PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Review

Cite this article: Chih-Lin I, Han S, Xu Z, Sun Q, Pan Z. 2016 5G: rethink mobile communications for 2020+. *Phil. Trans. R. Soc. A* **374**: 20140432.
http://dx.doi.org/10.1098/rsta.2014.0432

Accepted: 8 August 2015

One contribution of 14 to a discussion meeting issue 'Communication networks beyond the capacity crunch'.

Subject Areas:
electrical engineering

Keywords:
5G, cloud radio access network, software defined network, software defined air interface, ultra-dense network, user-centric network

Author for correspondence:
Shuangfeng Han
e-mail: hanshuangfeng@chinamobile.com

5G: rethink mobile communications for 2020+

I. Chih-Lin, Shuangfeng Han, Zhikun Xu, Qi Sun and Zhengang Pan

Green Communication Research Center, China Mobile Research Institute, Beijing 100053, People's Republic of China

The 5G network is anticipated to meet the challenging requirements of mobile traffic in the 2020s, which are characterized by super high data rate, low latency, high mobility, high energy efficiency and high traffic density. This paper provides an overview of China Mobile's 5G vision and potential solutions. Three key characteristics of 5G are analysed, i.e. super fast, soft and green. The main 5G R&D themes are further elaborated, which include five fundamental rethinkings of the traditional design methodologies. The 5G network design considerations are also discussed, with cloud radio access network, ultra-dense network, software defined network and network function virtualization examined as key potential solutions towards a green and soft 5G network. The paradigm shift to user-centric network operation from the traditional cell-centric operation is also investigated, where the decoupled downlink and uplink, control and data, and adaptive multiple connections provide sufficient means to achieve a user-centric 5G network with 'no more cells'. The software defined air interface is investigated under a uniform framework and can adaptively adapt the parameters to well satisfy various requirements in different 5G scenarios.

1. 5G vision and requirements

With the global commercialization of the fourth generation (4G) long-term evolution (LTE) standard, the wireless community is now looking forward to the 5G mobile network, which is expected to be launched in 2020 [1]. Worldwide initiatives on 5G research have been extensively carried out, starting with an investigation on user demands, application scenarios, technical trends

THE ROYAL SOCIETY PUBLISHING

Table 1. KPIs for 5G networks.

KPI items	KPI for 5G networks	definitions
peak data rate	≥10 Gbps	maximum achievable data rate for a user
minimum guaranteed user data rate	≥100 Mbps	minimum experienced data rate by a user
connection density	1 million connections km^{-2}	number of connected devices per unit area
traffic density	≥10 Tbps km^{-2}	total network throughput per unit area
radio latency	≤1 ms	duration between a packet being available at IP layer in BS and the availability of this packet at IP layer in terminal
end-to-end latency	millisecond level	duration between transmitting a data packet from source node and successfully receiving at destination node
mobility	up to 500 km h^{-1}	relative velocity between receiver and transmitter

and potential solutions, e.g. the work of METIS in Europe, 5G Forum in South Korea and IMT-2020 Promotion Group (PG) in China.

Mobile Internet and Internet of Things (IoT) are the two main drivers of future mobile networks and will span a broad prospect for 5G due to their diverse nature. As a consequence, 5G will touch many aspects of life in the future, such as home, work, leisure and transportation. The 5G scenarios include at least dense residential areas, office towers, stadiums, open-air gatherings, subways, highways, high-speed railways and wide-area coverage. These scenarios, which are characterized by ultra-high traffic volume density, ultra-high connection density or ultra-high mobility, raise extreme challenges for 5G.

Typical services, such as augmented reality, virtual reality, ultra-high-definition (UHD) video, cloud storage, Internet of Vehicles, smart home and over-the-top services, will be provided in these scenarios. The performance requirements for 5G are derived for each scenario, according to the predicted distribution of users, percentage of different services and service requirements such as data rate and latency. The key performance indicators (KPIs) for 5G include user experienced data rate, connection density, end-to-end latency, traffic volume density, mobility and peak data rate. The KPIs proposed by IMT-2020 PG are shown in table 1, which include, for example, over 100 Mbps user experienced data rate, 1 million connections per square kilometre, 1 ms end-to-end latency and tens of gigabits per second peak data rate [2]. To meet the extremely challenging user demands driven by Mobile Internet and IoT in a highly efficient way, the 5G network should be designed carefully.

In this paper, the vision and design methodologies of 5G are presented, from the perspective of China Mobile. In particular, the key characteristics of 5G, i.e. super fast, soft and green are analysed in §2. The main 5G R&D themes are further elaborated in §3, which include five fundamental rethinkings of the traditional design methodologies. In §4, the 5G network design considerations are discussed, with cloud radio access network (C-RAN), ultra-dense network (UDN), software defined network (SDN) and network function virtualization (NFV) examined as key potential solutions towards green and soft 5G network. The paradigm shift to user-centric network operation from the traditional cell-centric operation is investigated in §5, where the decoupled downlink and uplink, control and data, and adaptive multiple connections provide sufficient means to achieve a 5G network with 'no more cells' (NMC). The concept of the software defined air interface (SDAI) is presented in §6, which is a promising approach to meet the diverse demands in 5G by reconfiguring combinations of the physical layer building blocks. The paper is summarized in §7.

2. Three key characteristics of 5G

(a) Super fast networks

The 5G network is anticipated to provide fibre-like access data rate, 'zero' latency user experience and ultra-high mobility, and is envisioned to approach immersive and tactile user experience in any extreme scenario. An immersive user experience can be achieved with further development of Mobile Internet with high-definition video-dominated applications including augmented reality and UHD 3D vision. To this end, a 1000-times greater network capacity is expected by 2020 with 10 Gbps peak data rate requirements per wireless link. Further exploration in spatial domain, wideband systems with up to 500 MHz bandwidth in higher frequency, multi-connection in UDN scenario and other areas will be considered.

Use case scenarios such as remote surgery, auto-pilot and online gaming need a tactile round trip response. An end-to-end latency smaller than 10 ms is expected for future networks with a smaller than 1 ms delay budget reserved for air interfaces. New frame structures, access schemes based on new waveform design should be pursued with this target. 5G should cover mobility up to 500 km h^{-1}, due to the wide deployment of high-speed trains in China.

(b) Soft network

The 5G network is expected to be soft, with reconfigurable SDN and air interface. A soft network is envisioned to bring agility into implementation of each network element from core network (CN) to access network, as well as the building blocks of the air interface. The network function and resource virtualization should be the core of a soft network. It decouples software and hardware, control and data, uplink and downlink to facilitate a converged network synergistic with information and communication technology (ICT) convergence, multiple radio access technology (RAT) convergence, radio access network (RAN) and CN convergence, content convergence and spectrum convergence. This enables a super flat architecture that achieves cost-efficient network deployments, operation and management. In a soft network, the computing, storage and radio resources are virtualized and centralized to reach dynamic and user-centric resource management, matching service features. Soft networks are expected to build on a telecom-level cloud platform to enable network-as-a-service with the features of open network capability and network sharing. This makes it possible to achieve network flexibility and scalability and provides users with massive variety of services and consistent quality of experience (QoE). Soft networks may achieve breakthroughs first in C-RAN [3], NFV [4] and SDN [5] with control and data decoupling.

The soft network concept should be extended to the air interface as well. Instead of a global optimized air interface which is a trade-off among many factors, a SDAI will be considered. The parameters including spectrum, bandwidth, waveform, and so on, are all tuneable such that the air interface can be optimized to each individual application scenario. This enables broad adaption of future networks to application scenarios with extreme diverse requirements.

(c) Green network

Green communication is a social responsibility to reduce energy consumption as well as an economic target for wireless communication industry. High spectrum, spatial, temporal, hardware, software resource efficiency, low power consumption and low cost are the basic requirements of a green 5G network. Green Networks will achieve a 1000-fold capacity increase with minimum burden of spectrum resources. Advanced signal processing to effectively explore spatial resources, centralized coordination to reverse harmful interference to the useful signal, joint baseband and RF processing to enhance the same spectrum duplexing, etc., are some of the key technologies to improve radio resource efficiencies.

Green networks will achieve 100 times energy efficiency (EE) improvement to reduce operating expense (OPEX) for sustainable operations. It requires a capability for end-to-end energy management and optimization so that the total energy consumption will be minimized while meeting service requirements. Green Networks enable network capacity migration and breathing to match service variations without a waste of network resources. Moreover, 'plug and play' and on–off nodes are also essential parts of a green network. These massive nodes work without network planning in advance. Thus, an advanced self-organizing network (SON) is actually important for dynamic network planning and topology, as well as near real-time network optimization. Green Networks are able to use renewable energy, such as wind and/or solar energy as alternative power supply for networks, and bioelectric, kinetic and/or thermal energy for terminals.

3. 5G R&D themes

To achieve the 5G network vision above, the future wireless network should be fundamentally rethought from the following aspects:

— *Rethink Shannon* to start a green journey of wireless systems,
— *Rethink Ring and Young* for no more 'cells',
— *Rethink signalling and control* to make network application and load awareness,
— *Rethink antenna* to make base station (BS) invisible and
— *Rethink spectrum and air interface* to enable wireless signal 'dress for the occasion'.

(a) Rethink Shannon

After decades of high-speed development, the scale of ICT, or particularly communication networks, is huge enough such that its power consumption is no longer a negligible factor in global energy consumption. Considering 1000 times capacity increase by the year 2020, the power consumption of future networks is not affordable if the network is designed with the current energy scaling rule.

Classic Shannon theory, a 'bible' in the communications technical domain, has been leading the development of communication systems for over half a century. The extension of Shannon theory from scalar to vector in the early 1990s triggered the invention of the Multiple-Input Multiple-Output (MIMO) system which brought another 20 golden years of wireless communication systems. The spectrum efficiency (SE) and EE relationship has recently been explored by rethinking the Shannon theory, with a simple mathematically manipulation, for guidance on development of future green communication system in the next decade. By only considering transmit power over the air which traditional Shannon theory dealt with, a monotonic trade-off between SE and EE always exists, which means that increasing SE will induce an EE reduction. That would not have been very interesting, or useful. However, in any realistic network operations, besides the transmit power, the circuit power consumed by equipment also takes an important part. This power accounts for a greater and greater share of the total power as the cell becomes smaller and smaller. After taking into account the circuit power, the relationship of SE and EE is no longer monotonic. There is actually a win–win region for EE and SE. This realization presents a broad R&D field for joint SE and EE optimization. It applies in future networks from each individual component technology to network wide performance evaluation, ranging from the equipment level to the network level.

Diverse traffic fluctuation in the temporal and spatial domains provides another opportunity to rethink Shannon theory and different scales of traffic characteristics can be well exploited to improve both SE and EE. Network architecture and deployment can be smartly optimized by taking advantage of spatial correlation properties. Resources can be more efficiently managed and allocated by using the small-scale variations of traffic volume. Transmission technology can be adaptively selected or combined in different scenarios to implement EE–SE co-design [6–8].

(b) Rethink Ring and Young

The concept of cellular systems was proposed in 1947 by two researchers from Bell Labs, Douglas H. Ring and W. Rae Young. Since the first generation of cellular standards, this cell-centric design has been maintained through every new generation of standards including 4G. Towards the timeline of 2020 with the introduction of a heterogeneous network (HetNet) and a UDN, multiple layers of radio network have come into being. Energy consumption, interference, and mobility issues are becoming more serious due to smaller inter-site distance. Diverse types of BSs with different coverage, transmit power, frequency bands, among others, are introduced. Traffic fluctuation is more significant than before, taking into account emerging millions of mobile data applications. Therefore, in practical deployments, it is clear that the traditional homogeneous cell-centric design of the mobile network does not match the anticipated traffic variations and diverse radio environments.

The design of user-centric 5G radio networks should start with the principle of NMC, departing from cell-based coverage, resource management and signal processing. It should be predicated on the spatial and temporal variation of user demand, rather than a fixed cell-bounded configuration. Dynamically for each user, the available radio resources from multiple access points could be jointly scheduled, and the selection of Control/User plane and downlink/uplink channels, respectively, could be done separately.

A macro BS, using LTE evolution or a new RAT at lower frequency, provides wide coverage and serves as a signalling BS, while small cells at higher frequency, such as millimeter wave (mmWave), aim for boosting throughput and offloading traffic. Furthermore, to reduce the capital expenditure (CAPEX)/OPEX of small cells, by considering smaller coverage, supporting fewer users with low mobility, more relaxed synchronization requirement, smaller time and frequency selective fading, 'Data only Carrier' [9], with minimum control overhead and without common broadcast signalling, can be implemented to reduce interference and energy consumption. Macro cells can help small cells regarding discovery, synchronization, measurement, etc.

Given a great deal of overlapped coverage in UDN, to alleviate interference, more radio channel information between radio access points nearby should be shared in real-time and more joint cooperation between neighbouring access points is required. With the emergence of C-RAN, many technologies towards realization of the concept of NMC can be facilitated.

(c) Rethink signalling and control

As the proliferation of Mobile Internet continues, new services and applications appear at a fast pace. Some had exhibit orders of magnitude higher overhead over-the-air than more traditional services, in the forms of its data-signalling ratio, for instance. Careful studies had revealed its root cause: the over-the-air signalling/control of current networks has remained truthfully 'connection-oriented' through all four generations.

In the 5G era, the user and traffic characteristics will be even more diversified and differentiated, and the resource contending environment will be more complex. Therefore, more intelligent and adaptive signalling/control mechanisms must be devised, to achieve low-cost transmission with high signalling efficiency. Thus, we propose that 5G over-the-air signalling/control must be an intelligent combination of both connection-oriented and connectionless mechanisms. It must be application aware, load condition aware and user status (e.g. mobility) aware.

Furthermore, the mobile networks will be capable of providing differentiated user and traffic characteristics with customized signalling/control, where the different network functions, mobility management, security control, etc., are totally on demand. For example, during the low load period, new and slim air interfaces can be configured to achieve low cost. Differentiated and customized signalling/control (network slices) will be designed for different contexts (user, service, network circumstance), and, very importantly, a network framework is required for signalling/control allocation and network function orchestration. SDN is extremely suitable for the role of such a signalling/control framework. It provides a flexible and centralized control framework, and its open programmable interfaces also make it scalable to support new services. Moreover, with the centralized SDN framework, more contexts will be collected, and the SON functions will be performed better.

(d) Rethink antenna

Targeting significant capacity enhancement in 2020, the 5G network is expected to be ultra-dense with massive antennas deployed either in a distributed or centralized manner. Theoretically, massive MIMO [10] is expected to significantly reduce the inter-cell and intra-cell interference and hence may enhance both the SE and the EE. However, to accommodate a few hundred antenna and transceiver chains all on one structure in a traditional cell site manner appears to be nearly impossible, given the existing challenges and increasing difficulties of site acquisition, unless moving up to the mmWave band. For massive MIMO in the more desirable, lower frequency bands, we propose to fundamentally change the future scenes of cellular networks: make BS invisible, by configuring the active antenna arrays in a flexible manner on the walls of city buildings and town houses. For example, the Chinese character ("中") in the China Mobile logo ("中国移动") on buildings may actually be the BS antennas in the future.

Traditional multiple antenna transmission schemes, signalling protocol and network structure may not be sufficient and efficient in 5G, thus mandating fundamental rethinking. The key considerations include, for example, theoretical and practical algorithms of massive MIMO, practical implementation with low power and low cost massive MIMO system (especially the transceiver design), flexible and adaptive installation of antenna arrays with irregular antenna configurations, and distributed or centralized signal processing. In the aspect of standardization, dramatic changes may be needed in reference signals design, transmit and receive scheme design, RF path calibration, channel estimation and feedback. Proper beamforming structures need to be carefully investigated to identify the optimum digital, analogue or hybrid beamforming to best meet the requirements. The much reduced power in each RF chain may bring novel RF chain design, e.g. making use of low power low cost terminal-grade RFIC. It would be desirable to provide 'SmarTile', a 2×2 or 8×8 active antenna module, as the building blocks of centralized massive MIMO. The global optimal utilization of system resources with distributed massive MIMO, on the other hand, would be greatly facilitated via C-RAN architectures.

(e) Rethink spectrum and air interface

To provide high data rate with the capability of all spectrum access, the 5G air interface should provide flexible configuration according to the diverse service requirements. The traditional 'one-size-fits-all' air interface paradigm needs to undergo a fundamental change as well.

The mmWave spectrum is considered crucially important as a choice of new spectrum because of its significantly abundant bandwidth. 3GPP has already probed into the intelligent joint utilization of the existing licensed and unlicensed bands. Unified duplexing and full duplex provide another efficient solution to the utilization of the existing symmetric and asymmetric spectrum. New spectrum regulations are contemplating licenced shared access.

To support the diverse scenarios in 5G, the next-generation air interface will need to access all available spectrum, be scalable to deliver massive capacity and massive connectivity, and be adaptable to support new and existing services and applications with extreme requirements. The software defined concept is expected to be one cornerstone of the 5G air interface framework.

Based on the above five rethinking themes, we will respectively discuss network architecture, user-centric network operation and SDAI in details in the following three sections.

4. 5G network design considerations

(a) Cloud radio access network

C-RAN centralizes different processing resource together to form a pool so that the resource could be managed and dynamically allocated on demand on a pool level, as shown in figure 1. C-RAN

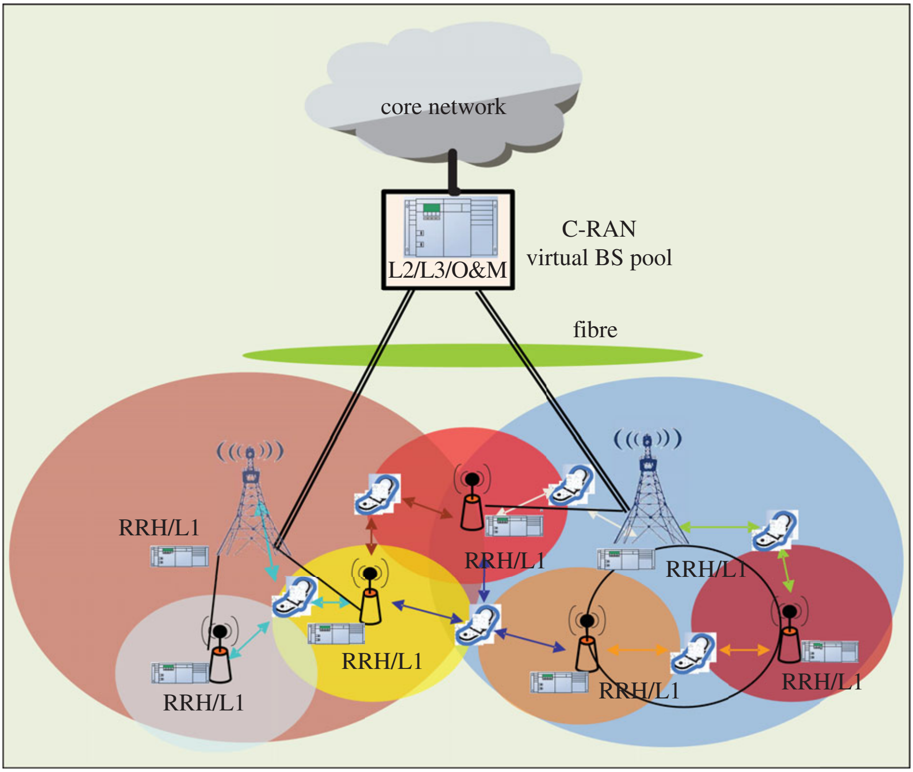

Figure 1. C-RAN. (Online version in colour.)

has demonstrated its advantages in total cost of ownership saving, network deployment speed-up, energy saving and, more importantly, C-RAN is a desirable architecture for 5G networks with facilitation of various 5G technologies. For example, one of the key visions is user-centric, or 'no-more cell', which means that users should be provided with not only much higher data rate services but also with less experience variation difference between the cell-centre and cell-edge regions. To achieve this goal, the severe interference suffered by the cell-edge users should be alleviated and thus coordination among multiple cells is necessary. C-RAN, on the other hand, thanks to the strong inherent central processing capability, provides an ideal structure to facilitate the implementation of coordination technologies with full or partial channel state information of all users available.

C-RAN is also an ideal match to the deployment of UDN which is deemed a promising solution to absorb highly dense user traffic. The design of UDN involves joint consideration of many issues, including the control and user plane decoupling, inter-site carrier aggregation and coordination, as well as interference mitigation in HetNet. In this case, C-RAN plays an important role with the internal high-speed low-latency switching mechanism and the central processing to implement those key technologies. Finally, C-RAN provides a unique opportunity to support multi-RAT with the adoption of the general purpose platform (GPP) and virtualization technology. In C-RAN, different RATs can be virtualized in forms of virtual machines (VM) and operate separately and independently on the same platform. Thanks to the highly efficient VM communication, C-RAN can further help with the multi-RAT coordination.

(b) Ultra-dense network

The main idea behind UDN is to increase the density of deployed stations by introducing abundant low-power small cells, so as to improve system capacity and network coverage

[11]. Obviously, UDN results in more overlapping areas among cells. Therefore, interference and mobility management become critical issues. Besides, it will be critical to reduce the CAPEX/OPEX. So advanced receiver and interference handling approaches, new architectures and protocols will be devised to handle the above issues: (i) interference management: while design of advanced receiver can improve the interference tolerance at both BS and user equipment (UE) sides, advanced interference management especially interference coordination shall be further considered to minimize interference. Further, accelerated recognition mechanisms can be achieved by introduction of special synchronization signals; (ii) mobility management: as mobility frequency will be extremely high in UDN, new connection and mobility handling schemes, especially assisted mobility methods, will be applied to guarantee the user's seamless mobility experience; simultaneous multiple connection, such as a UE connecting with more than two access points, can be used to improve the mobility experience of the user and improve data rate in the cell edge; (iii) CAPEX/OPEX reduction: ease of access node deployment will be critical for UDN, which requires low-cost nodes, easy backhauling (wired or wireless) and automatic configurations. Meanwhile, flexible and automatic network procedures will be designed for efficient operation, considering the diversity and complexity of UDN infrastructures. Besides, UDN also brings the BS and UE closer. It will make the communication link tend to be Ricean, especially in case of high-frequency communication with massive antennas deployed. This will further introduce new designs and reduce the complexity, compared to the conventional dense deployment.

(c) Software defined network and network function virtualization

The 5G network should be flexible and open with tight integration of multiple RATs. It provides a variety of new services and consistent service experience for end users. However, the existing mobile networks aim at limited categories of network services, which are 'one-size-fits-all' signalling/control frameworks and lack of flexibility, scalability and programmability. Actually, they are usually built on specialized equipment and dedicated platforms with weak flexibility and scalability, which is hard to meet the requirements of higher cost efficiency and shorter time-to-market of future mobile Internet applications innovation. Besides, the control and the data planes are tightly coupled. The data plane is too concentrated and results in delay, while the control plane is too scattered resulting in the low efficiency of radio resource management.

The proposed 5G mobile network architecture based on SDN is as follows [5]. First, the functions of the network are modularized, so the most appropriate signalling/control mechanisms can be orchestrated to satisfy different services, even the new services, under different network conditions. Second, the capability of the network is opened. A simplified representation of the network can be presented to applications via APIs. It enables the innovation of services and applications. Last but not least, SDN is mainly characterized as the decoupling of the control and data planes. The control plane can be centralized in the SDN controller, and the data plane gets down close to the users. Logically, the network architecture becomes flatter. It is noted that SDN is also extended to the radio side. Some of the radio control functions are centralized to perform the joint optimization and improve user experience. Both the radio control and edge network control functions can coexist in a physical SDN controller located at BBU pooling in case of C-RAN architecture.

As for SDN implementation, NFV is a good infrastructure. It aims to decouple network functions from dedicated hardware and implements network functions in software that can run on standard hardware, which locates in the network as required, without the installation of new equipment [4]. NFV adoption can reduce equipment costs and power consumption, speed up time to market and encourage more innovation. By decoupling mobile network functions from dedicated hardware, operators can flexibly choose network resources from different locations, which means that network services and resources can be shared and migrated on demand.

In practice, there are still lots of challenges to realize SDN and NFV in 5G mobile networks. For example, the function definition of the SDN controller, the standardization of the interfaces and protocols of the controller and forwarding hardware, and the security threats and solutions brought by SDN are still under investigation. Whether standard hardware can meet the performance and reliability requirements of RAN should be carefully evaluated. As a starting point, most of baseband processing has been virtualized over GPP servers in C-RAN, while part of PHY functionalities with extreme real-time requirements have been implemented by dedicated accelerators.

5. No more cell: paradigm shift from cell-centric to user-centric cell operation

(a) Efforts towards enhanced performance of cell-centric network

The concept of homogeneous cellular systems was proposed in 1947 and this cell-centric design has been maintained through every new generation of mobile communication standards including 4G. A typical cell-centric HetNet is illustrated in figure 2*a*, where the downlink and uplink, control and user planes have only one BS, e.g. UE1, UE3 and UE4. The nature of cell-centric designs is that cell planning and optimization, mobility handling, resource management, signalling and control, coverage and signal processing are all assumed to be done either for or by each BS uniformly. However, this system does not match with traffic variations and diverse environments in practical deployment, because the resources are conventionally allocated semi-statically from the standpoint of network.

To counteract the above issues, relay, distributed antenna systems (DASs) and CoMP [12] have been implemented as short-term solutions. While relay and DAS are mainly used for coverage extension, CoMP is aimed for capacity improvement for cell-edge users, as shown by UE2. In particular, CoMP has been intensively investigated by academia, industries and standard bodies like 3GPP and WiMax, in which inter BS joint processing and coordination are sought for enhanced cell-edge and cell-average performance. Note in the above initial efforts, the cell-centric network operation hardly changed.

(b) Efforts towards user-centric network

Given the above drawbacks of cell-centric network, user-centric has been pursued recently. The key features of a user-centric network include decoupled control and user plane, decoupled downlink and uplink, and multiple connections, for enhanced coverage, mobility support, EE and SE [13].

(i) Decoupling of control and user plane

The concept of decoupled control and data plane was recently proposed in Beyond Cellular Green Generation (BCG2) [14], liquid cells, soft cells and phantom cells. These new radio access architectures depart substantially from the conventional cell-based coverage, resource management and signal processing. One typical scenario is that a macro BS with wider coverage is responsible for signalling while small cells with smaller coverage are dedicated for data. New users can access the macro cell, and then the macro cell can coordinate with the small cell for possible data transmission. With decoupled signalling and data, the mobility robustness can be improved since handover signalling overhead is reduced with a more stable signalling connection with macro signalling BS, while the small cell deployment becomes much easier since no careful cell planning is required anymore. Also, spectrum utilization in small cells will be significantly enhanced due to the much relaxed requirement of control information and reference signals transmission from small cells.

As shown in figure 2*b*, the cell responsible for control and data can be flexible. For example, for UE1, the control is from the macro cell, while the data is from the small cell. For fast-moving UE3, both control and data are from the macro cell. For the static UE4, the control and data are from the nearest small cell.

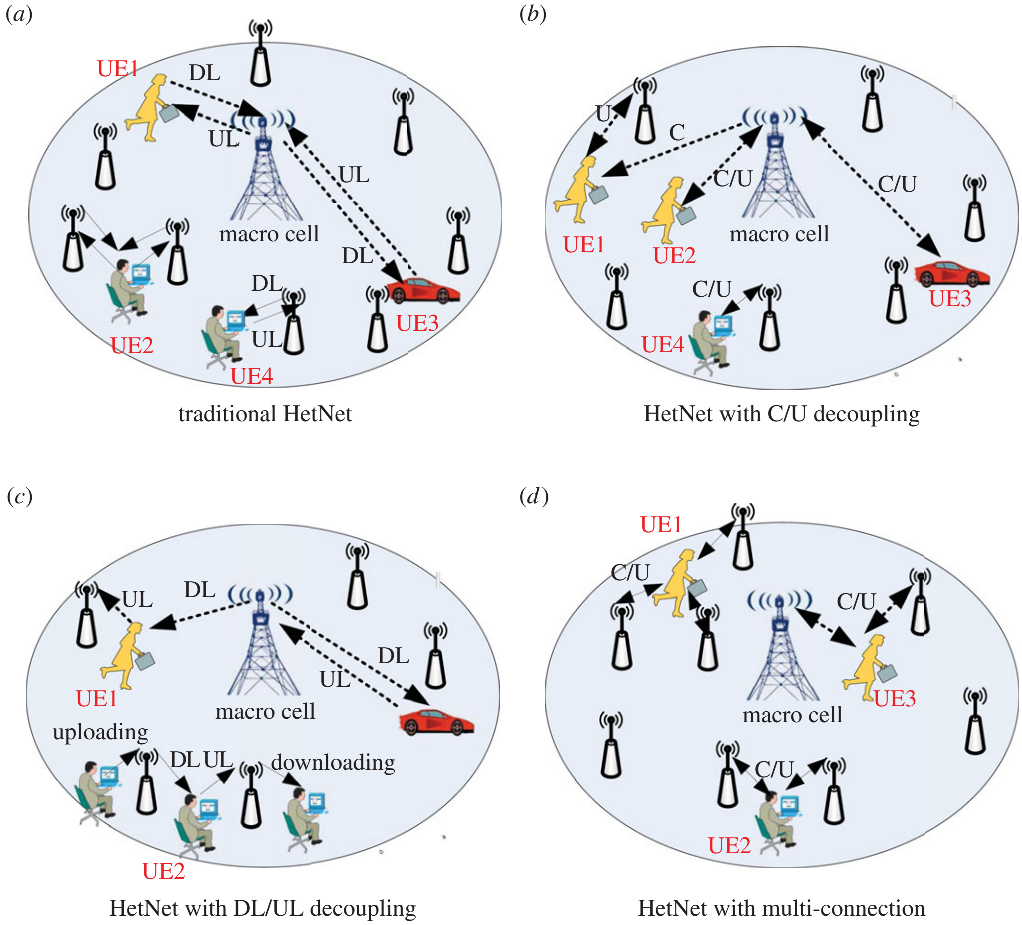

Figure 2. (*a*–*d*) HetNet scenarios. (Online version in colour.)

(ii) Decoupling of downlink and uplink

Another key feature of user-centric design is the decoupling of the downlink and uplink, which can facilitate flexible resource allocation between cells. In the traditional cell-centric network, the downlink and uplink connections are established with the same BS. However, in HetNet deployment, the nearby small cells with smaller downlink reference signal power may possibly provide a better uplink connection. Therefore, the downlink and uplink of one UE may well be established with different BSs. As depicted in figure 2*c*, we consider two cells where cell 1 is heavily loaded in the downlink, while cell 2 is overloaded in the uplink. In the traditional cell concept, if a UE device is located at the cell boundary with symmetric downlink and uplink data requirements, either its downlink or uplink requirement may not be satisfied, no matter which cell is accessed as the serving cell. However, with the decoupled downlink and uplink, this UE may naturally establish its downlink with cell 2 and its uplink with cell 1, as the case of UE2.

Obviously, global resource optimization in user-centric design involves optimal selection of downlink and uplink connections for both control and data flows of all users. This optimal multi-connection issue is not feasible in traditional RAN, because too much inter BS information sharing will be incurred, including dynamic user channel state information and scheduling information, etc. Fortunately, with the emergence of C-RAN, many technologies towards realization of user-centric network can be facilitated.

(iii) Small cell design simplification

Because of the potential tremendous increase of CAPEX and OPEX in deploying and maintaining a huge number of small cells, it is well motivated to investigate the simplification of small cells. Currently, the design principle of 3GPP is to apply carrier design from macro to small cells such that small cells keep most of functionalities of macro cells, such as measurement, synchronization, access, handover, etc. The Lean Carrier [15] is an attempt to minimize the control channel and reference signal overhead of the small cell, to increase resource utilization and reduce interference, thereby increasing spectral efficiency.

The capability of control and data decoupling in HetNet has phenomenal impacts on the design of small cells. For example, the cell common control information can only be transmitted from the macro cell. This indicates that small cells can be designed in extreme case (i.e. the most simplified design) not to have the following capabilities:

— synchronization signals like the primary synchronization sequence and secondary synchronization sequence, because only the macro cell is responsible for system entry;
— cell-specific reference signal, which was designed in LTE for detection of synchronization signals, channel measurement and control information demodulation. Only UE-specific reference signals for demodulation are needed for small cells; and
— cell-specific system information, which is same for all small cells within the coverage of the macro cell, e.g. the physical broadcast channel, main information block and system information block. UEs can acquire these information very conveniently from the macro cell.

With C-RAN architecture, the small cells can be further simplified. For example, as shown in figure 1, the traditional small cell BS can be sufficiently replaced by low-power RRH, with all the user scheduling, resource allocation and signalling processing being accomplished at the BBU pool. The macro cells are replaced with high-power RRU. Accordingly, 5G HetNet can, in essence, be realized in the form of distributed massive MIMO.

(iv) Adaptive multiple connections

In the user-centric HetNet with decoupled control and data, downlink and uplink, any information (control or data) can be flexibly transmitted to each user from one or multiple points, as shown in figure 2*d*. The optimal transmission points selection needs to consider the traffic load of each point, quality of service or QoE, user's mobility status, energy consumption of transferring of the related information, channel information, and signalling overhead involved. CoMP with coordinated scheduling and joint processing can be used to improve the decoupled control and user plane, downlink and uplink channel. The joint consideration of the above parameters is made possible with C-RAN framework.

6. Software defined air interface

The SDAI [16] will meet the diverse demands in 5G by reconfiguring combinations of the physical layer building blocks, including frame structure, duplex mode, waveforms and multiple access scheme, modulation and coding and spatial processing scheme, etc., as shown in figure 3.

(a) Flexible frame structure

In order to realize SDAI, the frame structure should be flexible enough: e.g. the time and frequency resources are allocated to different users with different service requirements, channel conditions, UE capabilities (multiple access support, full duplex mode, feature or smart phones), mobility, frequency bands, etc. In different resource blocks, different air interface solutions with different multiple access schemes, TTI parameters, waveforms, and duplex mode, pilot signals, etc., can be

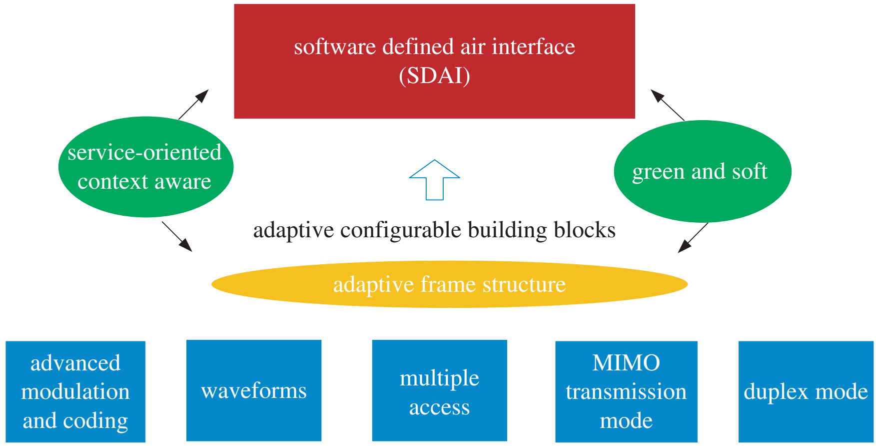

Figure 3. A framework for SDAI. (Online version in colour.)

defined. This is very challenging, since the inter-subcarrier band interference between different resource blocks needs to be carefully mitigated.

(b) Flexible waveforms

Waveforms, as the vital components of the air interface, are expected to be more flexible in SDAI. Beyond the perfectly synchronized and orthogonal OFDM signals designed for high volume data transmission, the non-orthogonal asynchronous waveforms, such as filter bank multiple carriers, Filter-OFDM, etc., emerge as promising solutions for sporadic traffic in the IoT applications in 5G. The flexible compatible framework for theses waveforms can be based on the carrier/waveform aggregation. Different waveforms located in different carriers can be aggregated in one air interface serving diverse 5G services. The waveform, sub-band bandwidth, subcarrier spacing bandwidth, filter length and cyclic prefix length in each wave can be flexibly chosen according to the dedicated scenarios and services.

(c) Adaptive multiple access

Non-orthogonal multiple access (NOMA) schemes have attracted considerable attention as a promising candidate for 5G systems since they can efficiently improve the spectral efficiency and accommodate the necessary scalability for massive IoT connectivity. However, the benefits of the typically identified candidates like sparse coding multiple access (SCMA), NOMA are usually achieved at the cost of higher signal processing complexity. Specifically, for SCMA, the decoding complexity dramatically increases with the overloading access users and modulation orders. The adaptation among different multiple access technologies or some parameters such as number of codeword of a SCMA codebook, spreading factor, maximum number of layers, number of non-zero elements of each codeword, based on different system requirements and network/UE capabilities such as coverage, connectivity, SE, EE, can be used to facilitate the SDAI adaptation [16].

(d) Configurable multiple-input multiple-output transmission

Different MIMO technologies perform differently in various channel conditions and antenna configurations. In 4G systems, the MIMO mode switch had already been introduced to achieve a consistent good performance. In the 5G SDAI, the MIMO mode switch will be further enhanced with more new MIMO modes, e.g. the spatial NOMA or the so-called enhanced MU-MIMO [17], and be compatible with the hybrid digital and analogue beamforming structure for massive MIMO antenna configurations.

7. Conclusion

Faced with the fierce challenges of mobile traffic in 2020s, the 5G network is envisioned to be green, soft and super fast, with user-centric cell operation. This paper gave an overview of China Mobile's 5G vision and potential solutions. Three key characteristics of 5G, i.e. green, soft and super fast were discussed, with five fundamental rethinking of the traditional design methodologies elaborated. The 5G network design considerations were also discussed. Further, the user-centric network operation was investigated. Finally, SDAI was presented to meet the diverse demands in 5G by reconfiguring combinations of the physical layer building blocks.

Competing interests. We declare we have no competing interests.
Funding. We received no funding for this study.

References

1. Chih-Lin I, Rowell C, Han S, Xu Z, Li G, Pan Z. 2014 Towards green & soft: a 5G perspective. *IEEE Commun. Mag*. **52**, 66–73. (doi:10.1109/MCOM.2014.6736745)
2. IMT-2020 (5G) Promotion Group (PG). 2014 5G vision and requirements. See http://www.imt-2020.org.cn/zh/documents (accessed May 2014).
3. CMRI. 2011 C-RAN: the road towards green RAN. See labs.chinamobile.com/cran (accessed October 2011).
4. Introduction White Paper. 2012 Network functions virtualization. 22–24 October 2012 at the 'SDN and OpenFlow World congress', Darmstadt, Germany.
5. ONF White Paper. 2012 Software-defined networking: The New Norm for Networks. April 2012.
6. Han S, Chih-Lin I, Xu Z, Rowell C. 2015 Large scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G. *IEEE Commun. Mag*. **53**, 186–194. (doi:10.1109/MCOM.2015.7010533)
7. Li GY, Xu Z, Xiong C, Yang C, Zhang S, Chen Y, Xu S. 2011 Energy-efficient wireless communications: tutorial, survey, and open issues. *IEEE Wirel. Commun.* **18**, 28–35. (doi:10.1109/MWC.2011.6108331)
8. Xu Z, Pan Z, Chih-Lin I. 2014 Fundamental properties of the EE–SE relationship. IEEE WCNC, 2014.
9. Zhang X, Zhang J, Wang W, Zhang Y, Chih-Lin I, Pan Z. 2015 Macro-assisted data-only carrier for 5G green cellular systems. *IEEE Commun. Mag*. **53**, 223–231. (doi:10.1109/MCOM.2015.7105669)
10. Rusek F, Persson D, Lau BK, Larsson E, Marzetta T, Edfors O, Tufvesson F. 2013 Scaling up Mimo: opportunities and challenges with very large arrays. *IEEE Signal Process. Mag*. **30**, 40–60. (doi:10.1109/MSP.2011.2178495)
11. METIS 2020 project. 2013 The 5G future scenarios identified by METIS: the first step toward a 5G mobile and wireless communications system. See https://www.metis2020.com/press-events/press/the-5g-future-scenarios-identified-by-metis/ (accessed September 2013).
12. Lee J, Kim Y, Lee H, Ng B, Mazzarese D, Liu J, Xiao W, Zhou Y. 2012 Coordinated multipoint transmission and reception in LTE- advanced systems. *IEEE Commun. Mag*. **50**, 44–50. (doi:10.1109/MCOM.2012.6353681)
13. Ishii H, Kishiyama Y, Takahashi H. 2012 A novel architecture for LTE-B: C-plane/U-plane split and phantom cell concept. IEEE Globecom Workshops, December 2012.
14. GreenTouch 2013–2014 Annual Report. See http://www.greentouch.org (accessed December 2013).
15. Hoymann C, Larsson D, Koorapaty H, Cheng J-F. 2013 A lean carrier for LTE. *IEEE Commun. Mag*. **51**, 74–80. (doi:10.1109/MCOM.2013.6461189)
16. Sun Q, Han S, Chih-Lin I, Pan Z. 2015 Software defined air interface a framework of 5G air interface. IEEE WCNC, 2015.
17. RP-140701, 3GPP TSG RAN Meeting #64. New study item proposal: enhanced MU-MIMO and network assisted interference cancellation.