Vision for Beyond 4G Broadband Radio Systems

(Invited Paper)

Bernhard Raaf, Wolfgang Zirwas, Karl-Josef Friederichs
Nokia Siemens Networks, Research – Radio Systems
St.-Martin-Str. 76, 81541 Munich, Germany

Esa Tiirola, Matti Laitila
Nokia Siemens Networks, Research – Radio Systems
Kaapelitie 4, 90620 Oulu, Finland

Patrick Marsch
Nokia Siemens Networks, Research – Radio Systems
Ul. Strzegomska 2-4, 53-611 Wrocław, Poland

Risto Wichman
Aalto University School of Electrical Engineering
P.O. Box 13000, FIN-00076 Aalto, Finland

Abstract— **Mobile communication systems have evolved over the past decades and each new generation brought new experience to the users enabled by technology innovations, while keeping some well established principles from previous generations. This trend continued up to LTE (Long Term Evolution) Advanced, the predominant 4th generation system which has just been standardized in 3GPP and is being rolled out soon. How will this trend continue to future systems which will be deployed in some 10 years from now which will be advanced enough to be called "Beyond 4G" (B4G)? This article presents how such B4G systems will look like and some key technologies they will rely on including versatile numerology, massive virtual MIMO from many base stations, both centralized and distributed architectures using fiber optics as backbone, advanced interference mitigation, cognitive self organization, and wideband RF radios.**

Keywords: B4G, Beyond 4G, LTE, LTE-Advanced.

I. Introduction

Mobile communication experienced a tremendous success, ever increasing availability and performance of mobile services, starting with voice but recently expanding into various data services. This trend is expected to accelerate due to increasing popularity of smart phones, super phones and tablets with powerful applications and multimedia capabilities up to pervasive 3D multimedia and other applications allowing a completely new mobile experience. While the number of subscribers may eventually saturate, proliferation of mobile broadband into billions of machine devices and consequent machine to machine (M2M) applications will add to the exponential increase in traffic volume and transactions.

Mobile broadband systems will have to become more powerful to meet the requirements which are outlined in Section II at reduced cost and energy consumption per bit in order to allow provisioning of ever more demanding services at roughly flat costs to subscribers. Candidate technologies which are essential to reach these requirements are introduced in section III, and section IV shows how they can concert together to shape a viable B4G system. Section V finally concludes the paper, indicating areas where significant research will still be needed to reach the required leap forward to guarantee the economic success of the wireless sector, from device and component manufacturers up to network and service providers.

Authors' email addresses: {bernhard.raaf, wolfgang.zirwas, karl-josef.friederichs, esa.tiirola, matti.laitila, patrick.marsch}@nsn.com, risto.wichman@aalto.fi

II. Requirements

Studies and extrapolations from recent developments predict a total mobile broadband traffic increase up to a factor of 1000 until 2020. This figure assumes a 10 times increase in broadband mobile subscribers and up to 100 times higher traffic per user (beyond 1 GByte/subscriber/day). Apart from laptops, smart phones and super phones will experience the fastest growth, demanding faster networking technologies so that they can have faster connections for videoconferencing, viewing HD and 3D video, and other purposes. Taking these needs into account, it is anticipated that the B4G system must bring at least 10x improvement in the peak and even more important the average data rate compared to LTE-Advanced, i.e. fulfill peak rate requirements beyond 10 Gb/s. Beside the overall traffic, the achievable throughput per user has to be increased significantly.

Another important target is latency minimization for unparalleled true "local feel" experience and support of cloud computing applications delegating part of the processing from the device to the network. Further, low latency also helps to recover from instances of degraded link-quality quickly. Similarly important is consistent performance guaranteeing data rates whenever and wherever users require them, including at cell edges (also related to fairness).

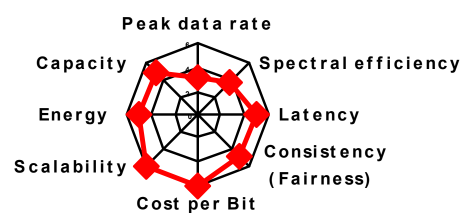

Figure 1. Relevant key requirements for B4G.

Moreover, as shown in Figure 1, some essential design criteria need higher attention than in today's systems. The most important design targets are expected to be cost per delivered bit and system scalability to different heterogeneous deployment scenarios (wide and local area) and frequency allocations (ranging from 1 to 5 GHz bands, under different licensing regimes) for providing “rock solid” ubiquitous connectivity.

A factor gaining importance in the overall design of wireless systems is the energy efficiency of the network deployment including its components such as base stations and access points. Reducing the environmental impact including the CO_2 emissions becomes essential for the ecosystem. Moreover, increased energy efficiency of the network reduces the operational expenses (OPEX), lowering the cost per bit. This measure is important considering the expected traffic and throughput growth until 2020.

III. TECHNOLOGY COMPONENTS

In order to achieve the requirements for the next decades as set out in the previous section, several key enablers which are introduced in the following sections will need to be investigated and eventually incorporated into the B4G system.

A. Versatile low latency frame structure

User plane latency is an important measure for end-to-end performance of many applications. The minimum round trip time (RTT) of current LTE-Advanced systems is about 10 ms. It is generally understood that latency must decrease at the same pace as the data rate increases. Hence, RTTs on the order of 0.1 – 1 ms need to be considered as the initial RTT target for B4G systems. It is obvious that very small latency on the radio level (such as 0.1 ms), is only beneficial for content that is very close, e.g. at home, in the office, or on the same campus.

It is difficult to achieve major latency improvements without major impact on the air interface. This is due to the fact that the latency relevant aspects such as frame structure, control signaling timing, and Hybrid Automated Repeat Request (HARQ) form the key building blocks of the air interface. Therefore, major latency improvements in user-plane (U-plane) imply considerable changes in the air interface. For example, an RTT requirement of 0.1 ms means that the Transmission Time Interval (TTI) length needs to be scaled down to 10-25 μs. This cannot be achieved with the current LTE-Advanced Orthogonal Frequency Division Multiple Access (OFDMA) numerology, but requires a more versatile approach to the numerology.

Overhead reduction is the easiest way to boost the radio performance. Even though LTE/LTE-Advanced has some scalability inbuilt in the system (e.g., cyclic prefix length, bandwidth, number of antenna ports) the total overhead is always considerable and increasing with the number of data streams. Scalable OFDMA design could provide means to optimize the radio performance for any scenario, for example local area or Coordinated Multi-Point (CoMP) schemes. Reduction of cyclic prefix (CP), reference signal and control signaling overhead could provide significant increase in the throughput. For example, reducing the CP length in LTE-Advanced down to 0.5 μs would reduce the CP overhead from 6.7% (normal CP length) close to zero. Optimizations for efficiency and latency will be partly conflicting, thus the system needs to provide adaptability to select configurations best supporting the required KPIs (Key Performance Indicators) for specific services and applications. This versatility can also be used to optimally adapt the system to varying deployment scenarios including size of the cell or the co-operating antenna pool which determines the multi path profile.

B. Modulation principle

OFDM has been the modulation format of choice in numerous wireless communication and broadcasting systems. OFDM facilitates simple equalization by FFT in time dispersive channels by inserting a guard interval in the form of a cyclic prefix between successive modulation symbols. Furthermore, the guard interval makes the link robust to timing errors and facilitates synchronization of different users and access points in multiple access systems.

Despite its undisputed advantages, OFDM has several shortcomings. A cyclic prefix that provides orthogonality between the symbols in time dispersive channels implies a loss in spectral efficiency as mentioned in the previous subsection. The efficiency is further degraded due to a guard band that is necessary to reduce adjacent channel interference caused by the poor frequency localization of the OFDM signal. Furthermore, OFDM signals are subject to a high peak-to-average power ratio, which imposes major challenges to power amplifiers, and the signals are sensitive to frequency impairments that cause inter-carrier interference.

Spectral containment of OFDM signals can be improved, but the overhead from guard band and guard interval remains. Another approach is to change the basis functions modulating the transmitted data to improve spectral efficiency [1]. Filterbank multicarrier waveforms do not require a guard interval and can be localized well in frequency. This is facilitated by sophisticated filter design that determines how the energy is distributed in time and frequency. In general, a good localization in frequency is achieved by extending and overlapping the multicarrier symbols in time.

Multicarrier modulation schemes partition the transmitted signal into subsequences in time and further into parallel subcarriers in frequency. The radio propagation environment disperses the signal in time and frequency and multicarrier modulation is a natural way to limit the effect of the channel distortion. Furthermore, localization of the waveforms in time and frequency facilitates efficient scheduling and interference management algorithms in multiple access systems. Thus, multicarrier modulation seems a reasonable choice for B4G systems.

Multicarrier modulation schemes trade off time and frequency localization in different ways depending on the prototype filter design. Therefore, the desired time and frequency resolution, latency and scalability requirements, number of transceiver antennas, etc., should be specified first before designing the modulation scheme. For example, filterbank multicarrier modulation [2] with 4 overlapping modulation symbols may improve the spectral efficiency in single-input single-output link by 15% compared to OFDM. This comes at the price of a 10-fold complexity in signal processing with respect to OFDM and a latency of 3 modulation symbols due to equalization. These factors may limit the applicability of the filterbank modulation from MIMO

(Multiple Input Multiple Output) deployment and latency point of view. On the other hand, good spectral containment of filterbank modulation is advantageous if different systems should share the same frequency bands.

In the context of CoMP, however, where multiple base stations jointly process the baseband signals connected to multiple terminals (see section C), the requirements regarding modulation schemes are again different. Given the structure of the OFDM signal, downlink joint transmission is strongly sensitive to imperfect synchronization in frequency [3], whereas CoMP in both uplink and downlink is sensitive to the timing delays introduced due to inter-cell signal propagation. While the first aspect can be solved with sophisticated synchronization techniques, the latter aspect can heavily limit the usefulness of CoMP in rural areas with larger inter-site distances. Hence, in this case an even longer cyclic prefix than in an LTE-A system is required, unless other means are applied to make cooperative transmission and reception more robust to these synchronization issues.

As a conclusion, changes in the modulation scheme are not expected to directly improve system performance significantly, but they can augment the more flexible frame structure approaches that are required to adapt key physical layer parameters according to the very diverse requirements in different contexts (e.g. local area, M2M, CoMP, spectrum sharing, etc.).

C. *Interference mitigation*

Since the beginning of mobile radio communications, inter-cell interference is one of the main issues to be solved. Some good progress has been made in the meantime, starting from large frequency reuse factors for 2G, frequency reuse one for 3G and some form of interference aware scheduling (inter cell interference coordination: ICIC) and interference-aware transceiver techniques such as interference rejection combining (IRC) for 4G. Within the 3GPP LTE Release 11 coordinated multi point (CoMP) study item, more powerful means like joint precoding over several cells are being investigated. Still, for the time being, performance is significantly limited by inter cell interference. Note that in theory effective cooperative transmission and detection schemes will even partly benefit from interference by exploiting spatial multiplexing, interference cancellation, array, and diversity gains [4].

Considerable progress has been made lately in the EU funded project Artist4G by defining a full framework for interference mitigation [5]. It combines practically feasible user centric clustering solutions like 'partial CoMP' [6] with UE-specific interference rejection combining and a novel interference floor shaping technique. Compared to ICIC the interference floor shaping is done per cooperation area (CA), exploiting new degrees of freedom, e.g. due to overlapping CAs. CA centric UEs are served with high power, while outward interference is minimized by optimizing antenna tilting and power reduction.

While joint transmission and detection algorithms are rather straightforward to design and have already been successfully tested in field trials, key remaining challenges connected to CoMP lie in the accurate estimation of interference links, and the feedback of this information to the transmitter sides at reasonable signaling overhead. This will be difficult to solve with the current simple codebook based schemes, while promising approaches like model based channel prediction [7] – a parameterized compression scheme based on detailed knowledge of the eNB surrounding – are not fully mature yet and need significant further research. In addition, these schemes rely on technology progresses like superior processing capabilities as well as much larger memory sizes. Despite these challenges, we expect that B4G will be capable to support accurate channel knowledge for multi-cell precoding and joint detection.

While currently for LTE Advanced spectral efficiencies of about *3-4 bit/s/Hz/cell* are considered, the Artist4G framework promises already about *10 bit/s/Hz/cell.* A natural path for further increases to some *30 bit/s/Hz/cell* will be opened by serving more users per cell simultaneously on the same physical layer resources. As a result, B4G will require a significant progress with respect to accurate channel estimation based on moderate pilot overhead for an increasing number of channel components with low to moderate feedback overhead.

In general, one can expect that the cooperation principle will spread into more and more other areas like e.g. Hetnet scenarios including picos, relay nodes and macros.

D. *Architecture*

To meet the demand of increased data volumes, future networks will need to be deployed much more densely than today's networks and, due to both economic constraints and site availability, will need to become significantly more heterogeneous than today. Networks are composed of layers of multiradio cells of different size. Diverse radio access technologies will need to be integrated, e.g., LTE/LTE-Advanced, UMTS/HSPA+, WiFi, future radio access technologies or any combination thereof.

The radio network architectures of the nodes will vary from stand alone base stations to systems with different degrees of centralized processing, depending on the available backhauling. The higher the degree of centralization becomes, the higher the transport capacity and the central processing power will need to be provided. Both fiber availability and transmission rates of optical systems will have increased substantially until 2020 and on the other hand digital processing power will continue its rapid development according to Moore's law.

It would be beneficial to push wide area cell efficiency by optimizing the system for inter-cell interference coordination together with centralized transmission schemes in order to fully utilize valuable wide area cell sites. Exploitation of cloud technologies, equipment/resource sharing and virtualization are key enabling technologies to cost efficiently reduce inter site (and therefore antenna cluster) distances and increase areas and degrees of coordination, a necessary prerequisite for higher traffic at lower cost.

Even though centralized Radio Access Network (RAN) architectures would apply also for local area access, for the sake of scalability and cost the distributed architecture will be more feasible for local area scenarios due to simplicity and flexibility. Local access naturally utilizes local content. But also globally offered delay sensitive and high volume services like video streaming could be beneficial to be offered locally utilizing advanced content delivery mechanisms and local caching in order to achieve true “local feel” experience. Direct access to local packet data networks and the Internet should also be supported to provide LAN connectivity and offload the Internet traffic (which brings less added value from the mobile operator point of view) from the transport network. User centric protocols are feasible in local area scenarios in order to lighten the signaling and Control-plane (C-plane) processing burden from small cell deployments. For highest flexibility, local area cells could also be exploited for offloading wide area access, wherever feasible.

In an intelligent future wireless network, the connectivity for delivering services should be designed also from an energy efficiency perspective. This means that the network should be intelligent enough to use the least amount of radio resources and energy as possible, still observing the maximum allowed latency. Hence, instead of offering a more or less uniform bit pipe with varying delivery capabilities, the services could be delivered intelligently to proactively take advantage of the operational environment, e.g. existence of alternative access networks, and to scale the quality according to available network and air interface resource capabilities.

E. Specific enhancements for local services

The LTE/LTE-Advanced air interface has been optimized for wide area environments where the transmit power difference between the UE and eNB is relatively large (up to 25 dB). As a consequence, both the modulation scheme (Single Carrier (SC-FDMA) vs. OFDMA) and the physical channel structure differ significantly between uplink (UL) and downlink (DL). UL has been designed maximizing the link budged for control and data channels. As discussed earlier, requirements for radio interface design are more versatile for a B4G system. For example, given the similar transmit powers of uplink and downlink and the limited coverage area required for indoor access points, it makes sense to maximize similarities between UL and DL. This implies that similar multiple access schemes could be used in both directions, simplifying system design and hardware implementations and providing better support for relaying including Device to Device (D2D) communications, i.e. direct connections or relaying amongst terminals. Thus, having a similar air interface in UL and DL will enable some new use cases and gain mechanisms, especially in Time Division Duplex (TDD) mode which is clearly one of the focus areas for the B4G air interface design. In this way, capacity can be provided efficiently and locally.

F. Further advanced technology components

If we venture far beyond 2020, and assume that by then advanced MIMO and CoMP techniques have already been expanded and exploited to the extent where complexity and cost start outweighing further performance gains, the question is where further data rate increases will come from.

In fact, at some point in the evolution of mobile communications systems it may not be possible to further increase spectral efficiency significantly at reasonable effort, simply because the theoretical degrees of freedom are already exploited and physical bounds are approached. If mobile data rate demand still increases exponentially at this point in time, the only way forward is to use more spectrum and significantly more base stations. This will of course put an even higher pressure on the cost and energy dissipation of access points and also call for further miniaturization. Similarly, the OPEX of running dense mobile infrastructures will need to be reduced, further calling for more and more sophisticated automated self-configuration, self-optimization and self-healing functionalities. It can hence be expected that in the very long term, major innovations will focus on a substantial decrease in cost per bit and also energy per bit, which could for example be possible in the following fields:

Innovative spectrum and infrastructure sharing concepts. In the long run, it may be possible to use limited radio resources and infrastructure more efficiently by reusing them among different network operators. It is expected that frequency licensing concepts will become more liberalized in the foreseeable future. On one hand, there are already new rules in place for secondary usage of primary assigned spectrum such as TV White Spaces, which provides access to unused spectrum suitable e.g. for traffic offloading. Moreover, there are also other regulatory models promoted e.g. the so called “Licensing Light” and “Authorized Secondary Access”, which refer to collaborative usage of a junk of spectrum under equal and fair terms or the shared use of licensed spectrum under an agreement with the licensee, respectively. It is believed that such new sharing models can be exploited through novel cognitive radio technologies and thus lead to better overall utilization and spectrum efficiency which is essential to utilize more than 1GHz of spectrum spread over all bands.

Re-definition of the concept of cells, and the infrastructure behind these. In order to strongly reduce the signaling overhead both over the air interface and also among base stations, it may be interesting to give up the current paradigm regarding cell concepts and assume that a terminal may be served by a fuzzy set of access points in the future instead. Different functionalities like control signaling, user data transfer in up- and downlink and mobility (including tracking of the location and paging), which are now all provided by the serving cell, may each be provided by a different set of “cells” or even more generally a different set of co-operating antennas. Such a scalable concept enables the introduction of much more massive MIMO deployments and advanced antenna concepts at some places, for some terminals or for part of the signaling, without necessarily increasing overhead for each terminal and each link everywhere. Also, innovative and potentially disruptive concepts regarding the interconnection and interaction of base stations, and the distribution of signal processing complexity over the network may be options to strongly reduce network operation costs. Eventually, the terminals are immersed in a sea of RF signals without being aware that they are sourced by a pool of myriads of antennas distributed all around and connected via fibers to flexibly shared digital processing units; a true representation of so-called liquid radio.

Novel air interface concepts. Besides the frame structure and modulation related air interface aspects discussed before, it may be interesting to consider a completely re-designed air interface, which could for example be layer-less. This might in particular be valuable in the context of M2M and D2D communication.

IV. Towards reaching B4G system performance

The well known ways to increase peak data rate are to provide more RF bandwidth, more MIMO branches and usage of higher modulation order. In addition, reduction of overhead through scalable air interface design helps in increasing both peak and average data rates.

Providing high data rates where users request them will mandate to augment macro cellular networks with many small cells. Thus, B4G high rate focus must be in heterogeneous deployments. It is noted that 10 Gb/s can be reached with 200 MHz bandwidth using eight parallel MIMO streams, 256 Quadrature Amplitude Modulation (QAM) per stream and with 20% total overhead (half the overhead compared to LTE-Advanced). Alternatively, with six MIMO streams and 64QAM per stream 10 Gb/s can be delivered more robustly relying on 350 MHz bandwidth.

SC-FDMA used in LTE/LTE-Advanced uplink already provides maximal coverage by maintaining low peak to average power ratio / cubic metric facilitating efficient power amplifiers in the devices. However, there is still room for power efficiency enhancements, not only in UL but more and more also in DL due to smaller base stations, relays and D2D, and utilization of extremely high bandwidths e.g. based on massive carrier aggregation.

Figure 2 shows more comprehensively which technology components contribute to reach the objectives set out in Section II. In short, utilizing some 10x spectrum 10x more effectively and increasing base station densities 10x supports the expected 1000x capacity and throughput increases.

V. Conclusion

Radio access technologies will utilize significant advancements allowing to meet the predicted requirements in 2020 of 1000 fold capacity increase under flattish revenues, and consequently flattish costs. The technology mix will combine a much more versatile radio numerology, a large number of heterogeneous base stations, some of them tightly co-operating using centralized architectures, others being deployed in a distributed manner, a massive use of advanced antenna technologies, self organizing networks enhanced with cognitive elements, true wideband RF radios and extensive use of available fiber optics networks for the back- & front-hauling. In the long run, when we reach the point where the spectral efficiency per access point may not be further improved at reasonable effort, we may also consider novel spectrum usage concepts, a re-definition of cell concepts or entirely re-designed air interfaces. We will have to drive re-search and development of the most economical and powerful technologies and solutions within industry and academia.

References

[1] B. Le Floch, M. Alard, C. Berrou, “Coded orthogonal frequency division multiplex [TV broadcasting], ” Proceedings of the IEEE , vol.83, no.6, pp. 982-996, Jun 1995

[2] V. Ringset, H. Rustad, F. Schaich, J. Vandermot, M. Najar, "Perfor-mance of a FilterBank MultiCarrier (FBMC) physical layer in the WiMAX context," Proc. Future Network and Mobile Summit, June 2010

[3] V. Kotzsch and G. Fettweis. “Interference Analysis in Time and Frequency Asynchronous Network MIMO OFDM Systems,” IEEE WCNC ‘10, Sydney, Australia, Apr. 18–21, 2010.

[4] P. Marsch and G. Fettweis (editors), “Coordinated Multi-Point in Mobile Communications – From Theory To Practice”, Cambridge University Press, June 2011

[5] W. Mennerich, W. Zirwas, “Reporting Effort for Cooperative Systems Applying Interference Floor Shaping,” PIMRC'11, Toronto, Canada, Sept. 2011.

[6] W. Mennerich, W. Zirwas, “User Centric Coordinated Multi Point Transmission,” Proc. IEEE VTC-fall, 2010

[7] W. Zirwas, Y. Tan, W. Mennerich, “Model based Channel Prediction,” InOWo'10, Hamburg, Germany, Sept. 2010

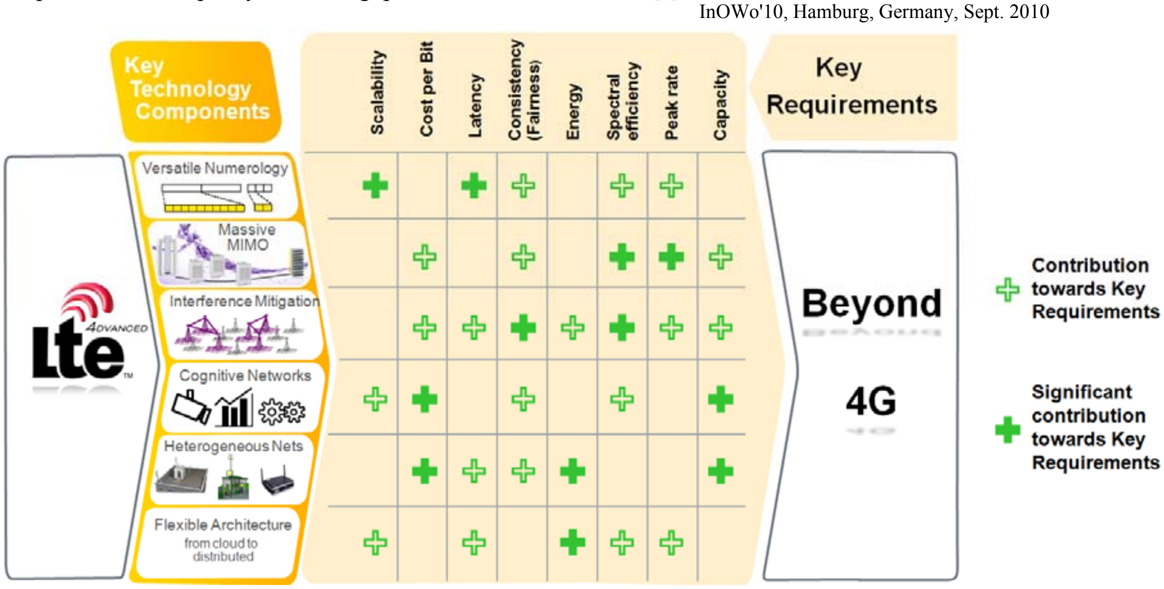

Figure 2. Key technology components contributing significantly towards the B4G key requirements introduced in figure 1.