IEEE
COMMUNICATIONS
SOCIETY
Cognitive Networks Technical Committee
http://committees.comsoc.org/tccn

TCCN Newsletter

Vol. 3, No. 2, November 2017

CONTENTS

Chair’s Message

Dear Fellow TCCN Members,

2017 has been a fruitful year for TCCN. We have made significant progresses on several frontiers:

- Establishment of new SIGs (Special Interest Groups) on emerging and hot research areas;
- Reactivation of TCCN Publication and Recognition Awards;
- Formalization of TCCN nomination procedures.

Next I will provide more details on each of the items above.

Establishment of new SIGs (Special Interest Groups): The scope of cognitive network is broad, and we have been encouraging colleagues to establish SIGs to promote specific research directions with the TC’s scope. Before 2017, TCCN already had two SIGs: *SIG on Cognitive Network Security* and *SIG on Cognitive Radio for 5G*. In 2017, we established four new SIGs by excellent technical leaders in the community:

- *SIG on Data-Driven Cognitive Networks:* chaired by Prof. Li-Chun Wang from National Chiao Tung University.
- *SIG on AI Embedded Cognitive Networks:* chaired by Prof. Kai Yang from Tongji University.
- *SIG on Social Behavior Driven Cognitive Radio Networks:* chaired by Prof. Li Wang from Beijing University of Posts and Telecommunications.
- *SIG on Cognitive Communications and Networking in Cyber-Physical Systems:* chaired by Prof. Xianghui Cao from Southeast University.

I would like to thank all the leaders (Chairs, Vice-Chairs, Advisors, and Key Members) of the SIGs for their excellent initiatives and contributions.

The information of all SIGs can be found at http://cn.committees.comsoc.org/special-interest-groups-sigs/. I strongly encourage you to take a look and join an SIG that matches your research interests best. It would be a great platform to meet new friends and advance career.

Reactivation of TCCN Publication and Recognition Awards: TCCN has two awards established several years back: *TCCN Publication Award* for those who are deemed to have made outstanding technical contributions to TCCN, and *TCCN Recognition Award* for who is deemed to have made significant and sustained contributions to cognitive network community. The awards have been given in 2011 and 2012, but have not been given for several years since then. This year, two award committees led by Prof. Geoffrey Li from Georgia institute of Technology and Prof. Ekram Hossain from University of Manitoba, respectively, evaluated many excellent nominations from the community and decided to give the awards to

- *Prof. Khaled B. Letaief* and *Prof. Wei Zhang* as the winners of 2017 IEEE TCCN Publication Award, for their outstanding contributions to theory and practice of cognitive radio technology
- *Prof. Ian F. Akyildiz* as the winner of 2017 IEEE TCCN Recognition Award, for his pioneering contributions to spectrum sensing, spectrum sharing algorithms and communication protocols for cognitive radio networks

My hearty congratulations to the winners, and many thanks to the volunteer work by the committee chairs and members. More information of the awards can be found at http://cn.committees.comsoc.org/awards/.

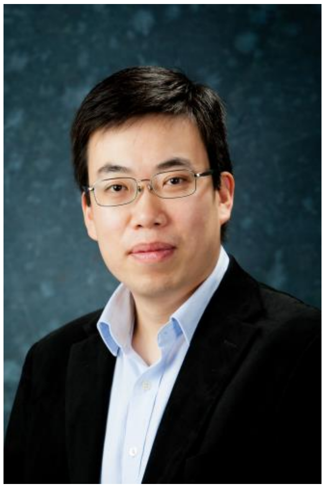
Formalization of TCCN nomination procedures: TCCN is responsible for nominating several important positions in ComSoc, such as ICC/GC Symposium Co-Chairs and IEEE ComSoc Distinguished Lecturers. To encourage the maximum participation from the community, we have streamlined the nomination procedure into two steps: (1) open call-for-nominations from the entire TCCN community, (2) voting by TCCN officers to determine the candidates to nominate, putting special emphasis on the existing and committed service contributions to TCCN. I would like to thank all nominators and nominees for your support of TCCN.

Last but not least, we are always looking for more volunteers to actively engage in various aspects of the TC, including but not limited to

- Organize Special Interest Groups (SIGs) (contact: Jianwei Huang, Lingyang Song)
- Organize special issues for the TCCN Newsletter (contact: Walid Saad)
- Participate in TCCN related standardization efforts (contact: Oliver Holland)
- Contribute to the publicity efforts of TCCN (contact: Yue Gao)
- Involve TCCN in ComSoc conference organization (contract: Lingyang Song)
- Involve TCCN in ComSoc journal special issues (contact: Jianwei Huang)

As always, I welcome any suggestions from TCCN members regarding how to make TCCN a better community. Please feel free to contact me at jwhuang@ie.cuhk.edu.hk if you have any suggestions.

Thanks and best regards,



Jianwei Huang
Chair, IEEE ComSoc TCCN

Professor and IEEE Fellow
IEEE ComSoc Distinguished Lecturer
Web of Science Highly Cited Researcher
Department of Information Engineering
The Chinese University of Hong Kong
http://jianwei.ie.cuhk.edu.hk/

Director's Message

Dynamic spectrum access and spectrum sharing lie at the heart of cognitive radio networks. As such, developing new multiple access techniques along with exploiting unconventional frequency bands are necessary steps needed to ensure the wide-scale deployment of cognitive radio networks. In this regard, this TCCN Newsletter will delve into two key emerging technologies in this regard: a) non-orthogonal multiple access (NOMA) and b) Millimeter wave communications (mmWave). Both NOMA and mmWave are considered as major components of emerging 5G networks and, as such, exposing their challenges and opportunities is essential. In order to do so, this second issue of the TCCN Newsletter of 2017 will bring together two feature topics on NOMA and mmWave. Within each feature topic, we review the state of the art and provide an in-depth exposition of some of the recent research contributions. For NOMA, we also provide a position paper from Dr. Zhiguo Ding's group who has been the driving force behind many of the key contributions in NOMA. In addition to these two feature topics, we discuss the theme of "Spectrum Scarcity", which has been driving much of the research in wireless networks, in general, and cognitive radio networking, in particular, over the past few years. Within the context of this theme, we get expert opinions from Drs. Akbar Sayeed and Mérouane Debbah along with two position papers from the groups of Dr. Danijela Cabric and Dr. George K. Karagiannidis, that expose various research challenges within this theme.

That said, we would like to thank our two feature topic editors: Dr. Daniel Benevides da Costa from UFC - Brazil and Dr. Omid Semiari, from Georgia Southern University, for their efforts in arranging the paper reviews, positions papers, and expert opinion. Moreover, we want to thank all authors and interviewees for contributing their significant research works to the two feature topics and sharing with us their useful experience and future outlook on the area. I would finally like to acknowledge the gracious support from the TCCN chair, Dr. Jianwei Huang and all TCCN officers. As always, if you have any suggestions, feel free to contact me at: walids@vt.edu. We hope that you enjoy the material provided here!

Walid Saad (S'07, M'10, SM'15) (walids@vt.edu) received his Ph.D degree from the University of Oslo in 2010. Currently, he is an Associate Professor at the Department of Electrical and Computer Engineering at Virginia Tech, where he leads the Network Science, Wireless, and Security (NetSciWiS) laboratory, within the Wireless@VT research group. His research interests include wireless networks, machine learning, game theory, cybersecurity, unmanned aerial vehicles, and cyber-physical systems. Dr. Saad is the recipient of the NSF CAREER award in 2013, the AFOSR summer faculty fellowship in 2014, and the Young Investigator Award from the Office of Naval Research (ONR) in 2015. He was the author/co-author of six conference best paper awards at WiOpt in 2009, ICIMP in 2010, IEEE WCNC in 2012, IEEE PIMRC in 2015, IEEE SmartGridComm in 2015, and EuCNC in 2017. He is the recipient of the 2015 Fred W. Ellersick Prize from the IEEE Communications Society and of the 2017 IEEE ComSoc Best Young Professional in Academia award. From 2015-2017, Dr. Saad was named the Stephen O. Lane Junior Faculty Fellow at Virginia Tech and, in 2017, he was named College of Engineering Faculty Fellow. He currently serves as an editor for the IEEE Transactions on Wireless Communications, IEEE Transactions on Communications, IEEE Transactions on Mobile Computing, and IEEE Transactions on Information Forensics and Security.

Feature Topic: Non-Orthogonal Multiple Access (NOMA)

Editor: Daniel Benevides da Costa

Department of Computer Engineering
Federal University of Ceará (UFC)
Sobral-CE-Brazil
danielbcosta@ieee.org

1. Introduction

As the latest member of the multiple access (MA) family, non-orthogonal multiple access (NOMA) has arisen as a promising MA technique for being applied not only in the fifth generation (5G) networks but also in existing and upcoming wireless systems, owing to its full compatibility with other communication technologies. Different from conventional multiple access (OMA), the key idea behind NOMA is to serve multiple users within the same orthogonal resource block (time slot, subcarrier, or spreading code), which provides the possibility to meet the heterogeneous 5G demands of ultra-low latency, ultra-high connectivity, and high throughput, in addition to result in a spectral efficiency gain over OMA.

Indeed, the NOMA concept refers to a general framework in which several recently proposed MA standards can be viewed as a variation of NOMA, such as power-domain NOMA, sparse code multiple access (SCMA), pattern division multiple access (PDMA), and lattice partition multiple access (LPMA). In addition, because of NOMA has shown to be compatible with conventional OMA, it has been recently proposed to be included in the 3GPP LTE-A standard, where it was termed as multiuser superposition coding (MUST). Furthermore, NOMA was recently included in the next digital TV standard, where it was referred to as layered division multiplexing (LDM).

Next, the benefits of NOMA in some emerging communication scenarios are illustrated, which will be followed by some discussions related to the main challenges inherent to its implementation. Finally, the organization of this Feature Topic is outlined.

2. Benefits of NOMA:

As aforementioned, the combination of NOMA with some emerging communication technologies has shown to be promising and has received great attention by the scientific community. In particular, cooperative NOMA arises as a natural choice since in NOMA systems some users know the information of the others (in order to perform successive interference cancelation (SIC)). Compared to conventional opportunistic scheduling schemes, which only serve the users with strong fading conditions, NOMA strikes a good balance between system throughput and user fairness. Specifically, NOMA can serve users with different channel conditions in a timely manner. In addition, the application of NOMA in millimeter wave (mmWave) networks provides an important tool to support massive connectivity as well as to cope with the rapidly growing demands for emerging data services, such as augment reality and virtual reality. It is noteworthy that some features of mmWave propagation (high directional antennas, for instance) facilitate the combination of both 5G technologies. NOMA has also shown promising for implementation in multiple-input multiple-output (MIMO) systems since the spatial degrees of freedom enabled by MIMO are crucial for meeting the performance requirements of 5G networks. Furthermore, the benefits of NOMA for heterogeneous are evident since more users can be served in a small cell by exploiting the NOMA principle. Moreover, NOMA can effectively support massive connectivity and internet of things (IoT) functionality so that its application in machine-to-machine, ultra-dense networks, and massive machine type communications is highly encouraged. NOMA has also arisen as an efficient and flexible MA technique for boosting spectral efficiency in visible light communication (VLC) systems thanks to its specific features.

Overall, the main benefits of NOMA for current and future wireless networks can be summarized as follows:

- Low latency
- High reliability
- Massive connectivity
- Improved fairness
- High throughput.

3. Challenges of NOMA

In order to guarantee that the achievable benefits predicted by theory can be realized in practice, NOMA faces some challenges that need to be carefully addressed before NOMA becomes a part of 5G in the future. Next, several important implementation challenges are briefly discussed.

Imperfect channel state information (CSI) is a key impairment in realizing the performance gain of NOMA in practice because whether an error occurs in SIC, all other user information will likely be decoded erroneously. Another issue regards to the cross-layer optimization. Practical designs of coding and modulation are important to realize the performance gain of NOMA at the physical layer, and it is crucial to study deeply how to feed these gains from the physical layer to the design of upper-layer protocols. Furthermore, since a hybrid MA scheme seems to be very promising in future wireless networks, it is important to study how to combine NOMA with other kinds of MA schemes, including not only the conventional OMA schemes but also the newly developed 5G MA techniques. Another issue is the proposition of low complexity user pairing/clustering approaches, which are effective means to reduce the system complexity since fewer users are coordinated for the implementation of NOMA. It is indeed very challenging to determine how best to dynamically allocate users to a fixed/dynamic number of clusters with different sizes. Most of the MIMO NOMA algorithms presented in the literature have high computational complexity. Therefore, there is an urgent need for research activities into complexity reduction.

Overall, the combination of NOMA and emerging communication technologies is a rich and promising research area. However, further comprehensive studies are needed to develop practical and low complexity schemes, in addition to ensure that the performance gains can be realized in practice.

4. Organization of this Feature Topic

In the next sections, we present one position paper and review two representative works on NOMA. The position paper has been written by a leading expert in the field, Prof. Zhiguo Ding, and deals with compatibility feature of NOMA, examining how it can be used with other multiple access techniques, advanced physical layer technologies, as well as other communication systems. I take this opportunity to take Prof. Zhiguo Ding for his precious contribution to this feature topic.
The first review considers the use of NOMA for enhancing mmWave multicasting, while the second review studies power-efficient resource allocation multicarrier NOMA systems.

Daniel Benevides da Costa (S'04-M'08-SM'14) was born in Fortaleza, Ceará, Brazil, in 1981. He received the B.Sc. degree in Telecommunications from the Military Institute of Engineering (IME), Rio de Janeiro, Brazil, in 2003, and the M.Sc. and Ph.D. degrees in Electrical Engineering, Area: Telecommunications, from the University of Campinas, SP, Brazil, in 2006 and 2008, respectively. His Ph.D thesis was awarded the Best Ph.D. Thesis in Electrical Engineering by the Brazilian Ministry of Education (CAPES) at the 2009 CAPES Thesis Contest. From 2008 to 2009, he was a Postdoctoral Research Fellow with INRS-EMT, University of Quebec, Montreal, QC, Canada. Since 2010, he has been with the Federal University of Ceará, where he is currently an Assistant Professor.

Prof. da Costa is currently Editor of the IEEE Communications Surveys and Tutorials, IEEE Access, IEEE Transactions on Communications, IEEE Transactions on Vehicular Technology, and EURASIP Journal on Wireless Communications and Networking. He has also served as Associate Technical Editor of the IEEE Communications Magazine. From 2012 to 2017, he was Editor of the IEEE Communications Letters. He has served as Guest Editor of several Journal Special Issues. He has been involved on the Organizing Committee of several conferences. He is currently the Latin American Chapters Coordinator of the IEEE Vehicular Technology Society. Also, he acts as a Scientific Consultant of the National Council of Scientific and Technological Development (CNPq), Brazil and he is a Productivity Research Fellow of CNPq.

Prof. da Costa is the recipient of three conference paper awards. He received the Exemplary Reviewer Certificate of the IEEE Wireless Communications Letters in 2013, the Exemplary Reviewer Certificate of the IEEE Communications Letters in 2016, the Certificate of Appreciation of Top Associate Editor for outstanding contributions to IEEE Transactions on Vehicular Technology in 2013, 2015 and 2016, and the Exemplary Editor Award of IEEE Communications Letters in 2016. He is a Distinguished Lecturer of the IEEE Vehicular Technology Society. He is a Senior Member of IEEE, Member of IEEE Communications Society and IEEE Vehicular Technology Society.

Review of: "Non-Orthogonal Multiple Access for Cooperative Multicast Millimeter Wave Wireless Networks", in IEEE JSAC, Aug. 2017

By Z. Zhang, Z. Ma, Y. Xiao, M. Xiao, G. K. Karagiannidis, and P. Fan

Objectives of the paper

The key objective of the paper is to employ non-orthogonal multiple access (NOMA) for enhancing millimeter wave (mmWave) multicasting. From a practical point of view, multicast mmWave wireless networks can provide small cells access for a large number of multicast groups to achieve high-speed and low-latency data transmission. In order to overcome the loss of data transmission rate caused by orthogonal transmission in multicast mmWave networks and increase spectrum efficiency, the application of NOMA in multicast mmWave networks arises as an interesting research topic and this is the aim of the paper. A cooperative NOMA multicasting scheme is also proposed assuming a two-tier mmWave heterogeneous network (HetNet), and several useful insights are highlighted.

Relevance to the feature topic

NOMA and mmWave transmission have been identified as essential enabling technologies for the fifth-generation (5G) wireless networks. Both technologies are motivated by the spectrum scarcity and aim to increase the spectrum efficiency. While mmWave utilizes the mmWave frequency bands, NOMA makes use of the power-domain (or code-domain) to serve multiple users within the same resource block. The use of NOMA in mmWave networks consists a promising research topic since NOMA provides an important tool to support massive connectivity and the internet of things (IoT) functionality of 5G. In addition, the boost of emerging data services will rapidly jeopardize the gains obtained from using mmWave bands and the use of NOMA can effectively improve the spectral efficiency of mmWave transmissions. It is noteworthy that NOMA is well aligned with the features of mmWave transmission and leads to substantially improved system throughput.

Major contributions

The major contribution of the paper is to study the use of NOMA for mmWave multicasting to improve the multicast performance. The paper has developed some protocols and analysis to this important design. The system modeling and performance analysis of mmWave-NOMA multicast are presented by using stochastic geometry. The impacts of data transmission rate and power allocation on the sum multicast rates are investigated by formulating them as optimizations problems and developing some new algorithms. Multicasting is further studied in a two-tier mmWave HetNet, in which a cooperative NOMA multicast is proposed to improve the system performance.

Novelty

The application of NOMA in multicast mmWave networks is the main novelty of this paper. The analysis is conducted by developing a tractable model for performance analysis, which had never been reported in the technical literature until that time. Network optimization on multicast rate transmission rate and NOMA power allocation are discussed.

Key results

In order to evaluate the performance of the proposed scenarios, some numerical results are plotted. The results show that NOMA multicast can achieve a significant gain (in terms of sum rates), compared with conventional multicast. Also, it is shown that the proposed cooperative NOMA multicast can further improve the performance of NOMA multicast.

Outlook

The paper investigates the use of NOMA in multicast mmWave wireless networks. Furthermore, multicasting in a two-tier mmWave HetNet consisting of one low frequency MBS tier and one mmWave small cell tier is studied, in which a cooperative NOMA multicast scheme is proposed.

To provide high quality-of-experience and data rate for 5G multicast services, mmWave communications and NOMA are important enabling technologies. Their joint use can provide high data rate and low-latency multicast services, in addition to support massive connectivity and improved fairness. The approach developed in the paper as well as the attained results can be used as a benchmark for future studies, helping significantly in designing further practical multicast mmWave-NOMA wireless networks.

As future works, it would be interesting to further explore cooperative schemes for improving the performance of NOMA in mmWave HetNets.

Review of: “Optimal Resource Allocation for Power-Efficient MC-NOMA with Imperfect Channel State Information”, in IEEE Trans. Commun, Sept. 2017

By Z. Wei, D. W. K. Ng, J. Yuan, and H.-M. Wang

Objectives of the paper

The key objective of the paper is to study power-efficient resource allocation for multicarrier (MC) non-orthogonal multiple access (NOMA) systems. The proposed framework takes into account the imperfection of channel state information at transmitter and quality of service requirements of users. To strike a balance between system performance and computational complexity, a suboptimal iterative resource allocation algorithm based on difference of convex programming is also proposed.

Relevance to the feature topic

MC-NOMA is a general framework where not only power-domain NOMA but also sparse code multiple access (SCMA) and pattern division multiple access (PDMA) can be included. In addition, MC-NOMA achieves a favourable tradeoff between system performance and complexity. Resource allocation design plays a crucial role in exploiting the potential performance gain of NOMA systems, especially for MC-NOMA systems. However, the imperfect channel state information at transmitter side (CSIT) may cause a resource allocation mismatch, which may degrade the system performance. It is interesting and more practical to design robust resource allocation strategy for MC-NOMA systems taking into account of CSIT imperfectness.

Major contributions

The paper considers a downlink MC-NOMA system with one base station (BS) and M downlink users, as shown in Fig. 1

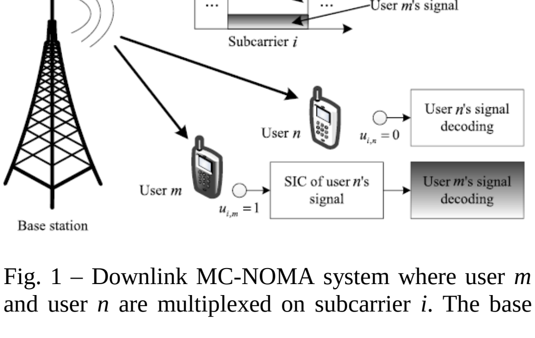


Fig. 1 – Downlink MC-NOMA system where user *m* and user *n* are multiplexed on subcarrier *i*. The base station transmits two superimposed signals with different powers. User *m* is selected to perform successive interference cancellation (SIC).

The major contribution of this paper is to study a power-efficient resource allocation design for downlink MC-NOMA systems under imperfect CSIT, where each user imposes its own quality of service (QoS) requirement. The joint design of power allocation, rate allocation, user scheduling, and SIC decoding policy is formulated as a non-convex optimization problem to minimize the total transmit power. An optimal resource allocation algorithm via the branch-and-bound approach is proposed. In addition, a sub-optimal iterative resource allocation algorithm is presented, which is shown to converge to a close-to-optimal solution rapidly.

Novelty

The main novelty of the paper is to propose the joint design of power allocation, rate allocation, user scheduling, and SIC decoding policy for power-efficient MC-NOMA under imperfect CSIT.

Key results

In order to evaluate the performance of the proposed algorithms, some simulation results are plotted. It is shown that the proposed resource allocation schemes enable significant transmit power saving and are robust against channel uncertainty via exploiting the heterogeneity of channel conditions and QoS requirements of users in MC-NOMA systems.

Outlook

The paper studies resource allocation for power-efficient MC-NOMA systems with imperfect CSIT. It is shown that the proposed schemes provide significant transmit power saving than that of conventional orthogonal multiple access schemes.

Green radio design has become an important research topic in both academia and industry due to the growing demands of energy consumption and the arising environmental concerns around the world. Owing to this fact, power-efficient resource allocation algorithms are important tools for alleviating the power consumption of future wireless networks. The algorithms developed in the paper as well as the attained results can be used as a benchmark for future studies, helping significantly in designing further practical MC-NOMA wireless networks.

In the paper, the analysis focuses on two-user MC-NOMA system. The generalization of the proposed algorithms to the case of serving multiple users on each subcarrier arises as an interesting open problem to be investigated.

Position Paper: On the Compatibility Feature of Non-Orthogonal Multiple Access

Zhiguo Ding[1], Yongxu Zhu[2], and Yan Chen[3]
[1]Lancaster University, UK, [2]Loughborough University, UK, [3]Huawei Technologies, Shanghai

1. Abstract

Non-orthogonal multiple access (NoMA) has recently emerged not only as a new design of multiple access techniques in cellular networks, but also as a general principle of network architecture for applications beyond cellular systems. This paper is to focus on the compatibility feature of NoMA, and examine how it can be used with other multiple access techniques, advanced physical layer technologies, as well as other communication systems other than cellular networks.

2. Introduction

Non-orthogonal multiple access (NoMA) has been recognized as a paradigm shift for the design of multiple access techniques in future wireless networks [1-3]. The essential idea of NoMA is to encourage radio resource sharing, which means that multiple users are served at the same frequency resource blocks and at the same time, instead of allowing a single user to solely occupy one block for a time as in orthogonal multiple access (OMA). It is worth pointing out that the NoMA principle is not new to wireless communications, and it is based on many existing ideas employing non-orthogonality. For example, the key components of NoMA systems, such as superposition coding, successive interference cancellation (SIC) and message passing algorithms, have been previously used for multi-user detection. However, intentionally multiplexing different users at the same time/frequency/code/spatial-direction has not been used in the previous generations of multiple access designs.

This paper is to focus on the compatibility feature of NoMA, which can be used to partially answer the following question: why the NoMA concept can gain huge popularity in the industrial and academic research community in such a short term. In particular, we will illustrate the compatibility of the NoMA principle from three aspects. Firstly, NoMA is compatible to all the other orthogonal multiple access techniques, and hybrid NoMA can be implemented in a straightforward manner without changing the fundamental blocks of other multiple access schemes while bringing additional benefits in improving the multi-user network performance. Secondly, NoMA provides a general network architecture, in which advanced communication techniques, such as flexible duplex, millimeter-wave (mmWave) transmission, multiple-input multiple-output (MIMO), etc., can be straightforwardly used. Finally, we show that the NoMA principle is not just applicable to cellular networks, but also to many other important communication systems.

3. A New Era of Hybrid Multiple Access

Hybrid multiple access is not new to cellular networks at all, and it has been employed in early generations of mobile networks. For example, GSM relies on the combination of time division multiple access (TDMA) and frequency division multiple access (FDMA). Particularly, the principle of TDMA is used to divide one GSM frame into 8 time slots, and hence 8 users can be supported at the same carrier frequency, which obviously is not sufficient to support a large number of users and motivates the use of FDMA in GSM. While in LTE, OFDMA is applied as another example of OMA where the time and frequency resources are gridded into orthogonal lattices and different users are scheduled on non-overlapping lattices exclusively. The future generation wireless networks are also expected to continue using such hybrid multiple access, and NoMA is envisioned to play an important role due to its superior compatibility. Particularly, NoMA can be used to improve spectral efficiency without any need for altering the fundamental resource blocks of other multiple access principles, where the integration of NoMA can be viewed as a smooth software upgrade.

While the principle of NoMA can be integrated with other multiple access techniques, there are various solutions to how such an integration can be done. In this paper, two popular solutions are introduced. *The first type of hybrid NoMA* is that the principle of NoMA, spectrum sharing, is implemented on each orthogonal resource element individually and separately. If the orthogonal resource element is an OFDM subcarrier within a given time unit, this type of hybrid multiple access has also been termed single-tone NoMA. The benefit of single-tone NoMA is simple in concept as well as the transceiver design, since the implementation of NoMA on one resource element is independent to those on the others, i.e., there is no need for joint NoMA encoding or decoding across different resource elements. The most well-known form of single-tone NoMA is power-domain NoMA. Particularly, power-domain NoMA invites multiple users to share the same resource element simultaneously, where the power domain is used for multiplexing, e.g., superposition coding is used at the downlink transmitter by allocating different power levels to users, and SIC is used at the receivers to remove co-channel interference. Cognitive radio inspired NoMA (CR-NoMA) is another important form of single-tone NoMA. The key difference between power-domain NoMA and CR-NoMA is that CR-NoMA recognizes the difference between users' quality of service (QoS) requirements, in addition to the users' channel difference. For example, one of the key features of future wireless networks is the heterogeneous traffic pattern, and such diverse QoS feature becomes more obvious after Internet of Things (IoT) is integrated with cellular networks. Compared to power-domain NoMA , CR-NoMA offers two advantages. One is that the principle of NoMA can be still implemented even if users' channel conditions are similar, since users are ordered accordingly their QoS targets. The other is that users' QoS requirements can be strictly guaranteed.

The second type of hybrid NoMA is to jointly implement the principle of NoMA across multiple orthogonal resource elements, which is thus referred as multi-tone NoMA. In such multi-tone NoMA, some types of joint coding across the multiple resource elements can be introduced to further improve NoMA performance. For instance, block based sequence spreading, scrambling, interleaving, or multi-tone joint modulation mapping can be applied. Joint decoding across multiple resource elements and multiple users is thus needed for the implementation of multi-tone NoMA. Compared to single-tone NoMA, multi-tone NoMA offers better reception reliability and throughputs, but may suffer from increased design complexity at both transmitter and receiver sides. In another aspect, because of the joint design over multiple resource elements, the performance of multi-tone NoMA is further impacted by the way of resource elements mapping or subcarrier allocation, which can also be taken as one dimension to be optimized. Currently, seeking the optimal solution for resource allocation with low complexity for multi-tone NoMA has become an important ongoing research direction.

Despite such increased design complexity, multi-tone NoMA has attracted a lot of attention from the industry, and many recently proposed industrial forms of NoMA are based on multi-tone NoMA. For example, uplink sparse code multiple access (SCMA) is implemented by asking each user jointly encodes its messages sent on multiple subcarriers, and using the message passing algorithm (MPA) at the base station for joint decoding. Regular SCMA has a strict constraint that each user can be allocated the same number of subcarriers, whereas irregular SCMA as well as pattern division multiple access (PDMA) use more relaxed requirements to the number of subcarriers allocated to each user. It is worth pointing out that many industrial forms of multi-tone NoMA are based on the open loop concept, i.e., users' channel information is not used for subcarrier allocation. While this open loop approach reduces the system complexity, the dynamic nature of users' channel conditions has not been used, which means that the performance of these industrial forms of multi-tone NoMA can be further improved by exploiting the users' channel information.

4. Compatibility of NoMA with Other Advanced Physical Layer Designs

Extensive existing studies have demonstrated that the principle of NoMA is compatible not only to other types of multiple access, but also to advanced physical layer techniques to be used in the future wireless networks. Some of these examples are provided in the following:

<u>*mmWave-NoMA:*</u> Both mmWave and NoMA have been recognized as key techniques to combat the spectrum crunch, i.e., there are not sufficient bandwidth for communications, although the solutions provided by the two techniques are

different. A common question from the research community is why to use NoMA for mmWave networks when there is plenty of bandwidth available at mmWave bands. This question can be answered by using the following example. In 1990s, when a movie is stored in a computer by using the VCD (MPEG-1) format, the size of such a file is around 200-500MB. When this was replaced by the DVD format, the size of a movie file is expanded to 2-3GB, and the size of a typical Blu-Ray file can be 20GB. Human become more and more demanding to the resolution and details, which means that the amount of information to be sent in the future wireless networks will also become larger and larger. Therefore, the gains we get from mmWave bands can soon hit its ceiling, and how to efficiently use the bandwidth from the mmWave bands will become a critical issue, which motivates the use of NoMA in mmWave networks.

Furthermore, existing studies of mmWave-NoMA have revealed that mmWave transmission exhibits some features which are ideal for the application of NoMA. For example, users in mmWave networks can have strongly correlated channels, even if the antennas of these users are separated much larger than half of the signal wavelength. While such correlation has been conventionally recognized as a harmful effect, the use of the quasi-degradation criterion reveals that this correlation results in an ideal situation for the application of NoMA. Another example is hardware impairments and limitations, e.g., the use of finite resolution analog beamforming, can also bring the opportunity for the integration of NoMA in mmWave networks.

MIMO-NoMA: During the last two decades, MIMO has been continuously in the spotlight of the communication research and industrial activities, mainly due to its superior spectral efficiency, i.e., high data rates can be supported without using extra spectrum bandwidth but by exploring the spatial domain. At a certain stage of the development of NoMA, there was confusion about the difference between MIMO and NoMA. The reason for this confusion is that using MIMO, we can also accommodate multiple users at the same spectrum at the same time, and hence yield the same non-orthogonality as NoMA. Actually, many convectional MIMO techniques aim to use the spatial domain and create spatially orthogonal channels between users, in order to avoid co-channel interference. Zero forcing and singular-value-decomposition based designs are typical examples to illustrate this orthogonal principle. On the other hand, the use of NoMA is to assume that multiple users share the same orthogonal resource unit, where one spatially orthogonal channel is just another example of such a resource unit. Or in other words, conventional MIMO allows users to use the same bandwidth, but tries to create multiple orthogonal spatial directions to differentiate multiple users, whereas NoMA further supports multiple users to share the same spatially orthogonal direction.

In the context of massive MIMO, there was a concern about the feasibility for the implementation of NoMA. The rationale behind this concern is that the quasi-degradation criterion reveals that NoMA is not preferable if users' channel vectors are orthogonal to each other, but in massive MIMO, users' channel vectors are asymptotically orthogonal. However, some existing studies have demonstrated that the use of NoMA is still important to massive MIMO, where the reason is that users' channels are not completely orthogonal in a practical scenario, because of channel correlation. For example, when implementing massive MIMO at a base station, most likely this base station will be amounted at a top of a high building, without many scatters around. As a result, users from one room in this building can have highly correlated channels, instead of orthogonal channels. As discussed in the mmWave-NoMA part, the correlation among users' channels does facilitate the implementation of NoMA. Moreover, it is very costly to get accurate channel state information for massive MIMO scenarios. In the case that the channel state information is not perfect due to limited feedback quantization, channel measurement latency, or user mobility, NoMA can help improve the system performance which will otherwise be degraded significantly.

Cooperative NoMA: The importance of cooperative diversity can be easily spotted by the fact that the paper by Laneman, Tse and Wornell has already attracted 12000+ citations, probably one of the most cited papers in the last two decades in communications. The use of cooperative transmission is important to NoMA since users with poor channel conditions in current NoMA can potentially suffer some performance loss, compared to the case with OMA. By using cooperative NoMA, the reception reliability of these users can be improved. Most existing designs of cooperative NoMA can be grouped into two categories. One category is to employ NoMA users with strong channel conditions as relays to help the other users, which is also known as user cooperation. The benefit of such cooperative NoMA is that the redundant structure of NoMA can be efficiently exploited. Particularly, these so-call strong users need to decode the messages to the users with poor channel conditions in order to decode their own information, and hence they are natural relays to help those weak users.

The use of dedicated relays is another important category of cooperative NoMA. In many communication scenarios, the number of mobile devices is large, but many of them are not active in transmitting or receiving. Therefore, these idle users can be used as relays to help the active users, and the number of such relays can be quite large in practice, which is the advantage of the second category of cooperative NoMA. Given the existence of multiple relays, distributed beamforming can be designed in order to efficiently utilize the spatial degrees of freedom offered by the dedicated relays, but the system overhead caused by the coordination among the relays needs to be carefully suppressed, particularly for the scenario with a large number of relays. A low-complexity alternative to distributed beamforming is relay selection, and recent studies reveal an interesting fact that relay selection in cooperative NoMA can be fundamentally different to that in conventional cooperative networks. For example, the max-min criterion which has been shown optimal in conventional cooperative networks is no longer optimal in cooperative NoMA. This explains why relay selection becomes a quite popular area in cooperative NoMA.

Network NoMA: Network MIMO has recently received a lot of attention, since the boundaries of cells are removed and base stations from different cells are encouraged to cooperate each other. There have been different forms of network MIMO, from distributed CoMP to jointly scheduled C-RAN. While the benefit of the NoMA principle in a single cell setup has been well recognized, its benefit to the multi-cell scenarios, such as CoMP and C-RAN, has not been fully exploited and investigated. Hence recently, a lot of efforts of the NoMA research community have been has been devoted to thoroughly examine the impact of NoMA on multi-cell scenarios, and the resulting novel designs of NoMA can be viewed as special cases of network NoMA, and these network NoMA designs clearly demonstrate that it is beneficial to use the NoMA principle in such network MIMO scenarios [4]. In particular, given its non-orthogonal nature, NoMA is expected to improve the CoMP transmission by relaxing the requirement of accurate time alignment between different transmit points and joint channel state information for precise beamforming, which prohibit the boom of CoMP applications in practical systems.

Without loss of general, take CoMP as an example, which is to ask multiple base stations to jointly serve a user which is at the cell boundaries. While such a design indeed helps the edge user, these base stations have to serve this user solely for a given bandwidth and time slot, which reduces the spectral efficiency since this user has poor connections to the base stations. After network NoMA is used, each base station can serve a near user while performing CoMP and helping the edge user. As a result, the overall system throughput as well as connectivity can be significantly improved. Advanced power allocation policies can be used to ensure that those near users are admitted without sacrificing the performance of the edge user. In the heterogeneous networks, the NoMA principle has also been shown to be very useful to improve the spectral efficiency as well as coverage, where not only more users can be served in each tier but also the cooperation among different tiers can be enabled.

5. Integrating NoMA into Systems Beyond Cellular Communications

The superior compatibility of the NoMA principle can be also demonstrated by the fact that NoMA has found a lot of applications beyond cellular communications, as illustrated in the following:

Wi-Fi Networks: While the concept of NoMA has been investigated for cellular systems, it can be straightforwardly applied to the next generation Wi-Fi systems. Conventional Wi-Fi networks still rely on orthogonal resource allocation. This leads to a difficult situation that some users cannot be admitted after all the limited orthogonal resource blocks are taken by other users. By applying the NoMA principle, more users can be simultaneously admitted, which is particularly important for the deployment of Wi-Fi in crowded areas, such as airports or sport stadiums.

VLC Systems: VLC has been recognized as an efficient method for the last mile connection in future communication networks. However, one disadvantage of VLC is that the use of narrow-band modulation and non-coherent detection limits the number of users which can be supported. As a result, the NoMA principle is naturally compatible to VLC, and its application ensures that VLC can be used not only for small-scale smart homes, but also large-scale scenarios, such as lecture halls.

TV Broadcasting: Digital TV broadcasting is another important application of NoMA. It is worth pointing out that the concept of NoMA has already been included into the next generation digital TV standard (ATSC 3.0), where it is termed Layered Division Multiplex (LDM). Particularly, a TV station will integrate several layers of video streams with different QoS requirements, and this superposition will be broadcasted to users. Each user will decode certain layers of the video streams according to its channel conditions.

Wireless Caching: The key idea of wireless caching is to proactively push content files to local caching infrastructure, before they are requested. As a consequence, users can fetch these files from their local caching infrastructure, without being directly served by the network controller. Existing studies have demonstrated that the NoMA concept not only helps to push the content files to local caching infrastructure timely and reliably, but also improves the spectral efficiency of content delivery from the caching infrastructure to the users.

Internet of Things (IoT): One key feature of IoT is the diverse traffic patterns of IoT devices. Particularly, some devices have demanding bandwidth requirements, e.g., environmental mentioning cameras deliver high-resolution images/videos, whereas the others need to be served with low data rates but timely, e.g., vehicles receive incident warning messages in intelligent transportation systems. NoMA can be naturally applied to handle such a challenging situation, by integrating devices with heterogeneous QoS requirements at the same bandwidth. Furthermore, recent studies have also demonstrated that the combination of encoding with finite block length and NoMA can be a promising solution for supporting IoT as well as ultra reliability and low latency communications (uRLLC). In particular, for contention based grant-free transmission, NoMA is the key solution to enable reliable communications while supporting massive connectivity.

6. Conclusions

The aim of this paper is to focus on the compatibility feature of NoMA. In particular, we have first examined how NoMA can be combined with other multiple access techniques. Then the compatibility between NoMA and various advanced physical layer techniques has been illustrated, and the applications of NoMA to communication systems beyond cellular networks have also been discussed.

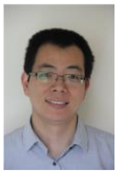
References

[1] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan and V. K. Bhargava, "A Survey on Non-Orthogonal Multiple Access for 5G Networks: Research Challenges and Future Trends," in IEEE J. Sel. Areas Commun., vol. 35, no. 10, pp. 2181-2195, Oct. 2017.

[2] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C-L. I and H. V. Poor, “Application of Non-orthogonal Multiple Access in LTE and 5G Networks”, IEEE Commun. Mag., 2017.

[3] L. Song, Y. Li, Z. Ding, and H. V. Poor, “Resource management in nonorthogonal multiple access networks for 5G and beyond”, IEEE Networks, 2017.

[4] Usa Vilaipornsawai, Hosein Nikopour, Alireza Bayesteh, and Jianglei Ma, “SCMA for Open-Loop Joint Transmission CoMP,” in Proc. IEEE VTC fall, 2015.



Dr. Zhiguo Ding is currently a Chair Professor in Signal Processing for Communications at School of Computing and Communications, Lancaster University, the UK.

Dr. Yongxu Zhu is currently a research associate at the Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, the UK.

Dr. Yan Chen is currently with Huawei Technologies Co., Ltd., Shanghai, as a Technical Leader on Huawei 5G air interface design.

Feature topic: Millimeter Wave Communications: Overview, Opportunities, and Challenges

Editor: Omid Semiari

Department of Electrical Engineering,
Georgia Southern University, Statesboro, GA, USA
Email: osemiari@georgiasouthern.edu

1. Introduction

Operating at high-frequency millimeter wave (mmWave) bands, spanning from 30 GHz to 300 GHz, is seen as a key enabler to address the capacity demands of the fifth generation (5G) wireless networks. MmWave technology has attracted substantial attention during past decades, mainly due to the availability of large bandwidth at the mmWave frequency bands as compared with the strained sub-6 GHz frequencies. Once properly deployed, mmWave technology can provide promising data rates, sufficient enough to support many of the emerging wireless services such as wireless virtual reality (VR) and Vehicle-to-Everything (V2X) networks.

Today, mmWave technology is deployed under different wireless local area network (WLAN) standards, including IEEE 802.11ad, to support ultra-high-definition video streaming to indoor wireless devices. In addition, mobile network operators have initiated channel sounding measurements and early deployments of mmWave communications for enhanced mobile broadband (eMBB) applications. These initiatives are coupled with recent advancements in silicon industry to manufacture low-cost mmWave chips for vast utilization in consumer devices. As such, it may no longer be ambitious to view mmWave technology as an integral element of 5G networks.

2. Applications

International Mobile Telecommunications (IMT) and the 3rd Generation Partnership Project (3GPP) have defined different use cases for 5G services, including: 1) eMBB; 2) Ultra-reliable and low-latency communications (URLLC); and 3) Massive machine-type communications. For these performance indicators (KPIs), mmWave technology can play a key role to achieve the target quality-of-service (QoS) metrics such as 20 Gbps data rate or less than 1 millisecond latency.

In fact, operating at mmWave frequencies allows to deploy small-size antenna arrays with large number of antenna element both at base stations and user equipment. Therefore, multi-antenna communication techniques can be used to achieve beamforming gains – to extend the communication range, or multiplex multiple stream of data to different users at the same time-frequency resource via spatial multiplexing.

Another key advantage of mmWave communication is that the available spectrum can be aggressively reused, since directional transmissions along with susceptibility of mmWave signals to blockage can substantially limit the interference across mmWave links. Furthermore, the large bandwidth can be used for in-band self-backhauling in small cells to provide wireless backhaul connectivity and reduce the cost of small cell deployment in 5G networks.

Nonetheless, despite these unique features, several main challenges must be addressed in order to reap the full benefits of mmWave technology

3. Challenges

One of main challenges of mmWave technology is the large path loss at mmWave frequency bands that mandates directional transmissions to achieve a reasonable transmission range. However, directionality of mmWave links introduce new challenges such as deafness (especially in ad-hoc deployments), the need for performing beam-training, and managing directional transmissions to multiple users. As such, mmWave communication will naturally require more overhead when compared with omni-directional control channels at sub-6 GHz frequencies. Such complexity scales with the number of devices and becomes a limiting factor specially for massive machine-type communications.

Additionally, mmWave signals are highly susceptible to blockage, because of the poor capability to penetrate obstacles which makes mmWave links intermittent. Therefore, it is challenging to support the required reliability and latency in URLLC over the mmWave frequencies. Finally, the Doppler effect is more severe at the high-frequency mmWave frequencies, thus, new radio interface is required to cope with fast channel variations.

In addition to these limitations that mainly stem from the characteristics of mmWave signals, addressing other challenges such as power consumption and digital/hybrid beamforming complexity, in particular at the user equipment, is central for vast proliferation of mmWave technology in 5G networks. In the subsequent section, we overview two recently published research papers that address some of the aforementioned mmWave challenges.

Omid Semiari is an Assistant Professor at the Electrical Engineering Department at Georgia Southern University. He received the BSc and MSc degrees in electrical engineering from the University of Tehran, in 2010 and 2012, respectively, and the PhD degree from Virginia Tech, in 2017. His research interests include wireless networks, millimeter wave communications, context-aware resource allocation, matching theory, and machine learning. In 2014, Dr. Semiari has worked as an intern at Bell Labs, in Stuttgart, on anticipatory, context-aware resource management in cellular networks. In 2016, he has joined Qualcomm CDMA Technologies for a summer internship, working on LTE-Advanced modem design. Dr. Semiari is the recipient of several research fellowship awards, including DAAD (German Academic Exchange Service) scholarship and NSF student travel grant. He has actively served as a reviewer for flagship IEEE Transactions and conferences and participated as the technical program committee (TPC) member for a variety of workshops at IEEE conferences, such as ICC and GLOBECOM.

Review of: "Achieving Ultra-Low Latency in 5G Millimeter Wave Cellular Networks" IEEE Communications Magazine, Volume: 55, Issue: 3, March 2017

by Russell Ford, Menglei Zhang, Marco Mezzavilla, Sourjya Dutta, Sundeep Rangan, and Michele Zorzi

1. Objectives

This work overviews some of the main challenges for achieving ultra-low latency and high reliability in millimeter wave (mmWave) communications. In addition, different potential solutions are introduced with regard to the medium access control (MAC) layer design, congestion control schemes, and enabling core network architectures to achieve the target performance in emerging mmWave cellular networks.

2. Relevance to the feature topic

Communications over the high-frequency mmWave bands is seen as an attractive solution to meet the stringent quality-of-service (QoS) requirements – 100 Mbps cell edge and 20 Gbps peak data rate – of emerging wireless networks determined by the International Mobile Telecommunications (IMT) 2020.

In addition to wireless services with high data rate requirements, various emerging applications will mandate extremely low latency and high reliability communications. Among these wireless services include immersive virtual reality, augmented reality, telesurgery, and real-time cloud/fog computing, that may necessitate less than 10 ms end-to-end (E2E) latencies. Other emerging use cases such as mission-critical machine-type communication (MTC) and autonomous vehicles will require less than 1 ms E2E latency.

Meeting these stringent latency constraints clearly falls beyond the capacity of existing cellular networks that operate at the congested sub-6 GHz frequencies. Leveraging the large available spectrum at the mmWave frequencies can reduce the over-the-air latency. Therefore, mmWave technology is viewed as an attractive solution, not only to achieve extremely high data rates, but also to enable delay intolerant applications.

3. Contributions and novelty

This work focuses on introducing key challenges of achieving ultra-low latency and proposing candidate mmWave solutions. Although mmWave communications can reduce the latency at the physical (PHY) layer, a fundamental paradigm shift from existing cellular architectures is needed to achieve the target E2E performance. To this end, the paper studies the challenges and potential solutions within three dimensions, namely: 1) Low-latency core network architecture; 2) Flexible MAC layer design; and 3) Congestion control.

The first challenge elaborated in this work is with regard to the core network architecture in existing cellular networks which is not suitable to support low-latency communications. This architecture, known as the Evolved Packet Core (EPC), is characterized by a small number of high-capacity, high-reliability network elements (NEs), such as serving gateways (SGWs), packet data network gateways (PGWs), and mobility management entities (MMEs), which rely on expensive application-specific hardware. These complex and expensive NEs are located in geographically dispersed areas, and thus, the existing core network architecture is highly centralized. Nonetheless, the paper clarifies why such a centralized architecture cannot meet the stringent latency requirements of emerging wireless services.

Another challenge to achieve the low-latency communications in mmWave cellular networks is the need for directional transmissions in both data and control planes. To date, most transceivers use analog beamforming which allows directional transmission to only one user at the time. Therefore, time-division multiple access (TDMA) is typically considered to schedule the users. However, TDMA is inefficient scheme to utilize the mmWave spectral resources, even if very short frames are employed. The existing solutions consider the decoupling of control and data planes to manage the control signaling over the omni-directional transmissions at sub-6 GHz frequencies, while mmWave spectrum is used to boost the capacity for data transmissions. Despite the importance of mmWave coexistence with sub-6 GHz communications in cellular network, the paper argues that decoupling control and data planes over different frequency bands may not be necessarily the only solution. In fact, some of the existing works show that more advanced transceivers with digital/hybrid beamforming can manage both control and data communications over the mmWave frequencies.

The last challenge that has been discussed in this paper is the latency associated with transport layer. Due to the intermittency of mmWave links, the stochastic link quality variations – resulting from different factors such as blockage – can result in a backlogged data at the downlink PDCP queue of the base station. That is because the existing TCP schemes are not fast enough to adapt the packet transmission rate when unexpected deep fade (blockage) occurs. Hence, the backlogged queue at the BS can result in an excessive latency.

The main contributions of this work are 1) Surveying main challenges in achieving low-latency communications over the mmWave frequency bands; and 2) Introducing enabling techniques to address these challenges. The paper categorizes the study into three parts, associated with the core network architecture, multiple access and the MAC layer design, and the congestion control.

To address the challenges of latency at the core network, one potential solution that has been proposed in this work is to bring the NEs closer to the edge of network by utilizing techniques from software-defined networking (SDN) and network functions virtualization (NFV). Furthermore, content and application servers can be moved to the edge of the network to further reduce the overhead and latency in the core network. The paper also discusses the potentials of distributed mobility management across virtual MMEs to manage the substantial traffic growth associated with handover in densified future cellular networks.

At the MAC layer, this paper advocates orthogonal frequency-division multiplexing (OFDM)-based systems that features three unique property compared with frame structure in existing cellular networks: 1) Short symbol periods (in order of few microseconds); 2) Flexible transmission time interval (TTI); and 3) low-power digital beamforming for managing control signals while using analog beamforming for data transmissions.

4. Outlook

The mmWave technology can be a promising solution to achieve extremely high data rates and low latency in cellular communications. Reaping the potentials of mmWave technology requires significant changes at different layers of the exiting protocol stack. In particular, major modifications are required to enable 1) flexible and scalable architecture in the core network to minimize the latency; 2) flexible MAC layer designs to achieve scheduling with minimum latency; and 3) efficient and fast congestion control algorithms to manage the stochastic and intermittent nature of the mmWave channel. Therefore, in addition to the prior art works, novel solutions must be introduced at different levels to guarantee the target E2E latency in future mmWave cellular networks.

Review of “Device-to-Device Millimeter Wave Communications: Interference, Coverage, Rate, and Finite Topologies” IEEE Transactions on Wireless Communications, Volume: 15, Issue: 9, Sept. 2016

by Kiran Venugopal, Matthew C. Valenti, and Robert W. Heath, Jr.

1. Objectives

This papers studies the opportunities for leveraging millimeter wave (mmWave) communication, as one of the key enabler for achieving multi-Gbps data rates for device-to-device (D2D) communications. This work provides comprehensive analysis of mmWave D2D communications, particularly for wearable devices in indoor environments, such as inside vehicles or airplane cabin, where the effect of blockage by human body can be severe. To this end, performance analysis has been done using the framework of stochastic geometry to analyze key features such as data rate and coverage.

2. Relevance to the feature topic

Wearable devices, such as smart watches, are becoming one of the integral component of everyday life. These devices collect/generate a large amount of data (from the human body or environment) and often require to send this information to one another or to a cloud. Therefore, a seamless wireless connectivity is essential to turn many of services offered at wearable devices into reality.

In this regard, leveraging mmWave communications is promising due to the following features: 1) MmWave frequency bands offer a substantially large bandwidth that can be exploited to support the required multi-Gbps data rates in D2D networks; 2) mmWave communication is suitable for directional transmissions which can limit the interference, especially in indoor environments with many users/wearable devices; 3) the mmWave RF components are small in size (due to operating at high frequency) and can be easily deployed in wearable devices.

3. Contributions and Novelty

The key challenge to analyze the performance of the mmWave D2D network is to derive trackable analytical expressions for the coverage and data rate, while considering characteristics of mmWave communications such as the directionality, antenna gain pattern, and blockage. For analytical tractability, existing works typically assume an infinite number of mobile devices spread over an infinite area. These assumptions allow to simplify the analysis (mainly by using the Campbell’s theorem) and derive analytical expressions related to the spatially averaged system performance. However, such assumptions may not be applied to an indoor environment with a finite number of devices.

Therefore, one of the contributions of this work is to analyze the performance of the mmWave D2D network under more realistic assumptions with finite area and number of interfering devices. This papers considers the human body as the main cause of blockage for mmWave communication among wearable devices and under certain assumptions, derives the spatially averaged coverage and data rate. To model the blockage in presence of human bodies in the network, different path loss and fading models are considered, depending on whether a link is blocked or not.

Comprehensive simulations are then provided to validate the assumptions and analysis. Simulations include repeated random placement of the users according to the modeled point process to find the spatial average of the system performance.

4. Key results

This paper analysis the coverage and data rate of mmWave D2D networks, particularly for communication among wearable devices in presence of human body blockage. In this regard, the impact of interference and blockage are analyzed with respect to the network size. Moreover, the results show that gain and directivity of the antenna’s main lobe can significantly impact the achievable data rate in mmWave D2D networks.

As an example, Fig. 1 shows the impact of different antenna configurations on the spatially averaged SINR coverage probability. In Fig. 1, N_t and N_r denote, respectively, the number of antenna elements at the transmitter and receiver. The results in Fig. 1 show that larger N_t is beneficial to improve the coverage. It is also interesting to note that the effects of the N_t and N_r on the performance is not symmetric.

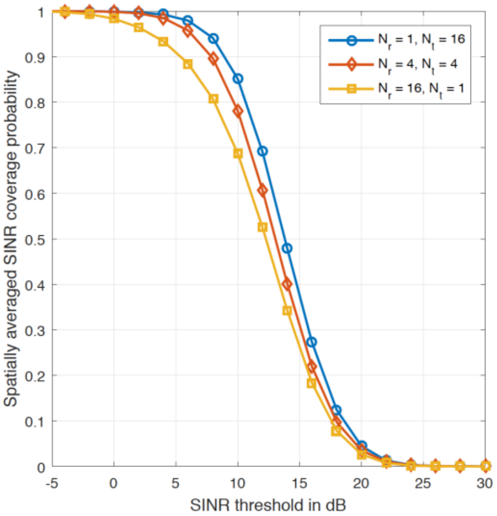


Figure 1- Effect of antenna configuration on the performance

The simulation results corroborate the analytical derivations and serve as an initial step in characterizing the performance of mmWave D2D networks via a stochastic geometry famework with a finite space and network size.

5. Outlook

The results presented in this work shows the capabilities of mmWave communications for achieving multi-Gbps data rate in D2D networks that involve wearable devices. The analytical derivations provide a tractable framework to design different system parameters without requiring to perform extensive simulations. The presented results in this work can be extended by developing a new model that can incorporate 3D locations for the devices. In addition, it is imperative to explicitly model the reflections of the mmWave signals from the boundaries of the finite network region. It is also interesting to extend the blockage model to consider the shadowing effect by user’s own body.

Feature Theme: “Spectrum Scarcity”

By Daniel Benevides da Costa and Omid Semiari

Over the last decades, wireless communications have undoubtedly undergone an explosive growth with myriad applications becoming an inevitable part of human life. In this view, broadband wireless technologies have enabled efficient voice and data communications and more lately a rather satisfactory mobile internet access. It is widely known that such telecommunication services require the transmission of high information content and thus high data rates. For example, wireless voice and video applications typically require data rates of several tens and hundreds million bits per second (Mbit/s), respectively. According to the foundations of communication theory, the data rate at which information can be wirelessly transmitted is directly related to the signal-to-noise-ratio (SNR) of the corresponding information signal and the channel bandwidth. As a result, the rapidly increasing demand for high performance broadband communications led to the development of numerous sophisticated technologies which are sought to meet the corresponding data rate requirements.

Nevertheless, these achievements were mainly accomplished at the expense of requiring substantially increased bandwidth (Hz) and energy (joules) resources. As a result, the corresponding engineering complexity and financial cost increased dramatically while a significant scarcity of the available spectrum resources became a notable burden. This motivated both industrial and academic sectors to focus their efforts in providing meaningful solutions to the aforementioned issues.

To date, many enabling techniques have been proposed to optimize the spectrum utilization. These approaches follow two main directions to 1) increase the spectral efficiency at the available spectrum (typical below 6 GHz), and 2) harnessing new frequency bands at centimeter and millimeter wave (mmWave) spectrum (above 6 GHz). Among the promising concepts for addressing the spectrum scarcity below 6 GHz include Multiple-Input Multiple-Output (MIMO) communications and more recently massive MIMO, device-to-device communications, and dynamic spectrum sharing schemes. Furthermore, substantial effort has been done to characterize the communications over the high-frequency mmWave bands to harness the large available bandwidth (which can be in order of multi GHz) to boost the capacity of wireless networks.

Nonetheless, despite the large body of research work that has been carried out so far, spectrum scarcity is still an ongoing challenge as new wireless services – such as virtual reality and connected autonomous vehicles – are emerging that impose new stringent quality-of-service requirements. As such, supporting the emerging wireless services may require novel solutions for: 1) further optimizing the spectral efficiency at the sub-6 GHz frequency bands; 2) enabling mmWave communications while providing a seamless integration with legacy sub-6 GHz wireless systems; 3) providing new models for enabling closer collaboration between academia and industry; 4) advancing the existing theoretical models and proposing proof-of-concept methods via new experimental research and development; 5) developing novel business-technical models to enable flexible and efficient spectrum utilization/sharing among operators and vendors.

For this newsletter, two position papers on the theme "Spectrum Scarcity" are presented. The first paper, entitled "Low-Cost Cognitive Radio Against Spectrum Scarcity" discusses the impact of hardware impairments in the spectrum sensing performance of low-cost cognitive radio.

Both the academia and the industry have several concerns regarding the limitations caused by the hardware imperfections and the self-interference leakage; especially in high-data rate systems, and the proposed study provides good insights for the technical literature. The second paper, entitled "5G Unlicensed Spectrum Access for Massive Machine Type Communications Enabled by Distributed Wideband Spectrum Sensing" presents a spectrum sensing based approach to enable massive machine-type communications over fifth-generation new radio (5G-NR) using unlicensed spectrum access. Some challenges, requirements, and key design components for distributed wideband spectrum sensing are presented and discussed. Furthermore, an interview is conducted with Drs. Sayeed and Debbah, two of the key experts in the field whom we asked about present and future research for addressing the challenge of spectrum scarcity.

The goal of the theme on "Spectrum Scarcity" is to provide an abstract, yet constructive and coherent overview of this timely research trend, especially to highlight the ongoing and potential research directions. We hope that this theme be of interest to the community, especially for those who are new to this area of research.

Finally, we would like to take the opportunity to thank all the contributors to this theme.

Position Paper: “5G Unlicensed Spectrum Access for Massive Machine Type Communications Enabled by Distributed Wideband Spectrum Sensing”

Ghaith Hattab and Danijela Cabric
University of California, Los Angeles, USA
ghattab@ucla.edu, danijela@ee.ucla.edu

1. Abstract

This position paper presents a spectrum sensing based approach to enable massive machine-type communications over fifth-generation new radio (5G-NR) using unlicensed spectrum access. We propose to enhance the cellular network architecture with a distributed wideband spectrum sensing system that provides fine spatial and frequency reuse of spectrum. We discuss challenges, requirements, and key design components for distributed wideband spectrum sensing and propose novel algorithmic framework for its implementation.

2. Introduction

Cellular networks are on the verge of breaking the traditional wireless access, which is limited to licensed channels, by exploring spectral opportunities over the unlicensed bands sub-6GHz. Indeed, the latest LTE Rel. 13 and Rel. 14 have introduced, respectively, *Licensed-Assisted Access (LAA)* and *enhanced-LAA* (eLAA) modes that allow an unlicensed carrier, at 5GHz, to be aggregated with a licensed one [1]. These modes are set to evolve with 5G-NR, as 3GPP has recently approved a study item of 5G-NR in the unlicensed spectrum without mandating a licensed anchor, i.e., a stand-alone unlicensed access [2].

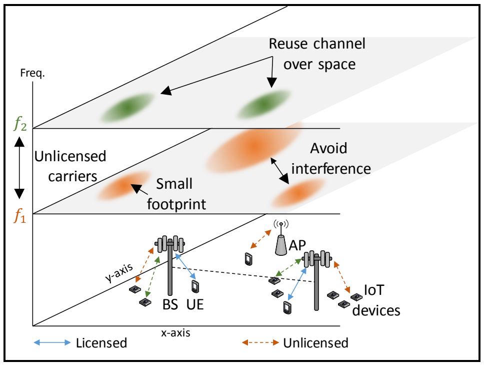


Fig. 1. Connecting a large number of machines over cellular networks requires aggressive reuse of unlicensed channels.

While enhanced mobile broadband is a key feature of 5G-NR, another critical use case is the native support of a massive number of machines and sensors also known as *massive IoT* [3]. The proposal to use unlicensed spectrum for IoT traffic is not new, e.g., LoRa networks [4], but existing solutions do not use cognition in spectrum access. They access fixed narrow spectrum bands, and rely on spread spectrum techniques for co-existence, but these do not scale well when the number of devices is high [4]. We argue that in order to connect a massive number of IoT devices over cellular networks, *these networks must aggressively reuse the unlicensed bands over space and frequency*. This requirement follows because many applications, e.g., deploying sensors across a city for utility metering and transportation management, have low-rate requirements [3], and thus instead of using a 20MHz channel at 5GHz for one device, the same channel can be divided into carriers, each of bandwidth, say 180KHz, serving in this case more than 100 devices. Second, these sensors transmit at low power, and thus they can be clustered over space, where each cluster occupies a small spatial footprint, motivating a reuse of the same channel across clusters. An illustrative example of the envisioned cellular network is shown in Fig. 1. The base station (BS) provides licensed access to UEs and unlicensed access to IoT devices. The BSs dynamically search for unoccupied unlicensed carriers to avoid interference from existing access points (APs), and the same channel can be reused over

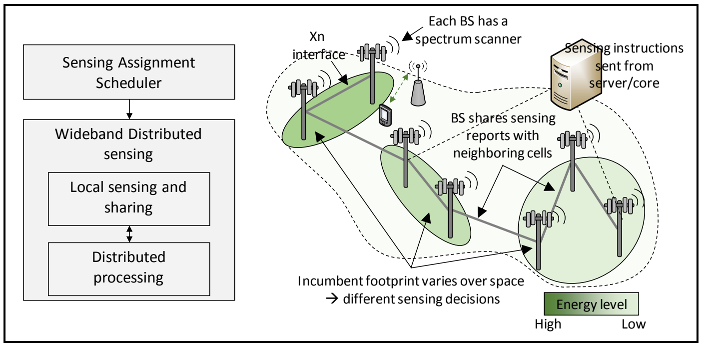


Fig. 2. The cellular network is enhanced by a sensing framework that consists of two key components: A sensing assignment scheduler and a wideband distributed sensing framework.

space to serve other IoT devices.

3. Challenges of Unlicensed Massive IoT Communications

The first step to unlicensed access for massive IoT is identifying channels with lower activity to minimize interference. To this end, spectrum sensing becomes central to the envisioned cellular architecture. Indeed, 5G-NR is expected to continue with the listen-before-talk (LBT) mechanisms, introduced in LAA, which are essentially simple sensing algorithms.

There exists a plethora of spectrum sensing algorithms, and the advantages and limitations of each have been thoroughly investigated since the inception of dynamic spectrum access over the TV white space [5]. However, massive IoT brings distinct challenges due to the large number of devices and their energy-constrained hardware. First, IoT devices transmit at low power, with a short communication range. Thus, it is imperative to develop sensing algorithms with high spatial resolution, to exploit the short transmission range and cluster devices to aggressively reuse the same channel over space. Such resolution can be achieved by allowing spectrum scanner to make a local decision about the occupancy of a channel. It is well known that single radio spectrum sensing is unreliable in fading channels [5]. Using centralized cooperative sensing, e.g., scanners reporting to a core network, improves sensing accuracy but loses the spatial resolution. This elevates the need of a truly distributed implementation of the sensing algorithm, where measurements and decisions are shared and processed by scanners in local proximity. Second, since the number of IoT devices is large, and their rate requirements are low, it is imperative to identify lightly loaded unlicensed carriers with fine frequency resolution, in the order of few tens to hundreds of KHz, over a wideband spectrum, e.g. 1GHz. Thus, deploying wideband spectrum scanners at each BS becomes necessary.

4. Proposed Architecture

To exploit the expected high density of small-cell BSs, e.g., femto cells, in 5G networks, the proposed architecture envisions equipping each BS with a spectrum scanner to enable fine spatial-sensing resolution. To improve sensing reliability, the architecture requires neighboring BSs to cooperate and share their processed sensing measurements. However, different from centralized cooperative processing, each BS may converge to a different decision, depending on the proximity of incumbent transmitters to each BS. Fig. 2 illustrates a high-level overview of the proposed *cognitive* cellular network architecture, which consists of two key components: 1) sensing assignment scheduler and 2) wideband spatio-spectral distributed sensing.

The sensing assignment scheduler is required to reduce the sensing burden on each BS. Specifically, it is desired to identify many spectral opportunities over a 1GHz of spectrum at 5GHz. Yet, instead of sensing $M \gg 1$ (e.g. M=5,000) channels at each BS, the scheduler at the core network assigns each BS a set of $P \ll M$ (e.g. P~100s) channels to sense. However, each BS will have missing information about the remaining M-P channels. Thus, the scheduler seeks to optimize the subset of sensing channels under constraints on the accuracy of the sensing decisions. The second component of the architecture is the distributed spatio-spectral sensing framework, which requires each BS to locally sense the assigned channels, process the sensing data, and share it with BSs in vicinity for further distributed processing. The framework will use the Xn interface, i.e., the standard interface that connects to 5G BSs [6], for sharing sensing data and decisions.

The sensing framework aims to overcome the limitations of the centralized (cooperative) and non-cooperative sensing frameworks. Specifically, the centralized one maximizes the sensing reliability and resolution over frequency, yet it arrives at a single decision, limiting its spatial resolution. In contrast, the non-cooperative framework maximizes the spatial resolution as each BS individually makes a decision, irrespective of other BSs. This, however, is unreliable in fading channels and can be costly due to the high sampling and power requirements needed for the wideband scanner. Fig. 3. summarizes the main properties of the different sensing frameworks, highlighting that the distributed one aims to balance the trade-offs between the spectral resolution and the spatial resolution. Next, we elaborate on the components of the proposed architecture.

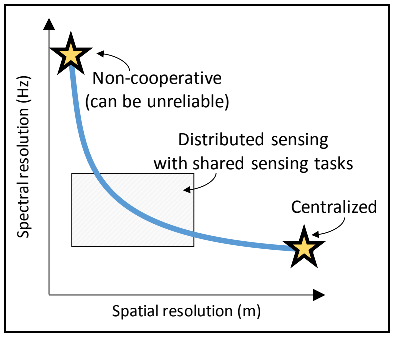


Fig. 3. Distributing the sensing tasks and sensing measurements locally can provide a better balance between the spatial and spectral resolutions.

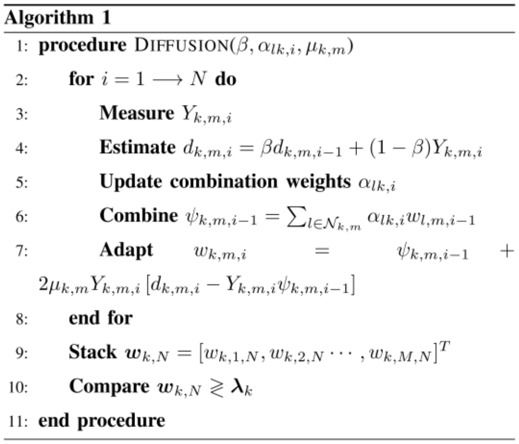
4.a) The Sensing Assignment Scheduler

By assigning each BS to sense a subset of channels, the sampling burden and sensing time on the spectrum scanner at each BS can be reduced. Nevertheless, this comes at the expense of inevitable spectrum sensing gaps over space. Thus, the objective of the proposed scheduler is to minimize these gaps. More formally, the scheduler seeks to find a sensing assignment that attains *an accurate radiomap for the entire geographic area across all channels, at each BS, given that each one merely senses* $P \ll M$ *channels.* Here, we define the radiomap as the aggregate received signal power at a given point in space. To this end, let **A** denote the sensing assignment matrix with rows corresponding to BSs and columns corresponding to channels, i.e., the *(k,m)*-th entry $a_{km} = 1$ denotes that the *k*-th BS is assigned the *m*-th channel. Then optimal scheduler is the one the minimizes the following cost function:

$$\begin{aligned} \underset{\mathbf{A}}{\text{minimize}} \; & \sum_{i \in \mathcal{J}} \sum_{k} \sum_{m} a_{km} x_{i,k,m}^{2} \\ \text{s.t.} \quad & \mathbf{A}^{T} \mathbf{1} = \boldsymbol{p} \;\; \& \;\; a_{km} \in \{0,1\} \end{aligned}$$

where the first sum is over the set of 2D grid points, the second sum is over the BSs in the network, the third sum is over all channels, and $x_{i,k,m}$ is the distance between the *i*-th grid point and the *k*-th BS when it senses channel *m*. The elements of ***p*** are integers, and the *m*-th entry determines how many BSs sense the *m*-th channel. The proposed integer-program is motivated as follows. The objective function seeks to find an assignment such that at a desired *i*-th location, the sum of the distances squared to all BSs that sense an *m*-th channel is minimized. Thus, a BS at the *i*-th grid point can still have information about the radiomap without sensing the channel. Finally, additional constraints can be introduced to the problem. For instance, the first constraint in the optimization framework is designed to update the sensing assignments for each channel, depending on the spectrum activity over time, e.g., if the traffic estimator finds higher activity over channel *m* then more BSs can be set to sense that channel to ensure reliable detection, i.e., higher value of the *m*-th entry of the vector ***p***.

4.b) Distributed Sensing

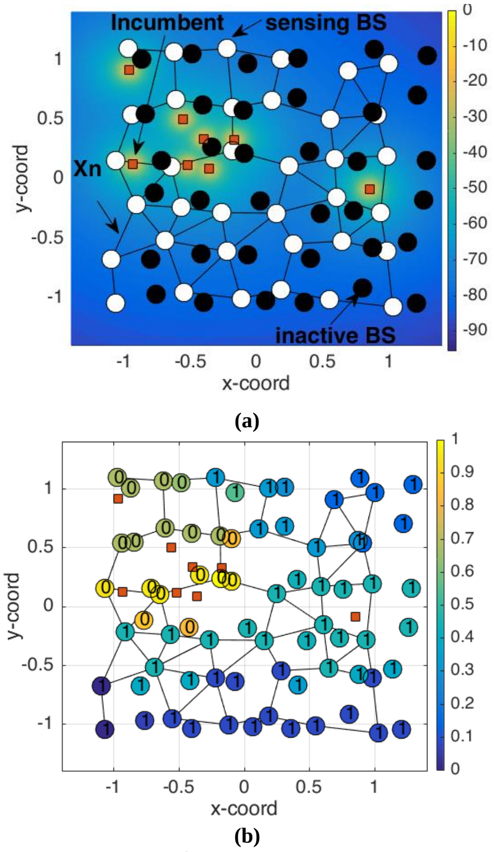


```
Algorithm 1
procedure DIFFUSION($\beta, \alpha_{lk,i}, \mu_{k,m}$)
    for $i = 1 \longrightarrow N$ do
        Measure $Y_{k,m,i}$
        Estimate $d_{k,m,i} = \beta d_{k,m,i-1} + (1-\beta) Y_{k,m,i}$
        Update combination weights $\alpha_{lk,i}$
        Combine $\psi_{k,m,i-1} = \sum_{l \in \mathcal{N}_{k,m}} \alpha_{lk,i} w_{l,m,i-1}$
        Adapt $w_{k,m,i} = \psi_{k,m,i-1} + 2\mu_{k,m} Y_{k,m,i} [d_{k,m,i} - Y_{k,m,i}\psi_{k,m,i-1}]$
    end for
    Stack $\boldsymbol{w}_{k,N} = [w_{k,1,N}, w_{k,2,N} \cdots, w_{k,M,N}]^T$
    Compare $\boldsymbol{w}_{k,N} \gtrless \boldsymbol{\lambda}_k$
end procedure
```

Fig. 4. Proposed diffusion-based sensing algorithm.

Once the sensing assignment is optimized, each BS locally senses the assigned channels, and then propagates the processed sensing data to neighboring cells for further distributed processing. The proposed distributed algorithm is based on adaptive diffusion-based signal processing that has been shown to be superior to LMS-like and consensus-based algorithms in multi-agent networks [7]. The proposed algorithm is shown in Fig. 4. First, the *k*-th BS computes the energy in its assigned channels, denoted as $Y_{k,m}$. Then, it cooperates with its neighbors to make a decision on the radiomap over the assigned channels. However, instead of fusing energy measurements, we propose an alternative message exchange. In particular, it is known that the energy detector is not robust in negative SNR regimes due to noise power uncertainty. Thus, the framework defines a BS-specific cost function, i.e., $J_k(w_{k,m})=\mathbf{E}[(d_{k,m} - w_{k,m}Y_{k,m})^2]$, where $d_{k,m}$ is a first-order filter approximation of $\mathbf{E}[Y_{k,m}]$, as shown in Fig. 4, where β a scaler close but less than one. Then, the BSs assigned to sense the *m*-th channel cooperate to distributively minimize the global cost function $J(w_m)=\Sigma_k J_k(w_{k,m})$. This is done in two steps: the combining step and the adaptation step. In the former, each BS shares its estimated $w_{k,m,i-1}$ with its neighbors and combines it to get $\psi_{k,m,i}$, where $\{\alpha_{k,l}\}$ are non-negative combining weights given to each neighbor [7]. In the adaptation stage, the BS processes the measurements and computes $w_{k,m,i}$, where $\mu_{k,m}$ is the step-size. Central to this algorithm is *an online clustering mechanism*, that is embedded in the algorithm by adapting the combination weights. In particular, the weighting mechanism finds similarities between the estimated $w_{k,m,i-1}$ and $w_{l,m,i-1}$, where a higher weight is given when these two values are closer to each other. Hence, as the algorithm progresses, cooperating BSs become clustered based on the similarities of their optimal solutions. After *N* iterations, $w_{k,m,N}$ is directly used as a test statistic, instead of energy measurement $Y_{k,m}$, i.e., $w_{k,m,N}$ is compared with a predetermined threshold $\lambda_{k,m}$ to make a final decision of the availability of the *m*-th channel. Then, the BS shares its sensing results with neighboring cells and collects the results of other channels to obtain a local view the wideband spectrum. An illustrative example is given next.

In this example, we consider a dense deployment of small cells over a grid, with a small random perturbation of the grid points. The inter-site distance between two BSs is, on average, 200m. We assume there are 8 incumbents, and they use



(a)

(b)

Fig. 5: (a) Radiomap of transmitting incumbents; (b) The value of $w_{k,m,N}$ and the sensing decision at each BS.

the same unlicensed channel, where they transmit at a fixed power of 20dBm. The normalized radiomap is shown in Fig. 5a (in dBm). A subset of BSs then implements the proposed sensing framework to sense this channel, where they share the sensing results using the Xn interface, denoted by solid lines. Fig. 5b shows the normalized value of the proposed statistic $w_{k,m,N}$ at each BS. We also show the sensing decision, for a normalized threshold of 0.68, where 1 denotes that the channel can be used for massive IoT access. It is observed that the variations of $w_{k,m,N}$ over space capture the variations of the incumbents radiomap. Also, it is evident that neighboring cells arrive at a similar statistic $w_{k,m,N}$, and hence make similar decisions, showing a clustering mechanism in the proposed algorithm.

5. Conclusion

The unlicensed spectrum access over cellular networks has the potential to connect thousands of machines and sensors, protecting legacy UEs from channel congestion over licensed carriers. To this end, it is imperative to explore spectral opportunities at a fine spatial resolution to aggressively reuse the same channel, and at fine spectral resolution to connect a massive number of low-rate devices. This requires innovative sensing assignment schemes as well as implementing distributive sensing algorithm.

References

[1] H. J. Kwon *et al.*, "Licensed-Assisted Access to Unlicensed Spectrum in LTE Release 13," in *IEEE Commun. Magazine*, vol. 55, no. 2, pp. 201-207, Feb. 2017.

[2] 3GPP, “Study on NR-based access to unlicensed spectrum,” 3rd Generation Partnership Project (3GPP), 38.425 , 2017.

[3] “IMT vision- framework and overall objectives of the future development of IMT for 2020 and beyond,” ITU-R, M.2083-0, Sep. 2015

[4] N. Sornin, M. Luis, T. Eirich, T. Kramp, and O. Hersent, “LoRaWAN specication," LoRA Alliance, Tech. Rep., Jul. 2016.

[5] G. Hattab and M. Ibnkahla, "Multiband Spectrum Access: Great Promises for Future Cognitive Radio Networks," in *Proceedings of the IEEE*, March 2014.

[6] 3GPP, “NG-RAN; Xn interface user plane protocol,” 3rd Generation Partnership Project (3GPP), Tech. Rep., 2017.

[7] J. Chen, C. Richard, and A. H. Sayed, “Diffusion LMS over multitask networks," IEEE Trans. Signal Process., Jun. 2015.

Biography

Ghaith Hattab is a PhD student at University of California, Los Angeles (UCLA). His current research focuses on algorithm development and optimization for wireless communication systems, including providing connectivity to massive IoT, 5G coexistence with incumbent systems, network densification, and dynamic spectrum access.

Danijela Cabric received the Dipl. Ing. degree from the University of Belgrade, Serbia, in 1998, and the M.Sc. degree in electrical engineering from the University of California, Los Angeles, in 2001. She received her Ph.D. degree in electrical engineering from the University of California, Berkeley, in 2007, where she was a member of the Berkeley Wireless Research Center. In 2008, she joined the faculty of the Electrical Engineering Department at the University of California, Los Angeles, where she is now Associate Professor. r. Cabric received the Samueli Fellowship in 2008, the Okawa Foundation Research Grant in 2009, Hellman Fellowship in 2012 and the National Science Foundation Faculty Early Career Development (CAREER) Award in 2012. She served as an Associate Editor in IEEE Journal on Selected Areas in Communications (Cognitive Radio series) and IEEE Communications Letters, and TPC Co-Chair of 8th International Conference on Cognitive Radio Oriented Wireless Networks (CROWNCOM) 2013. She is now Associated Editor of IEEE Transactions of Cognitive Communications and Networking. Her research interests include novel radio architecture, signal processing, and networking techniques for cognitive radio, 5G and massive MIMO systems.

Position Paper: "Low-cost Cognitive Radios against Spectrum Scarcity"

Dr. Alexandros-Apostolos A. Boulogeorgos[1]*, Member, IEEE,*
Prof. George K. Karagiannidis[2]*, Fellow, IEEE*
[1]*Department of Digital Systems, University of Piraeus, Greece*
(e-mail: al.boulogeorgos@ieee.org)
[2]*Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Greece (e-mail: geokarag@auth.gr)*

1. Abstract

The next generation wireless networks are envisioned to deal with the expected thousand-fold increase in total mobile broadband data and the hundred-fold increase in connected devices. In order to provide higher data rates, improved end-to-end performance, low latency, and low energy consumption at low cost per transmission, the fifth generation (5G) systems are required to overcome various handicaps of current cellular networks and wireless links. One of the key handicaps of 5G systems is the performance degradation of the communication link, due to the use of low-cost transceiver in high data rate. Motivated by this in this paper, we discuss the impact of transceiver front-end hardware imperfections on the spectrum sensing performance of cognitive radios.

2. Introduction

Radio frequency (RF) wireless spectrum is one of the most tightly regulated communication resources. From the early days of wireless communications, regulatory bodies were concerned about the interference that will be caused by different uses of the wireless spectrum. These concerns lead to the "doctrine of spectrum scarcity", which assigned each piece of spectrum with certain bandwidth to specific wireless systems [1]. With the proliferation of wireless services, in the last couple of decades, in several countries, most of the available spectrum has been fully (or almost-fully) allocated, which results in the spectrum scarcity problem. On the other hand, several studies have revealed that an important amount of spectrum experience low utilization (see e.g., [2] and references therein). Therefore, in order to maintain sustainable development of the wireless communication industry and market, rethinking of the spectrum allocation policies is necessary.

In this context, cognitive radios (CRs) are envisioned as one of the key enablers to deal with the RF spectrum scarcity issue. CRs are intelligent reconfigurable wireless devices capable of sensing the conditions of the surrounding RF environment and modifying their transmission para-meters accordingly for achieving best overall performance without interfering with other users. As a result, CR have recently been adopted in several wireless communication standards, such as long term evolution advanced (LTE-A), wireless fidelity (WiFi-IEEE 802.11), Zigbee (IEEE 802.15.4), and worldwide interoperability for microwave access (WiMAX-IEEE 802.16) [3].

One important task of CRs is spectrum sensing, i.e., the identification of temporarily vacant portions of spectrum. Spectrum sensing allows the exploitation of the under-utilized spectrum; hence, it is considered to be the main countermeasure against the spectrum scarcity problem. Moreover, it is an essential element in the operation of CRs. From technological point of view, in order to enable multiple-frequency band spectrum sensing, radio transceivers need to be flexible and software re-configurable devices. By definition, flexible radios are characterized by the ability to operate over multiple-frequency bands, and to support different types of waveforms, as well as various air interface technologies of currently existing and emerging wireless systems [4]. In this sense, the terms multi-mode, multi-band, and multi-standard are commonly used. The flexibility of transceivers is in-line with the software define radio (SDR) principle, which is considered to be one of the key technologies that enables the use of CRs [5].

From an economical point of view, the

advantages in integrated circuit technologies and the adoption of low-complexity transceiver structures, such as the direct-conversion radio (DCR) architecture, allowed improvements in manufacturing efficiency and automation that resulted in reducing the cost-per-device. Moreover, the use of low-complexity transceiver structures enable the reduction of the power consumption in battery-powered devices, without sacrificing too much performance. However, these advantages come with a cost in the device's hardware quality.

3. Hardware imperfections in Low-cost CR devices

In general, the demands for multi-standards operation, flexibility in order to deal with the spectrum scarcity problem, and higher data rate, as well as the constraints of product cost, device size, and energy efficiency, lead to the use of simplified radio architectures and low-cost radio electronics [6]. In this context, the DCR architecture provides an attractive front-end solution, since, as illustrated in Fig. 1, it requires neither external intermediate frequency filters nor image rejection filters [7]. Instead, the essential image rejection is achieved through signal processing methods. DCR architectures are low cost and can be easily integrated on-chip, which render them excellent candidates for modern wireless technologies [8]. However, direct-conversion transceivers are typically sensitive to front-end related impairments, such as in-phase (I) and quadrature (Q) imbalance (IQI), local oscillator (LO) phase noise and amplifiers nonlinearities, which are often inevitable due to components imperfections and manufacturing defects. Motivated by this, this section is focused on presenting the impact of hardware imperfections in spectrum sensing.

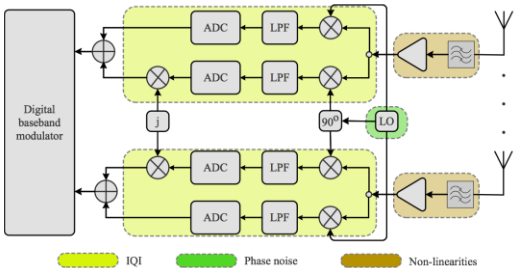


Fig. 1: DCR receiver architecture.

Amplifiers nonlinearties: Amplifiers nonlinearities cause spurious signals, which play the role of interference in adjacent channels. When the amp-lifier is used simultaneously by a number of carriers, intermodulation products are generated, which result to distortion in the desired signals. According to Bussgant's theorem, the amplifiers non-linearities results to an amplitude/phase distortion and a nonlinear distortion noise. Fig. 2.b intuitively presents the impact of amplifiers nonlinearities in spectrum sensing. In more detail, we consider a multi-channel spectrum sensing scenario in which the CR decides whether the channel k out of the K channels (in this case K=6) is occupied. We observe that due to the non-linearities, the signal power in an occupied channel is amplified, i.e., if a channel is busy, the accuracy of the correct is increased. On the other hand, if the channel is idle, the nonlinearities cause a noise power increase; hence, the detector can falsely decide that the channel is occupied. As discussed in [9], in order to mitigate the impact of false detection, due to the impact of amplifiers nonlinearities, the CR designer should appropriately adjust the spectrum sensing threshold.

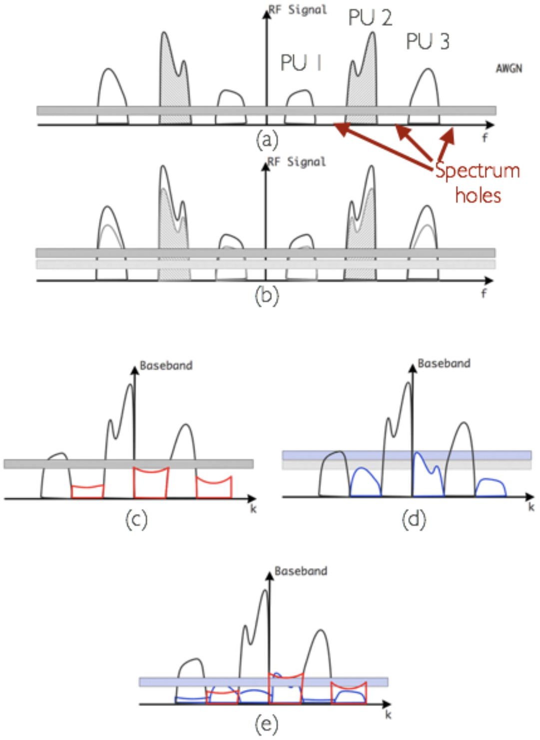


Fig. 2: Spectra of the received signal: (a) before the low noise amplifier (LNA)

(passband RF signal), (b) after LNA (passband RF signal), (c) after down-conversion (baseband signal), when local oscillator's phase noise is considered to be the only RF imperfection, (d) after down-conversion (baseband signal), when IQI is considered to be the only RF imperfection, (e) after down-conversion (baseband signal), the joint effect of LNA nonlinearities, phase noise and IQI. In this figure PU stands for the primary user.

LO Phase noise: Noise is of major concern in LOs, because introducing even small noise into a LO leads to dramatic changes in its frequency spectrum and timing properties. This phenomenon, peculiar to LOs, is known as phase noise or timing jitter, and it was identified as one of the major performance limiting factors of communication systems in several studies (see for example [9] and references therein). Generally, the disturbance of the amplitude of the oscillator output is marginal. As a result, most influence of the oscillator imperfection is noticeable in random deviation of the frequency of the oscillator output. These frequency deviations are often modelled as a random excess phase, and therefore referred to as phase noise. Phase noise will more and more appear to be a performance limiting factor especially in the case of multi-carrier and multi-channel communications, when low-cost implementations or systems with high carrier frequencies are considered, since, in these cases, it is harder to produce an oscillator with sufficient stability. As illustrated in Fig.2.c, phase noise causes adjacent channel interference in channel k from the channels k-1 and k+1. As a result, an idle channel might be identified as busy, due to the power leakage from a neighbor occupied channel. As illustrated in Fig. 3, this phenomenon can significantly limit the spectrum sensing capabilities of the CR [9], [10]. In more detail, assuming that the signal to noise ratio (SNR) for all the K channel is the same, we observe that, for a fixed SNR, as the 3 dB bandwidth of the LO, β, increases, the interference from adjacent channel increases, and the spectrum sensing capabilities of the CR decreases in comparison with the corre-sponding capabilities in ideal RF front-end case.

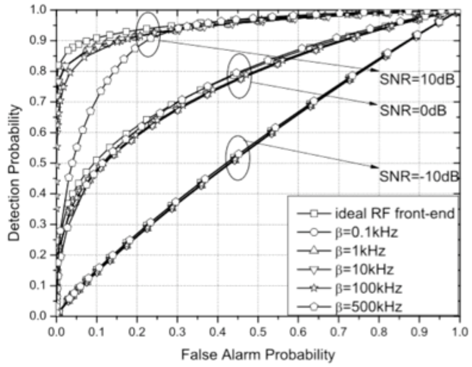


Fig. 3: Receiver operation curves (ROCs) for different values of β [9] .

IQI: It stems from the unavoidable amplitude and phase differences between the physical analog in-phase (I) and quadrature (Q) signal paths at the up- and down-converter of the TX and RX, respectively. In particular, IQI occurs due to the error in the nominally 90^o error shifter and the mismatch between the amplitudes of the LO I and Q outputs. This problem arises mainly due to the finite tolerances of the capacitors and the resistors used in the implementation of the analog front-end components. As depicted in Fig. 2.d., in the case of multi-channel spectrum sensing, IQI results in mirror-channel interference, which causes an energy reduction on the occupied channel and a corresponding energy increase in the idle mirror-channel. Fig. 4 numerically quantifies the impact of IQI in the spectrum sensing capabilities of the CR. Again, we assume that the SNR in all the K channels is the same. Fig. 4 indicates that as the level of this imperfection increases, i.e., as the image rejection ratio (IRR) decreases, the interference

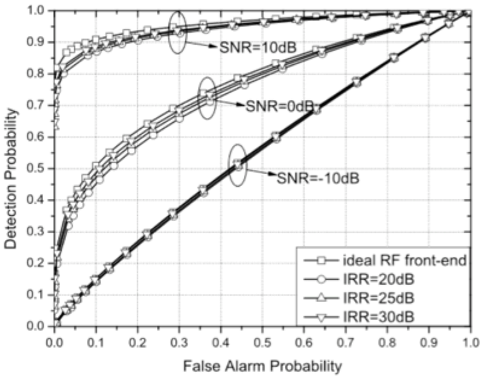


Fig. 4: ROCs for different values of image rejection ratio (IRR) [9].

of the mirror channel increases; hence, the spectrum sensing capability of the low-cost CR decreases.

Joint impact of RF impairments in spectrum sensing: The RF imperfections result in not only amplitude/phase distortion, but also neighbor and mirror interference, as demonstrated intuitively in Fig.2.e. amplifier's nonlinearities cause amplitu-de/phase distortion and an additive nonlinear distortion noise, whereas phase noise causes interference to the received base band signal at the k − th channel, due to the received base band signals at the neighbor channels k − 1 and k + 1. The joint effects of phase noise and IQI result in interference to the signal at the k-th channel by the signals at the channels -k-1, -k, -k+1, k-1 and k+1. Furthermore, the joint effects of LNA nonlinearties and IQI result in additive distortion noises and mirror channel interference. Fig.2.e. clearly demonstrates that LNA nonlinearities, IQI and phase noise results in an amplitude and phase distortion, as well as interference to channel k from the channels -k-1, -k, -k+1, k-1 and k+1, plus a distortion noise. If channel k is busy, the received signal's energy at channel k is increased, due to the interference of the neighbor and mirror channels, hence, the decision will be more accurate. However, if channel k is idle, the received signal's energy at channel k, due to the interference and the noise, may be greater than the decision threshold, and the detector will wrongly decide that the channel is busy. Consequently, the interference due to hardware imperfections plays an important role in the spectrum sensing capabilities; therefore, it should be quantified and taken into consideration when selecting the detection threshold.

4. Conclusion

In this paper, we presented the impact of hardware impairments in the spectrum sensing performance of low-cost CRs. Both the academia and the industry have several concerns regarding the limitations caused by the hardware imperfections and the self-interference leakage; especially in high-data rate systems. Therefore, several studies were focused on quantifying their impact in spectrum sensing and revealed that RF imperfections can significantly limit the CR performance and capability to identify spectrum holes. On the other hand, spectrum sensing solutions that take into account the impact of hardware imperfections of the CR's RF chain has not yet fully investigated and is a subject for future research. Such solutions are expected to drastically increase the spectrum utilization and deal with the spectrum scarcity issue in a more efficient manner.

References

[1] G. Staple and K. Werbach, "The end of spectrum scarcity [spectrum allocation and utilization]," IEEE Spectrum, vol. 41, no. 3, pp. 48–52, Mar. 2004.

[2] D. Datla, A. M. Wyglinski, and G. J. Minden, "A spectrum surveying framework for dynamic spectrum access networks," IEEE Trans. Veh. Technol., vol. 58, no. 8, pp. 4158–4168, Oct. 2009.

[3] M. Sherman, A. Mody, R. Martinez, C. Rodriguez, and R. Reddy, "IEEE standards supporting cognitive radio and networks, dynamic spectrum access, and coexistence," IEEE Commun. Mag., vol. 46, no. 7, pp. 72– 79, Jul. 2008.

[4] L. Anttila, M. Valkama, and M. Renfors, "Circularity-based I/Q imbalance compensation in wideband direct-conversion receivers," IEEE Trans. Veh. Technol., vol. 57, no. 4, pp. 2099–2113, Jul. 2008.

[5] M. Mueck, A. Piipponen, K. Kalliojarvi, G. Dimitrakopoulos, K. Tsagkaris, P. Demestichas, F. Casadevall, J. Perez-Romero, O. Sal- lent, G. Baldini, S. Filin, H. Harada, M. Debbah, T. Haustein, J. Gebert, B. Deschamps, P. Bender, M. Street, S. Kandeepan, J. Lota, and A. Hayar, "ETSI reconfigurable radio systems: status and future directions on software defined radio and cognitive radio standards," IEEE Commun. Mag., vol. 48, no. 9, pp. 78–86, Sep. 2010.

[6] H. H. Cho, C. F. Lai, T. K. Shih, and H. C. Chao, "Integration of SDR and SDN for 5G," IEEE Access, vol. 2, pp. 1196–1204, Oct. 2014.

[7] A.-A. A. Boulogeorgos, V. M. Kapinas, R. Schober, and G. K. Karagiannidis, "I/Q-imbalance self-interference coordination," IEEE Trans. Wireless Commun., vol. 15, no. 6, pp. 4157–4170, Jun. 2016.

[8] T. T. Duy, T. Duong, D. Benevides da Costa, V. N. Q. Bao, and M. Elkashlan, "Proactive relay selection with joint impact of hard- ware impairment and co-channel interference," IEEE Trans. Commun., vol. 63, no. 5, pp. 1594–1606, May 2015.

[9] A.-A. A. Boulogeorgos, N. D. Chatzidiamantis, and G. K. Karagiannidis, "Energy detection spectrum sensing under RF imperfections," IEEE Trans. Commun., vol. 64, no. 7, pp. 2754–2766, Jul. 2016.

[10] A. Gokceoglu, Y. Zou, M. Valkama, and P. C. Sofotasios, "Multi-channel energy detection under phase

noise: Analysis and mitigation," ACM/Springer Mob. Netw. Appl., vol. 19, no. 4, pp. 473–486, Aug. 2014.

Dr. Alexandros-Apostolos A. Boulogeorgos: IEEE Member, Researcher at University of Pire-aus, University of Thessaly and Visiting Lecturer at University of Western Macedonia, Greece.

Prof. George K. Karagiannidis: IEEE Fellow, Professor at Ari-stotle University of Thessaloni-ki, Greece. 2015, 2016 and 2017 Highly Cited Researcher.

Expert Opinion by Dr. Akbar M. Sayeed "Spectrum Scarcity: Can 5G Address the Spectrum Crunch?"

Department of Electrical and Computer Engineering,
University of Wisconsin-Madison, Madison, WI 53706 USA
Email: akbar@engr.wisc.edu

1) Over the past decades, many advanced technologies and techniques – such as MIMO, small cells, device-to-device communications, among others – have been introduced to increase spectral efficiency in cellular networks. Which of these technologies do you think has (or will have) the greatest impact on addressing the spectrum scarcity?

AMS: There are two fundamental ways in which spectral scarcity can be addressed: i) by making better use of a given frequency band (increasing spectral efficiency – bits/s/Hz), and ii) by using additional frequency bands (increasing bandwidth). Increasing bandwidth is the most straightforward approach to addressing spectrum scarcity, and new technologies operating in higher frequency bands (> 6 GHz), including centimeter-wave (6-30 GHz), millimeter-wave (30-300 GHz), and Terahertz (>300GHz) will facilitate that. However, several technological challenges remain in realizing their potential which are being currently being pursued.

MIMO, small cells, and device-to-device communications fall into the category of techniques that increase spectral efficiency through spatial multiplexing and/or spatial reuse of the spectrum. MIMO was invented in the mid-1990s and is finally beginning to have impact on spectral efficiency. Massive MIMO techniques are poised to significantly increase spectral efficiency at sub 6 GHz frequencies through spatial multiplexing. Small cells and device-to-device communication will further enhance spectral efficiency.

Operation at higher frequencies, such as mmWave bands, through the use of narrow beams, and in conjunction with small cells, is likely to have to the most significant impact in terms of both increasing the available bandwidth and increasing spectral efficiency through spatial/beamspace multiplexing. At these high frequencies, due to the highly directional nature of propagation, MIMO techniques are naturally better understood and analyzed in "beamspace". However, exploiting beamspace massive MIMO at mmWave and higher frequencies has significant technology challenges that are focus of intense current attention in industry and academia.

2) After all of these advancements, is the "spectrum crunch" still a valid concern?

AMS: The near-exponential growth in mobile network traffic that we have been witnessing in the past decade suggests that we will continue to need more spectrum and to utilize the available spectrum more efficiently. Given the central role of cloud-based computing going forward, with the implicit assumption of ubiquitous and high-rate wireless connectivity to the cloud, the demands on the available spectrum are only going to increase. From a physical and technological perspective, the usable spectrum is essentially limited. Thus, we will need to continually innovate to sustain the wireless communication needs of the information economy and society.

3) Millimeter wave (mmWave) communication is seen as a cornerstone technology to boost the capacity of next-generation wireless networks. Could you please elaborate how you view the potential of mmWave technology for overcoming spectrum scarcity?

AMS: Millimeter-wave wireless is emerging as a key technology for achieving multi-Gigabits/s data rates and millisecond latency requirements in 5G cellular networks and beyond. The ability of mmWave technology for alleviating spectrum scarcity and achieving high data rates rests on two fundamental characteristics: i) the availability of orders of magnitude larger bandwidth (on the order of GHz), and ii) the highly directional, quasi-optical nature of communication through narrow beams. Thus, beamspace (multiuser) MIMO – multiplexing data simultaneously through multiple dynamic beams – is a key operational functionality. There has been tremendous recent progress in basic research, technology development, and pre-commercialization industrial trials that support the viability of mmWave technology in alleviating spectrum scarcity and enabling a new class of high-rate/low latency use cases, such as augmented reality and autonomous vehicles.

4) Could you please explain the most pertinent spectrum sharing techniques/scenarios for 5G?

AMS: In sub 6 GHz technologies, dynamic spectrum sharing will be essential for efficient utilization of the limited available spectrum to support the growing data requirements. This will be achieved in conjunction with techniques such as MIMO, small cells, device-to-device and cooperative communications to increase spectral efficiency. Cloud-based network resource allocation will also likely play an important role.

For higher frequencies, such as mmWave frequencies, needed for high rate and low latency applications, spectrum sharing with lower bands could be critical for reliability, connectivity, and efficient network operation. This is due to the fact that mmWave propagation is predominantly limited to line-of-sight (LoS) and single-bounce multipath propagation (due to significant losses incurred in multiple bounces), and is also more likely to be blocked by humans and other objects compared to lower frequencies. Thus, the mmW link characteristics and connectivity can vary dramatically in mobile scenarios. Heavy rain and snow can also cause significant attenuation over longer links. Thus, it is likely that to ensure connectivity, mmWave links will have to be augmented with secondary links at lower frequencies, which could also be exploited for control information to improve network efficiency.

5) Given that ultra-high reliability and low latency are among key requirements for many emerging 5G applications, do you see the spectrum sharing techniques be able to support these services?

AMS: I think spectrum sharing techniques would play an important role in simultaneously delivering ultra-high reliability and low latency. In particular, low latency will generally require high bandwidth at the link layer, such as that enabled by mmWave frequencies. However, given the challenging propagation characteristics at mmWave frequencies, especially in mobile environments, reliability is more challenging and even connectivity may not be guaranteed in some extreme cases. In such scenarios, technologies that share higher and lower frequency bands to simultaneously satisfy reliability, rate and latency requirements would be needed. If high rates (needed for low latency) are to be achieved at lower frequency bands, this would necessarily require dynamic spectrum sharing for bandwidth aggregation over multiple available frequency bands.

6) Do you think that new economic models are required for 5G networks, on top of the engineering advancements, to incentivize/enable operators and vendors to share the spectrum more flexibly?

AMS: I do think that new business models will evolve for 5G and beyond as operators and equipment manufactures grapple with the development and deployment of 5G technology. For example, it is not clear what would be the first 5G use case for mmWave that will take traction. A lot of people think it will be fixed broadband wireless access (as an alternative to fiber). However, a business case has not been fully developed or demonstrated.

Another important aspect is connecting the remote, rural and unconnected parts of the US. With significant investments in the Connect America Fund, the reach of the fiber backbone will significantly increase in the coming years. However, the “last mile” local delivery will remain a big challenge to address cost effectively, and would benefit from new business models developed with collaboration between local and national service providers and equipment vendors.

Finally, new models are needed for collaboration between academia and industry to fully unleash the innovation in the US workforce. Currently – primarily due to IP related issues – this relationship is sub-optimal to put it mildly.

7) Could you please briefly introduce the most recent research project that you have done for addressing the challenges related to spectrum scarcity? (Please explain the key idea and interesting findings)

AMS: My group has been heavily involved in the past several years in the development of basic theory, new transceiver architectures, and prototypes and testbeds for mmWave wireless technology. In basic theory, building on our earlier work on modeling and analysis of multipath MIMO channels, we have pioneered the concept of beamspace MIMO for system design and analysis at mmWave and higher frequencies. The basic beamspace MIMO framework is being leveraged by our group and many other researchers for developing new approaches for beam tracking and management, beamspace channel estimation, precoding and receiver processing in multiuser scenarios, and new applications.

While multi-beam forming and data multiplexing is a key operational functionality, realizing it in practice is challenging due to both hardware and computational constraints. This is because the size of the antenna arrays needed to overcome the free-space path loss at mmWave frequencies is large (100s to 1000s of array elements). Conventional MIMO architectures based on digital beamforming are not feasible since a dedicated transmit/receive RF chain (mixer, amplifier, filter, ADC/DAC) is needed for each antenna element. Thus there has been significant interest in hybrid beamforming architectures that combine mmWave (analog) beamforming and low-dimensional digital processing to manage complexity. This is feasible due to the expected sparsity of mmWave channels in beamspace.

The most common and well-known hybrid architecture uses a phased array for beamforming. However, phased array beamforming is limited to a single beam per array, due to the hardware complexing of the phase shifting network, which significantly limits the beamspace multiplexing capability. Leveraging beamspace MIMO theory, our group has pioneered an alternative hybrid architecture – called continuous aperture phased (CAP)-MIMO – that uses a lens array for multi-beamforming. A lens array consists of a mmWave lens for spatial focusing and an array of antennas on the focal surface of the lens -- different antennas on the focal surface transmit or receive beams from corresponding directions. Importantly, a lens array can excited multiple simultaneous beams from the same array.

In the last few years, we have successfully developed CAP-MIMO prototypes at 10GHz and 28GHz for proof-of-concept demonstration. Building on this success, we are now developing a state-of-the-art testbed of multiple CAP-MIMO nodes, each capable of supporting at least four beams, and a flexible and reprogrammable FPGA-based architecture for baseband processing to enable real-time experimentation on the testbed network, spanning the physical layer, medium access control layer, and the higher network layers. The testbed also serves as a state-of-the-art channel sounder for measuring mmWave propagation characteristics with unprecedented spatial resolution. For example, our existing 28 GHz prototype has a 6” lens array with a spatial resolution of about 4 degrees. In contrast, a phased array with over 600 elements (26 in azimuth and elevation) would be needed to achieve the same spatial resolution!

8) Beyond your own work, are there any resources that you would like to recommend, especially to those who are new in this field and want to learn more about spectrum scarcity challenge in 5G networks?

AMS: For mmWave research as it relates to 5G and beyond, I would recommend getting involved in the NSF-sponsored mmW Research Coordination Network (http://mmwrcn.ece.wisc.edu). The mmW RCN is part of a new NSF Advanced Wireless Research Initiative (see below) and is aimed at bringing together researchers from academia, industry, and the federal government to develop a technology roadmap for mmWave wireless, and to tackle the critical research and technology development challenges that need to be address for realizing the potential of mmWave wireless.

The RCN is also actively pursuing a cross-disciplinary research agenda spanning hardware design, communication and signal processing techniques, networking protocols, and the development of prototypes and testbeds. Links to important resources related to mmWave technology are also provided on the RCN webpage, including the 5G webpages of all major telecom companies, such as Qualcomm, Samsung, Ericsson, Nokia Bell-Labs, Verizon, and AT&T.

In recent years, special issues of several magazines and journals have been dedicated to to emerging 5G technologies, including small cells, massive MIMO, devive-to-device communications, cooperative communication, spectrum sharing, and mmWave technology.

IEEE 5G Forum is another useful resource; http://ieee-wf-5g.org/

The National Institute of Standards and Technology (NIST) is leading a 5G Channel Modeling and Measurement Alliance, with a focus on mmWave frequencies; https://www.nist.gov/ctl/5g-mmwave-channel-model-alliance

DARPA spectrum collaboration challenge is a useful resource for spectrum sharing research and technology development; https://spectrumcollaborationchallenge.com/

Finally, NSF launched a new Advanced Wireless Research initiative in 2016 aimed at tackling challenges in wireless technology for 5G and beyond;
https://www.nsf.gov/cise/advancedwireless/

In particular, Platforms for Advanced Wireless Research (PAWR) is a joint effort between NSF and a wireless industry consortium that is a key component of the Advanced Wireless Initiative; https://www.advancedwireless.org/ . The PAWR program will fund and facilitate the development of about four city-scale experimental platforms for advanced wireless research and technology development. This is a unique and unprecedented opportunity for collaboration between researchers from academia and industry to unleash the full potential of US innovation in addressing hard and significant challenges in wireless technology for 5G and beyond.

9) What are the most important open problems and future research directions towards addressing the spectrum scarcity?

AMS: As noted above, spectrum scarcity is an ongoing challenge that will require sustained research, innovation, and technology development for: i) more efficient use of the existing spectrum (below 6 GHz), and ii) harnessing higher frequency bands (above 6 GHz), including cmWave, mmWave, and THz frequencies to increase the available bandwidth.

Given the central role of wireless infrastructure and connectivity in the information economy, any and all technologies that have the potential of impact must be pursued.

New and significant use cases may drive some of the innovation and open problems.

For example, the area of autonomous vehicles is considered by some to be the “killer app” of mmWave technology due to the high rate and low-latency requirements.

Similarly, new communication and spectrum sharing approaches for networking the billions of wireless devices in the “internet of things” is an important open problem.

Given that new frequency bands are opening up as older technologies become obsolete, there is a case to be made for taking a “clean-slate” and holistic approach for best utilization of the electromagnetic spectrum. This would necessarily require a multi-scale communication framework: lower frequencies for longer ranges, and higher frequencies for shorter ranges. If we were to design the communication infrastructure from scratch, given the current state of technology (and anticipated developments), what would it look like?

Given the recent advances in quantum information science, quantum computing, and quantum communication, the nexus between classical and quantum communication networks will become more important.

Some other general trends and patterns are emerging and may be helpful in guiding future research and development to fully unleash our innovation potential:

- The need for cross-disciplinary research; e.g., across hardware designers, signal

processors, communication and information theorists, and networking researchers.
- The need for integrated theoretical and experimental research and development.
- The need for new models and closer collaboration between industry and academia.

Biography: Akbar M. Sayeed is a Professor of Electrical and Computer Engineering at the University of Wisconsin-Madison, and leads the Wireless Communications and Sensing Laboratory. He received the B.S. degree from the University of Wisconsin, the M.S. and Ph.D. degrees from the University of Illinois, and was a postdoctoral fellow at Rice University. He is a Fellow of the IEEE, and has served the IEEE in a number of capacities, including as a member of Technical Committees, Guest Editor for special issues, Associate Editor, and as Technical Program Co-chair for workshops and conferences. His research interests include wireless communications, channel modeling, statistical signal processing, communication and information theory, time-frequency analysis, machine learning, and applications. A current research focus is the development of basic theory, system architectures, and testbeds for emerging 5G wireless technologies, including millimeter-wave and high-dimensional MIMO systems. He also leads the NSF Research Coordination Network on Millimeter-Wave Wireless.

Expert Opinion by Dr. Mérouane Debbah "Spectrum Scarcity: Can 5G Address the Spectrum Crunch?"

CentraleSupélec and Mathematical and Algorithmic Sciences Lab Huawei R&D
Paris, France
Email: merouane.debbah@huawei.com

1) **Over the past decades, many advanced technologies and techniques – such as MIMO, small cells, device-to-device communications, among others – have been introduced to increase spectral efficiency in cellular networks. Which of these technologies do you think has (or will have) the greatest impact on addressing the spectrum scarcity?**

MD: MIMO has been there for more than 20 years and for several reasons, which would be too long to detail here, the technology was not mature enough. I think there is no doubt that massive antenna densification will play a key role in 5G and spectrum will be traded for space. For me, Massive MIMO and small cells are just extreme operating points of the massive antenna densification trend where either all the antennas are co-located or distributed (with or without coordination). In between, we can come up with any combination of antenna configuration depending on the backhaul capillarity which feeds the antennas as well as the latency requirements of the fronthaul. The massive antenna densification approach (Cloud RAN, Massive MIMO, Small Cells, etc) will be for sure the key enabler to achieve the Mobile Broadband requirement of 5G.

2) **After all of these advancements, is the "spectrum crunch" still a valid concern?**

MD: The spectrum crunch will always be there when the coverage issue comes in the equation. This is for me one of the biggest challenges (with the latency of 1ms requirement) of 5G to ensure high data rate with good coverage (where the indoor coverage everywhere at 10 Gbs is even more difficult). As we know, good coverage requires low bands which are not available anymore. Already, the actual discussions around the sub-6Ghz C-Band Massive MIMO scenario show that we need more spectrum to achieve the target rates. Unfortunately, going into higher bands require more huge investments to ensure the coverage.

3) **Millimeter wave (mmWave) communication is seen as a cornerstone technology to boost the capacity of next-generation wireless networks. Could you please elaborate how you view the potential of mmWave technology for overcoming spectrum scarcity?**

MD: The most straightforward and realistic application of Millimeter wave communication is the WTTX (Wireless To The X, where X can be Home for example) scenario. In this case, MIMO Beamforming permits to point the transmission towards the receiver in a Line Of Sight environment and provides a real wireless fiber experience. In other cases, things are more complicated and turn out to be a headache: managing Non Line Of Sight, mobility or uplink transmissions are typical examples.

4) **Could you please explain the most pertinent spectrum sharing techniques/scenarios for 5G?**

MD: When 5G will start, many operators will still have LTE bands running and would like to have a starting 5G network using those LTE bands. I think that any technology that could provide a co-existence of 5G NR within the LTE bands (without ANY interference) will be welcome. Without spectrum sharing, this has already started in 5G where we are now deploying a 5G Non-Stand Alone network, using the core network of LTE.

5) Given that ultra-high reliability and low latency are among key requirements for many emerging 5G applications, do you see the spectrum sharing techniques be able to support these services?

MD: Sharing techniques are adequate for massive connectivity and we already see it with Narrow-band IoT in 4.5G in the mode of in-band transmissions. For the ultra-reliable and low latency case, this is more critical and for the moment, the actual existing techniques do not meet the extreme requirements we are facing.

6) Do you think that new economic models are required for 5G networks, on top of the engineering advancements, to incentivize/enable operators and vendors to share the spectrum more flexibly?

MD: New economic models will be a necessity for 5G networks, irrespective of spectrum sharing or not. As you know, the main driver will not be providing data but the services as well as all the verticals. From that point of view, operators will provide more and more on the fly and flexible services with new business models. For example, one could think of small rural infrastructures to become neutral hosts for anyone who wants to use them and get the payment in real time. Here, the blockchain could play an important role. The softwarization of the network will also provide more 5G as a service use cases.

7) Could you please briefly introduce the most recent research project that you have done for addressing the challenges related to spectrum scarcity? (Please explain the key idea and interesting findings)

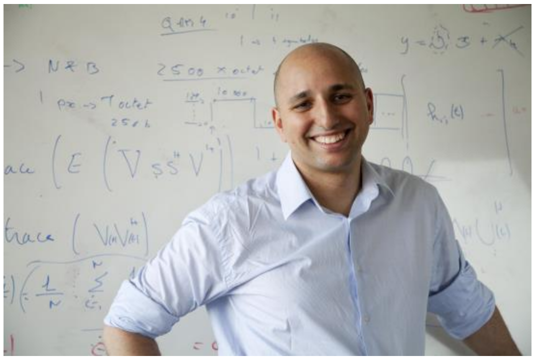
MD: We have been working these last 3 years on a couple of key technologies for 5G: Massive MIMO, Filtered OFDM, Polar Codes, SCMA (Sparse Code Multiple Access), Grant Free Signaling and Full Duplex Radio. The combination of Filtered OFDM, Polar Codes, SCMA (Sparse Code Multiple Access) and Grant Free Signaling provide a spectral efficiency gain of roughly 3 in the downlink and 5 in the uplink. Full Duplex Radio nearly doubles the throughput. As far as Massive MIMO is concerned, we have shown some recent demos where we are able to meet the requirements of 5G (10 Gbs in the uplink and 20Gbs in the downlink) with only 200 Mhz of band. We are now focusing our attention on the complexity/consumptions issues of these technologies.

8) Beyond your own work, are there any resources that you would like to recommend, especially to those who are new in this field and want to learn more about spectrum scarcity challenge in 5G networks?

MD: These recent years, I have had the pleasure to work with some outstanding researchers in the field of 5G such as Luca Sanguinetti, Emil Björnson and Jakob Hoydis. They have recently written a book “Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency” that compiles everything a 5G researcher should know and is a real piece of art. The book is available at Foundations and Trends in Signal Processing. I strongly recommend it.

9) What are the most important open problems and future research directions towards addressing the spectrum scarcity?

MD: There are many open important topics for the future such as accurate positioning in Non-Line of Sight Environments (for beamforming), wireless caching (where we trade memory for bandwidth), 3D drone cellular networks, TeraHertz communications, Ultra-Reliable Low Latency Communications and Large Intelligence Surfaces. These are just a few examples. We also tend to forget that the wireless dream can only happen if the backhauling is sufficient and new breakthroughs will be required in the cable, fiber and copper access. For example, the recent work in our community on the Non-Linear Fourier Transform for Fiber, which shows that there is no capacity crunch provides new research opportunities in the field. Finally, the design of these technologies will also require more advanced mathematical tools ranging from Machine learning, stochastic geometry, mean field games, random matrix theory, compressed sensing, nonlinear signal processing or distributed optimization, just to name a few.



Biography: Mérouane Debbah entered the Ecole Normale Supérieure Paris-Saclay (France) in 1996 where he received his M.Sc and Ph.D. degrees respectively. He worked for Motorola Labs (Saclay, France) from 1999-2002 and the Vienna Research Center for Telecommunications (Vienna, Austria) until 2003. From 2003 to 2007, he joined the Mobile Communications department of the Institut Eurecom (Sophia Antipolis, France) as an Assistant Professor. Since 2007, he is a Full Professor at CentraleSupelec (Gif-sur-Yvette, France). From 2007 to 2014, he was the director of the Alcatel-Lucent Chair on Flexible Radio. Since 2014, he is Vice-President of the Huawei France R&D center and director of the Mathematical and Algorithmic Sciences Lab. His research interests lie in fundamental mathematics, algorithms, statistics, information & communication sciences research. He is an Associate Editor in Chief of the journal Random Matrix: Theory and Applications and was an associate and senior area editor for IEEE Transactions on Signal Processing respectively in 2011-2013 and 2013-2014. Mérouane Debbah is a recipient of the ERC grant MORE (Advanced Mathematical Tools for Complex Network Engineering). He is a IEEE Fellow, a WWRF Fellow and a member of the academic senate of Paris-Saclay. He has managed 8 EU projects and more than 24 national and international projects. He received 17 best paper awards, among which the 2007 IEEE GLOBECOM best paper award, the Wi-Opt 2009 best paper award, the 2010 Newcom++ best paper award, the WUN CogCom Best Paper 2012 and 2013 Award, the 2014 WCNC best paper award, the 2015 ICC best paper award, the 2015 IEEE Communications Society Leonard G. Abraham Prize, the 2015 IEEE Communications Society Fred W. Ellersick Prize, the 2016 IEEE Communications Society Best Tutorial paper award, the 2016 European Wireless Best Paper Award and the 2017 Eurasip Best Paper Award as well as the Valuetools 2007, Valuetools 2008, CrownCom2009, Valuetools 2012 and SAM 2014 best student paper awards. He is the recipient of the Mario Boella award in 2005, the IEEE Glavieux Prize Award in 2011 and the Qualcomm Innovation Prize Award in 2012

TCCN Newsletter Editorial Board

TCCN NEWSLETTER DIRECTOR

Walid Saad
Virginia Tech, USA.

FEATURE TOPIC EDITORS

Daniel Benevides da Costa Department of Computer Engineering Federal University of Ceará (UFC) Sobral-CE-Brazil.
Omid Semiari, Georgia Southern University, USA.

TCCN Officers

CHAIR

Jianwei Huang
The Chinese University of Hong Kong, Hong Kong.

VICE CHAIRS

Walid Saad
Virginia Tech
USA
(TCCN Vice-Chair Americas)

Lingyang Song
Peking University
China
(TCCN Vice-Chair Asia Pacific)

Oliver Holland
King's College London
UK
(TCCN Vice-Chair Europe/Africa)

SECRETARY

Yue Gao
Queen Mary University of London
UK