Huawei Innovation Research Program

HIRP JOURNAL

June 2015 Issue 1

5G Inaugural Issue

-- 5G research and Innovation

SCMA

A Promising Technology for 5G Wireless Networks

F-OFDM

Flexible Air Interface with Filtered OFDM

Polar Code

A 5G enabling FEC scheme

HUAWEI

Huawei is dedicated to open innovation and is committed to working with industry counterparts to address future challenges and opportunities to create more values for the world.

The Huawei Innovation Research Program (HIRP for short) is a sponsorship program targeted at R&D talents and teams worldwide, helping them innovate in new technologies and services as well as solving traditional problems using new methodologies.

HIRP aims to enhance collaborative innovation to achieve a win-win situation. Along with making achievements through collaboration, Huawei will continue to strengthen knowledge transfer and sharing to drive technological advancement and industrial development, so as to build a better connected world.

To know more about HIRP, visit us at

http://www.huawei.com/en/hirp/

OPENINGFOREWORD

Dr. Peiying Zhu

—Huawei Fellow—

Huawei Global 5G
Research Program Leader

Senior Director

Wireless Technology Lab
in Huawei Canada

A Brief Introduction of 5G

Huawei Innovation Research Program (HIRP) is one of Huawei's research collaboration platforms with other industries, universities and research institutes. HIRP Journal is one of HIRP publications aiming to disseminate the output of HIRP and share Huawei's internal research progresses and interests.

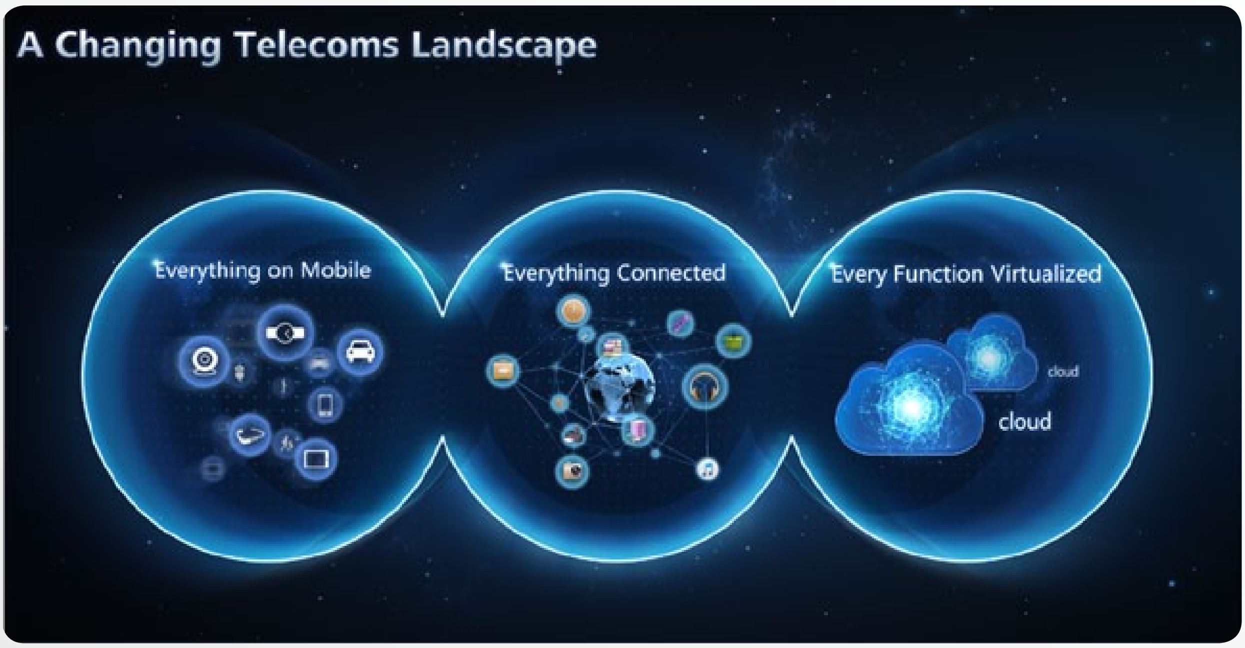
5G radio access research is selected as the theme for this debuting issue. I am extremely honored to provide the introduction for this special issue. I believe we will have more collaboration with R&D talents and teams worldwide though this HIRP platform and drive technological advancement and industrial development.

Huawei started 5G researches in year 2009 with the initial focus on improving spectrum efficiency and increasing overall system capacity. A much more comprehensive research program has been developed since then. The overall 5G technology vision white paper was published in 2013 [1], in which 1000 times capacity improvement, 100 times data rate increase, 1ms extreme low latency and 100 times more massive connectivity were identified as the key 5G requirements. ITU-R WP5D has been discussing 5G vision and requirements [2] and it is expected that the final agreement will be reached in the middle of the year. In parallel, NGMN has been working on 5G vision, requirements and use cases from the operator's perspective and published its white paper [3]. Various other 5G related organizations also published corresponding 5G white papers, such as IMT-2020 from china [4], 5GMF from Japan [5], 5G Forum from Korea [6], METIS [7], 4G America [8] etc.. The consensuses are that 5G will provide a rich set of services from extreme mobile broadband to machine type communication and ultra low latency and high reliable services.

In addition ITU has agreed on the 5G timeline from standard and regulatory perspective [2], which signaling the beginning of the 5G standardization activity. 3GPP is the key 5G standard body, it started 5G use case study item called SMART [9] and recently announced that a 5G RAN workshop will be held in Sept 2015 [10].

While the consensus are emerging on requirements and use cases, the key technologies for 5G are yet to be decided. 3GPP 5G RAN workshop will provide venue for such discussion. Within Huawei, the 5G research team has studies many promising technologies. The radio technology white paper [11] has been published recently, providing an overview on several potential technologies. In this Journal, more technical details related to these technologies will be discussed.

1 Why we need 5G?



• 5G will have fundamental impact on the ICT industrial transformation and human life. As to be a key enabler of the future digital world, 5G is to be an ultimate platform for a connected world to enable new ways of innovation and collaboration and to create new opportunities.

• Mobile broadband will permeate all areas of society in the future and the users would expect a better blueprint for the networked world. Huawei has proposed “Everything on Mobile, Everything connected, Every function virtualized” as the future trend of telecoms, which is acknowledged commonly.

Everything on Mobile

We hope we can access to the network, 24 hours a day, 7 days a week, to share the information, photos, and documents with our friends, colleagues and customers. The communication is anywhere anytime and can be well done on mobile devices.

Everything connected

It is estimated that by 2020, 6.5 billion people worldwide will use mobile networks for data communications and that 100 billion additional ‘things', such as vehicles, home appliances, and medical devices, will also be connected via mobile networks.

Every function virtualized

NFV/SDN now is the hot topic and cloud storage, cloud computing is more and more popular. Massive traffic is generated from cloud usage.

2 What is the deference between 5G and 2G/3G/4G?

5G is a truly converged network will support a variety of natural seamless deployment of new networks, including the deployment of ultra-high-density wireless networks, backhaul, devices and communications equipment, dynamic spectrum integration and wireless access infrastructure sharing.

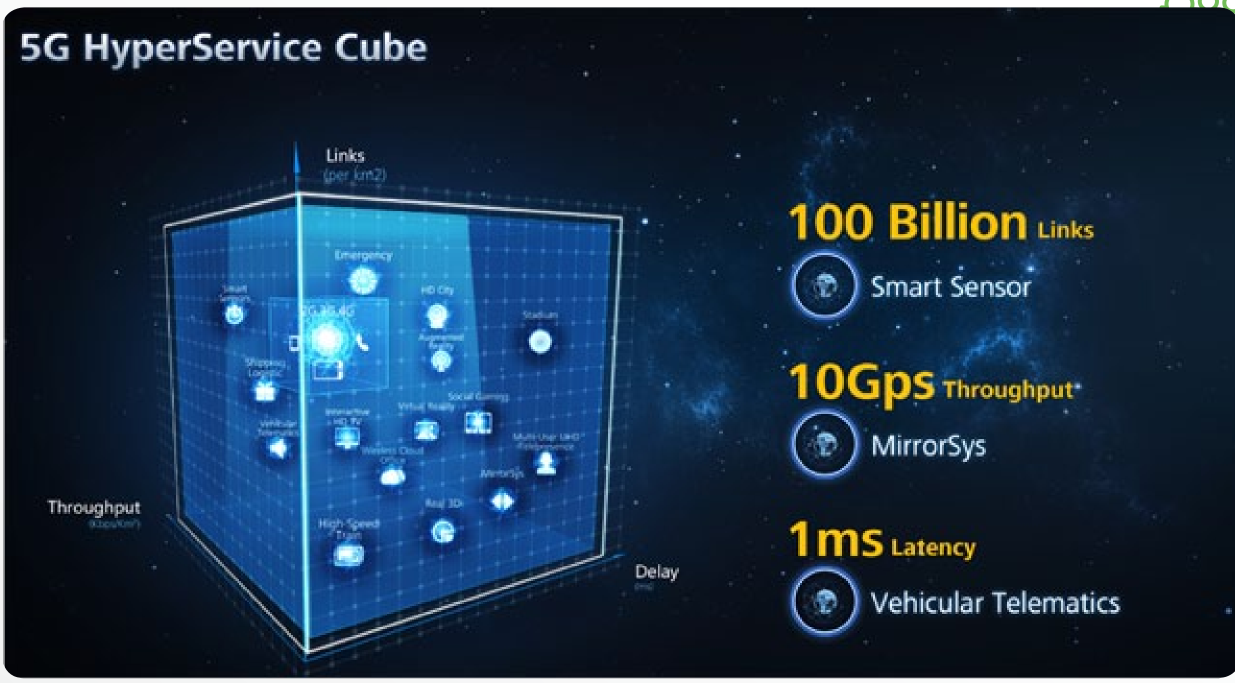
The characteristics of previous generations of wireless network access parameters are fixed and spectrum block, 5G network will allow the use of any spectrum with any access technology to provide the best communication services.

5G air interface and RAN system needs a revolutionary innovation designed to adapt to the new mobile access modes have large capacity, high speed connectivity and other features.

Compared to 2G, 3G and 4G, in addition to the high-capacity demands, 5G will introduce a new dimension to the large number of connections, and ultra-low latency.

3 What kind of experience can 5G bring us?

5G technology will bring a new level user experience.



• Huawei defines HyperService Cube, 15 typical application scenarios, such as Internet of things, ultra high-definition video, virtual reality, smart sensors and other things interconnected vertical industries such as automotive.

• Besides, HyperService Cube shows three dimensions of the requirement of future communications: latency, throughput and number of connections (1ms latency, 10Gbps of user rate, and 100 billion connections).

4 What is the key innovation breakthrough to achieve from a technological point of view?

Huawei has achieved network architecture, spectrum, new air interface and base station innovation breakthrough.

• Low band (<6GHz) is for voice & data access, high band (>6GHz) is for higher data access and also can be used as wireless backhaul.

Spectrum

• All spectrums, sub-6GHz (<6GHz) as core spectrum and 6GHz beyond as complementary part.

New air interface

• Air interface is adaptive for efficiently supporting service diversity in multiple scenarios, AI components include SCMA (Sparse Code Multiple Access), F-OFDM (Filtered-Orthogonal Frequency Division Multiplexing) and Polar Code, etc.

• The above will be complemented by massive multiple antenna arrays and MIMO techniques, which improve coverage, reduce interference and boost throughput. The demand for data rates beyond of 1Gbit per second per user enforces the usage of high frequency bands, e.g. millimeter waves.

Network technologies

• Service oriented network with unified connections, security, mobility and routing management limited to only three level of controllers, i.e. device controller, edge controller(s) and orchestrator controller meeting the "zero latency" requirement, with a large degree of reliability and flexibility (agile and dynamic deployment of network applications and services), without the use of tunnelling protocols, in in-coverage, out-of-coverage (cluster-head) and relay-control, in transparent and non-transparent modes.

• For supporting various vertical industry applications, the service- oriented network is necessary, and the virtual RAN and network technology will be sliced based on service diversity to achieve the dedicated virtual network for dedicated service.

• No Cell: this is another Huawei innovation, which is one of the foundations of UDN (Ultra Dense Network) architecture. The logical cell identification is kept unchanged while users cross multiple physical cells. The user experience improved because of the joint of multiple physical links and handover free.

Base Station (Ultra NodeTM)

• Huawei showcased the world first 100Gbps 5G small cell prototype, reached industry highest 115Gbps throughput under the circumstance on 5G high band.

- Support one million above sensor links per site

- Support 200+ users per site with 4~8K HD video communication

• Huawei innovated 100Gbps on Demand small cell, which integrated high & low band frequency modules, low band (<6GHz) is core for voice & data access, high band (>6GHz) is for higher data access and also can be used as wireless backhaul.

In the following of this first Issue of HIRP Journal, we introduce some of our air interface technologies.

References:

[1] http://www.huawei.com/5gwhitepaper/

[2] http://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2020/Pages/default.aspx

[3] https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf

[4] www.imt-2020.cn/zh/documents/download/25

[5] http://5gmf.jp/en/outline/organization/

[6] www.5gforum.co.kr

[7] https://www.metis2020.com/

[8] http://www.4gamericas.org/en/

[9] http://www.3gpp.org/ftp/tsg_sa/WG1_Serv/TSGS1_69_Sanya/docs/S1-150300.zip

[10] www.3gpp.org , SP-150149, "5G" timeline in 3GPP

[11] http://www.gsacom.com/downloads/pdf/5Gz_Huawei_White%20Paper_5G_Radio_140515.php4

Contents

HIRP - 5G

5G: New Air Interface and Radio Access Virtualization 09

SCMA: A Promising Technology for 5G Wireless Networks 18

F-OFDM: Flexible Air Interface with Filtered OFDM 32

Polar Codes: A 5G enabling FEC scheme 44

Massive MIMO: A Key Enabler for 5G Cellular Systems 52

Full-Duplex Technology for 5G 62

Radio Access Virtualization 70

5G mmWave Communications 83

5G: New Air Interface and Radio Access Virtualization

1 Introduction

5G is the next frontier of innovation for the wireless industry and the broader ICT ecosystem . It is common consensus that 5G will focus on the breakthroughs to support the expansion and enhancement of mobile internet and Internet of Things (IoT). The application of 5G in IoT and vertical industries will bring more market space and present business opportunities to operators. In addition, expanded and enhanced

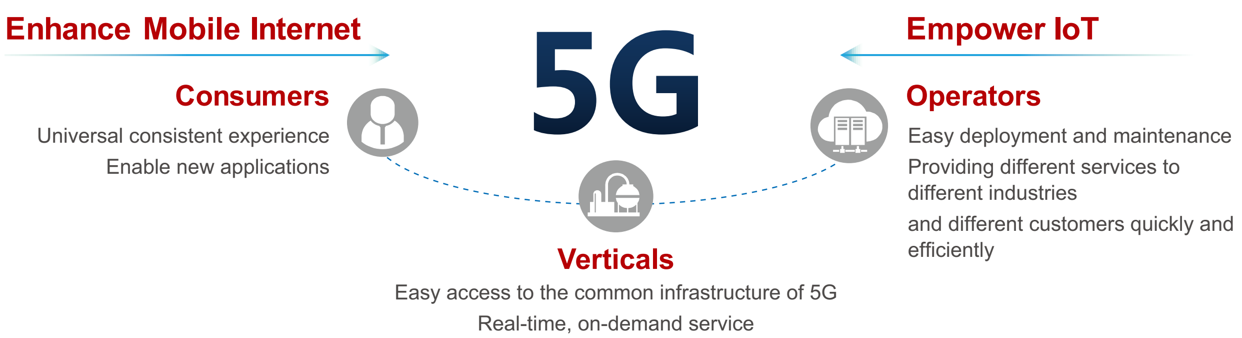


Figure 1: 5G will Carry Many Industries and Benefit Stakeholders

mobile internet services will help further improve the consumers experience, strengthen user stickiness and guarantee operators' revenues and profits, as Figure 1 shows.

To adequately support the development of mobile Internet and IoT, 5G networks will increasingly become the primary means of network access for person-to-person and person-to-machine connectivity. This means that 5G will need to match the diversity of service requirements and service characteristics. Examples include extreme broadband, ultra-low latency, massive connection and ultra-high reliability[1] etc., along with the ability to accommodate various use cases. The strong requirement of a service oriented network to provide better user experience in a flexible, efficient way is raised.

[1] Huawei Whitepaper, "5G: A Technology Vision," Nov. 2013. http://www.huawei.com/ilink/en/download/HW_314849

2 5G Performance Requirements

"Wireless networks will need to match advances in fixed networking in terms of delivered quality of service, reliability and security. It is expected that the 5G system design will support three orders of magnitude higher capacity per km2, a hundred times higher data rate, latency of less than 1 ms across the radio access link, a hundred times more connections (links) and three orders of magnitude lower energy consumption than the current generation of wireless network. Although these performance targets do not need to be met simultaneously, they provide the basis for the Gbit/s user experience for 5G networks. The improved performance targets of the overall 5G system are shown in Figure 2."

Figure 2: 5G Performance Targets

Based on the requirements, two major challenges should be addressed for the design of the 5G system:

- The 5G system should be capable of flexible and efficient use of all available spectrums from low band to high band and licensed to unlicensed bands.
- The 5G system should be adaptable to provide efficient support for the diverse set of service characteristics, massive connectivity and massive capacity. Flexible network design is required to improve spectral efficiency, increase connectivity and reduce latency.

The requirements and challenges will impact on the design of 5G air interface and the network architecture. In this white paper, the flexible 5G new air interface is explored in details. The viewpoint of network architecture is to be presented in subsequent white papers.
The 5G air interface framework is built upon two major concepts: software defined flexible air interface and radio access virtualization. In terms of air interface, it should be optimized in the way to support versatile application scenarios. In terms of radio access virtualization, it encompasses self-organization and coordination algorithms that utilize the features, protocols and interfaces to avoid the limitations of the geographic "cell" construct.

3 5G Spectrum

The growing traffic demand necessitates increasing the amount of spectrum that may be utilised by the 5G systems. High frequency bands in the centimeter wave (cmWave) and millimeter wave (mmWave) range will be adopted due to their potential for supporting wider channel bandwidths and the consequent capability to deliver high data rates.

The new spectrum below 6GHz is expected to be allocated for mobile communication at the World Radio Conference (WRC) 2015, and the band above 6GHz expected to be allocated at WRC 2019, as shown in Figure 3.

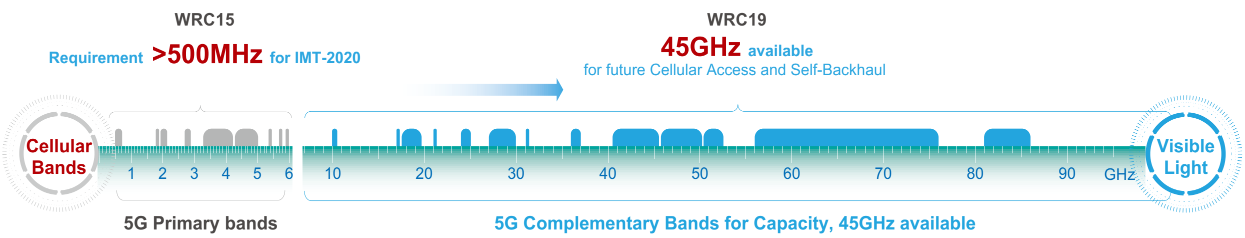


Figure 3: 5G will Aggregate Sub 6GHz and the Bands above 6GHz

5G network is a heterogeneous network which enables the cooperation between lower-frequency wide-area coverage network and high-frequency network. The consensus is higher frequency bands are the complementary bands to 5G whereas low frequency bands (<6GHz) are still the primary bands of 5G spectrum.

High frequency also enables unified access and backhaul since the same radio resources is shared. It is expected to use a unified air interface and a hierarchical scheduling for both radio access and backhaul which enables flexible backhauling and low-cost ultra dense networking (UDN).

Future radio access may also employ bands with different levels of access regulation including exclusive licensed, non-exclusive licensed and unlicensed bands. The 5G system treats both the licensed and unlicensed spectrum in a flexible, unified air interface framework.

4 5G Flexible New Air Interface

Different application requirements for air interface technology is complex and diverse, a unified new air interface with flexibility and adaptability is proposed to meet these requirements. New air interface consists of building blocks and configuration mechanisms such as adaptive waveform, adaptive protocols, adaptive frame structure, adaptive coding and modulation family and adaptive multiple access schemes. With these blocks and mechanisms, the air interface is able to accommodate the future wide variety of user services, spectrum bands and traffic levels.
Key technology components, as shown in Figure 4-1, include a new waveform technology Filtered-OFDM (Filtered-Orthogonal Frequency

Different application requirements for air interface technology is complex and diverse, a unified new air interface with flexibility and adaptability is proposed to meet these requirements. New air interface consists of building blocks and configuration mechanisms such as adaptive waveform, adaptive protocols, adaptive frame structure, adaptive coding and modulation family and adaptive multiple access schemes. With these blocks and mechanisms, the air interface is able to accommodate the future wide variety of user services, spectrum bands and traffic levels.

Key technology components, as shown in Figure 4-1, include a new waveform technology Filtered-OFDM (Filtered-Orthogonal Frequency Division Multiplexing), a new multiple access technology SCMA (Sparse Code Multiple Access), a new channel code Polar Code, the full-duplex mode and massive MIMO technology. The new air interface design can effectively improve spectral efficiency, increase connectivity, and reduce latency, thus facilitating the deployment of customized scenarios applied to the IoT and for high bandwidth-consuming scenarios such as virtual reality.

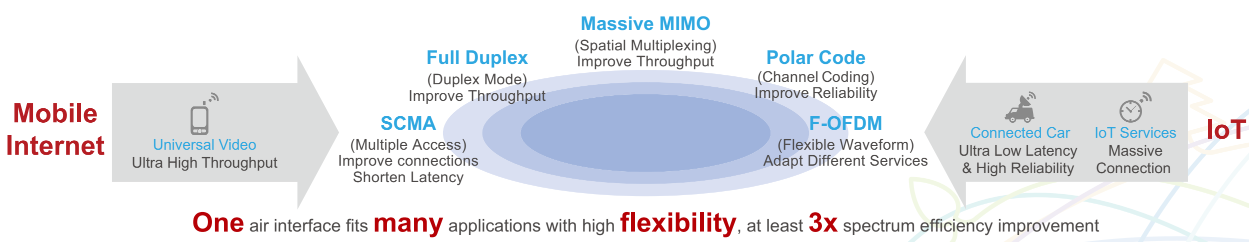


Figure 4-1: New air interface components

The new air interface exploits two-level non-orthogonality to maximize the spectrum efficiency, the number of connected devices and to provide flexibility to support vastly diverse services. Filtered OFDM allows inter-subband non-orthogonality while SCMA enables intra-subband non-orthogonality.

Filtered - OFDM

Filtered-OFDM is one element of fundamental waveform technology to support different waveforms, multiple access schemes and frame structures based on the application scenarios and service requirements simultaneously. It can facilitate the co-existence of different waveforms with different OFDM parameters as shown in Figure 4-2. In this figure different sub-band filters are used to create OFDM sub-carrier groupings with different inter-sub-carrier spacing,OFDM symbol durations and guard times. By enabling multiple parameter configurations, filtered-OFDM is able to provide a more optimum parameter choice for each service group and hence better overall system efficiency.

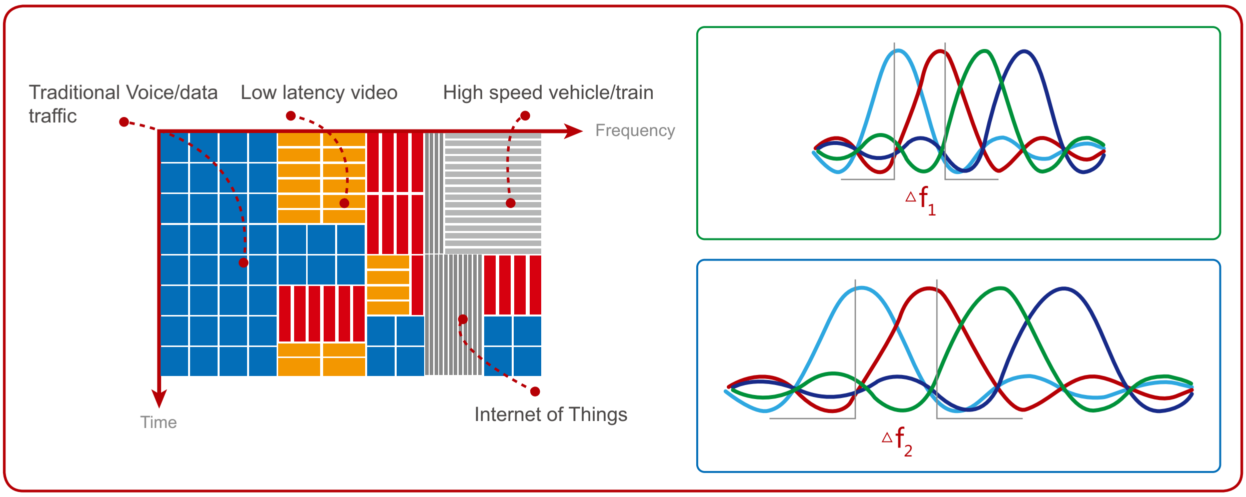


Figure 4-2: Filtered-OFDM enables flexible waveform parameters

Sparse Code Multiple Access

Sparse code multiple access (SCMA)[2] is another waveform configuration of the flexible new air interface. This non-orthogonal waveform facilitates a new multiple access scheme in which sparse codewords of multiple layers of devices are overlaid in code and power domains and carried over shared time-frequency resources. Typically, the multiplexing of multiple devices may become overloaded if the number of overlaid layers is more than the length of the multiplexed codewords. However, with SCMA, overloading is tolerable with moderate complexity of detection thanks to the reduced size of the SCMA multi-dimensional constellation and the sparseness of SCMA codewords. In SCMA, coded bits are directly mapped to multi-dimensional sparse codewords selected from layer-specific SCMA codebooks. The complexity of detection is controlled through two major factors. One is the sparseness level of codewords, and the second is the use of multi-dimensional constellations with a low number of projection points per dimension[3]. An example of device multiplexing with a low projection codebook and the resulting constellation mapping is shown in Figure 4-3. A device's encoded bits are first mapped to a codeword from a codebook. In the example, a codeword of length 4 is used. The low projection codebook has a reduced constellation (from 4 points to 3 points). Furthermore, each point (e.g.

"00") has non-zero component only in one tone. A codebook with one non-zero component is a zero-PAPR codebook.

Furthermore, a blind multi-device reception technique[4] can be applied to detect device activities and the information carried by them simultaneously. With such blind detection capability, grant-free multiple access[5] can be supported. Grant-free multiple access is a mechanism that eliminates the dynamic request and grant signaling overhead. It is an attractive solution for small packets transmission. SCMA is an enabler for grant-free multiple access. Due to these benefits,SCMA can support massive connectivity, reduce transmission latency and provide energy saving.

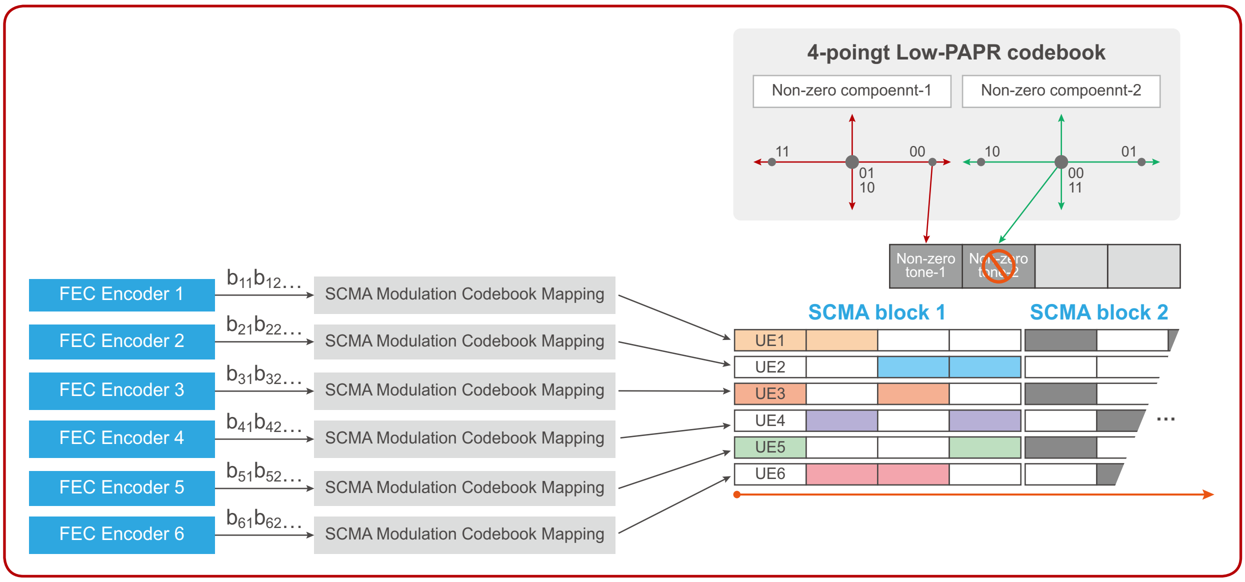


Figure 4-3: SCMA multiplexing and low projection codebook constellation

Polar Codes

Polar codes are a major breakthrough in coding theory. They can achieve Shannon capacity with a simple encoder and a simple successive cancellation (SC) decoder when the code block size is large enough. Polar codes have brought significant interests and a lot of research work has been done mainly on code design and decoding algorithm. One of the most important decoding algorithms is the SC-list decoding which can perform as well as the optimal maximum-likelihood (ML) decoding with a list size of 32 for moderate code block sizes. A lot of performance simulations show that Polar codes concatenated with cyclic redundancy codes (CRC) and an adaptive SC-list decoder can outperform turbo/LDPC (Low Density Parity Check) codes for short and moderate code block sizes. Polar code has better performance than all the codes currently used in the 4G LTE systems, especially for short code length, thus it is considered as a perfect candidate for the FEC (Forward error correction) module in 5G air interface design.

[2] H. Nikopour and H. Baligh, "Sparse code multiple access" IEEE 24th PIMRC, pp. 332 – 336, 2013.
[3] M. Taherzadeh, H. Nikopour, A. Bayesteh and H. Baligh, "SCMA codebook design," IEEE VTC Fall 2014.
[4] A. Bayesteh, E. Yi, H. Nikopour and H. Baligh, "Blind detection of SCMA for uplink grant-free multiple access," IEEE 11th ISWCS, pp. 853 – 857, 2014.
[5] K. Au, L. Zhang, H. Nikopour, E. Yi, A. Bayesteh, U. Vilaipornsawai, J. Ma and P. Zhu, "Uplink contention based SCMA for 5G radio access," IEEE Globecom Workshop on Emerging Technologies for 5G Wireless Cellular Network, pp. 900 – 905, 2014.

Massive-MIMO

Massive MIMO makes a clean break with current practice through the use of large-scale antenna systems over network and devices. As one of the most promising ingredients of the emerging 5G technology, massive MIMO is a commercially attractive solution since 100x higher efficiency is possible without installing 100x more base stations.

The progress on omni-directional beam with low power, low PAPR and flexible beam adjustment for mobility UE tracking is enabling the theoretical concept to commercial deployment reality in diverse scenarios, such as macro, micro, suburb and high-rise.

Full Duplex

Full-Duplex breaks the barrier of today's communications by supporting bi-directional communications without time or frequency duplex. By transmitting and receiving at the same time and on the same frequency, Full-Duplex has the potential to double the system capacity and reduce the system delay.

5 Radio Access Virtualization

5G networks using radio access virtualization strategies and advanced computational platforms will exploit network densification. The virtual cell concept removes the traditional cell boundary for the device and provides a consequent reduction of the detrimental “cell-edge experience” by the device.

Elimination of cell boundaries

Traditionally devices associate with a cell as a consequence the link performance may degrade as a device moves away from the cell center. In a virtualized device centric network, the network determines which access point(s) are to be associated with the device. The cell moves with and always surrounds the device in order to provide a cell-center experience throughout the entire network. The elimination of the device’s view of the cell boundary is illustrated in Figure 5.

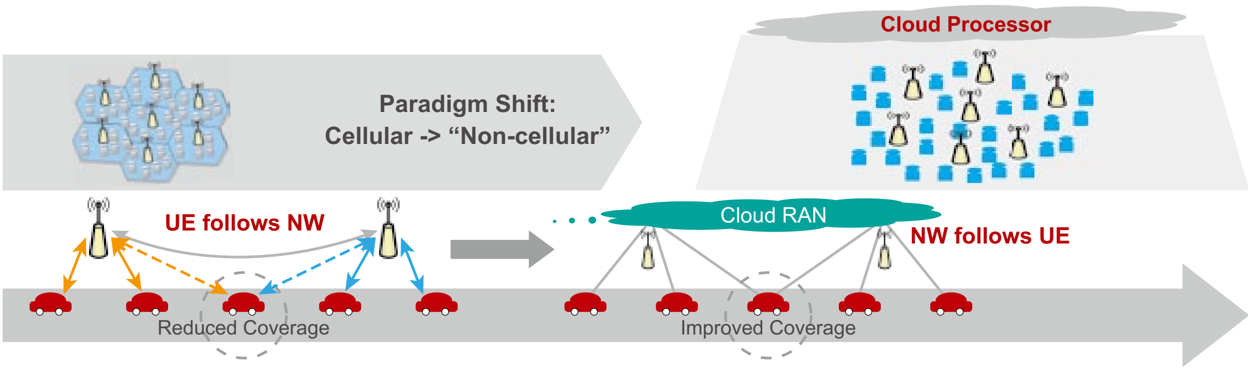


Figure 5: Elimination of cell boundaries

Device-centric Access Point Optimization

Each device is served by its preferred set of access points.The actual serving set for a device may contain one or multiple access points and the device’s data is partially or fully available at some or a small set of potential serving access points. The access point controller will accommodate each device with its preferred set and transmission mode at every communication instance while considering load and Channel State Information (CSI) knowledge associated with the access points.

Network-assisted device Cooperation

An important factor in determining and updating potential and actual serving access point sets is the possibility of cooperation among neighboring devices and the nature of such cooperation. The density of neighboring devices and the capability for device to device (D2D) connectivity provides the opportunity for device cooperation in transmission/reception.
The access point controller can schedule the devices benefiting from the device cooperation and manage factors such as cooperation collision, security and privacy restrictions, and cooperation incentive. A network assisted device cooperation results in better virtualization by providing more possible transmission paths from network to the target devices.

6 Conclusions

The future network will focus on the different business applications and user experience other than just the pursuit of the greater bandwidth and volume. This will raise the requirement to build service oriented networks to quickly and efficiently respond to user needs, as well as to offer consistent and high-quality services for different use cases.

This paper has outlined an overview of Huawei's 5G air interface design including the key concepts of air interface adaptation and radio access virtualization. Radio access virtualization technologies can provide the best transmit and receive conditions to users while flexible new air interface selects the best sets of air interface technologies on the wireless links. These two components together can bring the best user experience in the 5G wireless network. The goal is to design an air interface that is adaptable to the diverse services, applications and devices of the future, scalable to support massive connectivity and massive capacity and intelligent to adapt to all the locally available spectrum .

SCMA: A Promising Technology for 5G Wireless Networks

SCMA

Abstract — Sparse code multiple access (SCMA) is a code domain non-orthogonal multiple-access technique introduced for future 5G wireless networks. With SCMA, multiple incoming data streams are directly mapped to codewords selected from multi-dimensional modulating codebooks, where each codeword represents a spread transmission layer. Multiple SCMA layers are combined to share the same time-frequency resources of OFDMA. This article describes the SCMA structure including the SCMA encoder, codebook design, and near-optimal reception techniques. Several application scenarios are introduced to demonstrate the potential gain of SCMA for future 5G networks, including uplink grant-free contention-based transmission, downlink open-loop multi-user transmission, and downlink open-loop coordinated multi-point (CoMP) for ultra-dense (UDN) and moving networks (MN). Based on the numerical and experimental results, SCMA can resolve some major issues of current wireless systems and establish itself as a strong candidate for 5G networks.

Keywords — SCMA, non-orthogonal multiplexing, MBB, MTC, massive connectivity, CoMP, UDN, MN.

1 Introduction

Fifth generation (5G) wireless networks are expected to support very diverse applications and devices.[1],[2] Recent research activities[3] emerge toward new technologies and provide solutions to meet the targets of the next generation of wireless communication networks in 2020 and beyond. Some of the important 5G targets and use cases can be expressed as in [4], [5] :

- Enhanced mobile broadband (MBB) with very high traffic volume, ultra high peak rate, and very high and consistent expedited user rate over the entire network even for very high speed terminals.
- Massive machine type communication (m MTC) with very large number of devices having low rate, low cost, ultra low power consumption, and very deep coverage.
- Critical MTC (c-MTC) with ultra reliable and ultra low latency requirement even for moderate mobile nodes.

A wide spectrum range along with diverse use cases and application scenarios cannot all be addressed by a single and rigid air interface design of the current long term evolution (LTE) standard[6], [7]. For example, grant-based nature of LTE uplink is unable to support massive connectivity or ultra low latency for MTC. Among the issues arisen due to grant-based transmission of uplink LTE, the cost of control signaling overhead, latency, and energy efficiency are prominent. The cost of such dynamic signaling for uplink is higher for small packets since the ratio of signaling overhead to useful payload is high. As an example, it is estimated in[8] that to transmit a small amount of data (e.g. 20 bytes of data using QPSK ½), the overhead ratio can be as high as 30%. Semi-persistent scheduling can be an option in order to reduce the dynamic signaling overhead[7]. However, such mechanism is more suitable for traffic arrival that exhibits some form of periodicity and predictability such as VoIP and is not good for bursty traffic with tight delay requirement as in c-MTC. Therefore, one of the features of the future wireless networks is to support grant-free multiple-access mechanisms that enable different modes of MTC.

Multi-user MIMO (MU-MIMO) is a well-known spatial domain multiple access technique used in the current LTE system[9] to increase the overall throughput in downlink. Despite the promising throughput gain and the simplicity of detection at user nodes, MU-MIMO as a closed-loop system suffers from some practical difficulties in terms of channel aging, channel estimation impairments, and high overhead to report channel state information (CSI) on users to a serving transmit point (TP) in terms of precoder matrix indicator (PMI) or channel sounding. If CSI is not well estimated, cross-layer interference practically limits the potential performance gain of MU-MIMO. The problem is even more challenging in the coordinated multi-point (CoMP) transmission or ultra dense network (UDN) setup where CSI has to be reported to multiple TPs. This requires solutions provided based on the open-loop transmission and low feedback overhead which are robust against mobility, CSI error, interference, and other impairments in future wireless networks in the MBB scenario.

To resolve the above mentioned problems, the sparse code multiple access (SCMA) [10] [11] is proposed as a promising multiple access solution for 5G. SCMA is a non-orthogonal multiple-access scheme in the code domain. As shown in Fig. 1, multiple incoming data streams are directly mapped to modulate codewords selected from layer-specific codebooks. Each codeword is a complex vector representing a spread transmission layer. Like CDMA, codewords of multiple layers are combined and carried over shared time-frequency tones of OFDMA.

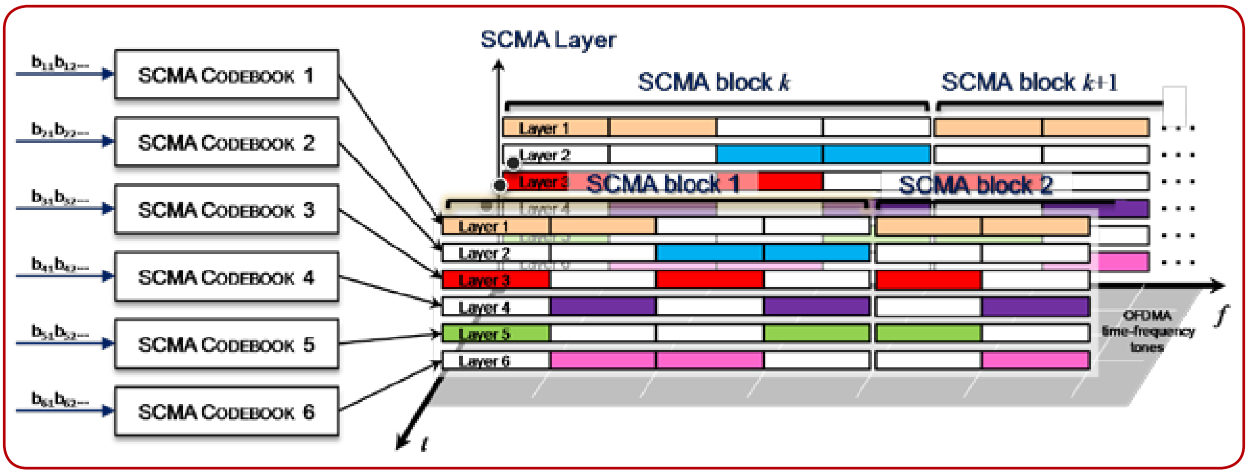


Fig.1. SCMA multiplexed layers carried over OFDMA tones. The 6 layers are multiplexed over 4 tones with 150% overloading. Every codeword contains two non-zero and two zero elements

Non-orthogonality, sparsity, and multi-dimensional constellation gain are key features of SCMA codewords. By relaxing the orthogonality constraint, the multiplexing system can be overload, that is, the number of multiplexed layers can exceed the length of codewords. It brings an advantage to SCMA system to enable massive connectivity through system overloading. In theory, overloading is achievable via non-orthogonal CDMA as well. However, to deliver a reasonably good and reliable performance, a non-linear joint detector is required to separate the combined layers at a receiver. The sparsity of SCMA codewords makes the near-optimal detection practically feasible through iterative message passing algorithm (MPA)[12]. Such relatively low complexity of multi-layer detection allows for a large codeword overloading which is not practically feasible with CDMA. On the other hand, multi-dimensional constellation gain provides further reliability of detection even when the system is overloaded. Therefore, non-orthogonal multiplexing along with sparsity and multi-dimensional codebooks are three key features of SCMA to enable either massive connectivity or reliable transmission for 5G MTC. Furthermore, SCMA is well-matched to downlink user multiplexing as we can allocate non-orthogonal code-domain layers to different users without need for PMI knowledge of paired users. With a very limited knowledge of CSI, a TP simply pairs user together. Compared to MU-MIMO, MU-SCMA system is more robust against channel variations and provides more consistent experience even for high mobility users as is a requirement for 5G MBB.

In this paper, we introduce various access modes and use cases for SCMA, including:
i) uplink grant-free contention-based transmission for MTC
ii) downlink open-loop multi-user transmission for MBB
iii) downlink open-loop CoMP and UDN

The rest of the paper is organized as follows. Section II defines the system model, SCMA system structure, encoder and detector. This section also describes the SCMA codebook design and major properties and benefits of these codebooks. Section III talks about the different access modes and use cases to highlight the potential gain of SCMA for future 5G systems. The promising experimental results and prototyping of uplink grant-free SCMA is reported in Section IV. The paper concludes in Section V.

2 SCMA: Model Description and Benefits

SCMA Encoding and Multiplexing

In SCMA, incoming bits of a layer are mapped to multi-dimensional codewords selected from a codebook. Every layer corresponds to a specific codebook. A layer-specific codebook contains a set of complex modulating codewords. The number of codewords in a codebook defines the codebook size.

SCMA multiplexing of 6 layers selected from 6 layer-specific codebooks. The size of every codebook is 8 and hence every codewords is labeled by 3 bits. The length of each codeword is 4 tones. The 6 sparse codewords are multiplexed but only 3 of them collide over each tone due to the sparsity of codewords.

As the example illustrated in Fig. 2, a codebook with 8 codewords is equivalent to an 8-point multi-dimensional constellation in which every codeword carries 3 bits of a data stream. The bit label of a codeword is defined by the codebook. According to the above description, the functionality of an SCMA encoder is equivalent to a QAM modulator. The only difference is that the QAM modulator is a one-dimensional complex modulator whereas the SCMA encoder is a multi-dimensional modulator defined based on a given codebook. After SCMA encoding, every element of an SCMA codeword is carried over an OFDMA tone. For the example illustrated in Fig. 1, the length of codewords is 4 and hence every codeword occupies 4 OFDMA tones.

Referring to Fig. 2, the output codewords of multiple SCMA encoders are combined together to realize SCMA code-domain multiplexing. Multiplexed codewords share the same OFDMA time-frequency resources. The layer multiplexing may happen at a transmitter as in downlink, or SCMA layers of multiple transmitters are combined over the air before approaching a receiver as in the uplink direction of a cellular network. To simplify the model, in uplink, we assume every SCMA layer represents a user terminal while in downlink a sub-set of available layers can be assigned to a user depending on the scheduler outcome.

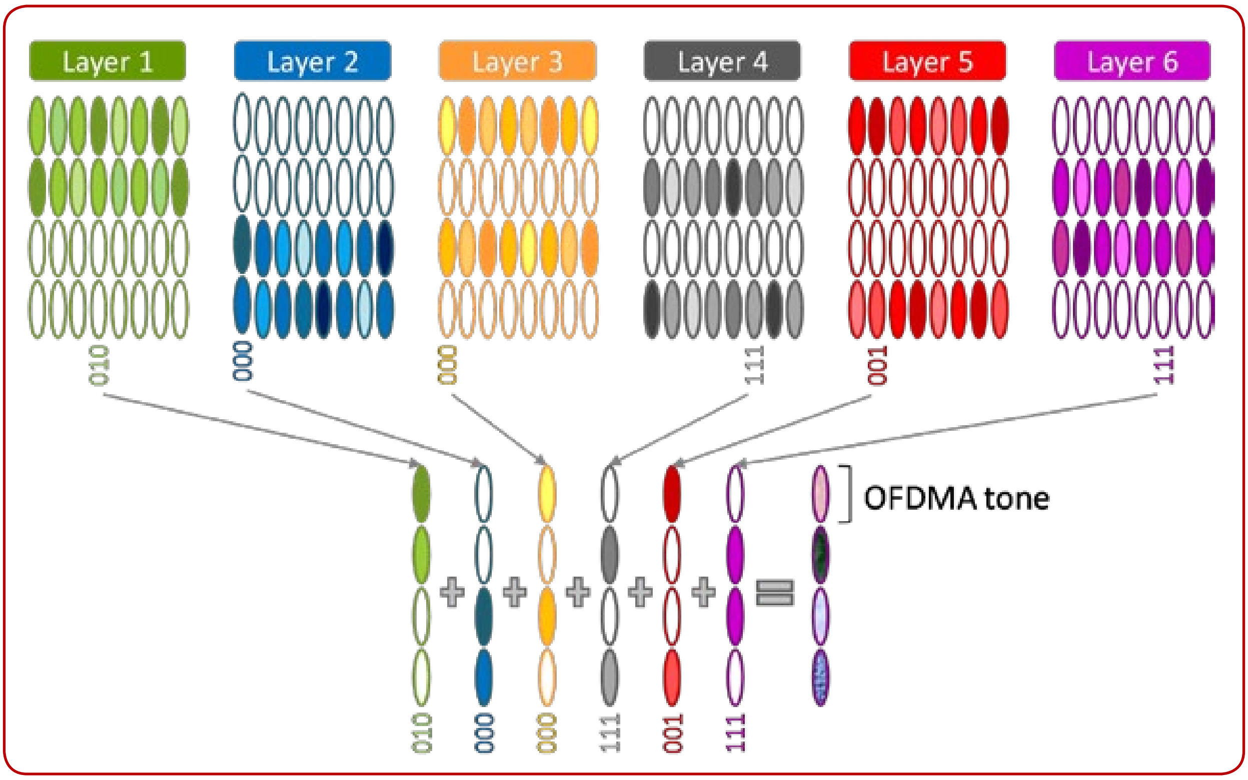


Fig. 2. SCMA multiplexing of 6 layers selected from 6 layer-specific codebooks. The size of every codebook is 8 and hence every codewords is labeled by 3 bits. The length of each codeword is 4 tones. The 6 sparse codewords are multiplexed but only 3 of them collide over each tone due to the sparsity of codewords

Sparsity and shaping gain of multi-dimensional codewords, overloaded non-orthogonal multiplexing, and spreading over multiple tones are some key features of SCMA codebooks further discussed in the sequel.

SCMA Codebooks, Complexity and, Performance

In SCMA, instead of spreading a QAM symbol with a signature as in CDMA, input bits are directly mapped to multi-dimensional codewords. One benefit of SCMA codebooks is the shaping gain of multi-dimensional constellations over repetition coding of CDMA spreading. According to[13], there is 1.53 dB asymptotic gap between QAM constellations and Shannon capacity in an ideal AWGN channel with Gaussian inputs. Following the principles of lattice constellation design,[11] proposes a systematic procedure to design SCMA codebooks with desirable properties to capture part of this gap and outperform CDMA and QAM OFDMA with a feasible complexity of detection.

Simulation results reported in [11], demonstrates the advantage of SCMA codebooks over CDMA spreading in AWGN and fading channels. According to these results, SCMA can tolerate more multiplexed layers while maintain the single-layer performance as if there is no cross-layer interference. In other words, SCMA is more robust against inter-layer interference even for large number of multiplexed layers. This is a key feature to support massive connectivity and reliability at the same time.

In addition, shaping gain of multi-dimensional codewords can increase the spectral efficiency and quality of detection of SCMA. Link performance evaluations reported in [11] show that SCMA outperforms CDMA, OFDMA, and single carrier FDMA (SC-FDMA) over a wide range of SNR operating points in frequency-selective fading channels. In some scenarios, the SNR gain can be more than 2 dB.

During the process of SCMA codebook design, the multi-dimensional constellations can be set in a way that leads to a lower number of projected points per dimension[11]. The low projection property helps the MPA receiver to exponentially reduce the number of hypotheses per dimension, while the total number of constellation points and overall transmit rate are not reduced. For example, a 16-point codebook can have 9 projections per non-zero tone of a codeword while still carries 4 bits[11]. This is one advantage of multi-dimensional constellations over QAM constellations where a large SCMA codebook is realized with lower detection complexity compared to a QAM constellation with the same size.

Multi-Layer Joint Detection Using MPA

As illustrated in Fig. 2, SCMA codewords are sparse with some zero elements within codewords. The sparsity pattern varies over codebooks to reduce the chance of codeword collision over each OFDMA tone. For example, in Fig. 2, 6 sparse codewords are multiplexed, but only 3 of them collide over each tone due to the layer-specific sparsity pattern of codebooks. MPA takes advantage of the sparsity feature of SCMA to make ML-like detection of multiplexed layers practically feasible.

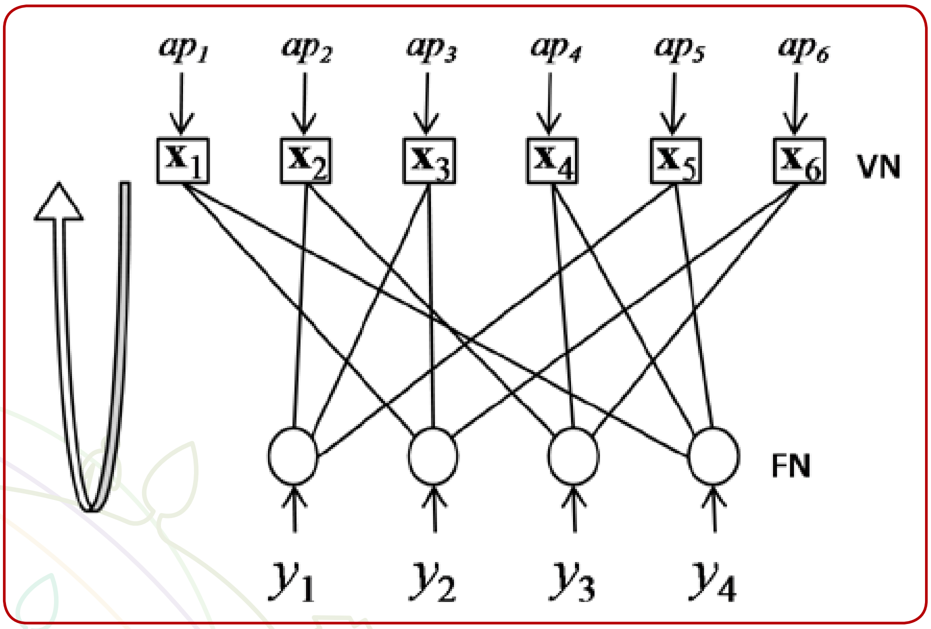


Fig. 3. Factor graph representation of 6 SCMA layers of length 4 tones for MPA iterations between VNs and FNs

As shown in Fig. 3, the structure of superimposed SCMA codewords can be represented by a factor graph. Every layer x_j,j=1,…,6 is represented by a variable node (VN) and every receiving OFDMA tone y_i,i=1,…,4 is represented by a function node (FN).

A VN is connected to an FN through a factor graph edge if and only if the corresponding element in the SCMA codebook is non-zero. The main advantage of SCMA is that it limits the number of colliding codewords at an FN to control the complexity of the non-linear joint detection. MPA operates based on the idea of message passing between VNs and FNs. Having the knowledge of the active codebooks and their corresponding channels, the MPA algorithm solves the joint detection problem of the multiplexed SCMA codewords in an iterative fashion.

The performance of SCMA detection can be further improved by combining MPA with turbo decoders or any other type of soft-input soft output (SISO) forward error correction (FEC) decoders. As shown in Fig. 4, a turbo-MPA takes advantage of an outer-loop to convert the extrinsic log likelihood ratios (LLRs) of the output of the FEC decoders to codeword probabilities and feeds them back to the MPA decoder as a priori information. The MPA detector uses the a priori (ap_j) information at the variables nodes as illustrated in Fig. 3. Note that the outer-loop can be early terminated to control the complexity of detection. Turbo-MPA shows its advantage especially when a large number of layers are superimposed and cross-layer interference is a dominant source of performance limitation. For example, for the scenario of Fig. 3, simulation results show more than 4 dB link performance gain.

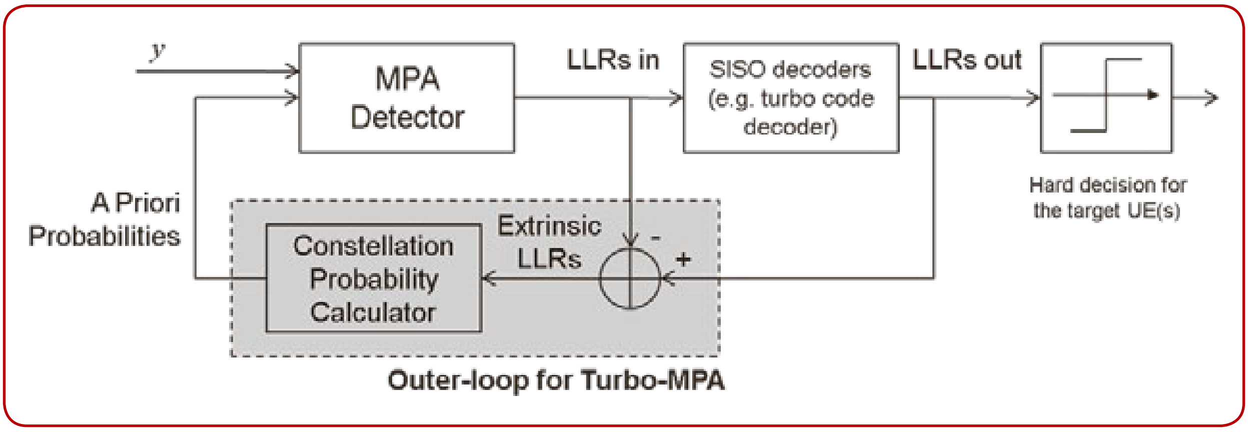


Fig. 4. Turbo-MPA with outer-loop to connect MPA and FEC decoders for better performance

Blind Detection

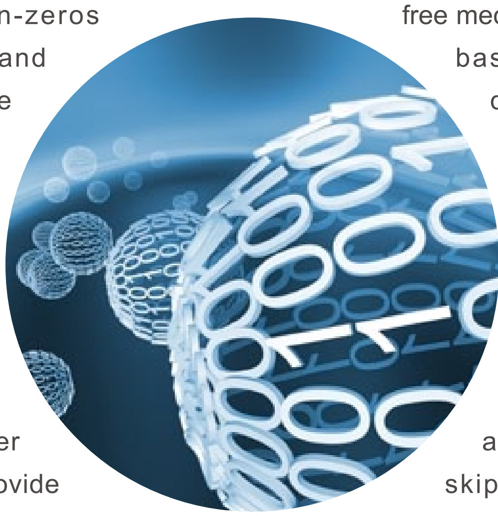
SCMA as a multiplexing scheme can be used to establish a contention-based uplink grant-free transmission to enable MTC. One of the key enablers of SCMA contention-based access is SCMA blind detection technique which can detect SCMA signal without the prior knowledge of the activity pattern of transmitted codebooks. The basic MPA requires the structure of factor graph and codebooks of every VN for its normal operation. In [14] [14], MPA is modified in order to simultaneously detect users' activity and their corresponding data. Simulation results confirm that the proposed blind detector can accurately identify active SCMA layers as if the full knowledge of active layers is available to a traditional MPA receiver.

3 SCMA Access Modes and USE Cases

Uplink Grant-Free Contention-Based Transmission for MTC

SCMA contention-based transmission is achieved by multiplexing different SCMA layers within a contention region. One or multiple contention regions are defined within the time-frequency plane of OFDMA. A user occupies the entire part of a contention region to transmit codewords of one SCMA layer. A user terminal is represented by its corresponding contention region, SCMA codebook, and pilot sequence. Each SCMA layer represents a user and is characterized by a specific SCMA codebook. To support m-MTC, the system can be overloaded where the number of users accessing a contention region simultaneously is more than the length of SCMA codewords. To even further increase the connectivity, codebook reuse is allowed among users since users are differentiated by their statistically independent random channels.

SCMA is scalable to support different levels of overloading for large number of connections. Size and number of the contention regions, length of codewords, sparsity pattern of codebooks (number of non-zeros tones of each codeword), and size of the codebooks are some of the parameters to trade off between number of supported users, traffic volume of users, complexity of detection, coverage, reliability of links, and system outage. For example, long codewords with large number of non-zero elements can provide better coding gain and hence better coverage, whereas more sparse codewords can tolerate further overloading to enable massive connectivity with feasible complexity of detection. A controller is required to adjust the system parameters semi-statically to the system load, traffic demand, and reliability for MTC communications.



Low dynamic signaling overhead and low transmission latency are achieved by a grant-free mechanism enabled by contention-based transmission and blind detection. A user transmits data in pre-configured resources that comprise of time, frequency, codebooks, and pilots without the dynamic request/grant procedure. Once a packet arrives, a user immediately wakes up and starts transmitting its data, skipping all the back and forth signaling required to establish a grant-based transmission as in LTE [7], [6]. The whole transmission can happen in a short period of time leading to much less energy consumption and latency which are critical issues in either m-MTC or c-MTC communications, respectively.

At the receiver side, an SCMA blind detector performs reliable joint user activity and signal detection without dynamic signaling overhead. The blind detection capability is realized with affordable complexity due to sparsity of SCMA codebooks. High reliability of MPA detection reduces the necessity of multiple retransmission which in turn leads to further energy saving of MTC terminals

System level simulation results demonstrate that an uplink contention-based SCMA system can support up to 3 times more devices [8] than a contention-based OFDMA system for delay sensitive small packet traffic under the 3GPP case 1 scenario [6].

Downlink Open-Loop Multi-User Transmission

In MBB, high data rate services are expected to be provided even for high mobility users such as car passengers. This wireless networking scenario falls into the moving network (MN) cluster as described by METIS [15]. The target is to provide mobility-robust and high-data rate communication links for vehicular users. As mentioned before, due to limitations of closed-loop MIMO multiplexing schemes such as MU-MIMO, open-loop downlink multiple access schemes can play an important role to facilitate MN for MBB.

Downlink multi-user SCMA (MU-SCMA) is an open-loop multiplexing scheme where different code domain layers are assigned to different users without the need of full CSI knowledge of the co-paired users. With a very limited need of channel knowledge in terms of channel quality indicator (CQI) [6], a TP simply pairs user together with appropriate power allocation among multiplexed layers. Compared to MU-MIMO, this system is more robust against dynamic channel variations in high speed scenarios. Code domain multiplexing and multi-user detection capability of MPA improves the robustness of SCMA multiplexing to CQI error. In addition, the problem of PMI measurement and report is removed for this open-loop multiple-access scheme.

On top of the advantages of MU-SCMA in terms of open-loop multiplexing and high spectral efficiency of multi-dimensional codewords, it provides more reliable link-adaptation from a system perspective.

The better link-adaptation of SCMA relies on the following factors:

i) in addition to the modulation level and coding rate as in QAM OFDMA systems, number of SCMA layers is an additional degree of freedom to finely tune transmission rates to channel conditions

ii) better resolution of codebook sizes. For example, one can design SCMA codebooks of size 4, 8, 16, 32, …,

iii) spreading of data across multiple OFDMA tones leads to interference whitening with robustness to time-variant interference environmentThe simulation results of [16] show more than 50% cell aggregate throughput gain over an open-loop OFDMA network. Moreover, performance gain is maintained at a high speed such as 120 km/h where MU MIMO fails due to channel aging. Therefore, downlink MU-SCMA scheme is suitable for MN scenarios with high data rate and mobility requirements.

Downlink Open-Loop CoMP for UDN and MN

UDN scenario posts challenges such as severe inter-cell interferences and mobility management, since a user sees a large number of TPs and suffers from interferences from those TPs. CoMP technique, where multiple TPs are coordinated, is a key technology to mitigate such interferences. Most proposed CoMP schemes are closed-loop and based on short-term CSI feedback from users to cooperating TPs. CSI reports can be challenging in UDN due to a large number of users and TPs, and in MN due to channel aging.

SCMA CoMP, with inter-TP layer assignment via a central scheduler, can provide an open-loop CoMP solution without knowledge of short-term multi-TP CSI. The benefits of SCMA CoMP scheme include: i) drastic overhead reduction of CSI acquisition, and ii) robustness to channel aging. In open-loop SCMA CoMP, different SCMA codebook sets are assigned to different TP antennas. Each transmit antenna uses a codebook set to multiplex users. A user terminal jointly detects the signals from multiple TPs within its CoMP cluster. As shown in Fig. 5, a neighboring TP can be either a cooperating or interfering TP. In the cooperating TP case of Fig. 5(a), the signal from a neighboring TP targets the same user and the open-loop joint-transmission CoMP is realized to improve the coverage for cell-edge users. The target user jointly detects the SCMA layers from multiple TPs. Note that the data of a user needs to be available at multiple TPs. Hence, it requires further backhaul traffic requirement which may not be available in every network infrastructure.

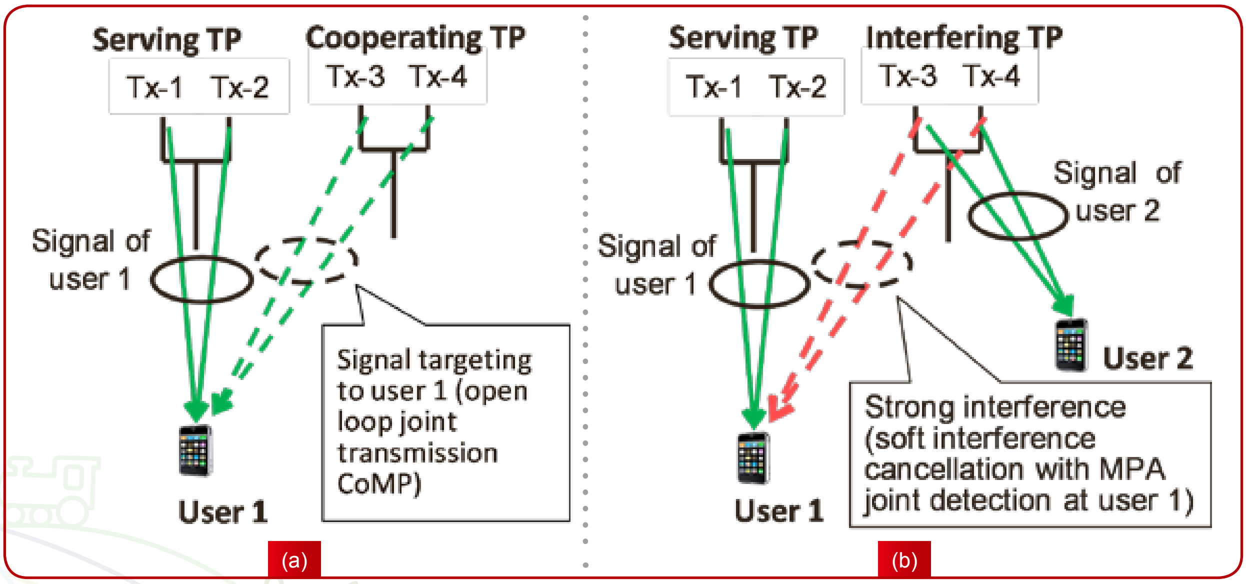


Fig. 5. SCMA CoMP, (a) open-loop join-transmission using SCMA CoMP, and (b) interference cancelation through joint detection

An alternative CoMP solution using SCMA joint detection for soft cancelation of interference is shown in Fig. 5(b), where TP 1 and 2 serve their own users. The cell-edge user receives a strong interference from its neighboring TP and conducts joint detection over its receiving signal and the interference for soft interference cancelation to provide better rate for cell-edge users.

To further enhance SCMA CoMP, it can be combined with multi-user SCMA to improve both the cell-edge and cell average throughput as illustrated in Fig. 6 for MN and UDN scenarios.

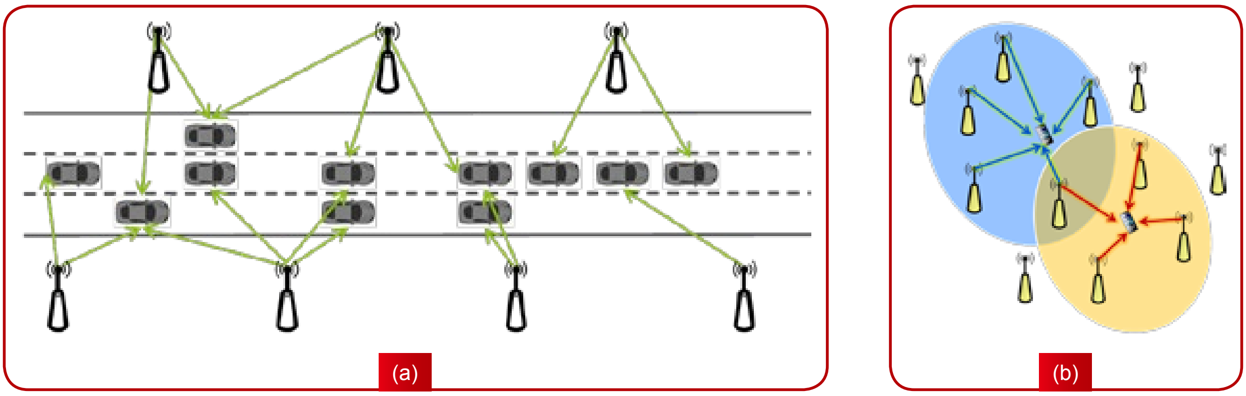


Fig. 6. SCMA CoMP and inter-TP MU-SCMA layer sharing for (a) MN, and (b) UDN scenarios

Multiple TPs serve a user through multiple SCMA links while SCMA layers are coordinately allocated to TPs. On the other hand, a TP may serve multiple users if they have overlapped CoMP clusters to enable user-centric CoMP as illustrated in Fig. 6. Multiple links to a user can facilitate seamless handover across UDN network in which frequent handover is a technical challenge. SCMA related parameters of neighboring TPs can be either blindly detected at a user, or signaled from a TP to scheduled users, provided that fast signaling exchanges are available among neighboring TPs. In [17], various algorithms of the combined schemes are presented. System level evaluations show that the proposed schemes provide more than 30% cell throughput and coverage gains over OFDMA for both pedestrian and vehicular users.

4 Uplink SCMA Prototype

In order to verify the SCMA technology and its advantages in real communication systems, we have developed an SCMA-based uplink multi-user system prototype on real-time hardware platforms [18]. Our demo system consists of 1 base station with 2 antennas for diversity combined receiving and in total 14 users each with 1 antenna for uplink access and data transmission. The basic system configurations of our demo system are set to align with the current LTE TDD system. In particular, we use LTE TDD frame structure configuration 1 and take the LTE physical layer with OFDMA orthogonal multiple access as the baseline for performance comparison.

The system diagram for SCMA uplink transmission is illustrated in Fig. 7 with the major difference to LTE system highlighted in different color.

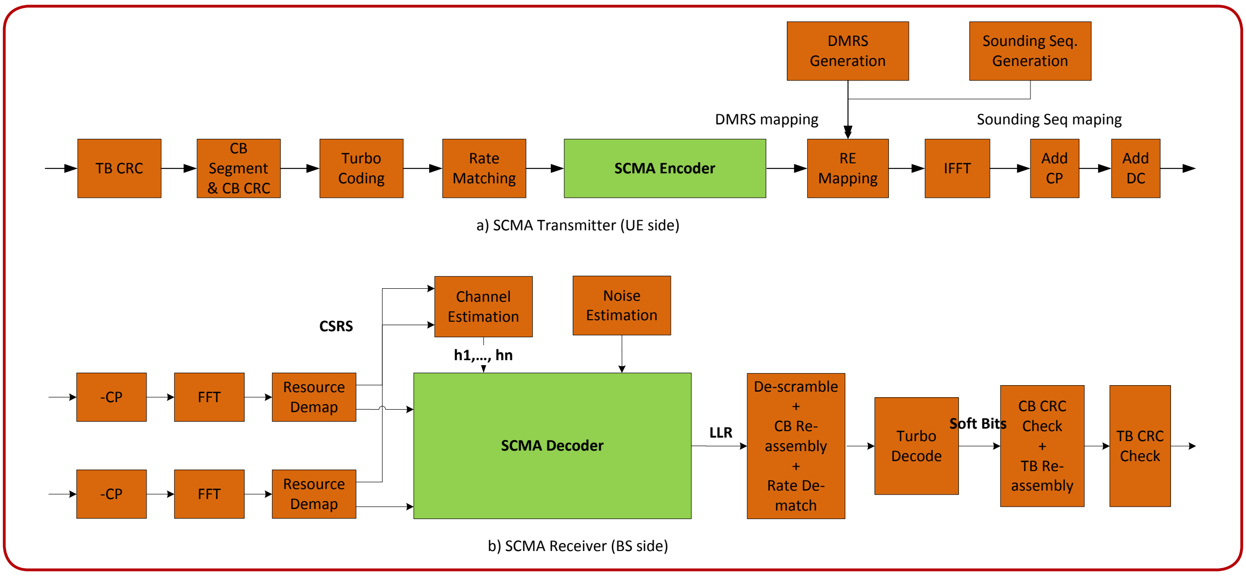


Fig. 7. System diagram for uplink SCMA system

The prototype can run in either mode, OFDMA or SCMA, separately, and support real-time switching from one to the other. To ensure a fair comparison, we keep the data rate of each user the same to guarantee the same quality of service. The major differences for the two systems are as follows:

· LTE OFDMA baseline: Multiple users transmit on separated resource blocks in an orthogonal way and are decoded in single user basis as implemented in current LTE OFDMA based systems.

· Proposed SCMA scheme: Multiple users transmit on all set of the resources blocks and overlap with each other in a low density way as explained in SCMA encoder while joint multi-user MPA detection is applied at the decoder to support larger number of co-current data stream transmissions.

The prototype system is built with soft baseband, that is, all the baseband processing is done by CPU instead of FPGA/DSP. At the base station side, one server (Huawei Tecal RH2288) is responsible for all the baseband processing, connecting with the standard commercial radio frequency components (Huawei product RRU3232). At the user side, the CPU of 1 laptop (MacBook Pro ME294CH/A) is used to model the processing of baseband for two users, which is then connecting to two mobile RF modules for testing. A user interface (UI) is developed to show the real-time throughput for each UE and all UEs, supporting also the real-time change of user status and system operation modes. Fig. 8 shows how the hardware demo looks like and the UI in real time.

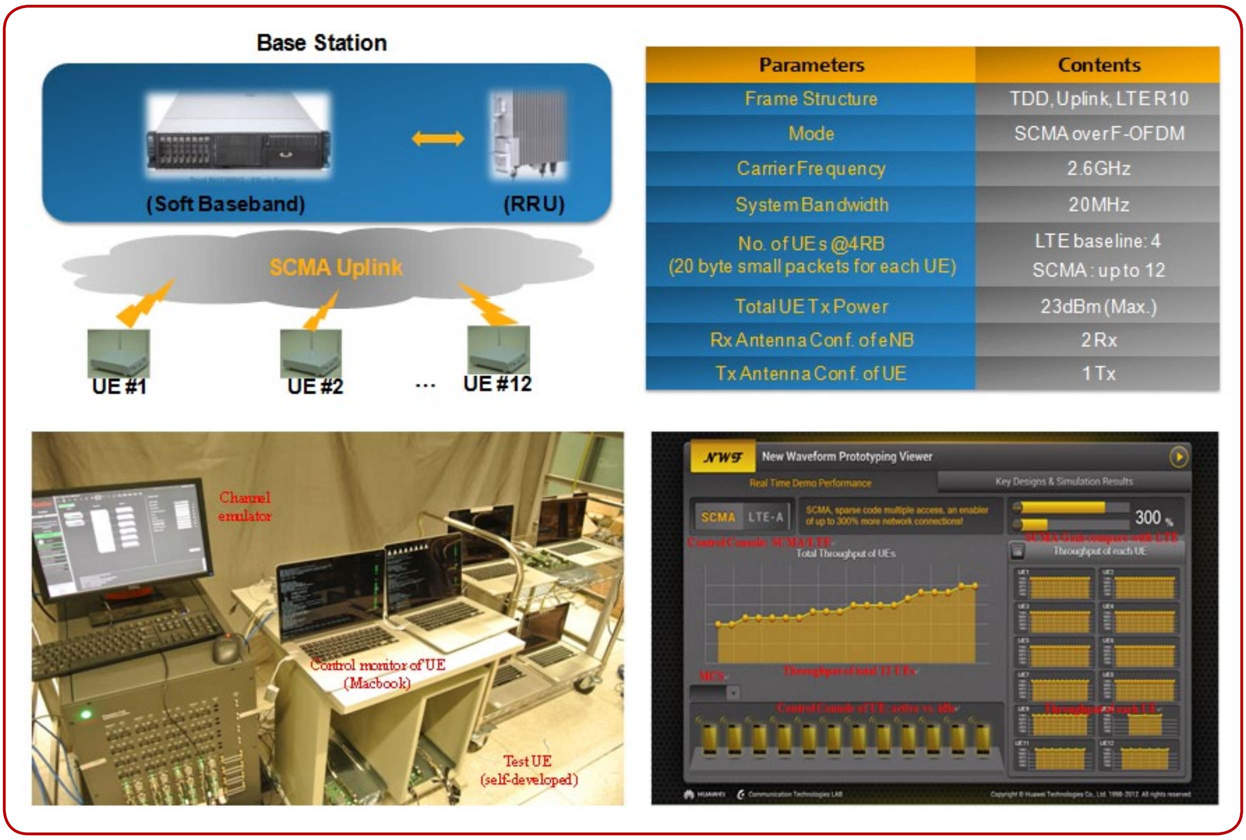


Fig. 8. The SCMA uplink overloading system demo (realtime)

We show by the prototype that with the application of SCMA technology over OFDMA, up to 300% overloading gain in the number of connections and network throughput are feasible, compared with the orthogonal multiple access baseline of 4G LTE. For instance, 150% overloading gain can be observed from the fact that given the data volume demand for each user to be 12 physical resource blocks (RBs), a system with a total of 48 RBs can serve at most 4 users using orthogonal LTE OFDMA. However, with SCMA, the codebook design supports 6 users with the same amount of data to share the 48 RBs simultaneously, thus the equivalent delivered amount of data is actually 12×6 = 72 RBs other than 12×4 = 48 RBs, resulting in the throughput gain of about 72/48 = 150%. The 300% gain is supported in a similar way but uses a different codebook with larger spreading factor and thus larger number of data layers. In our prototype, we use a 24-by-8 SCMA codebook to allow 12 users each with 2 data streams to access and transmit simultaneously with SCMA, while for LTE OFDMA, only 4 users out of 12 can transmit.

Moreover, we may further combine SCMA with F OFDM to improve the spectrum mask and to reduce the out-of-band emission. It has been proved through the prototype that the integration of SCMA and F-OFDM not only provide better localized spectrum than traditional OFDM, but also support asynchronous access between different SCMA groups.

5 Conclusion

In this paper, SCMA is introduced as a promising technology for the next generation of wireless networks. First, the SCMA structure, encoding, multiplexing, codebook design and low projection codebooks for lower detection complexity, and reception techniques are described. Then, SCMA features and advantages in terms of non-orthogonal multiplexing, overloading for massive connectivity and better spectral efficiency, sparsity of multi-dimensional codewords for low complexity of joint detection with high performance reliability, robustness of link-adaptation specially in time-variant interfering networks, and scalability to adapt the air interface to network condition and capabilities are explained. Inspired by these benefits, different application scenarios are introduced to demonstrate the potential gain of SCMA for 5G networks, including uplink grant-free contention-based transmission for MTC, downlink open-loop multi-user transmission for MBB, and downlink open-loop CoMP for UDN and MN. Based on the analysis and simulation results, SCMA is shown to resolve the major issues of the current wireless system and establish itself as a strong candidate for future 5G networks.

References:

[1] "5G: A Technology Vision," Huawei Technologies Co., Ltd., Shenzhen, China, Whitepaper Nov. 2013. [Online]. www.huawei.com/ilink/en/download/HW_314849

[2] 5GNOW. [Online]. http://www.5gnow.eu/

[3] Mobile and wireless communications Enablers for the Twenty-twenty Information Society (METIS). [Online]. https://www.metis2020.com/

[4] "NGMN 5G white paper," NGMN alliance, Feb. 2015. [Online]. https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0_01.pdf

[5] "4G Americas' recommendations on 5G requirments and solutions," 4G Americas, Oct. 2014. [Online]. http://www.4gamericas.org/files/2714/1471/2645/4G_Americas_Recommendations_on_5G_Requirements_and_Solutions_10_14_2014-FINALx.pdf

[6] 3GPP TR 36.814 V9.0.0, "3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release 9)," 3GPP, 2010-03.

[7] Erik Dahlman, Stefan Parkvall, and Johan Skold, 4G: LTE/LTE-Advanced for Mobile Broadband, 2nd ed. Waltham, MA, USA: Elsevier, 2014.

[8] K. Au et al., "Uplink contention based SCMA for 5G radio access," in Globecom Workshops (GC Wkshps), 2014, pp. 900-905.

[9] D.J. Love, R.W. Heath, V.K.N. Lau, D. Gesbert, and M. Andrews, "An overview of limited feedback in wireless communication systems," IEEE Journal on Selected Areas in Communications, vol. 26, no. 8, pp. 1341-1365, October 2008.

[10] Hosein Nikopour and Hadi Baligh, "Sparse code multiple access," in IEEE 24th PIMRC, 2013, pp. 332-336.

[11] Mahmoud Taherzadeh, Hosein Nikopour, Alireza Bayesteh, and Hadi Baligh, "SCMA Codebook Design," in IEEE VTC fall, Vancouver, 2014, p. Submitted.

[12] Reza Hoshyar, Razieh Razavi, and Mohammed AL-Imari, "LDS-OFDM an efficient multiple access technique," in VTC-spring, 2010, pp. 1-5.

[13] G.D., Jr. Forney and G. Ungerboeck, "Modulation and coding for linear Gaussian channels," IEEE Transactions on Information Theory, vol. 44, no. 6, pp. 2384-2415 , 1998.

[14] Bayesteh Alireza, Yi Eric, Nikopour Hosein, and Baligh Hadi, "Blind detection of SCMA for uplink grant-free multiple-access," in ISWCS, Barcelona, Spain, 2014, pp. 853-857.

[15] Petar Popovski and et al., "Initial report on horizontal topics, first results and 5G," METIS, ICT-317669-METIS/D6.2, April 2014. [Online]. https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D6.2_v1.pdf

[16] Hosein Nikopour et al., "SCMA for Downlink Multiple Access of 5G Wireless Networks," in IEEE Globecom, Austin, TX, 2014.

[17] Usa Vilaipornsawai, Hosein Nikopour, Alireza Bayesteh, "SCMA for Open-Loop Joint Transmission CoMP," submitted to VTC-fall, 2015.

[18] Zhang Shunqing et al., "Sparse Code Multiple Access: An Energy Efficient Uplink Approach for 5G Wireless Systems," in IEEE Globecom, 2014, pp. 4782-4787.

F-OFDM: Flexible Air Interface with Filtered OFDM

F-OFMD

Abstract — In this paper, a software defined flexible air interface is introduced to address the diverse services of the 5G wireless network. The air interface can be configured in a flexible way based on deployment scenarios, network and device capabilities, and traffic types. Software configurability and modularity allow future-proof and backward compatibility. A waveform constitutes an important component of the 5G air interface and plays a key role in enabling flexibility. Orthogonal frequency division multiplexing (OFDM) offers several advantages such as backward compatibility with LTE, ease of hardware implementation, low latency transmission with time localization, and multiple-input multiple-output (MIMO) friendliness. Due to the high spectral side lobes, large guard bands will be needed to accommodate different sets of numerology in the same system bandwidth in order to efficiently support diverse services. Filtered-OFDM (f-OFDM) is proposed to address this issue. A windowed-sinc filter is proposed that offers a good balance between time- and frequency-localization. Simulation results show that with properly designed filters, the number of guard tones required between subbands is minimal. Using a windowed-sinc filter with Hann window as an example, the number of guard tones required is between 0 and 2. With such a spectrally localized filter, scalable numerology based on the LTE backward compatible sampling frequency is proposed. Illustrative sets of numerology with subcarrier spacings of 7.5, 15, 30, and 60 kHz, cyclic prefix lengths as short as ~1 μs, and transmission time interval (TTI) lengths of 1, 0.5, 0.25, and 0.125 ms is presented for scenarios such as broadcasting/multicasting, vehicle-to-vehicle (V2V) communications and mobile broadband. F-OFDM allows the new air interface to better support diverse services, reduce or eliminate the guard band overhead, support asynchronous transmissions between subbands, and enable forward and backward compatibility.

Keywords — flexible air interface; filtered OFDM; OFDM; software configurable air interface.

1 Introduction

The rapid evolution of smartphones and tablets over the past several years has resulted in an explosive growth of data traffic over the mobile communications network that exceeds the growth seen in previous generations. With the proliferation of smart terminals communicating with servers and each other via broadband mobile networks, new applications are emerging to take advantage of the ubiquitous connectivity. As the computing power of mobile devices increases, people are evolving away from personal computers and laptops to mobile smartphones and tablets as they go about their daily work. At the same time, many "things" that were not previously connected, now have wireless connectivity. Examples range from home appliances, e-health, smart city, and smart cars to industrial automation.

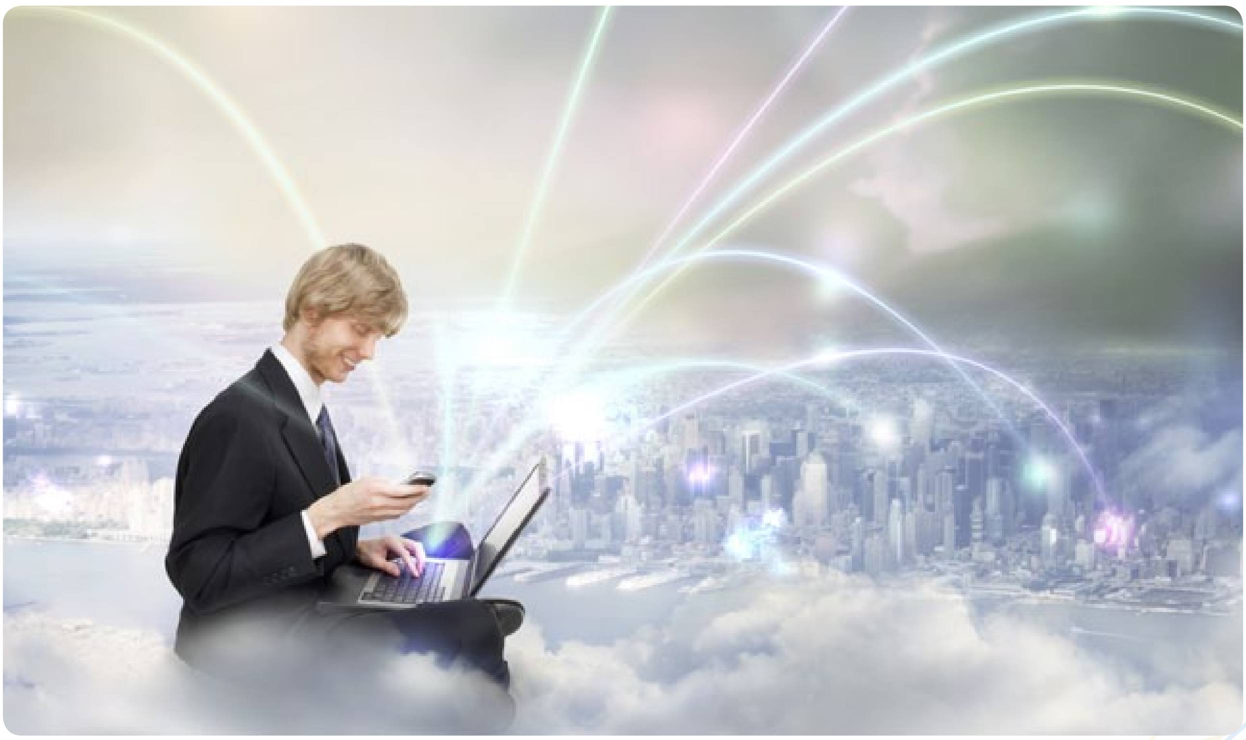
As a result of the explosive growth in the number of devices and applications, the 5G air interface needs to meet a wide-ranging set of requirements. Among them, the 5G air interface needs to support 1) diverse traffic characteristics such as QoS requirements, packet sizes and traffic patterns from a wide variety of applications, 2) diverse deployment scenarios from rural areas to ultra dense networks, 3) diverse types of transmit and receive points with different capabilities and 4) diverse application scenarios from human communications to machine type communications (MTC).

Current wireless systems (e.g. LTE) have been designed to support the requirements of applications envisioned at the time of the system's design. Furthermore, the scope of these requirements has traditionally been narrow enough to address by treating all traffic in nominally the same way. Going forward, the one-size-fits-all strategy will not be efficient to support the overall future traffic scenarios. The challenge for the 5G system design is to efficiently provide the wide variety of existing and future services, in a dense heterogeneous network setting.

This paper is organized as follows: The concept of a software defined flexible air interface for 5G is introduced in Section II. Filtered orthogonal frequency division multiplexing (f-OFDM) as an enabler of the new air interface is discussed in detail in Section III, including filter design, numerology discussion and system-level benefits of f-OFDM for supporting diverse services. Finally, conclusions are drawn in Section IV.

2 Software Defined Flexible Air Interface

It is a challenge to design a single fixed air interface to meet both future and legacy requirements together. In order to address the variety of requirements more efficiently, software configurability of the 5G air interface is essential to provide the required flexibility. A software defined flexible air interface is envisioned for 5G. It consists of a number of building blocks such as waveforms, frame structures, multiple access schemes, coding/modulation schemes and protocols. The new air interface is realized via adaptation and enhanced configurability of various air interface building blocks.



A software defined flexible air interface can be customized to best serve different applications under different transmission and reception conditions at the same time in a most efficient way. Moreover, the modular design of the new air interface offers the benefits of backward compatibility with existing radio access technologies and forward compatibility for future enhancements.

The procedures to realize an operating air interface configuration are shown in Fig. 1. Different scheme/technology candidates for each building block are predefined. Based on a set of input parameters such as the transmission content (e.g. traffic types), transmit/receive conditions (e.g. channel parameters, capabilities of transmitter and receiver) and the operating spectrum band, the air interface configuration process selects the optimized configuration. An air interface configuration is a set of schemes/ technologies chosen from one or more building blocks. Multiple configurations can be defined to support different scenarios. These configurations occupy different radio resources, but can co-exist in the same system and in the same or different operating spectrum bands. Moreover, the radio resource allocation for these configurations can be adjusted based on demand.

There are a number of ways the air interface configuration module (see Fig. 1) adapts the air interface. The new air interface can be a customized air interface for a vertical application. Based on the application scenario, a predefined configuration is applied. An example application is the air interface for MTC. Moreover, devices (and potentially some networks) that support only a particular vertical application can be configured for only one air interface configuration. Another adaptation mechanism is a multi-level adaptation where some air interface building blocks are configured statically or semi-statically while some building blocks can also adapt the air interface dynamically. An example of static configuration is based on transmit point capability. A low cost transmit point may support only a waveform with low Peak to Average Power Ratio (PAPR) and so would limit the range of its air interface configuration. In some scenarios such as intelligent spectrum utilization, semi-static configurations of waveform and frame structure building blocks are chosen depending on the operating frequency bands (e.g. to support devices or MTC sensors that are allocated for the band). An example scenario for dynamic configuration is content-aware services. For example, devices such as smartphones that support multiple applications with diverse characteristics (e.g. video, background traffic) may benefit from dynamically configuring building blocks based on their active applications.

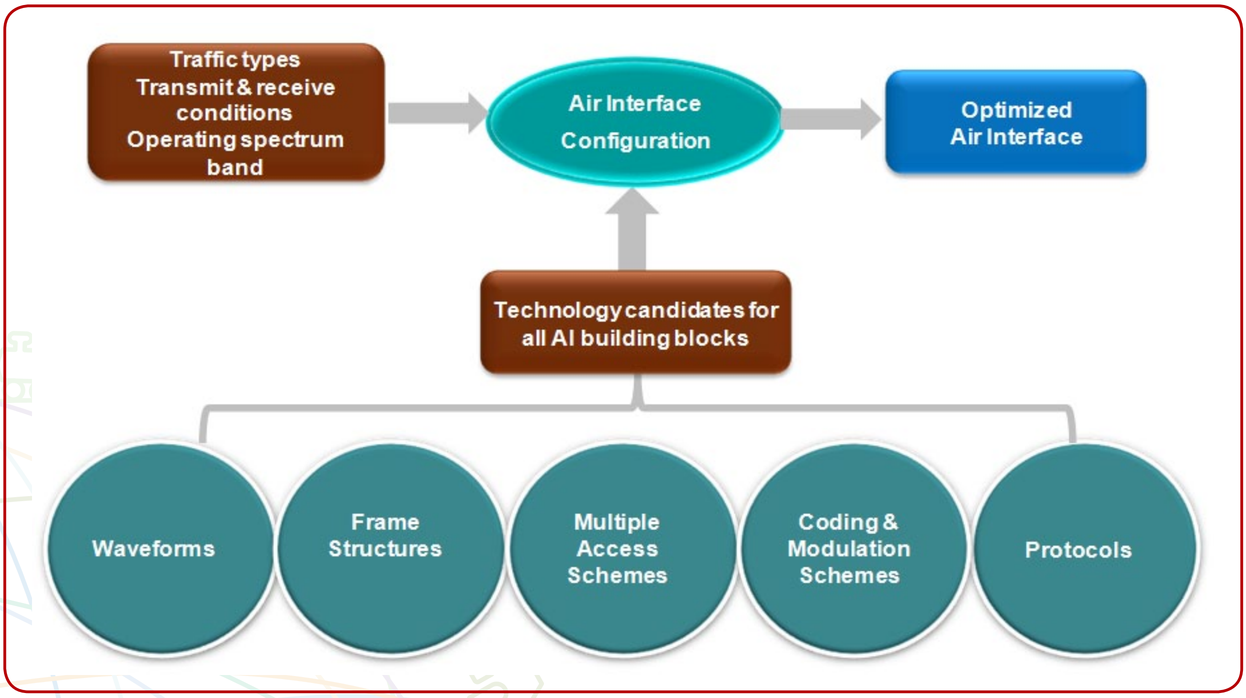
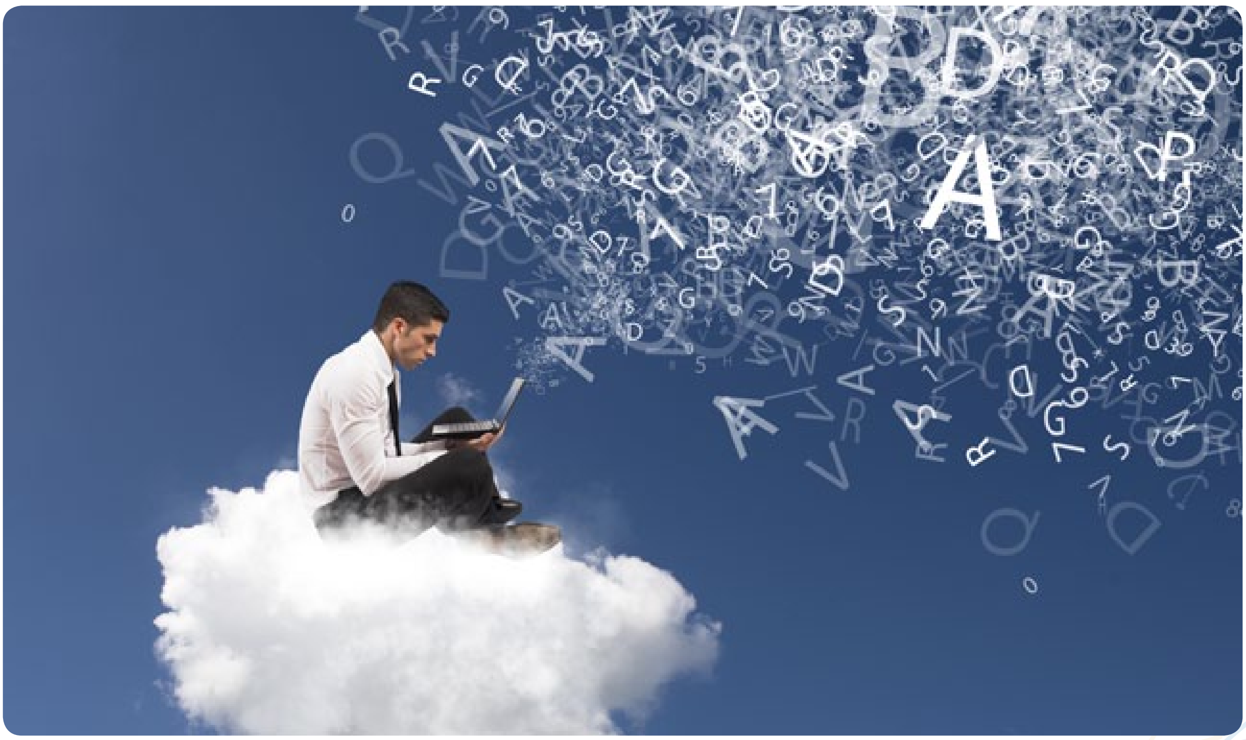


Fig. 1. Software defined flexible air interface is enabled by configuration of different building blocks

3 Filtered OFDM

Overview

Waveform constitutes an important component of the 5G air interface. Among the various waveforms studied, OFDM offers several advantages such as backward compatibility with LTE, ease of hardware implementation, low latency transmission with time localization and multiple-input multiple-output (MIMO)-friendliness. These benefits mean that OFDM will remain an important waveform of choice in the 5G air interface.



On the other hand, in pursuit of orthogonality for OFDM, a one-size-fits-all approach has been adopted in LTE. A unified waveform numerology, i.e., subcarrier spacing, length of cyclic prefix (CP), and transmission time interval (TTI), has been applied across the bandwidth provided in LTE. While easy to implement, the biggest issue for such a balanced and fixed design is the lack of flexibility. For instance, for critical MTC applications, the TTI duration should be made short to achieve ultra low round-trip latency. Besides, to support wide coverage multi-media broadcasting/multicasting, a long CP length is required to cope with potentially rich multi-path scattering, and thereby a longer symbol duration is preferred to balance the control signaling and CP overhead. In short, these typical 5G services cannot be supported efficiently under a unified

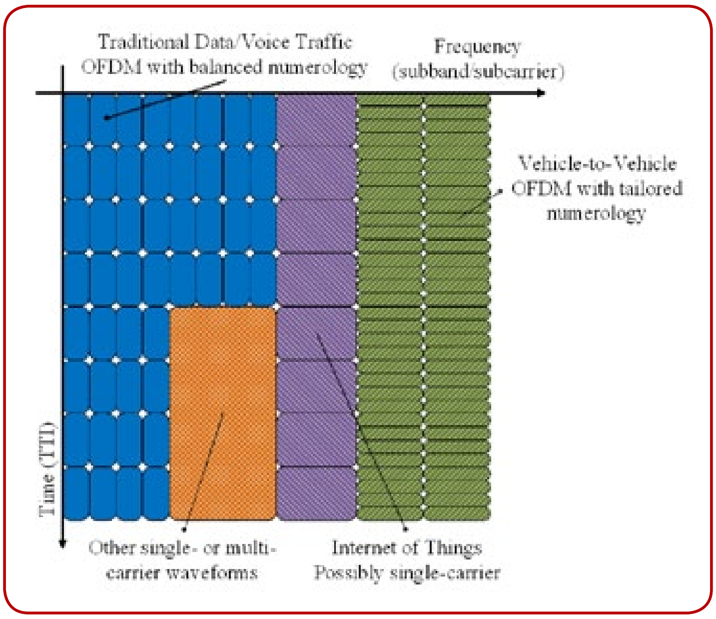


Fig. 2. Flexibility and coexistence of different waveforms by f-OFDM

OFDM numerology. Moreover, due to the high spectral side lobes of the OFDM sinc pulse in frequency domain, large guard bands will be needed to accommodate different numerologies in the same system bandwidth.

All in all, OFDM alone appears to be not suitable to provide the flexibility required by 5G, with which diversified types of services and channel characteristics are expected. In order to address the drawbacks while maintaining the advantages of OFDM, a new flexible waveform enabler named filtered-OFDM (f-OFDM) is proposed.

While it is expected that a larger bandwidth will be allocated to 5G, e.g. 100-200 MHz, it may not be efficient to apply a unified numerology to the entire system bandwidth. Instead, f-OFDM splits the assigned bandwidth into several independent subbands. In each subband, a conventional OFDM system (and possibly other waveforms) is tailored to suit the needs of a certain type of service and the associated channel characteristics. Subband-based filtering is then applied to suppress the inter-subband interference. By doing so, the strict time-domain orthogonality between consecutive OFDM symbols in each subband is given up intentionally in trade for a lower out-of-band emission. Therefore, subbands can be allocated with minimal guard tone overhead. With properly designed filters, the performance loss due to the filter introduced inter-symbol-interference (ISI) is marginal, as will be discussed later.

Fig. 2 gives a demonstration of the flexibility and coexistence of different waveforms and numerologies enabled by f-OFDM. Instead of a single numerology as employed by OFDM in LTE, the time-frequency arrangement/allocation of f-OFDM is much more flexible. Different types of services can be accommodated in different subbands of the same carrier with the most suitable waveform and numerology, leading to improved overall spectral efficiency. For instance, in order to provide ultra low latency and high reliability for vehicle-to-vehicle (V2V) communication, the TTI duration is shortened while the subcarrier spacing is enlarged. In general, different waveforms and numerologies can be incorporated under the flexible framework of f-OFDM, and the time-frequency arrangement may change with time, adapting to the changing service requirements over time.

Details of f-OFDM Design

1) Filter Design Based on Windowed-Sinc

Equiripple filter minimizes the maximum error between the desired and the actual filter response. It is optimal in the sense that, for a given set of design criteria, the length of the filter kernel is the shortest. However, equiripple filters are designed for time-domain implementation. In addition, with a sharp narrow transition region, the filter kernel of an equiripple filter is not well localized in the time domain.

In f-OFDM, a tradeoff between time and frequency domain localization of the filter is required. Windowed-sinc filter provides a convenient means to this end [1], wherein the prototype filter has a rectangular frequency response, i.e. sinc impulse response, with an appropriate bandwidth. This filter is ideal in the sense that it causes no distortion in its pass-band, while providing total out-of-band rejection. An appropriate time-windowing mask is applied on the impulse response of the prototype filter, and then, the resulting filter is shifted in frequency to be centered at the desired frequency. The windowing mask has smooth transitions to zero so that it avoids abrupt jumps at both ends of the truncated filter, and hence, avoids frequency spillovers. Moreover, the windowing provides a reasonable time-localization in the impulse response, and therefore, keeps the induced ISI in the resulting f-OFDM signal within an acceptable limit. A candidate for the windowing is the Hann window of an appropriate duration, e.g. half the duration of an OFDM symbol.

2) Differentiation Over UF-OFDM

A multicarrier waveform, called universal-filtered OFDM (UF-OFDM) or universal-filtered multicarrier (UFMC), was proposed in [2] based on subband filtering of the zero prefix (ZP)-OFDM. The proposed f-OFDM has two main differences from UF-OFDM:

• The filter length in UF-OFDM is limited to the length of ZP in order to avoid ISI. In contrast, the filter length is allowed to exceed the CP length in the proposed f-OFDM to achieve a better balance between the frequency- and time-localization. In particular, on the one hand, the extended filter length bestows the flexibility to design a filter that offers a desirable frequency-localization as well as an acceptable pass-band distortion for bandwidths as narrow as a few tens of OFDM subcarriers. On the other hand, using windowed-sinc, the filter can be designed such that the ISI incurred due to the extended filter length is very limited since the main energy of the filter is confined within its main lobe in time domain. The filtering performance is achieved without adding to signaling overhead

• Due to the use of CP, the channel equalization in f-OFDM has lower complexity than in UF-OFDM.

The main differences between the proposed f-OFDM and UF-OFDM are highlighted in Table I.

	f-OFDM	UF-OFDM
OFDM prefix	CP	ZP
Filter length	Beyond CP	Within ZP
Filter design	Windowed-sinc	Dolph-Chebyshev

Fig. 3 compares the power spectral density (PSD) of f-OFDM with that of LTE CP-OFDM and UF-OFDM. It shows that f-OFDM has a superior frequency localization compared with the other two waveforms. Fig. 4 shows the BLER performance of the asynchronous uplink filtered orthogonal frequency division multiple access (f-OFDMA) scheme, proposed based on f-OFDM, and compares it with that of the asynchronous UF-OFDM and also synchronous OFDMA. The performance of f-OFDMA is very close to that of synchronous OFDMA, while UF-OFDM has non-negligible performance loss.

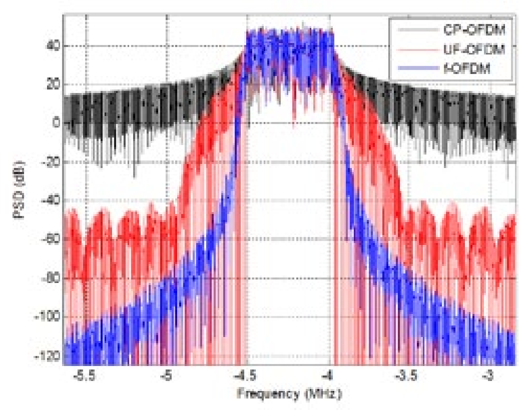


Fig. 3. PSD of f-OFDM compared with CP-OFDM and UF-OFDM

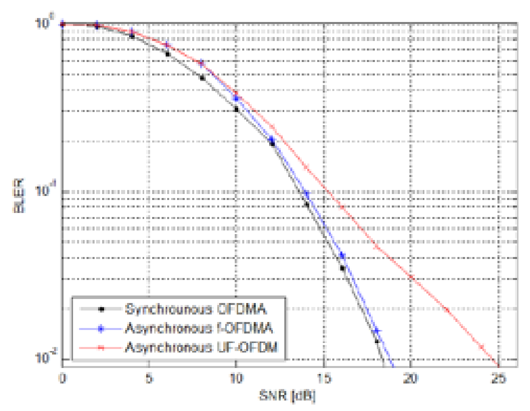


Fig. 4. The BLER performance of the proposed asynchronous f-OFDMA with 16QAM and FEC rate 1/2 over TU channel

Adaptive Parameters Configuration

As discussed earlier, with f-OFDM, different numerologies can be applied in different subbands, adapting to the needs of different types of services and the associated channel characteristics. Table II presents several possible parameter configurations, scaled from that of the LTE systems i.e., the second column. Based on this table, customized numerology can be provided to suit different needs. For instances, the first column, i.e., the OFDM numerology with narrow subcarrier spacing of 7.5 kHz, long CP of around 9.5 μs and TTI duration of 1 ms, can be applied to provide a wide coverage broadcasting/multicasting services. The fourth column, which consists of large subcarrier spacing of 60 kHz and short TTI duration of 0.125 ms, is particularly suitable for V2V communications. The second and third columns could be used for traditional mobile broadband services. With such a strategy to implement f-OFDM, i.e., different numerologies scaled from a common ancestor, the sampling frequency can be maintained in a uniform and LTE backward compatible manner, which will be easier for practical implementation. It should be noted that this is an example solely for illustrative purposes. The numerologies can be further optimized for each type of service.

The adaptation of f-OFDM can also include subband division and relocation. That is, the subband's bandwidth allocated to different types of services can be dynamically adjusted, based on a continuous tracking of the active users in each subband. In this way, a higher flexibility can be achieved for f-OFDM, leading to better spectrum utilization.

TABLE II. SCALABLE NUMEROLOGY OF F-OFDM.

Subcarrier spacing (KHz)	7.5	15	30	60
Symbol duration (μs)	133.33	66.67	33.33	16.67
CP length (μs)	9.54/9.44	5.2/4.7	2.4/2.38	1.2/1.17
# of symbols per TTI	6/1	1/6	1/6	4/3
TTI (ms)	1	0.5	0.25	0.125
CP overhead	6.7%	6.7%	6.7%	6.7%
Bandwidth (MHz) / FFT size	2.5/512	20/2048	20/1024	20/512
Sampling frequency (MHz)	3.84	30.72	30.72	30.72

To accommodate the inter-subband interference brought by non-orthogonal numerologies and asynchronous transmission, guard tones are needed between adjacent subbands. The question is how large the number of guard tones should be. The following discussion provides some insight into the guard tones required in different situations.

With properly designed filters such as the windowed-sinc filter with Hann window, the number of guard tones between subbands can be set to a minimum level to maximize the spectrum utilization. Our simulation results (Fig. 4) have indicated that:

• With equal transmit power in adjacent subbands:

○ For low to medium modulation orders (e.g., QPSK and 16QAM), no guard tone is required (see Fig. 4).

○ For high-order modulation (e.g., 64QAM), up to 2 guard tones are needed.

• Even if the transmit power in adjacent subbands is lifted by 10 dB, 2 guard tones are sufficient for suppressing the inter-subband interference for most practically used modulation/coding scheme (MCS) levels.

• If combined with proper scheduling, e.g., putting the high MCS user equipments (UEs) away from the subband edges, the required guard tones can be reduced to zero.

In summary, adaptive and differentiated waveform parameter configuration can be supported by f-OFDM, and the guard tone overhead between subbands has been verified to be marginal.

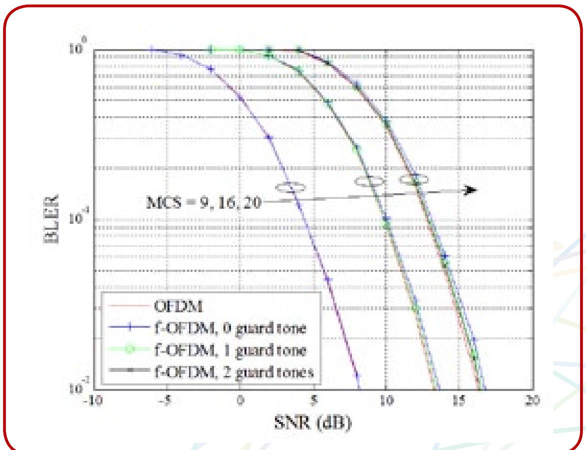


Fig. 5. BLER in a subband of f-OFDM with different numbers of guard tones between adjacent subbands

Benefit Analyses

The benefits of f-OFDM are not only the flexibility to support various types of services, but also include:

• Reduced out-of-band emission through filtering, and thus much smaller guard band consumption, and thereby improving spectrum utilization.

• Relaxed requirement on global synchronization and support for inter-subband asynchronous transmission, thus reducing synchronization overhead.

• Both backward and forward compatibility on the evolution path from 4G to 5G and even beyond 5G.

Experimentations have been carried out to verify the benefits promised by f-OFDM. In the first verification case, the 20 MHz bandwidth used in LTE systems is split up into 3 subbands through f-OFDM: 1st and 3rd with 1MHz bandwidth in the edges and the 2nd with 18MHz bandwidth in the middle. The standard OFDM numerology in LTE is applied across the three subbands, and a half symbol timing difference is introduced to model asynchronous transmission. Simulation results have indicated that with elaborately designed subband filters, through f-OFDM, these two 1MHz subbands can be efficiently utilized for data transmission without interfering with the existing LTE service in the middle 18MHz subband. These observations have also been verified in prototyping and field testing, as shown in Fig. 6.

In the second case, initial efforts have been taken to investigate the throughput gain due to numerology adaptation for various application scenarios. In this simple experiment, four types of scenarios with different channel dispersion characteristics are considered: 1. Pedestrian, 2. Urban, 3. Highway, 4. V2V. The differences in channel characteristics among these four scenarios are reflected in the numerologies chosen by the f-OFDM system. In addition, with f-OFDM, the guard tone overhead is minimized. The system bandwidth available is evenly split up into four subbands to accommodate the aforementioned four cases. OFDM with a uniform numerology is taken as the benchmark. However, the numerology of OFDM is chosen according to the worst case scenario, e.g. extended CP is used to combat rich multi-path scattering in urban environment, and 10% of the bandwidth is reserved as guard band. As shown in Fig. 7, significant throughput gains can be achieved in each subband and in total. The throughput gain comes from not only the savings on guard band, but also the numerology adaptations to the channel characteristics, i.e., reduced CP length for smaller multi-path delay spread and reduced subcarrier spacing for stronger frequency selectivity. In this toy example, the total throughput gain of f-OFDM over OFDM is up to 46%, which is very appealing.

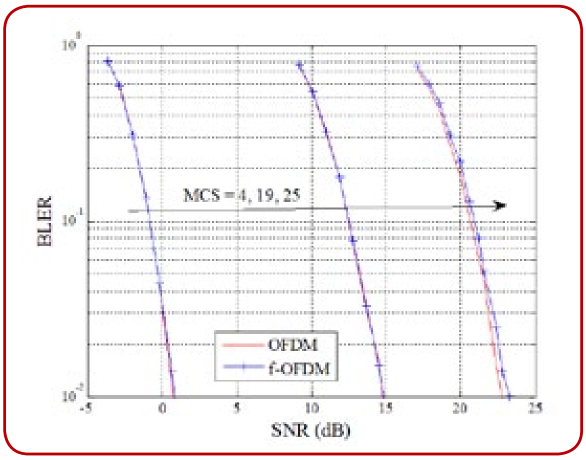


Fig. 6. Prototype BLER in the 2nd subband with (blue) and without (red) data transmission in the 1st and 3rd subbands

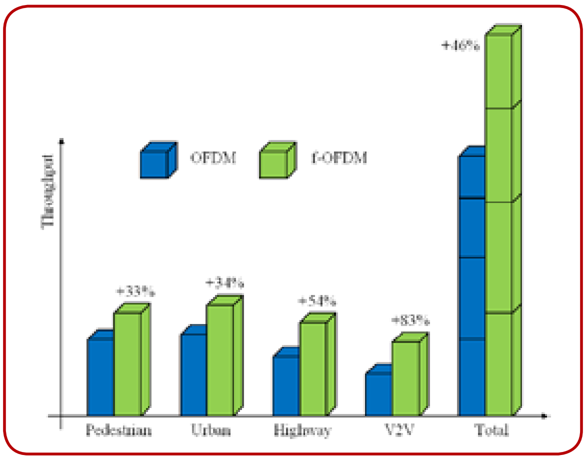


Fig. 7. Throughput gain of f-OFDM over OFDM

In our study, f-OFDM has demonstrated excellent capability for improving the spectrum utilization by reducing guard band consumption, and outstanding suitability for accommodating various application scenarios with higher spectrum efficiency. Nevertheless, more performance improvements can be expected from f-OFDM, if combined with other new multiple access and resource allocation schemes.

4 Summary and Conclusions

In order to support the very diverse services of the 5G wireless network, a flexible air interface is needed. In this paper, a software defined flexible air interface is introduced to provide such flexibility. The new air interface consists of a number of building blocks such as waveforms, frame structures, multiple access schemes, coding/modulation schemes and protocols. Within each building block, technology candidates are defined. An air interface configuration is formed by selecting technologies from each building block. The air interface can be configured in a flexible way based on deployment scenarios, network and device capabilities and traffic types. Software configurability and modularity allows future-proofing and backward compatibility.

A key enabler of such a flexible air interface is f-OFDM. With f-OFDM, a spectrally localized subband filter is applied in the system. Therefore, different waveforms and numerologies can co-exist in the same system bandwidth. A windowed-sinc with balanced time and frequency localization is described in this paper.

A scalable numerology based on LTE backward compatible sampling frequency is proposed. An illustrative set of numerologies with subcarrier spacings of 7.5, 15, 30, and 60 kHz, cyclic prefix lengths as short as ~1 μs and TTI lengths of 1, 0.5, 0.25, and 0.125 ms, is presented to address diverse services such as broadcasting/multicasting, V2V communications and mobile broadband.

Simulation results have shown that with properly designed filters, the number of guard tones required between subbands is minimal. Specifically, our evaluations have shown that with a windowed-sinc filter with Hann window, the number of guard tones required is 2 when there is a transmit power difference of 10 dB between subbands. For equal transmit powers, guard tones required are 2 for 64 QAM and no guard tone is needed for QPSK and 16 QAM. With proper scheduling of users, the system can therefore be designed with very small (or even zero) guard tones between subbands.

Some initial evaluations are also performed to estimate potential throughput gain of an f-OFDM system with optimized numerology parameters vs. a one-size-fits-all set of OFDM parameters with guard bands. It shows a gain of up to 46%.

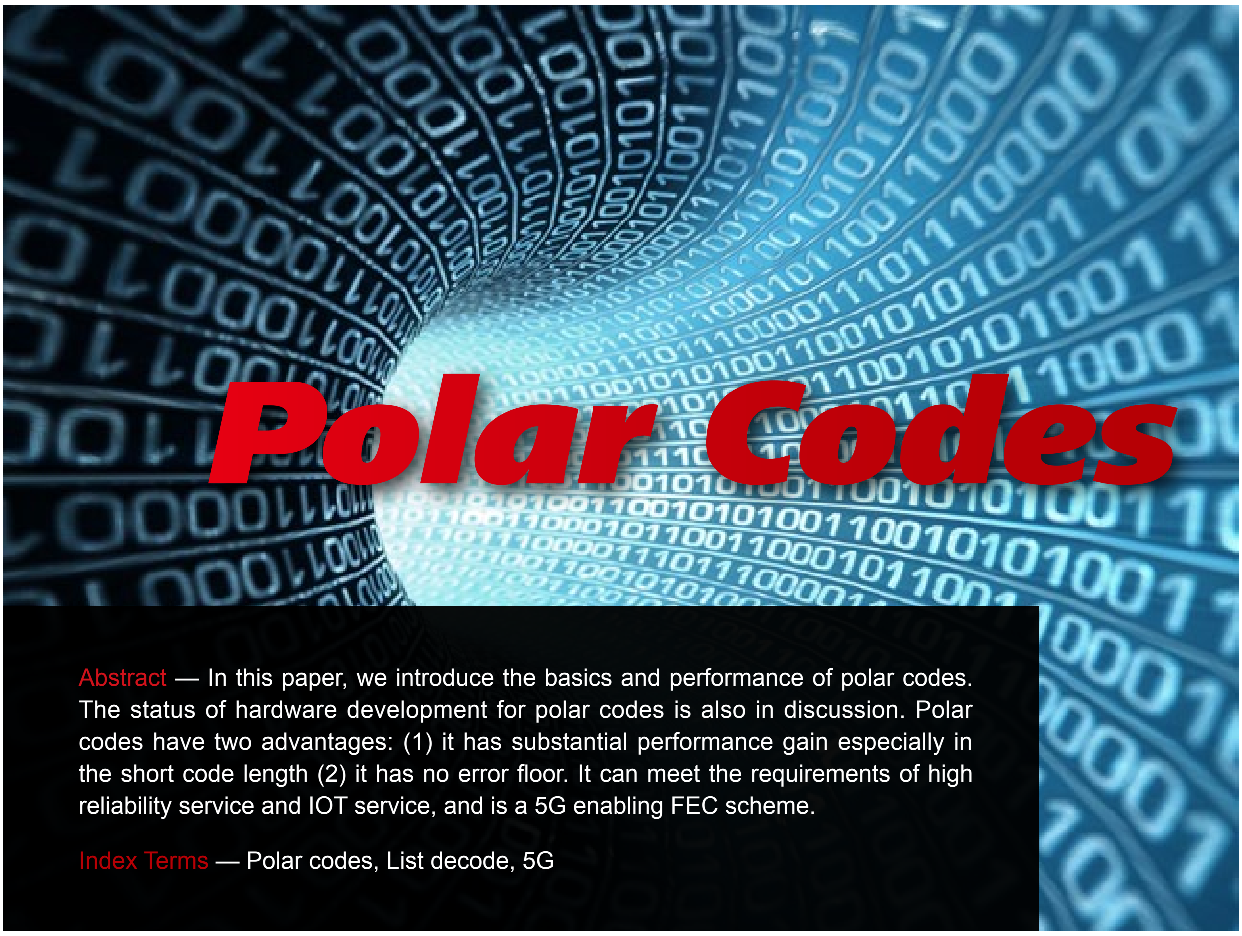
F-OFDM allows the new air interface to better support diverse services, reduce or eliminate the guard band overhead, support asynchronous transmissions between subbands and enable forward and backward compatibility. It is a key enabler for the 5G software defined flexible air interface.

References:

[1] J. Abdoli, M. Jia, J. Ma, “Filtered OFDM: A New Waveform for Future Wireless Systems,” to appear in IEEE Internationl Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2015.

[2] V. Vakilian, T. Wild, F. Schaich, S. Ten Brink, and J.-F. Frigon, “Universal-filtered multi-carrier technique for wireless systems beyond LTE,” in IEEE Globcom Workshops (GC Wkshps), 2013, pp. 223-228.

Polar Codes: A 5G enabling FEC scheme



Abstract — In this paper, we introduce the basics and performance of polar codes. The status of hardware development for polar codes is also in discussion. Polar codes have two advantages: (1) it has substantial performance gain especially in the short code length (2) it has no error floor. It can meet the requirements of high reliability service and IOT service, and is a 5G enabling FEC scheme.

Index Terms — Polar codes, List decode, 5G

1 Introduction

Polar codes are a major breakthrough in coding theory [1].

Polar codes are the first family of codes known to achieve channel capacity with explicit construction. They can achieve Shannon capacity with a simple encoder and a simple successive cancellation (SC) decoder, both with low complexity of the order of , where N is the code block size. Notwithstanding great advances in the literature on LPDC/Turbo codes and iterative decoding, no other family of codes with this property was known before polar codes came along — not even for the simple case of the binary symmetric channel.

In comparison with the current state-of-the-art error-correcting codes, polar coding offers several distinct advantages. Polar codes are the first-ever deterministic family of capacity-achieving codes; no element of randomness is involved in their design. Combined with the extremely regular and recursive structure of polar codes, this distinct feature greatly facilitates their implementation, in both software and hardware. Moreover, polar codes come with an explicit guarantee on the probability of decoding error. This makes these codes immediately applicable in situations where the required error rates are difficult to verify by simulation, and where the error-floor phenomenon inherent in turbo and LDPC codes presents a problem. Finally, the polarization paradigm that underlies polar codes extends beyond point-to-point transmission scenarios, to situations where LDPC or turbo codes are inadequate. Other examples where polar codes provide a capacity-achieving solution include multi-user channels, lossless and lossy data compression, Slepian-Wolf coding, randomness extraction, and sparse recovery.

2 Introduction of Polar Codes

Polar Codes

Let $F=\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}$, $F^{\otimes n}$ is a $N\times N$ matrix, where $N=2^n$, $\otimes n$ denotes nth Kronecker power, and $F^{\otimes n}=F\otimes F^{\otimes(n-1)}$. Let the n-bit binary representation of integer *i* be $b_{n-1},b_{n-2},\dots,b_0$. The n-bit representation $b_0,b_1,\dots,b_n$ is a bit-reversal order of *i*. The generator matrix of polar code is defined as $G_N=B_N F^{\otimes n}$, where B_N is a bit-reversal permutation matrix. The polar code is generated by

$$x_1^N=u_1^N G_N=u_1^N B_N F^{\otimes n} \quad (1)$$

where $x_1^N=(x_1,x_2,\dots,x_N)$ is the encoded bit sequence, and $u_1^N=(u_1,u_2,\dots,u_N)$ is the encoding bit sequence. The bit indexes of u_1^N are divided into two subsets: the one containing the information bits represented by *A* and the other containing the frozen bits represented by A^c. The polar code can be further expressed as

$$x_1^N=u_A G_N(A)\oplus u_{A^c} G_N(A^c) \quad (2)$$

where $G_N(A)$ denotes the submatrix of G_N formed by the rows with indices in *A*, and $G_N(A^c)$ denotes the submatrix of G_N formed by the rows with indices in A^c. u_A denotes the information bits, and u_{A^c} denotes the frozen bits.

Polar codes can be decoded with a highly efficient SC decoder, which has a decoding complexity of $O(N\log N)$ and can achieve capacity when *N* is very large.

SC-List Decoder for Polar Codes

Though the SC decoder approaches Shannon capacity, it does not perform well for polar codes with small and medium block lengths. Therefore, a more powerful SC-List decoder is proposed in [2] and performs much better than the SC decoder. Instead of keeping only one survival path as in the SC decoder, the SC-List decoder keeps L survival paths. At each decoding time, the SC-List decoder splits each current decoding path into two paths attempting both $\hat{u}_i = 0$ and $\hat{u}_i = 1$ (if u_i is an unfrozen bit). When the number of paths grows beyond a predefined threshold *L*, the SC-List decoder discards the worst (least probable) paths, and only keeps the *L* best paths. Simulations show that the SC-List decoder with *L*=32 performs much better than the SC decoder, and it can achieve the same performance as the optimal maximum likelihood (ML) decoder at high SNR for the polar code with coding rate R=1/2, and *N*=2048 and 8192 [2].

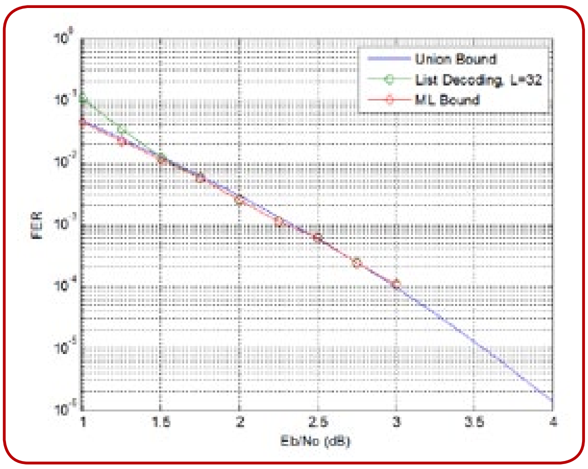


Fig.1. The FER performance of the SC-List decoding with *L*=32 for polar code (2048, 1024)

Fig.1 shows the frame error rate (FER) of the polar code (2048, 1024) using the SC-List decoder with *L*=32, where the additive white Gaussian noise (AWGN) channel is used and the signal is BPSK modulated. The ML (maximum likelihood) bound is also simulated in the same way as that used in [2]. We plot a union bound using the minimum distance d_0=16 and the second least distance d_1=24, and the numbers of codewords with d_0 and d_1 are N_{16}=11648 and N_{24}=215040 respectively. The parameters d_0, d_1, N_{16} and N_{24} are obtained as follows. As we know, the list decoder with *L*=32 can achieve the same performance as the ML decoder at high SNR. Therefore, we set very high *SNR*→∞ and use very large list size *L*=1,000,000 for the list decoder. We assume that the error patterns, which are obtained from the incorrect survival paths of the decoder output, generate codewords with the least weights. It is found that N_{16}=11648 and N_{24}=215040 error patterns generate weight-16 and weight-24 codewords respectively, and the rest of error patterns generate weight-32 codewords. We approximate the union bound on the frame error rate using the first two terms as

$$P_e \le \frac{N_{16}}{2}\,erfc\left(\sqrt{\frac{d_0 E_b}{N_o}}\right) + \frac{N_{24}}{2}\,erfc\left(\sqrt{\frac{d_1 E_b}{N_o}}\right).$$

In order to improve the minimum distance of polar codes, the concatenation of polar codes with CRC [2]was proposed, and the performance of polar code concatenated with CRC can be improved further.

Adaptive SC-List Decoder for Polar Codes

With the increase in list size, the performance of polar codes concatenated with CRC can be improved further. However, list decoders with large list size have high complexity. In fact, we observed that for most of the received frames, the SC-List decoder with very small L can successfully decode information bits, and few frames need large L for successful decoding. Therefore, in order to reduce the decoding complexity, we propose an adaptive SC-List decoder for polar codes with CRC. The adaptive SC-List decoder initially uses very small L, and then iteratively increases L (if there is no survival path passing CRC), until L reaches a predefined number L_{max}.

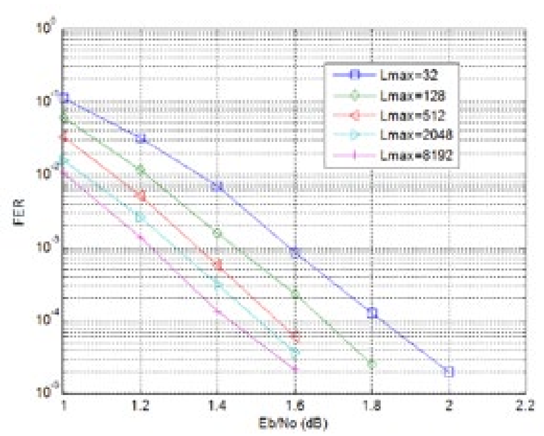


Fig. 2. The FER performance of the polar code (2048, 1024) with 16-bit CRC using the adaptive SC-List decoder with different L_{max}

Fig.2 shows the FER performance of the adaptive SC-List decoder for the polar code (2048, 1024) with a 16-bit CRC. The channel is an AWGN channel, and the signal is modulated by BPSK modulation. It is shown that there is about 0.4dB gain of L_{max}=8192 over L_{max}=32 at FER=10^{-3}. The false alarm probability that the incorrect paths pass CRC is much lower than the target FER=10^{-3}. If the adaptive list decoder with $L<L_{max}$ contains the correct path, then the non-adaptive list decoder with $L=L_{max}$ also contains the correct path, and both decoders can successfully decode the polar code; If the adaptive list decoder with $L<L_{max}$ does not contain the correct path, it will increase L until $L=L_{max}$, and this leads to that both decoders use the same $L=L_{max}$ and perform the same as each other. Therefore the performance of the adaptive SC-List decoder with L_{max}=32 and L_{max}=8192 is the same as the non-adaptive SC-List decoder with constant L=32 and L=8192 respectively. TABLE I shows the mean L for different E_b/N_0 and different L_{max}. With the increase of E_b/N_0, the SC-List decoder is more likely to successfully decode the received frames with the same L, and therefore the mean of L becomes smaller for the adaptive SC-List decoder. Since the complexity of the SC-List decoder is linear in the list size, the SC-List decoder with constant L has a complexity of the order of $O(LN\log N)$ and our adaptive SC-List decoder has an average complexity of the order of $O(\bar{L}N\log N)$. It is seen that under L_{max}=32, the mean of L is $\bar{L}=2.04$ for E_b/N_0=1.6dB; this is about 16 times complexity reduction but with the same performance compared with the constant L=32. The mean of L under L_{max}=8192 is $\bar{L}=2.47$ for E_b/N_0=1.6dB; this is about 3316 times complexity reduction but with the same performance compared with the constant L=8192.

It is interesting to mention that when we simulated the adaptive SC-List decoder with very large L_{max}=262144 for the polar code (2048, 1024) with a 24-bit CRC, the mean $\bar{L}=818.5$. We found that this concatenated code can achieve FER≤10^{-3} at E_b/N_0=1.1dB, which is about 0.2 dB from the information theoretic limit at the same block length for the BAWGN channel [7]. This performance is much more difficult to simulate by the SC-List decoder with constant L=262144.

TABLE I. THE MEAN OF L OF THE ADAPTIVE SC-LIST DECODER

E_b/N_o(dB)	1.0	1.2	1.4	1.6	1.8	2.0
L_{max}=32	16.64	8.03	3.86	2.04	1.39	1.14
L_{max}=128	35.31	12.16	4.52	2.17	1.41	
L_{max}=512	70.41	19.14	5.45	2.27		
L_{max}=2048	133.40	30.80	6.64	2.36		
L_{max}=8192	271.07	52.59	7.88	2.47		

3 Performance of Polar Codes

A lot of performance simulations show that polar codes concatenated with cyclic redundancy codes (CRC) and an adaptive SC-list decoder can outperform turbo/LDPC codes for short and moderate code block sizes. For large code block sizes, polar codes are slightly better. With the increase in list size, the performance of polar codes concatenated with CRC can be improved further.

Fig.3 describes the performance in the case: N=256, K=88, R=K/N; QPSK Modulation. K equals turbo internal interleaver length. For simplification, only CB is considered. For 24 bit CRC, effective code rate R_eff=(K-24)/N. Eb/N0 is calculated by R_eff.

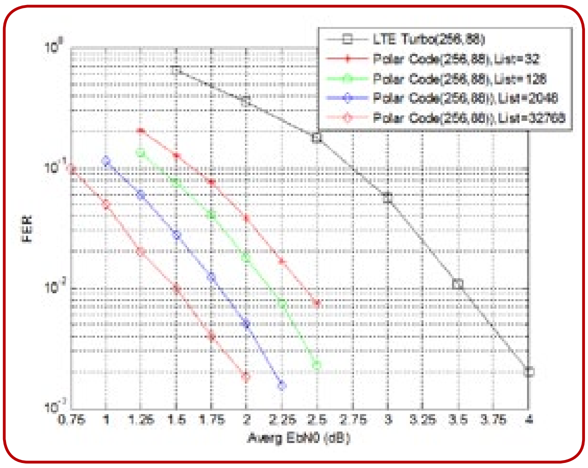


Fig. 3. The FER performance of the polar code (256, 88) with 24-bit CRC

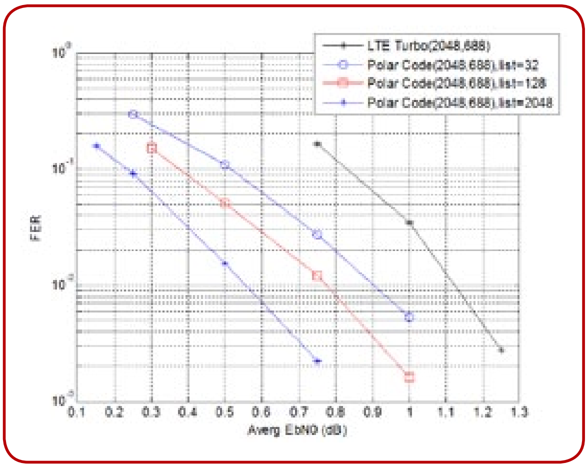


Fig. 4. The FER performance of the polar code (2048, 688) with 24-bit CRC

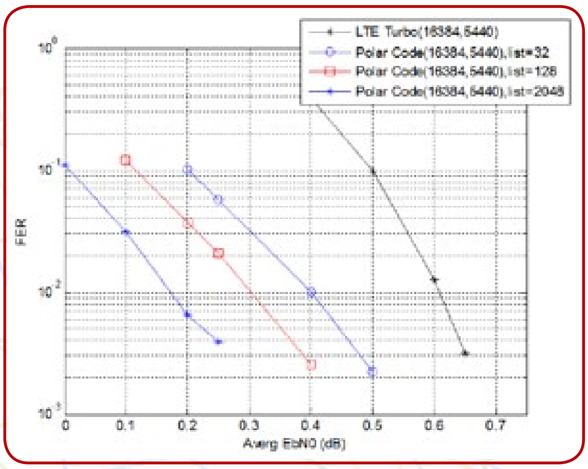


Fig. 5. The FER performance of the polar code (16384, 5440) with 24-bit CRC

In LTE, some control channels, such as PBCH channel, PDCCH and some PUSCH (UCI, non-periodic large CQI report [great than11 bits]), use Tail biting Convolution code (TB CC). In fact, the polar code in control channel has substantial performance gain. The key point is to design the Rate matching scheme of polar codes in control channel. We propose the sorted congruential Rate Match scheme. The performance is as follows:

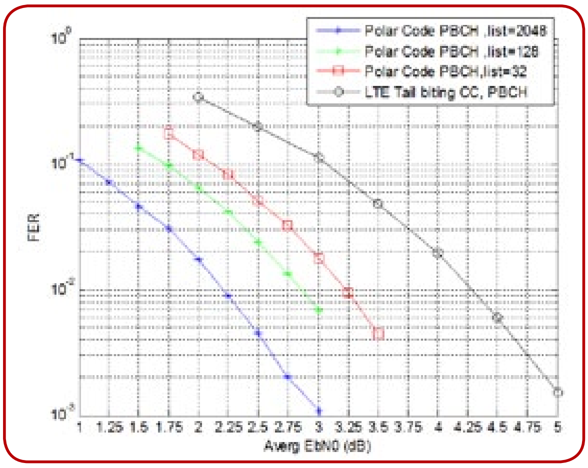


Fig. 6. The design of the polar code in PBCH channel

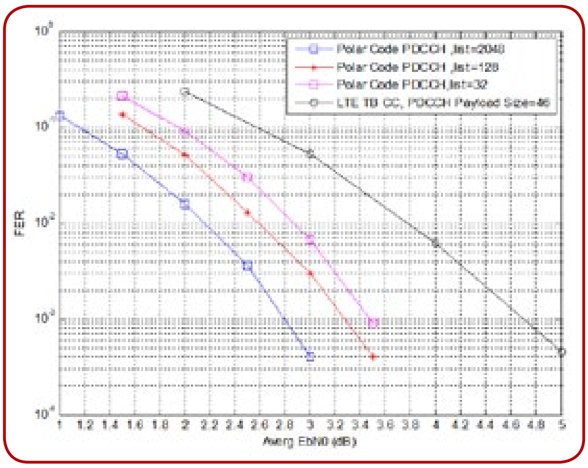


Fig. 7. The design of the polar code in PDCCH channel

Fig.6 and Fig.7 show that the polar code with list size 32 in LTE control channel has at least 0.9 dB performance gain compared with tail biting convolution code.

Finally, the hybrid automatic repeat request (HARQ) schemes based on polar codes will be a key technique in the link adaptive transmission, especially for future wireless communications. Although there are some advances in researches on HARQ of polar code, more efforts should be paid in this area.

4 Hardware Development Polar Code

There are a few different kinds of implementation architecture for polar codes, depending on the decoding algorithm. Among the various decoding algorithms, successive cancellation (SC) decoding algorithm has the lowest computational complexity; belief propagation (BP) has the elegant nature of parallel decoding; while SC-list decoding algorithm has best error correcting performance. In current stage, a high throughput, low latency architecture for low complexity SC decoding is being developed. Using TSMC 28nm CMOS process, the decoder achieves more than10Gbps throughput, while consumes 102mW power and occupies 0.30mm2 silicon area. The corresponding energy efficiency is 8.96pJ/bit and the area efficiency is 37.52Gbps/mm2. As a comparison, the implementation of turbo code using 40nm CMOS has the energy efficiency of 408 pJ/bit and area efficiency of 0.42 Gbps/mm2. As for the implementation of LDPC code, using 65nm process, the energy efficiency and area efficiency are 62.4 pJ/bit and 3.70Gbps/mm2. Even if the process difference is taken into consideration, the implementation of the polar code still have gains several times over those of turbo codes and LDPC codes. These results show that the implementation of polar codes is one of high speed, low power consumption, high energy efficiency and low complexity. The implementation of polar codes is also carried out on Xilinx FPGA platform. The error correcting performance of the polar code, especially the error floor performance, is tested, using the Xilinx FPGA. The test results show that polar codes do not have error floor, which is consistent with the theoretical analysis.

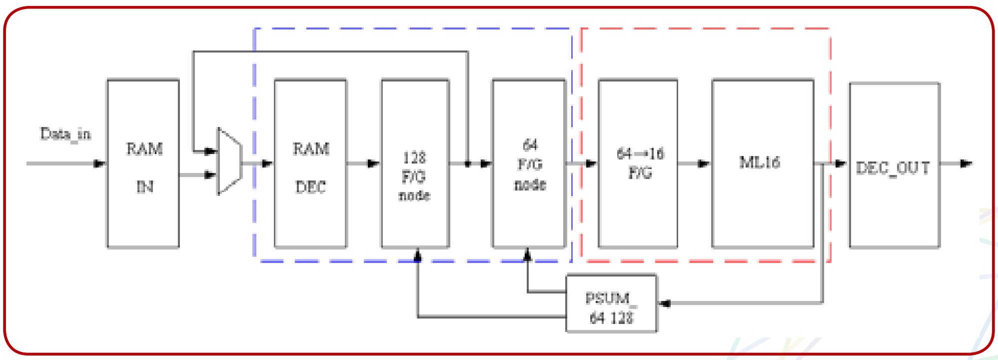


Fig.8. Block diagram of Polar code decoder

At current stage, Polar code SC decoding algorithm is chosen for ASIC and FPGA implementation. TSMC 28nm CMOS process is used in the ASIC implementation, and Xilinx Virtex-7 XC7VX690T-2FFG1761C chip is used in the FPGA implementation. The same block architecture is used for both ASIC and FPGA implementation, as shown in Fig. 1.

The main results of the polar decoder implementation such as clock frequency, throughput, area and power consumption are summarized in this section. The clock frequency of the ASIC implementation is 1GHz; of the FPGA implementation, 200MHz. When the decoder is running at above clock frequency, the throughput for ASIC and FPGA design is 11.38Gbps and 2.28Gbps, respectively. As the resources occupied by the FPGA design is less than 6% of the FPGA chip resources, as much as 15 decoders can be implemented in a single FPGA chip. Thus, a total throughput of 20 or 30 Gbps per FPGA chip can be achieved.

Recently, Polaran[8], a startup company, released the high throughput FPGA design of list decoder (1 cycle decoded 2 information bit when list size is 32).

However, it still remains an open problem to design and implement a practical list decoder with high throughput and low power consumption, for polar codes.

5 Conclusion

In summary, polar codes have two advantages: (1) it has substantial performance gain especially in the short code length (2) it has no error floor. It can meet the requirements of high reliability service and IOT service of 5G. Compared with the LTE turbo codes and LDPC codes, polar codes have some potential advantages. Although there remain some open issues, it is believed that polar coding will be a competitive channel coding solution in 5G wireless system in the future.

References:

[1] E. Arıkan, "Channel polarization: A method for constructing capacity achieving codes for symmetric binary-input memoryless channels," IEEE Trans. Inform. Theory, vol. 55, pp. 3051–3073, July 2009.

[2] I. Tal and A. Vardy, "List Decoding of Polar Codes," available as online as arXiv: 1206.0050v1.

[3] E. Arıkan, "Polar Coding: Status and Prospects", Plenary Speak of IEEE International Symposium on Inform. Theory, Saint Petersburg, Russia, 2011.

[4] S. B. Korada, E. Sasoglu, and R. L. Urbanke, "Polar codes: Characterization of exponent, bounds, and constructions," IEEE Trans. Inform. Theory, vol. 56, pp. 6253–6264, 2010.

[5] E. Arıkan and E. Telatar, "On the rate of channel polarization," in Proc.IEEE Int'l Symp. Inform. Theory, Seoul, South Korea, 2009, pp. 1493–1495.

[6] I. Tal and A. Vardy, "How to construct polar codes," in Proc. IEEE Inform. Theory Workshop, Dublin, Ireland, 2010.

[7] Y. Polyanskiy, H. Vincent Poor, and S. Verdú, "Channel Coding Rate in the Finite Blocklength Regime", IEEE Trans. Inform. Theory, vol. 56, No. 5, May 2010.

[8] http://polaran.com/

Massive MIMO: A Key Enabler for 5G Cellular Systems

Abstract — Massive MIMO is well regarded as one of the key enablers of 5G cellular systems due to its prominent performance advantage compared to regular MIMO systems with a moderate number of antennas, e.g., eight antennas for 4G LTE-A configuration. Turning theoretical performance gain, especially multiplexing gain and array gain, into implementable 5G features, massive MIMO is still confronted with some issues, such as coverage with low-power power amplifiers, high mobility support, and high overhead/complexity for channel state information (CSI) acquirement. The paper introduces how to address these issues from a practical point of view.

Huawei has also developed the M-MIMO prototype, thereby validating the multiplexing performance of this technology. For demonstration, Huawei unveiled a 32T32R prototype to support up to 24 transmission layers simultaneously and to achieve up to 1Gbps with 20MHz bandwidth.

Keywords — Massive MIMO, 5G, coverage, high mobility, CSI acquirement, prototype.

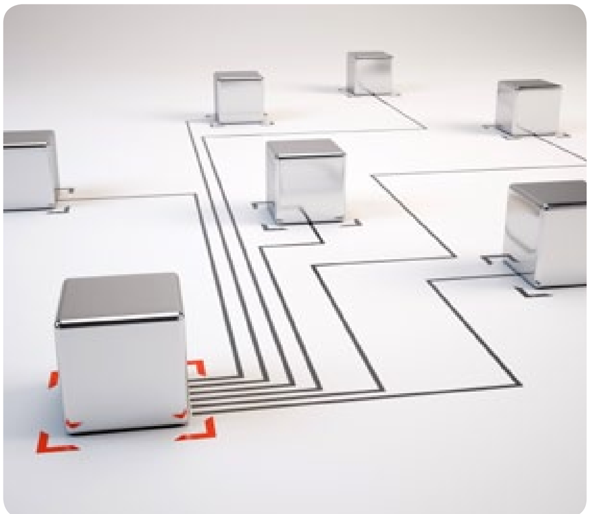
1 Introduction

Since MIMO technology was introduced in 3GPP long term evolution (LTE) Release 8 [1], wireless cellular systems have seen an unprecedentedly fast developing period. With spectral efficiency lying in the primary pursuit of cellular vendors, MIMO technology has played an essential role in current 4G, and will continue to do so in future 5G cellular systems. To respond to 5G high performance requirements, massive MIMO [8] is proposed and advocated across academia and industry. There are mainly two aspects of benefits in adopting massive MIMO technologies. With massive antenna elements on base station, a high degree of freedom (DOF) becomes feasible. In practice, such a high DOF can turn into either high multiplexing gain or high array gain. The former is the key enabler of providing high spectral efficiency, while the latter would improve the quality of received signals.

Multiplexing gain, which is, at the beginning, the main reason to advocate massive MIMO, requires accurate-enough CSI for precoding purpose at base station [6]. However, obtaining CSI is a formidable task if the old scheme in [1], consisting of reference signal (RS) and feedback, is directly employed. First, RS and feedback overhead will be substantial for massive MIMO if the old scheme is used. Second, computational complexity for obtaining CSI due to the increase of channel dimension is also formidable.

Array gain comes from power concentration of beams naturally generated by massive antennas. There are two aspects of problems that need solving. First, massive MIMO beams must have the capability of flexible array gain by dynamic beam width, which is important to cell coverage. Second, array gain by massive MIMO, via delicate design, should be able to cover and track mobility UEs, which is a huge improvement over spatial diversity schemes in [1].

The rest of the paper addresses several aspects of implementing massive MIMO in successive chapters. Section II provides some system-level evaluation results for massive MIMO with perfect channel state information (CSI). Section III provides Huawei's key technological solutions in practical M-MIMO, which include key schemes for high capacity, coverage, and beam tracking. Section IV shows the prototype work for massive MIMO. The conclusion is drawn in the final part.



2 M-MIMO System Performance

In this section, we evaluate the system performance of M-MIMO with perfect CSI, which makes upper bound achievable in system-level simulations. As shown in Fig. 1, M-MIMO can achieve a substantial gain over the baseline, which represents LTE downlink with 4X2 MIMO. Given the 256Tx32Rx configuration, M-MIMO can provide the 17x cell average gain and 37.5x cell edge gain over the baseline. The simulation configuration is provided in the appendix.

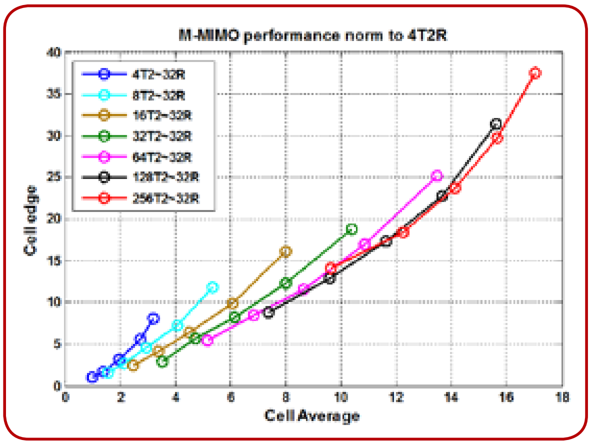


Fig.1. System-level performance compared to 4x2 MIMO

From both literature and our evaluation results, we observe that the more antennas, the more gains using M-MIMO. However, to achieve the full gain by massive MIMO, there are still some challenges to address in practical systems, which include low overhead reference signal (RS) and feedback, common control channel and broadcast channel design, support of UEs with high mobility, low PAPR for high power efficiency, etc. In the following sections, we will provide some key solutions for practical massive MIMO systems.

3 M-MIMO Key Technology

When Massive MIMO, as a key 5G enabler system, is put into practice, practical solutions regarding coverage, high capacity and beam tracking are the main pursuit in wireless industry. Huawei has taken the leading position by providing complete solutions of these aspects, which are introduced in the following chapters.

High capacity

Massive MIMO is capable of offering considerable multiplexing gain. As shown in theoretical analysis results in [10-12], throughput of massive MIMO increases linearly with both Tx/Rx antennas number [7]. Moreover, this conclusion can also be reached in CoMP scenarios.

However, the ability to realize high system throughput in a practical 5G system may be limited by mechanisms which convey CSI to BS. When the concept of massive MIMO was proposed originally, most schemes were assumed to work only in TDD systems, in which CSI at BS can be obtained through channel reciprocity. In recent literature, CSI acquisition for FDD systems has also been proposed as key techniques for practical massive MIMO systems. For both TDD and FDD massive MIMO systems, low overhead RS and feedback schemes are preferred since saving time-frequency resources for RS transmission and feedback means more payload transmission. There are two kinds of typical CSI acquisition methods for massive MIMO, including: 1) beam scanning or beam switching, which use the predefined beam pattern to scan all the UEs within a cell in time, frequency to find some potential transmission beam subsets [14]; 2) Compress sensing based CSI acquisition [15]. Scheme 1) saves overhead by using predefined beam patterns, but sacrifices too much multiplexing gain. The implementation of scheme 2) requires a strong correlation between different UE channels, which, however, is hardly assured in practice.

Here, we propose a novel two-level RS CQI acquisition scheme, capable of achieving high multiplexing gain and low overhead RS and feedback and without stringent requirement on channel scenarios. This is a huge improvement compared with the existing schemes 1) and 2). What appears to be the most appealing aspect of our CSI acquisition scheme is its effective spatial compression technique. With the proposed scheme, most CSI acquisition is based on a reduced beam space, instead of original spatial space. To achieve this, we design 1st-level RS (i.e., statistical RS in Fig. 2) which is responsible for obtaining 1st-level common beam space precoder for all UEs queued for scheduling. The 2nd-level RS is used for obtaining a 2nd-level precoder using instantaneous channel projected on 1st-level beam space precoder. Since 1st-level RS is sparse in both time and frequency domains and 2nd-level RS only oriented for paired UEs, the overall overhead is much lower than LTE-like CSI acquisition schemes, in both TDD and FDD. Meanwhile, the feedback quantity and computational complexity at BS are also significantly reduced thanks to space dimension reduction. Fig. 3 shows that, with 256Tx reduced to 48 spatial dimensions, dynamic beam space provides less than 4 dB performance loss compared to LTE-A like 1-level precoding.

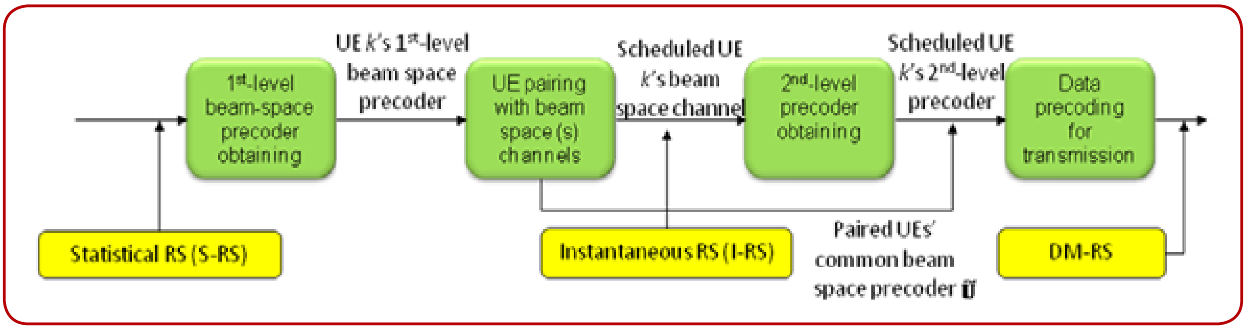


Fig. 2. Two-level CSI acquisition flow chart for dynamic beam space

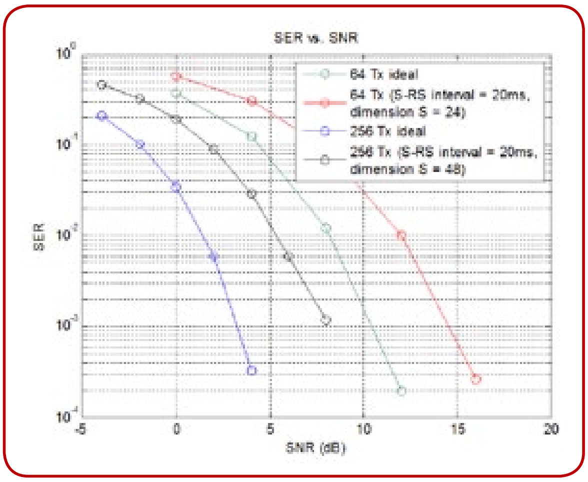


Fig. 3. Symbol error rate vs. SNR for dynamic beam space MU-MIMO (2 paired UEs)

Coverage for massive MIMO

Massive MIMO can provide high spectrum efficiency for data channel, due to high array gain with narrow and high directive beam. But the narrow beam is not suitable for common control channel and broadcast channel, which should support all the users in a cell. Broadcast requires omni-directional beam to cover the whole sectorized area. We obtain the desired beam by 1) enlarging transmission power on all desired radiation directions and 2) smoothing power variation for all desired radiation directions.

· Broadcast beam with uniform linear array (ULA)

Fig. 4 shows the basic omni-direction effect by our optimization. In the left sub-figure, optimization aims to reduce power variation among different directions, while transmission power is enlarged as much as possible for the right sub-figure. Upper bound is generated by using full power on a single antenna. We can observe that 1) the more transmit antennas, the less array gain loss and the less power variation and 2) the optimization can be more focused on either removing power variation as in the left sub-figure, or enlarging transmission power as in the right sub-figure. With 64Tx, there is only less than around 2 dB loss compared to the upper bound power even for the power-variation-removal case, which implies a high power utilization rate.

· Broadcast beam with uniform planar array (UPA)

UPA is more suitable for future 5G massive MIMO deployment due to its more compact size compared to ULA. The additional elevation dimension for UPA needs optimization. With the 3GPP base station height and coverage configurations adopted, Fig. 5 shows the broadcast beam after optimization, which maximizes transmission power. UPA almost achieves the same good power variation effect as ULA, which must be compensated by sacrificing some array gain.

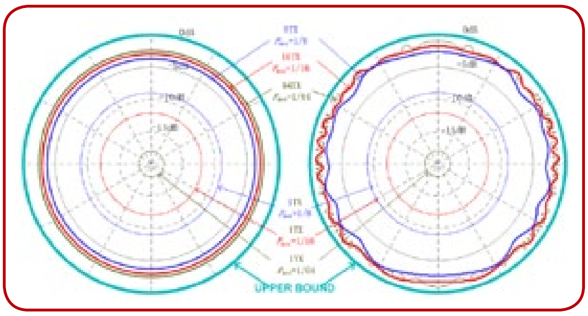


Fig. 4. Massive MIMO ULA broadcast beam

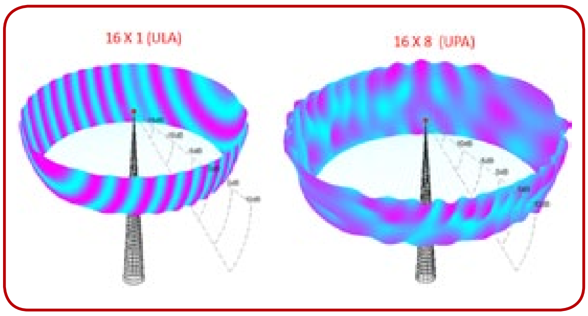


Fig. 5. Massive MIMO ULA multicast beam

Beam Tracking

In the future network, high mobility will be an important application scenario. Massive MIMO, as a potential key technology for 5G, should also provide high performance for the users with high mobility, such as 120~500km/h. Beam tracking for massive MIMO will be a potential technique in high mobility scenario. There are three types of beam tracking methods proposed: 1) beam tracking with adaptive beam width; and 2) beam tracking with adaptive beam direction.

• Beam tracking with adaptive beam width

The key procedures of beam tracking with adaptive beamwidth can be summarized as follows:

• BS estimates the DOAs of UEs;

• UEs iteratively measure signal strength and decide the proper beamwidth, respectively;

• Based on the beamwidth and DOA, BS iteratively generates the beam for intended UEs.

• Beam tracking with adaptive beam direction

The key procedures of beam tracking with adaptive beam direction can be summarized as below.

• BS estimates the DOAs of UEs;

• BS iteratively generates the beam for data transmission with the estimated DOA, plus the pilot for beam with potential direction after adjustment;

• UE iteratively decides the proper beam direction and feedback

We evaluate new beam tracking schemes, in which system ITU-UMI channel model, 64 BS antennas and 4 UE antennas are employed. As can be seen in Table V, in the presence of DOA estimation error, the beam tracking schemes achieve better performance compared with DOA transmission. Also, beam tracking with adaptive beam direction achieves the best performance, with 42.4% increase in CE compared with DOA transmission and 170% increase in CE compared with broad beam transmission.

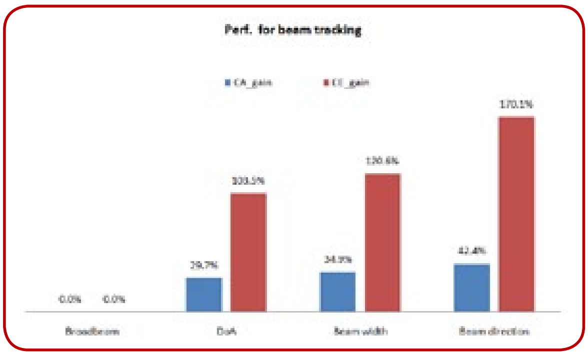


Fig. 6 performance for beam tracking in massive MIMO

Like the conventional MIMO in LTE system, PAPR reduction is still a challenge in M-MIMO, which is mainly caused by both OFDM and multi-user MIMO. To solve the high PAPR issue, Huawei considers low PAPR schemes in M-MIMO for practical 5G systems:

Firstly, for traditional PAPR reduction, some peak cancellation signal will be generated to cancel the signal peak. This kind of PAPR reduction will cause EVM loss due to the lost peak signal component. The existing PAPR reduction method is used in time domain and conforms to 64QAM EVM loss requirement, that is 8% in LTE standard, hence could not distinguish various QAM signals. That is to say, whether QPSK, 16QAM or 64QAM, all QAM signals have the same EVM loss although QPSK and 16QAM could allow bigger peak cancellation signal due to their relaxed EVM loss tolerance. Hence, Huawei proposed PAPR reduction method, based on the EVM tolerance difference for various modulation levels. Different peak cancellation signals will be used for different QAM signal according to their corresponding EVM loss tolerance to reap the further PAPR reduction.

Secondly, most PAPR reduction algorithms use the unused degrees of freedom in the system either directly or indirectly to reduce the PAPR. They may also cause some level of the constellation error at the target user or some level of out-of-band radiation. Tone reservation and "clipping and filtering" [13] are two examples of such algorithms. In tone reservation, some OFDM tones are reserved as unused degrees of freedom to transmit fake signals to get rid of the big spikes in the time domain signal. In clipping and filtering, the signal is clipped to some predefined amplitude and then is filtered to get rid of the out-of-band radiation. This algorithm adds some error to the signal constellation points. However, the degrees of freedom in the case of massive MIMO is huge. That opens the door for effectively reducing the PAPR.

With many antennas at its disposal, a massive MIMO system has plenty of unused spatial dimensions to be used for PAPR reduction.

$$Y=Hx+n$$

$$X_{lowPAPR}=\arg\min_x \|x\|_\infty \ .$$

To eliminate MUI, the transmit-vector x has to satisfy the precoding constraint.

$$s=Hx$$

Since the number of transmitter antenna is far bigger than the receiver antenna number, the equation $\boldsymbol{s=Hx}$ is underdetermined, which implies that there are infinitely many solutions $\boldsymbol{x}$ satisfying the precoding constraint. One can note that ZF is just $\boldsymbol{X_{ZF}=\arg\min_x} \|x\|_2$ with the same criteria as above. Therefore, any low PAPR algorithm (or in fact any algorithm) with zero leakage has higher total transmit power than ZF.

With clipping, one can project the clipping error signals into other spatial dimensions, in order that the new clipping error after precoding (we call it compensation signals), together with the clipped transmission signals, has very low PAPR. On the receiver side, the EVM loss is either very small or even none, since the clipping error is also received in the receiver. The above clipping and project process could be repeated to further reduce the PAPR.

If small deviation from the ideal compensation signal is allowed as long as it is much smaller than the tolerable noise for the underlying MCS, the PAPR could be further reduced.

In this simulation example, we have 128 antennas at the transmitter side and 5 single antenna users with AWGN channel. And the 16QAM signal uses 200 tones out of 256 FFT size.

	Mean power (dBm)	10^-4 peak power(dBm)	10^-4 PAPR (dB)
	16QAM signal with 0 error tolerance		
Original	-42.12	-32.7	9.42
1-level clipping	-41.87	-34.6	7.27
2-level clipping	-41.86	-34.9	6.96
3-level clipping	-41.84	-35	6.84
	16QAM signal with -15dB error tolerance		
1-level clipping	-42.83	-37.5	5.33
2-level clipping	-42.97	-38.1	4.87
3-level clipping	-43.02	-38.3	4.72

4 Prototype for M-MIMO

In this section, we present Huawei's M-MIMO prototype works. We have successfully developed the SU-MIMO and MU-MIMO prototypes of M-MIMO. The hardware cart of the SU-MIMO prototype is shown as in Fig.7, in which the RF units of UE are composed of 4 RF modules and 32 antennas. Based on our new hardware and physical layer architecture, a SU-MIMO prototype is designed to reveal decoding potential of SU on real-time air channels. Demonstrated by the prototype, the spatial multiplexing performance can reach 10Gbps of sum rate due to massive antennas at both BS and UEs. In terms of physical layer algorithm, sparse DMRS pattern, advanced channel estimation and novel MIMO decoding algorithm are proposed to support much higher layer transmission.

Fig. 7. Hardware cart of SU-MIMO prototype

Fig. 8. Antenna Unit for MU-MIMO prototype

On April 23, 2014, Huawei and KT announced that it succeeded in outdoor demonstration of massive MIMO for 5G communication. During the recent demonstration, 128 antennas, 16 data channels, and 3D beam forming were combined with each other to quintuple commercial LTE terminals' capacity.

It can provide performance gains 5 times more than LTE systems. The antenna unit of the MU-MIMO prototype is provided in Fig.8. With our proposed creative reference signal scheme and advanced channel estimation algorithm, the MU-MIMO prototype can support ultrafine UE-specific narrow beam and 2D beamforming in order to drop radio power accurately into the users in the network, which can dramatically reduce interference between multiple users.

5 Conclusion

In the course of developing new conceptual transmission schemes for 5G massive MIMO systems, we have proposed solutions covering spatial multiplexing for high throughput, cell coverage of control and data channel, and tracking technologies for high mobility. The high performance of massive MIMO is further demonstrated in the real-time prototype. It is highly expected that these new schemes will play a significant role in future 5G massive MIMO systems, and become an essential part of Huawei 5G solution.

6 Appendix

The parameters adopted in our system-level evaluation are given in Table 1.

Table 1 system-level simulation configurations

Parameter	Assumption
Cellular Layout	Hexagonal grid, 19 sites, 3 cells/site,
ISD	200m –ITU UMi
Load	Average 10 UE per cell
Max Feedback layer per UE	4 layers per UE
Duplex mode	TDD
Uplink-downlink configurations in TDD	Configuration 1 -- 2:1:2
	11 symbols for DwPTS; 1 symbol for GP, 2 symbol for UpPTS
channel model	ITU UMi
UE speed	3km/h
Antenna configuration	X-pol for BS and UE
Traffic model	Full Buffer
link to system interface	MI-ESM
Scheduler	Greedy PF
Frequency scheduling granularity	Subband Band (5RB)
Precoding method	Beamforming/zf /global zf
HARQ	Chase Combining (CC), Max 4 Transmission
Receiver algorithm	Full IRC

Reference:

[1] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, Massive MIMO for Next Generation Wireless Systems, IEEE Commun. Mag., vol. 52, no. 2, pp. 186-195, Feb. 2014.

[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, Scaling up MIMO: Opportunities and Challenges with Very Large Arrays, IEEE Signal Proces. Mag., vol. 30, no. 1, pp. 40-46, Jan. 2013.

[3] ICC 2013 tutorial on Massive MIMO by Erik G. Larsson and Fredrik Tufvesson, part I and part II.

[4] E. Björnson, E. G. Larsson and T. L. Marzetta, Massive MIMO: 10 Myths and One Grand Question, IEEE Commun. Mag., Mar. 2015, submitted.

[5] 3GPP TS 36.213: "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures". 2010.

[6] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems, IEEE Trans. Commun., vol. 61, no. 4, pp. 1436-1449, April 2013.

[7] J. Hoydis, S. ten Brink, and M. Debbah, Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?, IEEE J. Sel. Areas Commun, vol. 31, no. 2, pp. 160-171, Feb. 2013.

[8] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas", IEEE Trans. Wireless Commun., vol. 9, no. 11, pp. 3590–3600, Nov.

[9] "R1-144437 - WF on Antenna modeling", 3GPP TSG RAN WG1 Meeting #78bis

[10] B. Lee, J. Choi, J. Seol, D. J. Love and B. Shim, Antenna Grouping based Feedback Compression for FDD-based Massive MIMO Systems, IEEE Transactions on Communications, submitted, Aug 2014.

[11] S. Nguyen and A. Ghrayeb,Compressive sensing-based channel estimation for massive multiuser MIMO systems, in Proc. IEEE WCNC, Shanghai, China, April 2013.

[12] J. Choi, D. J. Love, and P. Bidigare, Downlink Training Techniques for FDD Massive MIMO Systems: Open-Loop and Closed-Loop Training with Memory, IEEE Journal of Selected Topics in Signal Processing on Signal Processing for Large-Scale MIMO Communications, 2013, submitted.

[13] Y. Rahmatallah, and S. Mohan, "Peak-To-Average Power Ratio Reduction in OFDM Systems: A Survey And Taxonomy", IEEE Communications Surveys & Tutorials, Vol. 15, No. 4, 4th Quarter 2013, pp.1567-1591.

[14] RP-143883, “High-level views on FD-MIMO and elevation beamforming”, 3GPP RAN#78bis, Samsung, 2014

[15] X. Rao and V. K. N. Lau, "IMO and elevation beamforming tio Reduction in OFDM Systems: A Survey And Taxonomy", IEEE Communications Surveys & Tutorials, Vol. 15, No. 4, 4th Quarter 2013, pp.1567-1591. dat

Full-Duplex Technology for 5G

1 Introduction

Full-duplex (FD) breaks the barrier of today's communications by supporting bi-directional communications without time, frequency or even spatial duplex. By transmitting and receiving at the same time, on the same frequency and in the same spatiality, full-duplex has the potential to double the system capacity and reduce the system delay [1]. Because of its potentials in PHY layers and other benefits in upper layers, full-duplex has been considered as one of the enabling technologies for the next generation 5G wireless communication systems [2]. The major challenge for an FD-capable device (e.g., an FD-capable BS or an FD-capable UE) is how to effectively cancel the self-interference (SI) that consists of the leakage and reflection of its own transmitting signal which can be more than 100 dB stronger than the sensitivity level of a receiver.

In the past few years, the SI cancellation techniques have attracted attentions from both industry and academia and remarkable progress in design and implementations has been made [3][4]. A design of orthogonal pilots for FD systems in reducing FD training overhead and improving SI channel estimation and cancellation performance has been proposed in [5]. A non-intrusive and yet backward compatible air interface design for the LTE standard has been introduced in [6]. Several FD prototyping systems have been reported with promising test results [3][4][7].

With the FD technology approaching its mature stage, the implication and impact of FD-enabled devices on the FD wireless network have been recently studied [8][9]. One issue that is particularly detrimental to system gain is the additional mutual interference (MI) among FD-enabled devices when all or some of them are operating in full-duplex mode. A network-wide interference cancellation scheme that will reduce both the SI and the MI in the receiver of an FD device has been proposed in [10].

In this white paper, an overview of some of the key aspects of the FD technology is provided. The focus is on the innovative solutions for the application of FD technology to the 5G networks.

2 SI Channel and Canceller

The channel impulse response (CIR) of the self-interference can be modeled as clusters of echoes as shown in Figure 1. The first cluster is composed of leakage, reflection and coupling between the transmit antennas and receive antennas. In an FD system with shared transmit and receive antennas, circulators are commonly used to isolate the transmit and the receive paths. The leakage of transmit signal to receiver through the circulator and the reflection of transmit signal back to receiver by the antenna are unique in this type of systems. The coupling between antennas can also be seen in this type of systems and the other type of systems where transmit and receive antennas are separated but collocated. The multipath echoes due to reflection of near-end fixtures make up the second cluster of the self-interference. The rest of the self-interference comes from the reflection of far-end objects which is usually much weaker but with wider delay spread than those in the first two clusters.

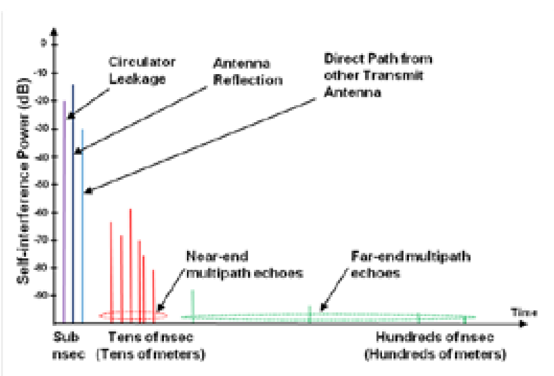


Figure 1: Typical Self-interference Channel Impulse Response

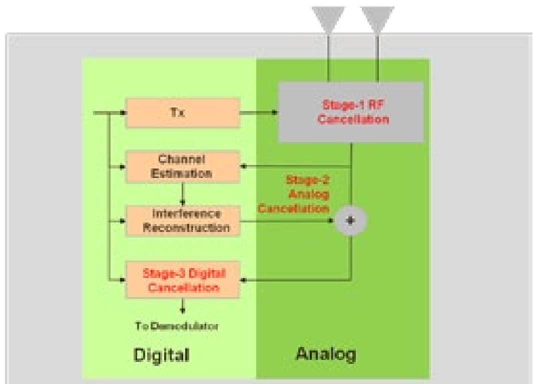


Figure 2: Diagram of a self-interference canceller

An example of a self-interference canceller illustrated in Figure 2 is based on a three-stage solution. The stage-1 is an RF analog cancellation with the self-interference reconstructed in analog domain. The stage-2 canceller reconstructs the SI in digital domain and up-converts it to RF for cancellation in analog domain. The stage-3 is implemented entirely in digital domain when the desired signal plus SI residuals is down-converted and digitized to baseband.

Each stage in the FD system may target different clusters in the aforementioned channel impulse response of SI shown in Figure 1. The stage-1 focuses on the cancellation of the circulator leakage, antenna reflection and antenna cross-talks, which are the components of the first cluster in the sub-nanoseconds region of the delay profile of the CIR. The stage-2 targets the strong near-end multipath echoes of the second cluster in the region extended to tens of nanoseconds in the CIR delay profile. The cancellation residuals of the first two stages plus some far-end multipath echoes are the targets the stage-3 will work on.

The architecture of the example solution may be advantageous in the achievable yet controllable wide dynamic range – capable of canceling strong self-interference and detecting week signal of interest (SOI) at the same time. The goal of stage-1 is to cancel the SI as much as possible such that the transmitter noise floor is close to or below the thermal noise floor in the receiver. The challenge is that the additional noise and nonlinear distortion generated by stage-1 must be minimized and clamped down to the receiver noise floor. The stage-2 is designed to cancel the SI to a level such that it is comparable to that of normally received SOI in half-duplex mode. Therefore the dynamic range of stage-3 can be optimized for both SI and SOI. The goal of stage-3 is to reduce the residual of SI left by the previous two stages close to or under the receiver noise floor.

3 SI Channel Estimation

In a typical FD system, the self-interference cancellation in a receiver, as shown in Figure 2, is composed of two main steps: the first step is the SI channel estimation in a training period where some pilot signals are transmitted to facilitate the MIMO SI channel estimation; the second step is the reconstruction of the self-interference based on the known transmitting data symbols and the SI channel estimation acquired in the first step. The training period in the first step is typically a half duplex time slot where only pilot signals are transmitted to facilitate the estimation of CIRs of a MIMO system without the interference of the SOIs. The design of the pilot signals transmitted in the training period is of great importance because it dictates the overhead and complexity and affects the performance of an FD system.

The self-interference in the *p-th* MIMO receiver can be formulated by

$$\mathbf{y}_p = \mathbf{X}\mathbf{h}_p + \mathbf{v}_p, \tag{1}$$

where $\mathbf{y}_p$ and $\mathbf{v}_p$ are the SI samples and additive white Gaussian noise (AWGN) at receiver (both $N \times 1$ in size); $\mathbf{h}_p$ is the CIR of SI with size $D\times 1$, and X is the transmitting symbol (of pilots or user data) matrix with size $N\times D$. The optimum estimation of $\mathbf{h}_p$ is the least-square (LS) estimation which is given by

$$\hat{\mathbf{h}}_p = (\mathbf{X}^H\mathbf{X})^{-1}\mathbf{X}^H\mathbf{y}_p. \tag{2}$$

The design of pilots in [5] makes pilot X be composed of orthogonal sequences across transmit antennas and CIR delay spread such that the LS estimator of (2) is reduced to a set of cross-correlations,

$$\hat{\mathbf{h}}_p = \frac{1}{N\sigma_x^2}\mathbf{X}^H\mathbf{y}_p. \tag{3}$$

And the cancellation mean square error can be minimized to

$$P_\varepsilon = \frac{D}{N}\sigma_v^2 \tag{4}$$

where σ_v^2 is the variance of the AWGN $\mathbf{v}_p$.

The benefits of the orthogonal pilots based channel estimation can be summarized as follows:

• Significant improvement of the numerical stability of the LS channel estimator due to the elimination of the matrix inversion. The improvement is translated to improved accuracy of the CIR estimation, which in turn contributes to a better cancellation performance with less cancellation errors. The numerical issue becomes more significant as the number of antennas grows in a MIMO system.

• Pilots for all the antennas can be transmitted simultaneously thanks to the orthogonality of the sequences across the antennas. The FD overhead is reduced due to shorter duration of the training period. In a typical LTE implementation, it only requires one OFDM symbol to complete the training for all antennas.

4 Air Interface for FD Operation

Although full-duplex has been considered as one of the enabling radio access technologies for the next generation (5G) wireless communication systems, its applications to the current generation (4G) systems have not been widely discussed and well addressed due to concerns in complexity and compatibility with potential changes required in the current system. In [6], the concept of a backward compatible full-duplex base station (FD-BS) is introduced to the current LTE TDD systems [10] where the FD-BS is able to work with the legacy half-duplex user equipment (HD-UE's), a new classes of FD-aware HD-UE's and the future full-duplex UE's (FD-UE's). The focus is particularly on the design of air interface in the full-duplex base station in the current LTE TDD systems. Specifically, a new radio frame structure is designed for the operation of full-duplex. The new frame structure, which requires minimal changes in the current LTE standard, is non-intrusive and backward compatible to legacy base stations (BS's) and user equipment (UE's) on the market.

Since SI channel estimation is crucial to FD operation, one of the goals of the proposal is to design a dedicated HD training period in the current LTE TDD frame structure with requirements of minimum intrusion and overhead and maximum transparency and compatibility to the system. There are a number of constraints we need to consider for the design of the training period:

• The training period shall be introduced in a nonintrusive way in the FD-BS such that the compatibility with the current LTE TDD system is maintained.

• The legacy UE's shall have minimal or no changes in order to work with either the current BS or the new FD-BS.

• The training period shall be guaranteed in HD mode and it shall be long and frequent enough for the SNR and adaptability requirements of channel estimation and, in the meantime, it shall be as short and as few as possible to reduce the overhead of FD operation.

It is proposed in [6] that training period (TP) for FD channel estimation is shared with the guard period (GP) in the special subframe of the LTE TDD frame structure. Extension of the TP to the DwPTS slot can also be optionally made. The orthogonal pilot signals discussed in the previous section will be used over the TP to minimize the overhead and improve the performance. The duration of the TP can be tailored to the configuration of the special subframe. The maximum length of the pilot signals can be up to the entire duration of the TP plus extension in DwPTS. The actual length can be varying depending on the performance requirement of (4), the timing error budget of the uplink and the special subframe configuration. The design is illustrated in Figure 3.

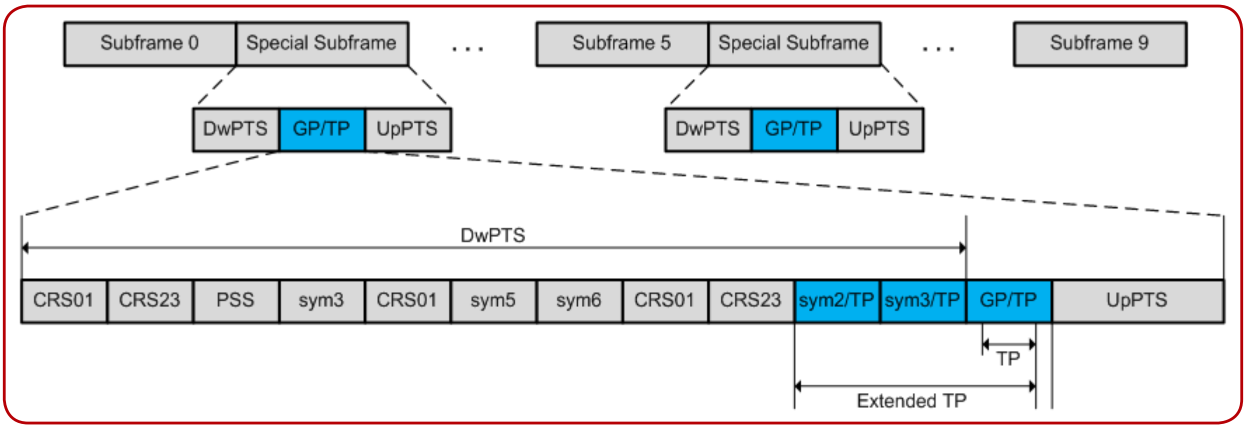


Figure 3: FD TP and extended with frame structure type 2 (5 ms switch-point periodicity)

Another goal of the proposal is to allow the base station to operate in full duplex mode while the UEs can still be in one of the three modes: the legacy half duplex mode, the new FD-aware half duplex mode and the new full duplex mode. The legacy UEs do not need to be aware of the FD operation in a base station. Thus an FD-BS can work transparently and compatibly with the current UE's in the current LTE TDD systems. To be able to schedule a DL UE and a UL UE on the same resource blocks (RBs), an overlay of FD configurations on the existing uplink-downlink configurations [10] is proposed. A new flexible allocation type 'F', in addition to the existing static types of 'U' and 'D', is introduced with the overlay FD configurations to allow an FD-BS to schedule UL UE's (one or more) and DL UE's (one or more) on the same subframes.

The overlay FD configuration shall be broadcast by the FD-BS, along with the original configuration, to the cell such that all the UEs served by FD-BS are able to read the configurations and detect the operation mode if necessary. For the existing (legacy) UEs, they continue to read and interpret the original configurations as usual, without having to be aware of the FD operations. New classes of FD-aware HD-UE's and FD-UE's shall be able to read and interpret both the original and the overlay configurations depending on the operation mode of its serving BS (legacy BS or FD-BS). The exact operation in the 'F' allocation subframe is determined dynamically by the latest assignment from a control channel. The concept is demonstrated in Table 1. It is also possible to define some customized overlay configurations with predefined 'U', 'D' or 'F' allocations for certain subframes to meet system or operator's requirements. An example is shown in Table 2 where the subframes adjacent to the special subframe are kept in their original allocations.

Table 1: Proposed FD configurations with 'F' allocation type overlaid on existing uplink-downlink configurations

UL-DL Config. #	Subframe Number													
	0	1			2	3	4	5	6			7	8	9
0	D	D	GP	U	U	U	U	D	D	GP	U	U	U	U
	F	F/TP	TP	F	F	F	F	F	F/TP	TP	F	F	F	F
1	D	D	GP	U	U	U	D	D	D	GP	U	U	U	D
	F	F/TP	TP	F	F	F	F	F	F/TP	TP	F	F	F	F
2	D	D	GP	U	U	D	D	D	D	GP	U	U	D	D
	F	F/TP	TP	F	F	F	F	F	F/TP	TP	F	F	F	F
3	D	D	GP	U	U	U	U	D	D			D	D	D
	F	F/TP	TP	F	F	F	F	F	F			F	F	F
4	D	D	GP	U	U	U	D	D	D			D	D	D
	F	F/TP	TP	F	F	F	F	F	F			F	F	F
5	D	D	GP	U	U	D	D	D	D			D	D	D
	F	F/TP	TP	F	F	F	F	F	F			F	F	F
6	D	D	GP	U	U	U	U	D	D	GP	U	U	U	D
	F	F/TP	TP	F	F	F	F	F	F/TP	TP	F	F	F	F

Table 2: An example of the proposed overlay FD configurations with predefined UL/DL allocations for certain subframes

UL-DL Config. #	Subframe Number													
	0	1			2	3	4	5	6			7	8	9
0	D	D	GP	U	U	U	U	D	D	GP	U	U	U	U
	D	D/TP	TP	U	U	F	F	D	D/TP	TP	U	U	F	F
1	D	D	GP	U	U	U	D	D	D	GP	U	U	U	D
	D	D/TP	TP	U	U	F	F	D	D/TP	TP	U	U	F	F
2	D	D	GP	U	U	D	D	D	D	GP	U	U	D	D
	D	D/TP	TP	U	U	F	F	D	D/TP	TP	U	U	F	F
3	D	D	GP	U	U	U	U	D	D			D	D	D
	D	D/TP	TP	U	U	F	F	F	F			F	F	F
4	D	D	GP	U	U	U	D	D	D			D	D	D
	D	D/TP	TP	U	U	F	F	F	F			F	F	F
5	D	D	GP	U	U	D	D	D	D			D	D	D
	D	D/TP	TP	U	U	F	F	F	F			F	F	F
6	D	D	GP	U	U	U	U	D	D	GP	U	U	U	D
	D	D/TP	TP	U	U	F	F	D	D/TP	TP	U	U	F	F

5 FD Network Interference

In a traditional LTE-TDD network with a frequency reuse of one, the network interference coming from the BS-UE co-channel interference can be observed on the UL slot in a BS (UE-BS IF) and on the DL slot of a UE (BS-UE IF). The inter-cell interferences from BS to BS (BS-BS IF) and from UE to UE (UE-UE IF2), and the intra-cell interference from UE to UE (UE-UE IF1) are largely avoided due to the requirement of synchronization on a common UL-DL configuration across the entire network. In an FD-enabled network (FD-Net) where the BS's in the network may operate in FD mode, the inter-cell and intra-cell interference pertaining to BS-BS IF and UE-UE IF's can be significant because the UL of one device (a BS or UE) can be on the same time slot and resource blocks (RB's) with the DL of another device. This situation is illustrated in Figure 4 where the two BS's operate in FD mode while the UE's operate in legacy HD mode. A scheme for FD network interference cancellation is proposed in [10]. In particular, the focus is on the cancellation of SI and BS-BS IF even though the solution can be extended to include the UE-BS IF and can be applicable to the UE's.

The general approach to the cancellation of BS-BS IF in a full-duplex enabled network is to leverage as much as possible the existing infrastructure for SI cancellation while taking into account the unique characteristics of BS-BS IF's from neighboring BS's. Additional considerations may be needed to accommodate the following factors:

• The propagation channel that a BS-BS IF undergoes is different from that of SI and may be different from other BS-BS IF. Specifically the propagation delay and the spread of the channel impulse response (CIR) for BS-BS IF will differ from one to another.

• The size of the extended MIMO setup including both SI and BS-BS IF could be significantly larger if multiple BS's in the neighboring cells are considered. Selection of a number of strongest interfering BS's has to be made adaptively and to be capable of tailoring to performance and complexity requirements.

• A backhaul (in the case of distributed BS's) or fronthaul (in the case of centralized BS's like C-RAN) data link is required for tunneling transmit data across BS's. We assume that the capacity of the data link is sufficient and the link is lossless for the data transfer.

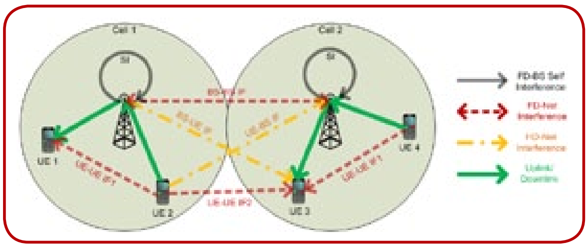


Figure 4: Interference map in an FD-enabled network

The cancellation is based on the techniques discussed in Section 3 and Section 4 and the extension of those from individual FD devices to an FD network. Specifically, the scheme is based on an extended MIMO treatment of the whole FD network. The network-wide interference cancellation only becomes feasible when the unique pilots design and training structure are in place network wide. The orthogonality of the pilot signals makes the complexity and stability of the canceller manageable. It also optimizes the selection of the number of strongest interfering

BS's from the neighboring cells. That is, the selection can be made microscopically to the antenna level for each of BS's involved.

It may be advantageous for a C-RAN based FD-Net to realize the solution in a centralized unit (CU) because the transmission information can be readily available at the CU and the interference cancellation can be done digitally in the centralized baseband unit (C-BBU). The solution can also be applied to the remote radio units (RRUs) simultaneously for hybrid analog and digital interference cancellation in RRUs where the required transmission plus channel information, or the reconstructed copies of the interference, can be passed from CU to RRUs with proper timing advance.

6 Conclusion

Some of the important aspects of the full-duplex technology, including channel estimation, training period and air interfaces, are discussed. The impact and implication of FD operation to the network are also briefed. Some new and innovative solutions are provided for the application of FD to the 5G networks.

The self-interference channel is a unique signature of the FD technology and a key to the performance of SI cancellation. More research and measurements are needed for more accurate characterization of the SI channel. Cancellation of nonlinear components of SI, which are introduced in various stages of the transceiver and the SI canceller, is another challenge for the application of FD in 5G networks.

References:

[1] S. Herath and T. Le-Ngoc, "Sum-Rate Performance and Impact of Self-Interference Cancellation on Full-Duplex Wireless Systems," IEEE 24th International Symposium on Personal, Indoor and Mobile Radio Communications: Fundamentals and PHY Track, 2013, pp. 881-885.

[2] S. Hong, et al, "Applications of Self-Interference Cancellation in 5G and Beyond," IEEE Communications Magazine, February 2014, pp. 114-121.

[3] M. Duarte, et al, "Design and Characterization of a Full-Duplex Multiantenna System for WiFi Networks," IEEE Trans. VT, Vol. 63, March 2014, pp. 1160-1177.

[4] D. Bharadia, and S. Katti, "Full Duplex MIMO Radios," Proceedings of NSDI'14, April 2014, pp. 359-372.

[5] H. Wu and E. Hum, "System and Method for Training Signals for Full-Duplex Communications Systems," US Patent Application 14/617598, April 2014.

[6] E. Hum, et al, "System and Method for Full-Duplex Operation in a Wireless Communications System," US Patent Application 14/617679, May 2014.

[7] E. Hum, et al, "Full-Duplex Canceller Integration Test Report," Internal Technical Report, Huawei Technologies Canada Co. Ltd., Feb. 2015.

[8] S. Goyal, et al, "Analyzing a Full-Duplex Cellular System, " IEEE 47th Annual Conference on Information Sciences and Systems (CISS), March 2013, pp. 1-6

[9] O. Simeone, et al, "Full-Duplex Cloud Radio Access Networks: An Information-Theoretic Viewpoint, " IEEE Wireless Communications Letters, Vol. 3, No. 4, August 2014, pp. 413-416

[10] H. Wu, et al, "Interference Cancellation for Full-Duplex Communication Networks," US Patent Application 62/151632, February 2015.

[11] 3GPP TS 36.211, "LTE Physical Channels and Modulation," Version 11.5.0 Release 11, January 2014.

Radio Access

Abstract — Future 5G access networks will provide a ubiquitous user experience by eliminating the cell edge boundaries especially in dense networks thanks to a shift in paradigm from a cell-centric radio access network to a user-centric and virtualized one. In such a network, a UE is surrounded by a virtual cell that moves with the user and dynamically adapts to user QoS requirements. To enable a virtualized radio access network, a new design for physical layer channels is required. In this article, several aspects of the network are studied and solutions to enable a virtualized network are proposed. These solutions include the concept of hyper cell and user-centric transmit point optimization, user cooperation in transmission and reception through regular or strategically dedicated devices, device-centric physical channels that are virtualized and disassociated from each other, user-centric connection signature and measurement channels. In all the proposed schemes, the new design is not associated to a specific node in the network.

Keywords — virtualized radio access network, user-centric physical channels, network cooperation, device cooperation.

1 Introduction

> Radio access virtualization creates a shift in the radio access network design paradigm to enable a ubiquitous user experience and to utilize very dense networks better. At the same time, it can dramatically improve the network energy efficiency.

In traditional wireless access networks, devices associate with a cell as a consequence of being in the cell's unique geographic coverage area. As a device moves away from the cell center, the link performance may degrade until, eventually, the device is handed over to a neighboring cell. This link quality variation is analogous to the changes in illumination observed as a car driving at night moves between streetlights. In a virtualized device-centric network, the coverage area is formed around the wireless device as the network determines which access point(s) are to be associated with the device. The cell moves with and always surrounds the device (like a stage spotlight follows a performer), in order to provide a cell-center experience throughout the entire network coverage area. The elimination of the device's view of the cell boundary is illustrated in Fig. 1.

Radio access virtualization enables improved network energy efficiency through its ability to activate and deactivate uplink and downlink transmission points in response to local traffic conditions. The separation of uplink and downlink flows enables only the necessary transmission resources to be activated to suit the demand. Separating the control and data planes also enables savings as the data plane transmission points only need to be activated when there is traffic for devices in the area.

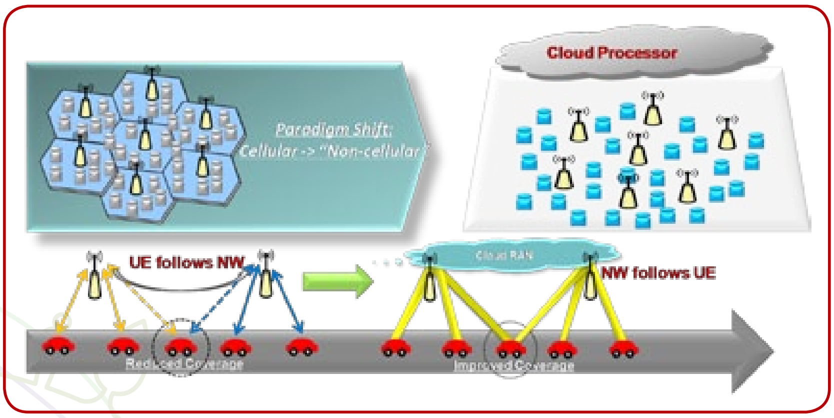


Fig. 1. Virtual radio access: elimination of cell boundaries

This paper is organized as follows. In section II, we study hyper cell and propose a user-centric transmit point (TP) optimization technique. Section III studies device cooperation and section IV discusses the user-centric physical channels. Sections V and VI are devoted to the user-centric connection signature and measurement channel. Section VII summarizes the paper.

2 Hyper Cell and UE-Centric TP Optimization

The dense network structure and the demanding performance requirements of 5G networks necessitates every user equipment (UE) to be dynamically served by an optimally selected set of cooperative transmit points (TPs). As the UE moves, the network side determines a new optimal serving TP set that follows the UE, essentially creating a "spotlight" around the UE at each point of time and every location. This dynamically optimized serving TP set may be transparent to the UE. In other words, each UE may receive service from some hyper-cells (HCs) in the network that may not be known to that UE. This fundamental structural difference between the cellular-based 3G/4G wireless networks and the future 5G networks calls for a change in 5G design principles to form collaborative TP sets. In particular, forming a "spotlight" around UEs necessitates abandoning the conventional TP cooperation approach in 3G/4G wireless networks wherein collaborating TP groups are formed based on the rigid cellular network structure [1], [2]. Instead, it requires using rigorous quantitative UE-centric metrics to partition the network into HCs at each time instance.

In [3], we have used UE-centric metrics to partition the network into multiple overlapping HC sets. The HC sets are formed so that no UE experiences an edge status in all HC sets. This guarantees that there is at least one HC set that can properly serve each UE. Fig. 2 depicts this idea. The upper part of Fig. 2 shows a partition of the network into a set of two HCs: I.HC1 and I.HC2. TP1 and TP2 in I.HC1 cooperatively communicate with the UEs that are associated with this HC, that is, UE1, UE2, and UE3, while TPs in I.HC2 cooperatively communicate with UE4 to UE7. In this subfigure, UE3 and UE4 are close to their associated HC border and may experience a substantial amount of interference from some TPs in their neighboring interfering HC. To avoid this problem, we provide other partitions to the network that are more beneficial to UE3 and UE4 and schedule these UEs when these more beneficial partitions are used as shown in the lower part of Fig. 2 where II.HC1 and II.HC2 are formed. It can be observed that UE3 is now cooperatively served by TP2 and TP4, while UE4 is cooperatively served by TP1 and TP3. Therefore, major interfering TPs to UE3 and UE4 in the first partition act as their serving TPs in the second partition. In the depicted simple network of Fig. 2, partitioning the network into two overlapping HC sets guarantees that no UE experiences edge status in all HC sets. In more complicated scenarios, creating more overlapping HC sets may be necessary to guarantee the requirement above. However, if the HC sets are created based on proper UE-centric metrics, only a few partitions would be sufficient to guarantee a required quality of service (QoS) to all UEs in at least one partition.

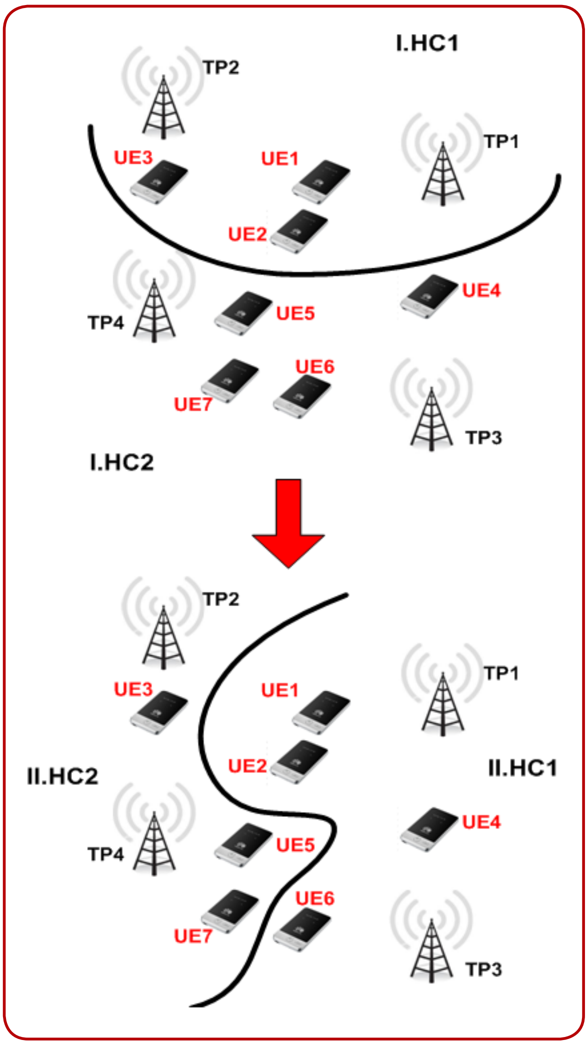


Fig. 2. Network partition into two HC sets I and II

The description above illustrates the idea behind the proposed HC sets formation. However, in practice, rigorous UE-centric metrics are used to form HCs [3]. To form HCs, we first calculate the impact weights among every TP pair in the network. The impact weight is a measure that determines the amount of interference that each TP inflicts on the UEs that are associated with the other TP in the pair. Once these impact weights are determined, calculating a suitable set of HCs boils down to partitioning a graph into a set of smaller subsets so that sum weights of the edges between the nodes (TPs) in two different subsets is minimized. This guarantees that TPs in an HC inflict minimal interference on the UEs that are served by other HCs. Once the first HC set is determined, the second HC set can be calculated by manipulating the weights of the UEs: The UEs that are guaranteed a good performance in the already-made HC set get lower weights than the UEs that are at the edge of the already-made HC set. This approach guarantees that the second HC set is mainly tailored to serve the UEs that have the edge status in the first HC. The procedure of weight updating and network partitioning can continue until a required performance metric is satisfied for all UEs in at least one HC.

Once all necessary HC sets are determined, a proper resource scheduling technique should be devised so that UEs would be scheduled with their appropriate partition. An example of a technique to select the best HC is to use a UE-centric metric and hypothetically schedule UEs in all HCs at each scheduling resource unit. The partition that generates the highest utility among all hypothetically-scheduled UEs will be used to actually schedule the UEs in the corresponding resource unit (RU). With this approach, each UE tends to be scheduled in a HC that serves that UE best in the scheduled RU and the "spotlight" follows the UE at every scheduling occasion.

In some practical scenarios, there may be a large discrepancy among traffic loads at different parts of the network. For such cases, a solution is developed in [3] that not only eliminates edge UE status, but also offers balanced load traffic among various TPs. The load-balancing solution is based on a new concept of soft UE-TP association where each UE is partially associated with multiple TPs and the associated weight is inversely proportional to the load of that TP and aims to measure the TP’s relative suitability to serve that UE. As such, if a TP is highly loaded, the algorithm reduces UEs’ soft association weights towards that particular TP and results in higher association weights of UEs towards less heavily loaded TPs, and a load balancing among TPs in the network is thus achieved. These soft UE-TP association weights can be taken into account in calculating inter-TP impact weights required to generate UE-centric load-balancing HC sets [3].

3 Device-Centric UE Cooperation

The density of neighboring devices and their capability of device to device (D2D) connectivity provide the opportunity for device cooperation in transmission/reception. As illustrated in Fig. 3, groups of devices (active or idle) in close proximity to each other may form logical/virtual nodes, acting as consolidated distributed transceivers assisting the network nodes. A virtual device consists of a set of cooperating devices and target devices. Cooperating devices help target devices communicate with the network. For each target device (TUE), the group of cooperating devices (CUE) could be chosen from the set of active or idle devices that are willing to cooperate in exchange for some incentive, or network aware dummy devices strategically deployed by users or network operators for the sole purpose of UE cooperation.

The access point controller can schedule the terminals benefiting from device cooperation and manage factors such as cooperation collision, privacy restrictions, and cooperation incentive. A network-assisted device cooperation results in better virtualization by providing more possible transmission paths from network to the target devices. In a very dense network, such path diversity is even more beneficial when the cooperating devices are exposed to different access point sets than the target devices. As such, more traffic-offloading and access point-muting opportunities arise for conserving more energy without compromising the QoS/QoE satisfaction.

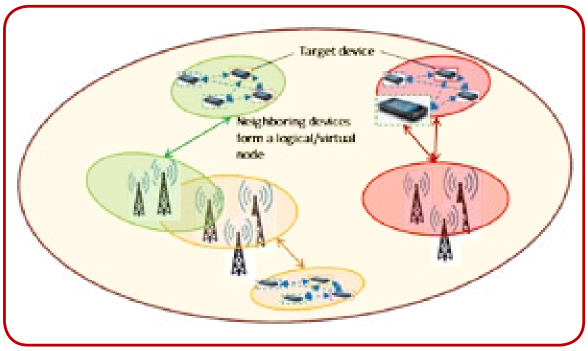


Fig. 3. Group-to-group communications between virtual hyper-transceiver nodes

Device cooperation provides diversity in space, time and frequency, thereby providing increased robustness against fading and interference. 5G networks will be able to track user locations and have accurate knowledge of their channel conditions and capabilities, hence providing them with adequate assistance in discovering and connecting with each other. Such knowledge can be leveraged during scheduling in order to maximize the benefits of UE cooperation. Network-assisted device cooperation can therefore be channeled towards achieving different design targets such as enhancing network throughput and coverage, overcoming cell edge limitation, and/or improving the overall energy efficiency of 5G networks owing to (i) inherent multi-user diversity induced by dissimilar channel and interference conditions experienced by cooperating terminals, (ii) enhanced antenna gains through distributed MIMO and advanced beamforming schemes, and (iii) enhanced interference cancellation capabilities using linear and non-linear transmission and reception techniques.

Focusing on the forward link, virtual multi-node receiver cooperation with half-duplex devices involves two phases:

• **Network broadcast phase:** Network multicasts a data packet to a group of UEs sharing a group identifier. Depending on the cooperation scenario, both TUEs and CUEs may listen to the said data packet during this phase.

• **UE Cooperation phase:** CUEs forward some information to the TUEs to help them decode the information broadcast by the network during the initial phase. It is assumed that CUEs operate over a dedicated frequency band, separate from that of the access link bandwidth, to relay information during the cooperation phase. The cooperation bandwidth can be spatially reused over different cooperation clusters.

Information sent by the CUEs during the cooperation phase depends on the cooperation strategy, e.g. amplify-and-forward (AF), decode-and-forward (DF), compress-and-forward (CF), joint reception (JR). Other cooperation strategies include (frequency-selective) soft-forwarding (SF) device cooperation as described in [4].

As illustrated in Fig. 4, the potential upper bound throughput and coverage gains that can be obtained from the combination of CRAN-based joint transmission (JT) and UE cooperation based JR techniques are quite impressive. However, the hyper transceiver complexity scales proportionately with the number of transmit and receive points and can quickly become prohibitive. Therefore, lower-complexity hyper-transceiver designs need to be considered.

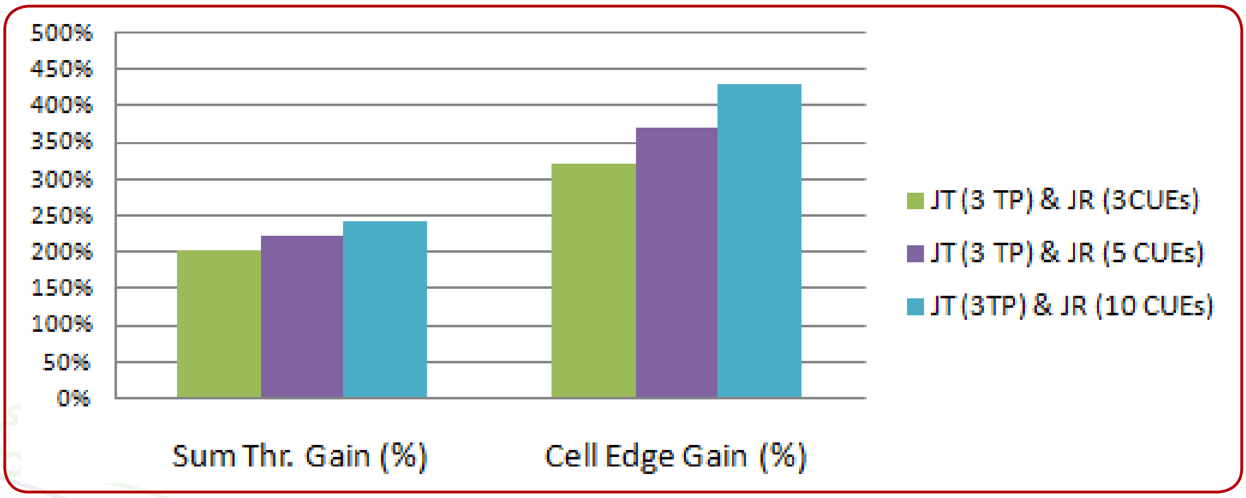


Fig. 4. Throughput and coverage gains with respect to a baseline scenario not employing joint transmission or joint reception techniques

The rest of this section introduces the network-assisted device-centric low-complexity hyper-transceiver design, device cooperation using network aware dummy devices for hotspot scenarios, and flexible UE cooperation using higher-layer fountain codes.

Cooperation Group Formation and Low-Complexity Hyper-Transceiver

In this section, we propose a low-complexity hyper-transceiver design based on a joint TP-UE selection mechanism which leverages the benefits of both CRAN-type dynamic point selection (DPS) and UE cooperation. As illustrated in Fig. 5, for each TUE in the network (NW), we determine a UE-centric cooperation candidate set (CCS) of potential CUEs. A cooperation active set (CAS), i.e. a small subset of CUEs belonging to the CCS, is then down-selected from CCS to simplify the design of the virtual UE while extracting most of the UE cooperation gains compared to the full overhead case where the CAS consists of the whole CCS.

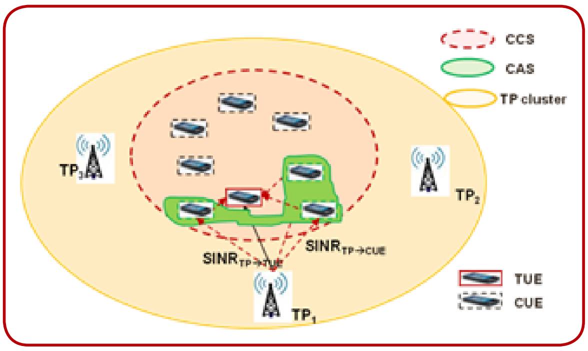


Fig. 5. Device-centric joint UE-TP selection in the presence of DPS [2]

The proposed low-complexity UE-centric virtual UE formation method leverages most of the UE cooperation gains and consists of a two-step process where NW and UE jointly optimize CCS and CAS. CCS is based on long-term information/ measurements/prediction and can therefore be updated semi-statically, whereas CAS can be updated more frequently. Another advantage of the proposed virtual UE formation method is that it allows the joint UE-TP selection to reflect both access link and D2D link qualities, which makes it very suitable to be used in conjunction with low-complexity device-centric transmission schemes such as DPS, as illustrated in Fig. 5. More details on low-complexity hyper-transceiver design and the simulation setup considered herein can be found in [5].

Fig. 6 shows the sum throughput and coverage gains when combining DPS (with different transmit point cluster sizes) with DF-based UE cooperation, respectively. The baseline reference for this figure is the performance of single-cell SU-MIMO 2 x 2 SFBC downlink without UE cooperation. As can be observed, the gains of DPS and UE cooperation are quite substantial. Besides, the benefits of the proposed low-complexity hyper-transceiver are clearly visible, because with only 2 CUEs, it can realize as much gains as with 10 CUEs, thereby achieving the target of a practically feasible low-complexity hyper transceiver suitable for virtualized radio access in 5G dense networks.

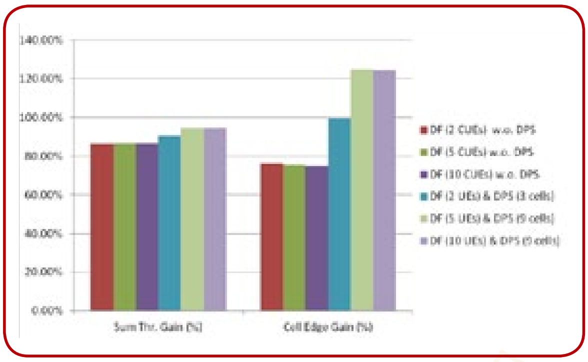


Fig. 6. Sum throughput and coverage gain comparison with respect to non-UE cooperation for DF with different DPS cluster sizes

Dummy UE (DUE) for Hotspot Scenarios

A DUE is a strategically deployed device in the network dedicated for cooperation with neighboring UEs. Such devices may have a powerful RF front-end and computational capability to support many different inter-user connections, radio access technologies (with or without human interface) and high level applications specialized for device cooperation. DUEs can be privy to the network and its information about the users, which facilitates user cooperation without breaching their privacy.

A dummy user can access the network through the same radio access as other ordinary users and utilize similar inter-user connection technologies for user cooperation. DUEs can be cost effective and either deployed by the network to relieve the stress caused by a hotspot scenario, or deployed by the end user to enhance the performance and coverage in its desired area. Some of the use case scenarios where a dummy user can be beneficial are listed as follows.

- Open air festival: METIS Test Case 9 [6] is a study case for an outdoor event describing a scenario where many users attend an outdoor event in a crowded venue that is otherwise not crowded and hence, the infrastructure for such venue is not provisioned to support such a event.
- Blind spot coverage.
- Personal area performance improvement: Dummy users can be implemented by end users to improve the service at their workplace or residential area.

In the following part, we study the potential performance of dummy users implemented in the open air festival scenario. To exhibit the scenario, we assume that the network is of a heterogeneous deployment that is not designed for an outdoor event. It consists of 63 wrapped-around hexagonal cells covered by 21 three-sector sites with inter-site distance of 500 meters and 200 picos uniformly distributed over the network. Each macro cell forms a hyper-cell with all the picos geographically located within the coverage area of that cell, and DPS transmission scheme is performed between the transmission nodes in the hyper-cell. The outdoor event is simulated by 1000 active users in a radius of 500 meters and 300 active users are located in the rest of the network.

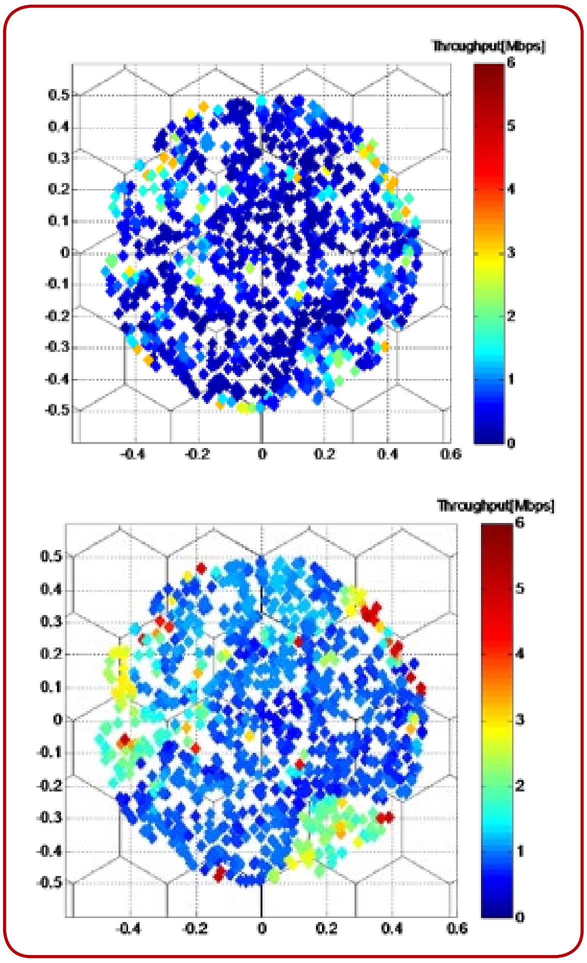


Fig. 7. User's rates in the hotspot with 1-cell DPS and no dummy users (left) compared to the same scenario with 400 dummy users (right)

Fig. 7 shows the user rate for the users within the hotspot created by the outdoor event. Since the infrastructure is not provisioned to support the hotspot created by the event, the service is very poor especially in the inner side of the hotspot. It is worth noting that some users at the edge of the circle experience a favorable performance due to being served by the cells outside the hotspot which are less crowded. Adding the dummy users can benefit the users in the heart of the hotspot.

Fig. 8 shows the performance of the users in the hotspot with different numbers of picos and dummy users. The figure shows the average throughput and coverage gains of the users in the hotspot compared to the same hotspot scenario without DUEs. The results show that introducing dummy users not only increases the overall throughput of the system, but also helps with the ubiquity of the service in the hotspot area and enhances fairness among the users. With the highest network cooperation group (9-cell cooperation) and 400 dummy users, the five percentile rate user can benefit from 50% of the rate for an average user in the hotspot.

The increase in the system throughput is due to a variety of factors: better access channel of the dummy users to the network, higher utilization of rank-2 transmission and better interference mitigation from the dummy users. The performance can be further improved through closed loop transmission and linear/non-linear multi-user MIMO to the dummy users.

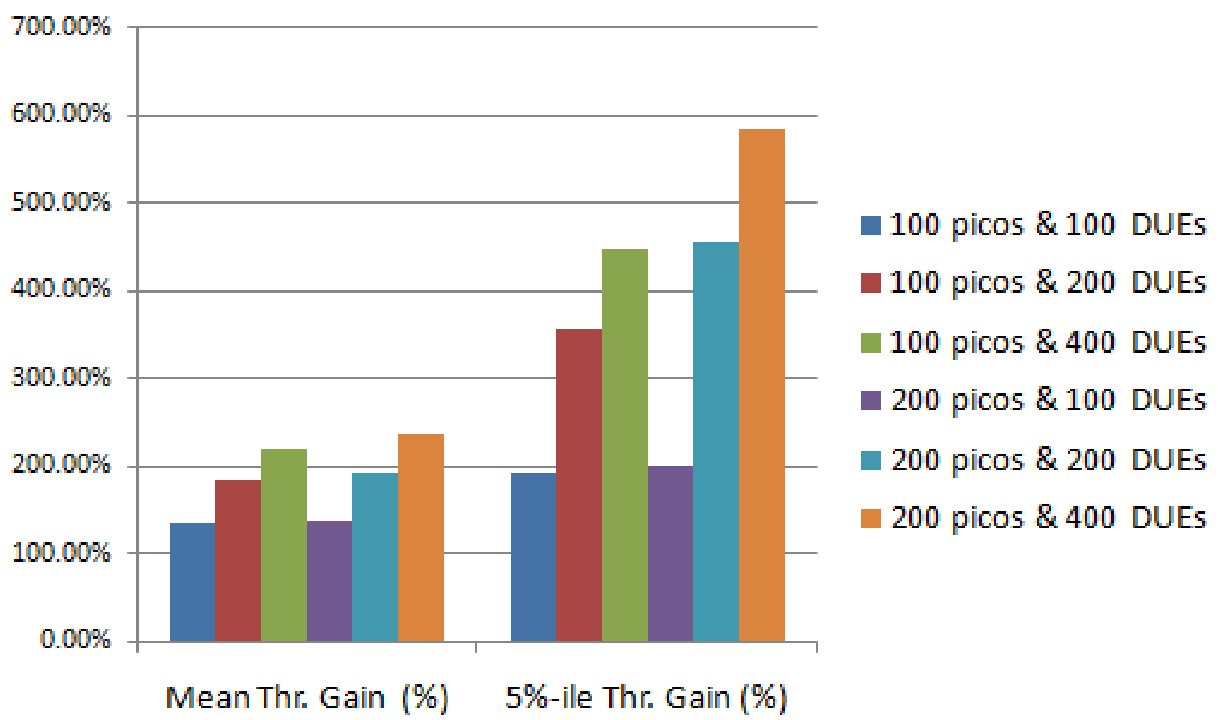


Fig. 8. Dummy UE simulation results

4 Higher Layer Fountain Code Enabled UE Cooperation

This section describes a flexible UE cooperation scheme based on the usage of application-layer fountain codes. The design has two considerations, namely requiring as little involvement of the system as possible and rendering UE cooperation more flexible by relying on group efforts instead of the commitment of a few assigned CUEs. As a result, CUEs are free to join and leave the cooperation groups at will.

For every transmitted data packet, a UE having successfully received the packet can send a MAC sub-layer ACK to the transmitter. In this case, if more than one UE have correctly received the packet, they may send an ACK independently of each other, which mimics multi-path transmissions. Since a CUE can send an ACK to the system and may occasionally fail to forward the decoded packet to TUE, forwarding a received packet is not guaranteed. In addition, even with PHY layer H-ARQ, residual error can occur. To ensure both flexibility and reliability, fountain codes are used in the higher layer (relative to the MAC sub-layer). With UE fountain cooperation, each CUE does the "decode-and-forward" on a best-effort basis. In other words, if a UE cannot help at a given point in time, it does not have to forward the data.

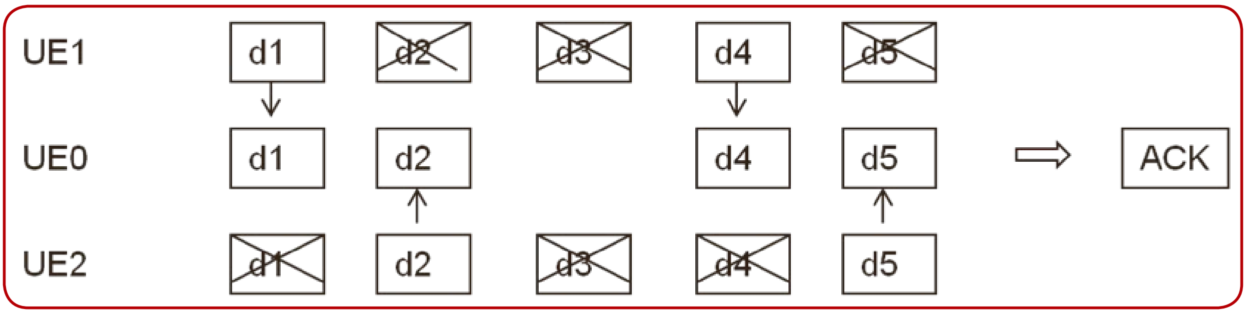


Fig. 9. Flexible UE cooperation enabled by higher layer fountain codes

In Fig. 9, UE0 is the TUE, UE1 and UE2 are CUEs, and d1 to d5 are the transmitted data packets. With fountain codes, UE0 needs to collect four error-free packets to recover the transmitted data. Packet d3 was lost in the process. After successfully collecting packets d1, d2, d4, d5, UE0 sends back an ACK to the corresponding high layer.

With the introduction of fountain codes in the high layer, PHY layer codes can be jointly optimized. Reflected in the scenario under consideration, H-ARQ design can take advantage of the existence of high layer fountain codes by reducing the maximum number of H-ARQ retransmissions. Fountain codes can also be used in conjunction with UE cooperation to enable efficient scalable video multicast in 5G networks [6].

A UE-centric sounding and UL measurement scheme is proposed to overcome the challenges imposed by network virtualization. The sounding scheme decouples a UE-centric SRS from any TPs in a wireless network area, and associates the UE SRS uniquely with UE in the network area.

5 Device-Centric Physical Channels

In current networks, broadcast control channel, UE-specific control channel and data channel are all associated with a serving cell and UE ID. With virtualization, the three channel types are redesigned, and disassociated from each other.

The design utilizes some or all nodes to transmit broadcast control channels, while the UE-dedicated channels in UL and DL are optimized dynamically and from different access-node sets. This is transparent to the UE.

Virtualization of the physical channels provides more scheduling flexibilities, improvement in data and control capacity, better energy saving and mobility management.

6 UE-Centric Dedicated Connection Signature

To decouple UE access resources with individual TPs, a UE-centric unified access scheme is proposed. It keeps the UEs always "connected" to the network or a region of it by allocating each UE a unique connection information, referred to as dedicated connection signature (DCS) in the network region.

A DCS for one mobile UE can associate a sequence from a predefined set of random access sequences to it or for a static or fixed UE can associate one of the random access sequences and its location information, resulting in a unique DCS in the associated network region. For example, by employing a large pool of well structured Zadoff-Chu sequences, the pool of the sequences can be divided between fixed and mobile users. DCS allocation strategy for static/fixed users with known positioning information is to maximize code reuse efficiency; while for mobile users, to minimize DCS switching between the network areas. The DCS will allow a UE for a dedicated usage to access the network anytime and anywhere, while for the UE being uniquely identified by any TPs in the wireless network area such as small city or large downtown area.

The DCS of a UE is independent of cells or TP/RP and can be assigned by the network during its initial entry. After a DCS is assigned to a mobile UE, the network will track the UE position and notify close-by TPs/RPs with the UE profile including its DCS such that the neighboring TPs can serve the UE seamlessly without any interruption.

A UE DCS can be used for access activities, including (i) seamless handover without signaling as well as reliable soft-switching; (ii) fast terminal state transition; (iii) quick connection setup from any link failure or retry; (iv) fast bandwidth request for any scheduling based traffic/application types; (v) support for grant-free transmission; and (vi) fast UL re-sync. Table 1 provides some benefits in a few use cases.

Table 1 Performance summary of UE centric access

Signaling Saving Benefits with Unified Access Scheme

Performance/Signaling		Current schemes	Unified access	Savings
Latency*		10ms	4ms	60%
Access messaging		4	2	50%
Air link signaling for (HO) connection sequence switching	for 3km/h UE in 500m-radius cells	max: 3 times/hr	max: 0.25 times/hr	92%
	for 3km/h UE in 50m-radius cells	max: 30 times/hr	max: 0.25 times/hr	99%

* On best scheduling scenario

7 Network Oriented Measurement

With the virtualization of the network, the channel measurement should transform from cell specific to network specific. We propose a network oriented measurement based on UE-centric SRS that is not tied to a specific TP that utilizes a sounding sequence transmitted over certain time frequency resources. The UE-centric sounding information is set up during the UE initial network entry, where the SRS information setup can be predefined, e.g., a mapping rule from the UE DCS, or directly sent by a TP in the network.

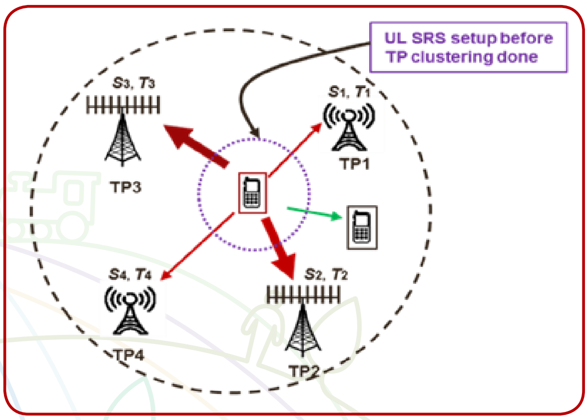


Fig. 10. UE centric sounding scheme

The mechanism of network oriented measurement and control solution is that each surrounding TP is able to monitor the sounding signals of any UE or keep-alive messages associated with the UE, and measure the UE's signal strength as well as timing information; it then periodically reports the measurement information of the UE to a network control node where UE-TP relationship table with channel information can be generated and maintained. Therefore, multiple TPs in neighborhood are related to each UE, and a subset of the TPs will serve the UE. In addition, UE SRS signals can also be sensed by its neighboring UEs as well for services such as D2D or MTC devices.

The UE-TP association map can be updated periodically, which will reflect UE mobility. Moreover, the proposed network oriented measurements can be applied directly for DL and UL functional controls such as both slow and fast DL link adaptation and power control.

The UE-centric sounding scheme is shown in Fig.10 where any TPs within a UE sounding signal coverage can detect and evaluate the UE UL channels, while uniquely identifying the UE, thus making the network track the UE completely and provide optimal services to it.

8 Summary and Conclusions

In this paper, the concept of radio access virtualization is presented where the cell edge experience is eliminated through cooperation of network nodes and other user nodes. To enable these features, several aspects of the cellular networks are examined and redesigned to disassociate them from a specific node in the network and builds a virtual cell around the users. Methods to optimize the cooperation sets among the network nodes and among the users are essential to achieve the ubiquitous user experience. The new design also includes solutions to various aspects of the network such as physical channel design, dedicated connection signature and measurement channels.

References:

[1] 3GPP, "Final report of 3GPP TSG RAN WG1 #74 v1.0.0," Barcelona, Spain, Aug. 2013.

[2] 3GPP, "Final report of 3GPP TSG RAN WG1 #75 v1.0.0," San Francisco, USA, Nov. 2013.

[3] K. Zarifi, H. Baligh, J. Ma, M. Salem, and A. Maaref, "Radio access virtualization: Cell follows user" IEEE PIMRC, Washington, DC, Sept. 2014.

[4] Y. Cao and A. Maaref, "Soft Forwarding Device Cooperation Strategies for 5G Radio Access Networks" IEEE 25th Annual international Symposium on Personal, Indoor and mobile Radio Communications (PIMRC), Washington, DC, Sept. 2014.

[5] A. Maaref, J. Ma, M. Salem, H. Baligh and K. Zarifi, "Device-centric radio access virtualization for 5G networks", IEEE Globecom International Workshop on Emerging Technologies for 5G Wireless Cellular Networks, pp. 887-893, Austin, TX, Dec. 2014.

[6] Y. Cao and A. Maaref, "Device Cooperation-Assisted Scalable Video Multicast with Heterogeneous QoE Guarantees" The Eleventh International Symposium on Wireless Communication Systems (ISWCS), Special Session on Empowering Technologies of 5G Wireless Communications, Barcelona, Spain, Aug. 2014.

[7] METIS D1.1, April 30, 2013.

5G mmWave Communications

5G mmWave

Abstract — Millimeter wave communication is deemed to be one of the promising solutions to meet the ever growing traffic explosion for 5G wireless communications [1-5]. The abundant available bandwidth in 6-100GHz makes it easy to deliver multiple to tens of Giga bps data rate to users. However, mmWave transmission meets many challenges from channel propagation and front-end limits. In this paper, we introduce the key technologies of mmWave communications to investigate its feasibility and future applications.

Keywords — millimeter wave communications, 5G, unified access and backhaul, uniform air interface; beam tracking.

1 Introduction

Recently, broadening 5G to millimeter wave (mmWave) bands is an emerging hot topic widely discussed in industry and academia [1-4]. The widely accepted term on the mmWave bands is 6-100GHz with a comparison to the spectrum bands below 6GHz for IMT systems, although the wavelength of 6-30GHz is actually in centimeters level. The mmWave bands have the advantage of an ultra-wide band available for transmission. A survey [4] shows that a total spectrum of 45GHz is available in 6-100GHz, which is tens of times of the available bands below 6GHz. Such a huge spectrum of bands makes it easy to achieve multiple to tens of Gbps data rate for transmission, and 1000 times throughput improvement over LTE systems [5][12][16].

However, mmWave suffers from larger propagation fading than lower frequency bands, particularly in NLOS and moving scenarios. Another challenge is that the transmit power decreases as carrier frequency increases due to the limits of components in front-end such as power amplifier [2]. Fortunately, pioneering researches show that there are lists of key technologies which make mmWave communications feasible for some scenarios [2][3][12]. Antenna array-based beamforming and tracking could partly compensate pathloss. Channel measurements show that 6-100GHz is possible to cover a range of small cell [2][9][10]. Ultra-dense network and self-backhauling could improve the network capacity at a reasonable cost [4][12][16]. Recently several companies announced their prototype verification for mmWave communications [17-20]. Samsung realized a peak data rate of 7.5Gbps at 28GHz frequency [17], DoCoMo realized a peak data rate of 10Gbps at 11GHz frequency [18]. Huawei and Nokia demonstrated a peak data rate of 115Gbps [19] and 10Gbps [20] at 72GHz bands, respectively. These researches have attracted further focus on 5G mmWave communications.

In parallel, industrial standards are widely discussed to pave the way for mmWave communications. WRC-15 is going to decide the spectrum for 5G in range of sub-6GHz, and a common view is that WRC-19 will decide the spectrums above 6GHz for 5G. ITU-R WP5D was initiated in 2012 aiming to standardize the IMT system that will be commercialized in 2020 (called 5G). Currently it has finished the standard timeline plan in ITU and will output a vision paper of 5G in this year [6]. In March 2015, 3GPP agreed to initiate 5G standards at the end of 2015, and create a study item working on the channel model of mmWave bands [7]. A widely accepted view is that 3GPP Release 14, 15 and 16 will be the period to standardize 5G systems. It is expected that 5G mmWave communications will be standardized in R15 and R16, later than 5G sub-6GHz standards. Regional discussions on mmWave communications include IMT2020 of China where Huawei is chairing mmWave communications topic, European projects such as METIS and 5GPPP, ARIB activities in Japan, etc. Most of them are expected to output their research results to 3GPP and ITU.

Huawei began the mmWave communications research in 2010. Researches cover spectrum, channel measurement, modeling, air interface algorithm, networking and front-end technologies. In early 2014, we published the prototype that realized a peak data rate of 115Gbps in E-band, which is the first prototype in the world that can achieve such a high data rate [19]. In IEEE Globecom'14, Huawei demonstrated the system-level capacity of mmWave communications based on Munich city map, new channel measurement and model in E-band and 28GHz. We also demonstrated the beam tracking technique for two users in field tests in E-band which shows real-time tracking ability.

In this paper, the authors attempt to give an overview of Huawei researches on 5G mmWave communications, and investigate the key technological solutions till the current stage. Section II summarizes Huawei's general views on 5G mmWave communications. Section III presents the channel measurement and model results. Section IV investigates unified radio and access architecture, and Section V studies beamforming and tracking techniques. Huawei's prototypes are introduced in Section VI. Finally we summarize the challenges of mmWave communication left to be dealt with in future researches.

2 Huawei's Viewson mmWave Communications

The mmWave bands considered in 5G system range from 6GHz to 100GHz, whose wavelengths are mostly located in millimeter wave range, thus the name mmWave communications. A survey [4] shows that the available spectrum in 6-100GHz is 45GHz which is tens of times of the spectrum band available in spectrum below 6GHz and hence mmWave communications can easily meet multiple to tens of Gbps data rate required by 5G[12][16].

mmWave bands are the complementary bands to 5G whereas low frequency bands (<6GHz) are still the primary bands of 5G spectrum. It is expected that a variety of spectrums will be used in 5G, so it is called all spectrum access (ASA). 5G ASA investigates above-6GHz (higher) and sub-6GHz (lower) frequency bands, paired and unpaired, licensed and unlicensed, contiguous and no-contiguous frequency bands. Fig.1. shows that E-band has available bandwidth of 10GHz, which is larger than any other bands, and suffers no impacts from satellite communications as met in 28GHz. Hence E-band is preferred as a candidate spectrum for 5G mmWave communications.

Fig.1. All spectrum access for 5G

ASA network is a heterogeneous network which enables the cooperation between lower frequency (LF) and higher frequency (HF) networks. For example, macro base stations guarantee a seamless coverage for all UEs, whereas mmWave communications is more suitable for hotspots, both in indoor and outdoor scenarios, to provide high-throughput and small coverage services. mmWave communications also enable unified access and backhaul (UAB) since both radio access and wireless backhaul could share the same radio resources in higher frequency bands. It is expected to use uniform air interface and a joint hierarchical scheduling for both radio access and backhaul. Moreover, mmWave is suitable for flexible backhauling and low-cost ultra dense networking (UDN) because of the large bandwidth and small coverage of mmWave signals. Channel propagation property in higher frequency bands is the fundamental issue to be studied in mmWave communications. We selected 4 typical frequencies in 6-100GHz for channel measurements, and a 3GPP-style channel model is recommended for future modeling. Beamforming and beam tracking technique based on antenna array is the key technique to enhance the coverage as well as to improve link performance, particularly in moving and NLOS scenarios. Front-end technology is to study the key components such as power amplifiers and high-gain antenna arrays to ensure a front-end of low power consumption, small-volume and high gains.

3 mmWave Channel Measurement and Model

Channel propagation property is a fundamental topic of mmWave communications. The channel model will affect the spectrum allocation in ITU and WRC as well as system design and performance evaluation in 3GPP standards.

Huawei's researches cover multiple spectrum frequency in 6-100GHz, including 15GHz, 28GHz, 60GHz, and 72GHz. Currently we have developed channel sounders with multiple antennas and for different frequency bands. Meanwhile, international cooperation is built with universities across Asia, North America and Europe, including joining the NYU Wireless for channel modeling study.

Ultra-Wide Band Channel Sounder

We developed an ultra-wide band channel measurement sounder for transmitters and receivers, ref. Fig.2. The front-end can be changed in order to match different carrier frequencies. The transmitter has a signal generator for signal generation, and the receiver uses a signal analyzer for measurement data acquisition. A TX and RX synchronized spectrum sweeping solution is adopted. In this way, 2GHz or even larger bandwidth can be measured. The signal generator at TX sites and the signal analyzer at RX sites are controlled remotely by a personal computer through network.

Fig.2. Channel Sounder at E-band, transmitter (left) and receiver (right)

Frequency bands from 71GHz to 73GHz were measured. Vertically polarized horn antennas with 25dBi gain and 10° half power beam width are used for both outdoor and indoor measurement. At the receiver, the channel information from 4 different neighboring angles of arrival can be caught by a four-channel RF-front together with four horn antennas.

Channel Measurements

In this section we present current measurements in E-band, although channel measurements are conducted in 15GHz, 28GHz, and 60GHz bands as well. The typical scenarios of mmWave communications are indoor hotspots and outdoor hotspots. These two scenarios have been measured in Huawei Chengdu branch.

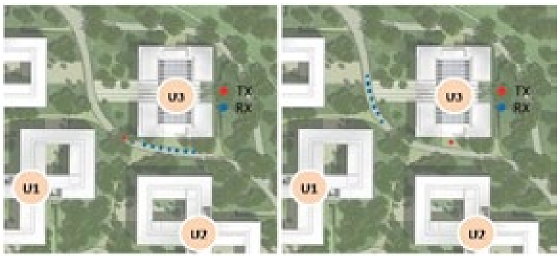


Fig. 3. UMi outdoor scenario, LOS(left) and NLOS(right)

Fig. 4. Indoor hotspot scenario, LOS(left) and NLOS(right)

The outdoor scenario is selected in dense office buildings, see Fig.3. Both the transmitter and the receiver were set at the height of 2–4 meters. In NLOS channel there are buildings and trees for blocking but there are reflected signals from the opposite building where glass windows and concrete walls provide reflections. Occasionally there were a few people going across the road in between, which causes body penetrate loss. The dining room was selected for the indoor hotspot scenario, see Fig.4. The dining room is 70 meters long and 25 meters wide. Food lines are located at both ends. Most part of the walls is made of glass; the rest, of concrete.

For the LOS TX-RX location, the TX and RX antennas were aligned by a laser pen. Then the TX antenna was fixed pointing at the aligned angle. The RX antenna adjusted elevation angle of the aligned line to -10°, 0° and 10°, respectively, with the antenna scanning around the entire 360° azimuth plane in 10° (i.e. antenna half power beam width) step.

For the NLOS Scenario, there are two steps to complete the measurement procedure. First, search the strongest path in each TX-RX location. In this step, both the TX and RX antennas were automatically rotated to find the strongest received power. Second, the pointing angles of the TX and RX antennas for maximum received power were recorded as the 0° azimuth and elevation angles for the TX and RX antennas, respectively. The RX antenna adjusts elevation angle of the aligned line to -10°, 0° and 10°, with the antenna scanning around the entire 360° azimuth plane in every 10° step.

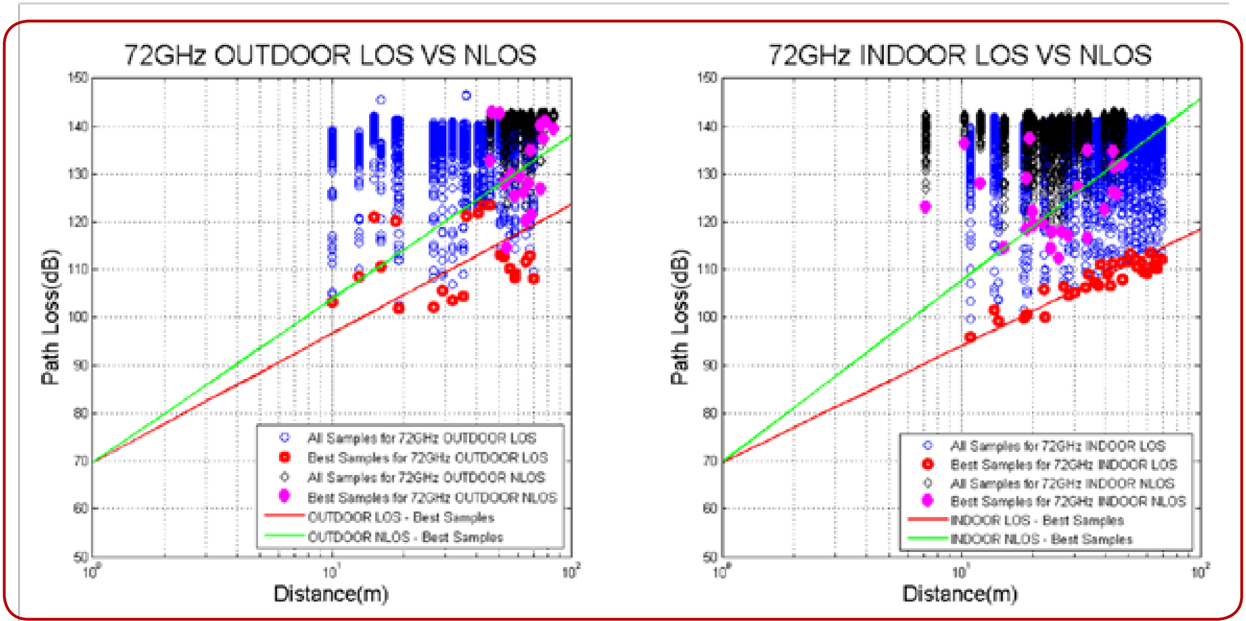


Fig. 5. E-band pathloss in UMi outdoor and indoor

Table I The pathloss parameters in 72GHz, close-in reference model with reference distance 1 meter is applied

Scenario	LOS/NLOS	PLE(β)	α	σ
Indoor	LOS	2.58	69.47	2.38
	NLOS	4.08	69.47	10.63
Outdoor	LOS	2.86	69.47	9.76
	NLOS	3.67	69.47	8.34

The measurement results are shown in Fig. 5. We take close-in reference model with reference distance 1 meter for pathloss since all the parameters have physical meaning [8], and the model is reasonable when the measurements cannot cover all distances and sites. Table I lists the parameters of close-in reference model based on the measurements. The NLOS PLEs of 72GHz is 4.08 for indoor and 3.67 for outdoor which are comparable to that of 2GHz in 3GPP model. However, a more accurate pathloss channel model and more measurement data are required to analyze channel propagation before making any conclusion.

Channel modeling

There exist a number of channel models that could be the references for mmWave channels, such as the map-based model in METIS project, S-V model in IEEE802.11ad, and 3GPP 3D channel model for sub-6GHz bands. Since 5G will be standardized in 3GPP and ITU, a 3GPP-like channel model is a practical approaching method, that is, it might be feasible to reuse the framework of 3GPP channel model while adding new features of mmWave channels.

The first new feature is that mmWave channel measurement is crucially dependent on the directional antenna used in channel sounder. Even at the same scenario, measuring with different horn antennas resulted in different pathloss [8][14]. How to decouple the channel model with antenna pattern is challenging. Secondly, measurement with directional antenna may meet the cases where the beams of TX and TX are not aligned even in LOS scenario which yields NLOS-like pathloss. So the3GPP definition on LOS/NLOS may lead to a larger pathloss for LOS channel. Thirdly, the clusters property in mmWave is crucially dependent on the antenna pattern for measurement. This will affect delay spread, such as AoA and AoD. Future researches will study how to include the new features in mmWave channel models.

4 Unified Access and Backhaul (UAB)

In this section we present unified access and backhaul (UAB) techniques under hybrid network. A hybrid network consists of Macro base stations (MBs) and millimeter-wave base stations (mBs), where MBs and mBs are connected with each other via backhauls, as illustrated in Fig. 6.

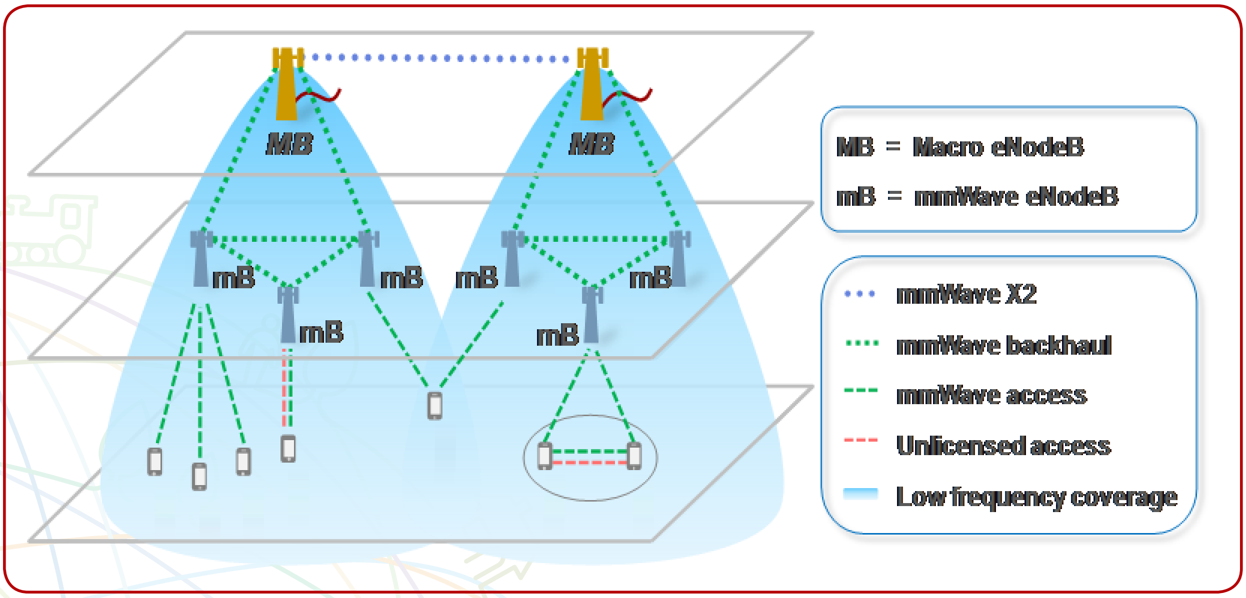


Fig. 6. Hybrid Network Architecture

The first layer is made of MBs which are working both in sub-6GHz frequency carriers as well as higher frequency (above-6GHz) carriers. An MB can communicate with all the mBs in its covered area through lower frequency. Since the bandwidth of sub-6GHz frequency is relatively narrow and has smaller propagation loss than higher frequency, it is designed to deliver important information such as control-plane data. Millimeter-wave in MB is used to deliver user-plane data to UEs as well as to backhaul data to neighboring mBs. Although lower frequency and millimeter-wave frequency share the same site in MB, their antennas and remote RF units are completely separated and may be installed in different height depending on coverage requirements.

The second layer is mBs, which are much denser than MB and current LTE small cell networks. Considering large propagation loss, the density of mBs may range between 6 and 500 mBs per square kilometer which corresponds to a cell radius of 25~200m. The data from or to core network are communicated through MBs to mBs and/or UEs. Each mB installs millimeter-wave frequency for both backhaul and radio access. A unified architecture is that both radio access and backhaul share the same platform including antenna array, IRF and baseband units. The advantage of the unified architecture is that backhaul and radio access can be jointly managed to schedule radio resources and antenna resources, and hence resources can be utilized in a more efficient way. For example, when backhauling has larger load than radio access, more frequency band or antenna beams can be allocated to backhauls rather than radio access, and vice versa.

The third layer is radio access where UEs could access MBs through lower frequency and millimeter wave, or to mBs through millimeter wave. Unlicensed spectrum, in the network architecture, is used only for radio access to deliver less important information, because links over unlicensed spectrum might suffer from unexpected interference. Millimeter wave is also suitable for device-to-device communication (D2D). Multiple UEs with D2D connections can do joint transmission and joint reception to improve transmission performance.

Load-Centric Backhauling (LCB)

Hybrid network enables adaptive backhauling. Beams can be adaptively generated to adjust backhauls. Adaptive self-backhauling is particularly important when traffic loads vary in different areas. Since traffic loads may distribute very un-uniformly in geography, it is expected that backhaul network can be adaptively adjusted to track traffic load changes in network, and that is called load-centric backhauling (LCB).

Therefore, we propose to use hierarchical Radio Resource Management (RRM) architecture to realize the load-centric networking. The RRM for backhauling (BH-RRM) performs the function to allocate radio resources for all the backhauls between nodes. The function locates in MB which can communicate with all its covered mBs via sub-6GHz frequency. Each mB performs the RRM of radio access (RA-RRM) as well as the function to execute backhauling at resource and configuration given by BH-RRM. Note that RA-RRM function is the same as the RRM in LTE eNodeB which allocates radio resources to local users.

The function splitting between BH-RRM and RA-RRM is crucially dependent on system architecture. If there is enough bandwidth between MB and mB, RA-RRM can actually move to MB, and a powerful RRM in MB may do scheduling for both backhaul and radio access. Such a centralized structure is also suitable for cooperative communications such as coordinated multipoint (CoMP). Joint transmission and joint reception among multiple distributed mBs can be successfully implemented in the centralized unit. Another extreme scenario is that BH-RRM is combined with RA-RRM and located in every mB. The network then becomes similar to a mesh network and every node performs scheduling in a distributed way. The advantage of the distributed architecture is that the centralized node is not required and this simplifies network deployment. However, the disadvantage of this approach is that it is hard to do adaptive backhauling and cooperative communication.

The authors in [4] presented an example of LCB under unbalanced traffic load among mBs where dynamic routing is applied. It is assumed that 60~90% of traffic is concentrated in 10% of mBs, and it is shown that network capacity gains by using LCB are 9.60%, 21.50%, 56.00% and 159.70% in cases of 60%, 70%, 80% and 90% traffic loads concentrated in 10% nodes. The trend is that when traffic loads distribute more non-uniformly in geography, adaptive backhauling may achieve larger capacity gain than fixed backhaul.

Dynamic Radio Resource Allocation for UAB [11]

Under UAB architecture, radio access and backhaul of each mB may share radio resource and the hardware such as baseband unit and front-end units. UAB makes it possible to do joint radio resource allocation between backhaul and access to improve spectrum efficiency.

In a hybrid network with MB for large area coverage and mB for small cell coverage, each mB has an mmWave backhaul connecting to MB. Each mB can support six pico cells with 60 degree each. Inside each pico cell, there will be a large number of UEs connecting to the mB. These UEs communicate with their associated mB through high frequency bands. For those UEs that are unable to connect to mBs, they can choose to communicate with MB through low frequency bands.

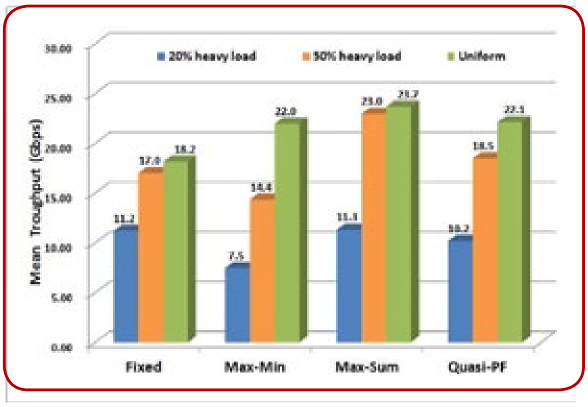


Fig. 7. UAB throughput with dynamic radio resource allocations between access and backhaul at E-band

In order to mitigate interference between radio access and backhaul, it is necessary to separate them either in time or in frequency resource blocks. Here we assume that the resource partition ratios are the same for all mBs in the network in order to avoid frequency band overlap between radio access and backhaul from other mBs. As an example, the frequency division is taken in this paper. The extension to time division is straightforward. We assume that the overall bandwidth is B, where backhaul is allocated with BBH and radio access is allocated with BRA. The throughput for backhaul and radio access is computed by multiplying backhaul rate and access rate with their bandwidths, correspondingly. As the overall throughput of one mB is limited by the minimal value of the two, the overall throughput versus backhaul bandwidth is a triangular function, where the ascending slope is the backhaul rate and the descending slope is the access rate. For multiple mBs, each triangular function has its individual backhaul rate and access rate, thus having different peak values. The goal is to choose a unified backhaul bandwidth for all mBs, based on the following three criteria:

- Maximization of the minimum throughput
- Maximization of the sum throughput
- Maximization of the satisfactory factor

Three dynamic resource allocation algorithms based on the above criteria have been proposed in [11]. The simulations are carried out in 72GHz with bandwidth B = 5GHz and using 3GPP HetNet topology setup. The results are illustrated in Fig. 7. The throughput gains of Max-Sum are 35.3% and 30.2% over the fixed allocation in the second and third case, respectively. Therefore, we conclude that UAB with dynamic resource allocation between radio access and backhaul is a promising solution for mmWave communications.

5 Beamforming and Tracking

The air interface of mmWave communications is featured with antenna-array based beamforming and tracking. Both mB and UE use antenna arrays to compensate the large pathloss of mmWave propagation. Such a scheme with high-gain narrow beams brings challenges for air interface design. In this section we present the concepts of uniform air interface and multi-mode beamforming, and then discuss whether MIMO is mandatory.

Uniform Air Interface

In all spectrum access (ASA) system of 5G, the mmWave communication is not an independent air interface but subscribes to several constraints from lower frequency air interface and UAB architecture. Hence the concept of uniform air interface is expected to realize with parameter configuration and to flexibly meet different requirements.

Firstly, there might be multiple mmWave bands allocated to 5G, and air interface should be uniform among different carrier frequencies, such as frame and PHY numerologies. It is recommended that mmWave could reuse the 10ms frame structure and 1ms subframe. Slot can be redesigned to be 100us for all mmWave bands. A continuous spectrum bands may have similar channel propagation property, for example, 6-20GHz, 20-50GHz, and 50-90GHz. Different PHY numerologies can be used for different frequency band groups.

Secondly, UAB architecture requires a uniform air interface design between radio access and backhaul. It is better that backhaul works like a special UE of mB where backhaul and UEs share the same radio resources. The difference is that backhaul will use special link configuration, e.g. beams, coding and modulation, to meet its QoS requirements. This will be further discussed in multi-mode beamforming.

Thirdly, mmWave communication allows efficient waveform design to meet different QoS requirements in different scenarios, such as power efficiency, latency, etc. Both OFDM and SC-OFDM waveforms used in LTE can be used for mmWave communications. Other candidate waveforms might be f-OFDM and SCMA currently proposed for 5G low frequency systems. A good combination of them is to use software-defined adaptive air interface to select different waveforms for different scenarios [1]. However, a common constraint is that the waveform is robust to phase noise which may dominate other distortions in front-end. PAPR is also an important issue to waveform selection. It is expected that efficient PAPR reduction algorithms are deployed to improve power amplifier efficiency particularly when using multiple carrier waveforms.

Multi-mode beamforming

Multi-mode beamforming (MM-BF) scheme is proposed in this section. It allocates antenna elements, IRF channels and baseband units to the beams of backhaul links and radio access adaptively. Antenna arrays are divided into sub-arrays, and the beamforming process is carried out in not only digital baseband but also IRF part. User data streams and backhaul streams are first put into Adaptive MIMO Mode Selection (AMMS), and AMMS will do MIMO processing of each stream based on the current channel environment, where MIMO mode can be spatial multiplexing (SM), space time coding (STC), space time beamforming, etc. After AMMS, the data streams are precoded in digital domain based on beam requirements and downlink channel estimation results, then the signal streams pass through DAC, IR+RF processing. Before the signal to feed to the antenna elements, the phase of signals should be shifted for each antenna elements based on the beams requirements and channel estimation results. Such a MM-BF structure can flexibly realize multiple antennas techniques including MIMO, beamforming, diversity and their hybrid forms.

In a mobile scenario with multiple users, the challenge is to design algorithms to align the narrow beams between mB and UEs. Overhead cost, complexity and tracking ability are the key merits of figure to evaluate performance. Two beam phases are proposed to finish the beam alignment: beam training and beam tracking. Beam training performs a rough beam alignment where both quasi-omni-directional beams and wide beams can be used for training. Since exhaustive beam search might bring a high cost in design pilots, there are potential methods to shorten beam training period and overhead. Hierarchical beam training method [15] is an efficient way which firstly use sector-level beam to do training and then use wide beams for searching. Frame design for beam training, for example, centralized training, can also shorten training time.

Beam tracking performs channel information update when there is no beam training. The updated channel information is typically Angle of Arrival (AoA) and Angle of Departure (AoD) for transmitter and receiver to do beamforming. There are two methods to solve the problem. One is to use reference signal and old channel information to predict the AoD and AoA for the next data transmission. The authors in [9] propose a new channel tracking technique based on sequentially updating the beamforming at transmitter and receiver. Numerical results demonstrate the superior channel tracking ability of the proposed technique over various baselines in the literature. The second method is to estimate the AoA and AoD based on reference signal. The challenge is that the cost of reference signal increases with the antenna size in BS and UE, and becomes prohibitive when the antenna size is larger than 16. An efficient method to overcome the problem is to use compression sensing (CS) technique which can exploit the sparse property of mmWave channel, and can significantly reduce overhead. The authors in [10] propose a joint channel estimation and beamforming method where CS is used to estimate AoA and AoD jointly. Results show that the overhead can be saved by up to 75% compared with traditional non-CS estimation method, e.g., least square algorithm, under practical scenarios. Furthermore, the proposed method in [10] has only 2~3dB loss compared with the method with perfect channel information.

Is MIMO mandatory?

In addition to the antenna array used in front-end, MIMO with digital steering in baseband is an efficient method to enhance throughput or performance. There are two factors limiting the MIMO realization, however. One is that high-speed ADC and DAC for mmWave are expensive and its power consumption is high. Since each digital chain needs a set of ADC and DAC, it is preferred that the number of chains is no more than 4. Computational complexity in baseband also prevents high-order MIMO precoding and detection, particularly for bands wide up to GHz.

In multiuser communications, an efficient solution is to separate users in space by beams and multiuser MIMO to improve transmission performance. Each user may own one or multiple data streams depending on channel status. The authors in [12] have analyzed performance of multiuser MIMO+beamforming. An interesting result is that with the same antenna aperture in mB and UE, higher frequency may suffer from smaller interference of neighboring beams, cells and sites. System level performance is compared at 72GHz and 28GHz, deploying 3GPP HetNet topology with small cell radius of 50-150 meters. The 72GHz system uses beams with half power beamwidth (HPBW) of 4 and 13in mB and UE whereas the 28GHz system uses 10 and 21. Each cell is assumed to support four beams for simultaneous transmission. It is found that the interference in 28GHz system may degrade average SNR by 30dB, whereas the degradation in 72GHz system is 5dB. An important reason is that the wider the HPBW, the larger interference may be produced. Wide sidelobes also play an important role in causing interference. Hence, we conclude that MIMO precoding and complex MIMO detection is not mandatory in case where the beam is narrow enough to avoid interference. Interference mitigation technique such as MIMO precoding may be not necessary for 72GHz, but it is mandatory for 28GHz.

System-Level Performance Evaluation

In this section, we evaluate the system performance of 72GHz and 28GHz systems. The bandwidths of 72GHz and 28Hz are 2.5Hz and 500MHz, respectively. This is because the total available bandwidths are 10GHz and 2GHz for 72GHz and 28GHz, respectively. Consider 3GPP HetNet topology where an MB with a radius of 500m and three mBs are distributed in a macro cell. Each mB has 6 cells with a radius of 50m. The antenna apertures are 66mm by 66 mm at each cell of mB and 16mm by 16mm at UE. All antenna elements have half wavelength separation. Exhaustive beam training is applied to align beams in mB and UEs. Phase noise with Wiener model and EVM are included for the OFDM-based systems.

In this section, we consider downlink only. Perfect channel estimation and perfect CQI feedback are assumed. Radio resources for multiple users are scheduled in space and time dimensions. There is no further division in frequency domain because smaller granularity, such as the PRB in LTE system, can bring 6.8% performance gain [12] which is little compared to the cost in front-end to support frequency division. So each user will occupy a slot with whole frequency band. This is suitable for mmWave communications since analogy antenna can form a beam pattern at a time which applies to all frequency bands. Authors in [13] have studied different scheduling algorithms to reduce interference and presented

SLNR-based and SINR-based PF algorithms. The idea is to select the beams with smaller interference with each other while keeping fairness. In this paper, we use PF for simplicity purpose and it will bring 20-30% throughput degradation compared to SINR-based algorithms [13].

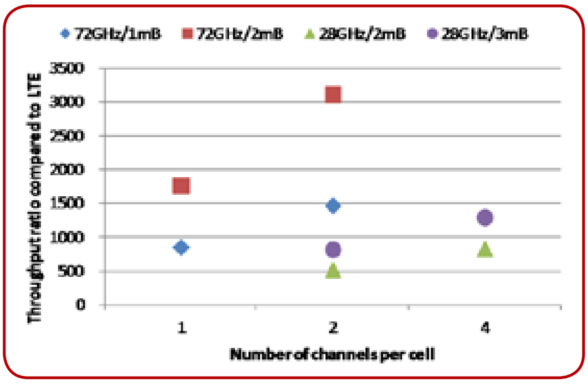


Fig. 8. System throughput compared to LTE, 72GHz and 28GHz

System throughput performance of downlink is shown in Fig. 8. The baseline is LTE system configured with 20MHz bandwidth, with ISD=500m, and 4x2 MIMO downlink. The throughput of the LTE baseline is 0.69Gbps/km2. We investigate the cases with 1, 2, 3 mBs per macro cell, and 1, 2, 4 channels per mB cell, respectively. For fairness 72GHz system has bandwidth of 2.5GHz and 28GHz system has 500MHz bandwidth. In order to reach a 1000 times throughput increase over LTE, it is shown that 72GHz system needs 1 channel per mB cell and 2 mBs per macro cell, or 2 channels per mB cell and 1 mB per macro cell. However, 28GHz needs 4 channels per mB cell and 3 mBs per macro cell. Compared to 72GHz, 28GHz system needs to increase node density and channels to 6 times. Therefore, we conclude that 72GHz can achieve a 1000x throughput enhancement with less channels and sparser node density and hence reducing the CAPEX and OPEX in networking.

6 Prototype Demonstration

In this section, we present 115Gbps prototype working in E-band and a mmWave platform supporting multiple bands.

115Gbps prototype

The 115Gbps prototype, see Fig. 9, aims to exploit the peak data rate by using total available bandwidth 10GHz in E-band. Point-to-point transmission is demonstrated in indoor LOS scenario. MIMO is configured with 2 polarized antennas in mB and 2 polarized antennas in UE. With LOS MIMO, two data streams are transmitted at the time. OFDM waveform and LDPC channel coding are used for link transmission. With 20% overhead considered, two streams, each with 64QAM, are successfully transmitted which reach data rate of 115Gbps.

Fig. 9. Huawei 115Gbps prototype at E-band, transmitter and receiver (left), real-time receiver performance (right)

To reduce the severe phase noise at E-band, a low-complexity two-stage estimation and compensation scheme is proposed for OFDM-MIMO systems [20]. It is a joint channel estimation and phase noise compensation method. Results show that the proposed algorithm is robust to phase noise even when noise model is uncertain.

mmWave Platform

The mmWave platform aims to support multiple mmWave bands, including Ka-band, V-band and E-band. All the spectrum bands share the same baseband which are realized by FPGA. Separate IRF (intermediate and radio frequency) units are realized for the three bands. So it is easy to realize prototype for different spectrum bands simply by changing the IRF units. Baseband and IRF are connected with I and Q digital samples. The ADC and DAC have sampling rate of 2.5Gsps which supports signal bandwidth up to 2GHz. Four antenna sub-arrays are supported, each with 64 antenna elements, reaching 256 antennas in total. Two UEs are realized with the same platform with mB.

A field trial is deployed in Huawei Chengdu branch. Fig. 10 illustrates the prototype in LOS/NLOS scenarios. Currently we have realized the proposed multi-mode beamforming and tracking algorithm in platform. Results show that mB can trace the two UE in a moving speed of 3km/h in NLOS channel well, and mB could finish beamtracking in no more than 49ms. The next step is to demonstrate the data communications and improve the link performance.

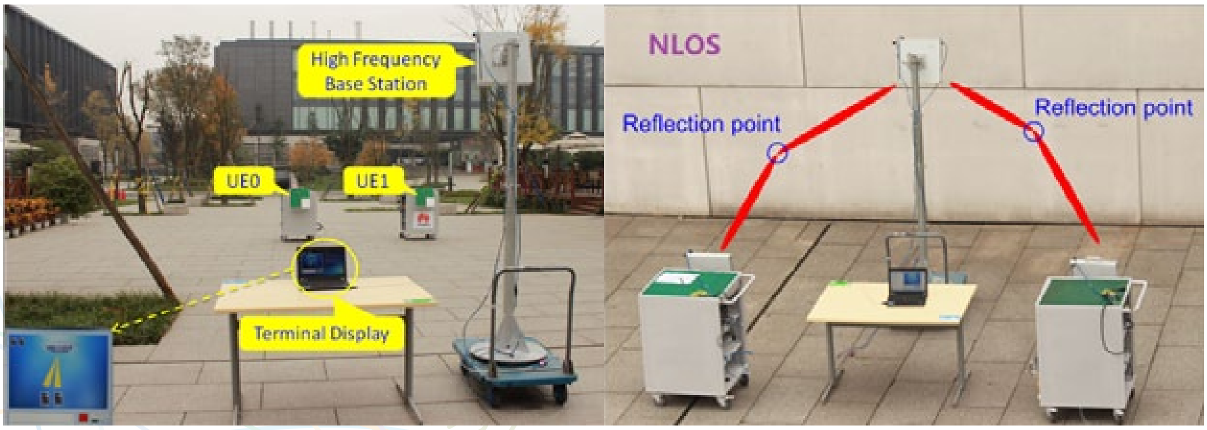


Fig. 10 Full-spectrum prototype in LOS (left) and NLOS (right) scenario at Huawei Chengdu branch

7 Summary and Future Challenges

In this paper, we have introduced Huawei researches on 5G mmWave communications. Based on all spectrum access concepts, new techniques are introduced for mmWave communications, including unified access and backhaul, load-centric backhauling, uniform air interface and multimode beamforming. Although multiple spectrum bands are suggested for 5G, E-band is recommended because of its large bandwidth available. Researches show that E-band suffers from smaller interference in MIMO+beamforming structure and MIMO precoding is not mandatory. Moreover, E-band system utilizes smaller node density and smaller number of RF channels in order to achieve throughput 1000 times more than LTE system. This will lead to lower cost in CAPEX and OPEX compared with other spectrum bands such as 28GHz.

However, there are still many open challenges in applying mmWave bands to mobile communications. The property of mmWave channel propagation is still not well understood. For example, reflection and diffraction property are uncertain in NLOS scenario; channel is not decoupled from measured directional antenna yet; penetration loss is not agreed yet; and a suitable channel model for mmWave band is uncertain. The second challenge is there is an area where UE may not get enough signal power for detection, which is called blind area. Blind area may seriously degrade user experience. The third challenge is the front-end. How to reduce the high power consumption in baseband, particularly in the MIMO case? How to increase the power efficiency of RF? How to form a good pattern in antenna array with low side lobes? These are all front-end challenges. The front-end of terminal is more challenging since terminal is limited by volume, power and cost. Future researches will focus on these challenges. All the key technologies need to be further studied and solidly verified before answering whether mmWave bands are feasible for mobile communications.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Wen Tong and Dr. Peiying Zhu for supervisory and general discussions. We would like to thank some colleagues of 5G HF project for their support in project researches. They are Stan Lu, Chang Cao, Jia He, Xianfeng Du, Zimin Yu, Keli Zou, Guangjian Jiang, Jingjing Cheng, Kunpeng Liu, and Tianxiang Wang.

References:

[1] W. Tong, "5G goes beyond smartphone", keynote speech at IEEE Globecom'14, Austin, USA, Dec. 2014.

[2] T. Rappaport, S. Sun, R. Mayzus, et al, "Millimeter wave mobile communications for 5G cellular: It will work!" IEEE Access, vol. 1,pp. 335–349, May 2013.

[3] W. Roh, J.Y. Seol, J.H. Park, B. Lee, J. Lee, Y. Kim, J. Cho and K. Cheun, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: theoretical feasibility and prototype results", IEEE Communications Magazine, Feb. 2014, pp. 106-113.

[4] Y. Wang, J. Li, L. Huang, J. Yao, A. Georgakopoulos, and P. Demestichas, "5G Mobile: spectrum broadening to higher-frequency bands to support high data rates", IEEE Vehicular Technology Magazine, Sept. 2014, pp. 39-46.

[5] Y. Wang, etc, Workshop on Mobile communications with higher frequency bands, in conjunction with IEEE Globecom'14, Austin, USA, Dec. 2014.

[6] ITU-R WP 5D, Document 5D/TEMP/548(Rev.3), "PRELIMINARY DRAFT NEW RECOMMENDATION ITU-R M.[IMT.VISION]", Jan. 2015.

[7] RP-150483, RAN Chairman, "Getting ready for 5G", 3GPP TSG RAN #67, Shanghai, China, Mar 9-12, 2015.

[8] T. Rappaport, R. W. Heath Jr., R.C.Daniels and J.N.Murdock, Millimeter wave wireless communications, Prentice Hall Book, Sept. 2014.

[9] J. He, T. Kim, H. Ghauch, K. Liu, and G. Wang, "Millimeter wave MIMO channel tracking systems",in Proc. IEEE Globecom'14 Workshop - Mobile Communications in Higher Frequency Bands, pp. 414–419, Austin, USA, Dec. 2014

[10] H. Huang, K. Liu, R. Wen, Y. Wang and G. Wang, "Joint channel estimation and beamforming for millimeter wave cellular system", Submitted to IEEE Globecom'15, San Diego, USA, Dec. 2015.

[11] Z. Shi, Y. Wang, L. Huang and T. Wang, "Dynamic resource allocation in mmWave unified access and backhaul network" submitted to IEEE Globecom '15, San Diego, USA, Dec. 2015.

[12] Z. Shi, Y. Wang and L. Huang, "System capacity of 72GHz mmWave transmission in hybrid networks" submitted to PIMRC Workshop 2015, Hongkong, China, Sep. 2015.

[13] H. Li, L. Huang and Y. Wang, "Scheduling schemes for interference suppression in millimeter-wave cellular network" submitted to. PIMRC Workshop 2015, Hongkong, China, Sep. 2015.

[14] G. MacCartney, J. Zhang, S. Nie and T.S. Rappaport, "Path loss models for 5G millimeter wave propagation channels in urban microcells", in Proc. IEEE Globecom '13, pp. 3948 – 3953.

[15] Wireless LAN medium access control (MAC) and physical layer (PHY) specifications, IEEE std 802.11ad., Dec. 2012.

[16] J. Zhang, A. Beletchi, Y.Yi and H.Zhuang, "Capacity performance of millimeter wave heterogeneous networks at 28GHz/73GHz", in Proc. Workshop on Mobile communications with higher frequency bands, in conjunction with IEEE Globecom'14, Austin, USA, Dec. 2014.

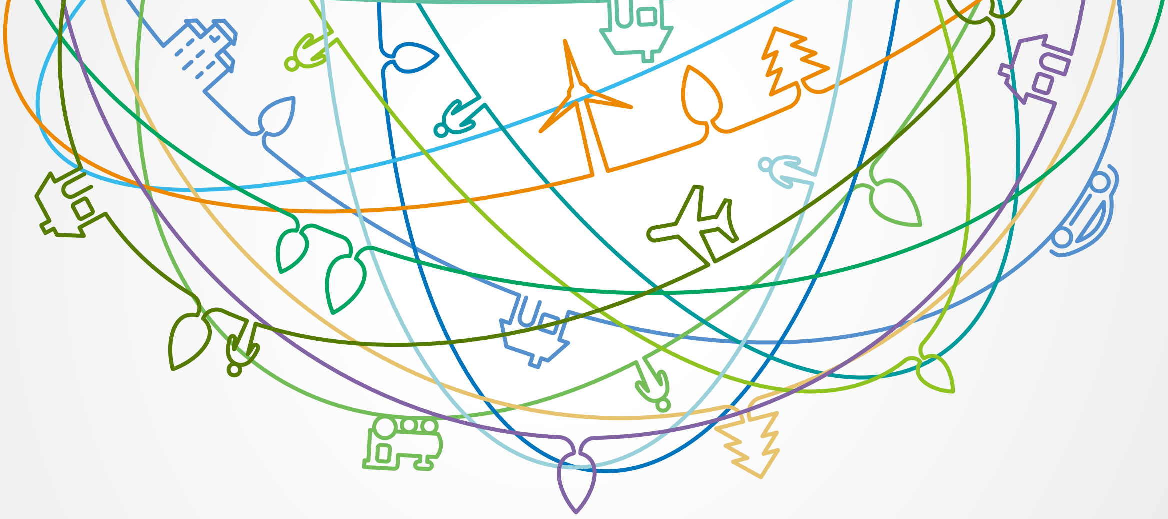
[17] Samsung," Samsung achieves 7.5Gbps transmission speeds on a 5G data network", *http://www.sammobile.com/2014/10/16/samsung-achieves-7-5gbps-transmission-speeds-on-a-5g-data-network/*

[18] S. Suyama, J. Shen, Y. Oda, H.Suzuki and K.Fukawa,"DOCOMO and Tokyo Institute of Technology Achieve World's First 10Gbps Packet Transmission in Outdoor Experiment", in Proc. IEEE PIMRC 2013

[19] Huawei,"Huawei Named Key Member of New 5G Association, Announces Faster than 100 Gbps Speed Achievement at Mobile World Congress 2014", *http://pr.huawei.com/en/news/hw-328622-ictmwc.htm*

[20] Nokia, "Nokia Networks showcases 5G speed of 10Gbps with NI at the Brooklyn 5G Summit", *http://networks.nokia.com/news-events/press-room/press-releases/nokia-networks-showcases-5g-speed-of-10gbps-with-ni-at-the-brooklyn-5g-summit*

[21] H. Huang, G. Wang and J. He, "Phase noise and frequency offset compensation in high frequency MIMO-OFDM system", accepted by IEEE ICC 2015, June 2015.



HUAWEI TECHNOLOGIES CO., LTD.
Bantian, Longgang District
Shenzhen518129, P. R. China
Tel:+86-755-28780808

www.huawei.com