (DFT-s–OFDM) modulation which emulates a single carrier signal with significant advantages in terms of PAPR [6].

Recently, we have proposed the usage of Zero-tail DFT-s-OFDM (ZT DFT-s-OFDM) as a solution for avoiding the drawbacks related to an hard-coded CP as utilized in current LTE and LTE-A radio standards [15]. In these RATs, the length of the CP is indeed fixed in the system numerology and its value is set as a compromise between the root mean square delay spread of the channel where the system is expected to operate and the necessity of fitting a predefined frame duration. This leads to unnecessary overhead in case of a system operating over channels having delay spreads which are significantly lower than the CP duration, or conversely to a block error rate (BLER) degradation in case of channels with larger delay spread.

The principle of ZT DFT-s-OFDM is to replace the CP with a low power tail which is obtained as a natural output of the Inverse Fast Fourier Transform (IFFT) and is expected to accommodate the delay spread of the channel. Such signals can be generated as a modified form of DFT-s-OFDM, specifically by adding a number of zeros at the input of the DFT block, as shown in Figure 5(a). The main difference with respect to the CP-based transmission, is that the length of the low power tail can be tuned dynamically depending on the estimated instantaneous delay spread of the channel, by simply varying the number of zeros at the DFT input. This allows avoiding the aforementioned inefficiencies of a hard-coded CP. Note that, besides the low power tail, also a short low power head is generated at the beginning of the time symbols (see Figure 5(b)). Though the low power head represents a system overhead, it has two advantageous properties; it avoids the power regrowth at the tail due to the cyclicity of the IFFT, and allows smoothening the transition between adjacent time symbols. In particular, the second property leads to a significant reduction of the OOB emissions (even 30 dB lower than baseline OFDM) [15]. It is worth emphasizing that such significant spectral containment does not come at the expense of any extra filtering or windowing, thus fully preserving the subcarrier orthogonality. From simulation results, we have shown that the low power head can be maintained extremely short (e.g., 2 samples out of 1200), thus representing a negligible overhead. For further details, we refer to [15]. We are currently evaluating the impact of non-idealities such as phase noise and non-linear power amplifier in order to assess the effective potential of ZT DFT-s-OFDM as 5G waveform.

V. REDUCING THE ENERGY CONSUMPTION

Long UE battery life will be a key performance indicator for a future 5G RAT, because it has a major effect on user satisfaction in conjunction with the aforementioned higher data rates and lower latency. As discussed in [16], several of our 5G design choices are also meant to reduce the UE energy consumption while satisfying the other requirements. For example, the usage of TDD induces that the costly duplexer component can be removed from the radio frequency Front-End, and therefore the insertion loss in both directions can be reduced by at least 3 dB. This results in a lower transmit power, with obvious benefits in terms of energy consumption. Further power saving comes from the fact that, by using TDD, Third-Order-Intercept-Point requirements can be relaxed, since the transmission harmonics will not occur at times of reception. Finally, the amount of channel feedback may also be reduced due to channel reciprocity.

Our optimized frame structure also provides improved energy savings. First of all the separation between the control and the data plane allows the UE to no longer receive and buffer non-dedicated data, since it is able to determine whether it is scheduled before the data is transmitted. The decreased frame length entails the transmission and reception of requests, grants, and acknowledgements will be completed faster, thus effectively reducing the UE’s total active time. A short active time is the key to long UE battery life because it directly translates to longer sleep mode.

Sleep mode is indeed the most effective way to prolong the battery life. Measurements on recent LTE smartphones have shown that the power consumption in sleep mode is at least 40 times lower than in active mode [17], and this is not expected...
to be different for a 5G UE. Therefore the use of sleep mode must be maximized in 5G.

In case synchronization signals are available at each frame, the UE can achieve very fast synchronization (possibly within 1 frame) when it wakes up from the sleep mode. This is a major improvement compared to LTE, where the recent smartphones achieve synchronization within 16 ms because of the timewise scattered synchronization signals [6]. Further, the possibility of achieving high data rate allows the UE to enter the sleep mode faster, and thereby improve battery life. Moreover, measurements [17] have shown that the energy required to transfer one bit decreases with increasing data rate; thus the targeted 5G high data rates are also beneficial.

To summarize the above discussion, Figure 6 shows an example of the achievable improvements in terms of battery lifetime when moving from LTE to 5G. For a Machine-Type-Communication device with one reception and transmission per second, the battery life will approximately improve from 4 months to 7 years [18].

Figure 6. Battery life time as a function of transmissions and receptions for LTE and 5G [18]

It must be noted that the 5G concept also introduces new challenges from a UE energy consumption perspective. Obviously advanced receivers such as IRC or SIC have higher complexity than baseline MRC, but on the other hand can significantly boost the data rate in interference limited conditions. Another critical component is the ADC. Maintaining the power consumption of the ADC at the current levels will be a major challenge, because the sampling frequency will have to increase as the cell bandwidth increases, which will cause increased switched capacitance and short circuit current draw. Furthermore the envisioned higher order modulation (256QAM) may require the ADC’s resolution, to increase causing increased power consumption. The total number of ADCs is also expected to increase because of the application of Carrier Aggregation and multiple spatial streams. Addressing the trade-off between increased complexity and power consumption is an object of our current research.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we have introduced our preliminary vision for a 5G cm-wave concept targeting ultra-dense deployment of small cells. We have presented our key technology proposals for achieving multi-Gbps data rates and ultra-low latency in such a strongly interference-limited scenario. The design of a novel flexible waveform is also presented, along with a discussion on the energy saving benefits of our envisioned concept.

Future research will address the open challenges in each of the presented areas; in particular, a joint optimization of rank adaptation with novel scheduling algorithms is to be investigated, along with further energy savings techniques.

REFERENCES