Beyond 5G - wireless data center connectivity

Citation for published version (APA):

Document license:
Unspecified

DOI:
10.1117/12.2506269

Document status and date:
Published: 01/02/2019

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:
• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.
Beyond 5G - wireless data center connectivity

Simon Rommel, Thiago R. Raddo, Ulf Johannsen, Chigo Okonkwo, Idelfonso Tafur Monroy

Event: SPIE OPTO, 2019, San Francisco, California, United States

Beyond 5G – Wireless Data Center Connectivity

Simon Rommela, Thiago R. Raddoa, Ulf Johannsenb, Chigo Okonkwoa, and Idelfonso Tafur Monroya

aInstitute for Photonic Integration, Eindhoven University of Technology, 5600 MB Eindhoven, Netherlands
bCentre for Wireless Technology, Eindhoven University of Technology, 5600 MB Eindhoven, Netherlands

ABSTRACT

With the digitalization of industry and society, data centers have grown into an essential key strategic infrastructure, centralizing the processing, storage and distribution of vast amounts of information. Through continuing centralization, their size grows ever larger, while at the same time they need to remain flexible and dynamic to adapt to the temporary nature and diverse requirements of many tasks. Modular data centers can fulfill this requirement, allowing quick deployment and provisioning, while being highly reconfigurable – however, in such data centers interconnectivity is a complex and difficult issue. Wired connections based on optical fibers are the standard in data centers, but come at a significant cost and lack reconfigurability. The introduction of wireless connectivity at millimeter and tera-Hertz frequencies offers similar capacities, while allowing dynamic and reconfigurable deployment and wireless or hybrid data center architectures have been suggested. In this context, the innovations in high capacity millimeter wave communications and in the convergence of optical and wireless networking developed for 5G mobile networks may offer a potential technology candidate for high-density and high-capacity data center network deployments. Such networks allow the layout and topology of the network to be changed on demand and to adapt to the changing needs of different applications, creating a data center network that matches the multi-purpose nature of the computation and storage hardware. In this paper, the recent trends in wireless technologies for data centers are reviewed and connected to the innovations of optical and wireless convergence seen in 5G networks, perceiving a data center network that is better able to cope with the demanding requirements in terms of network reconfigurability, installation and running cost, as well as power consumption and cooling efficiency.

Keywords: wireless data center networks, beyond 5G technology, millimeter wave communications, THz communications, optical wireless communications

1. INTRODUCTION

A digital revolution is foreseen for the coming years in which mobile and cloud will play key roles along with emerging technologies such as artificial intelligence (AI), fifth generation mobile (5G), robotics, and cyber security. The revolution of the digital ecosystem promises to disrupt current business models and society, enabling cutting-edge technologies to deliver services and solutions to an exploding universe of distributed devices and people. As data volumes grow substantially due to new multimedia applications such as super high-resolution streaming, augmented and virtual reality, and interactive e-learning, new data centers and innovative solutions are expected to attend the growing volume of data storage.

Furthermore, new digital scenarios like self-autonomous driving applications, tactile Internet, industry 4.0, and internet of things (IoT), to name a few, are driving the amount of data to be stored and processed to a higher level, making current data center solutions unsatisfactory as they do not satisfy this data storage demand.
in a cost-efficient way. Only IoT for example, has immense volumes of data to be collected from distributed IoT terminals and edge nodes that need to be stored, processed, analyzed, and managed in a data center based on the service. Loads of machine learning and deep learning tasks require extensive resources including storage for different parallel computing operations like derivatives, logarithms, and matrix multiplication with floating numbers. Even if these operations extract valuable business insights and train prediction models based on inputted raw data opening new business opportunities, innovative solutions are still required to handle such a massive amount of generated data.

Data centers are considered the core hub of the digital revolution and may play a crucial role in the continuous development of next-generation mobile systems, where the latter can be useful to help with many issues faced today’s data center technology.\textsuperscript{6–9} Indeed, data center vendors are looking for hybrid architectures based on wireless and mobile technology to ensure future businesses can run at the highest cost-effectiveness. The new technological trend will also reshape infrastructure, where, millimeter wave, optical beamforming, and multiple-input multiple-output (MIMO) techniques, will lead to new hybrid data center solutions, boosting capacity, data processing, energy efficient, and maintenance.

Conventional data centers, based on wired networks, entail high wiring costs, suffer from performance bottlenecks, and have low resilience to network failures. Limitations of space and power, along with the enormous complexity of managing a large data center can be now addressed with wireless mobile technologies. Moreover, wired topologies render the network static and inflexible with poor reconfigurability, which impacts on the energy efficiency and scalability of an overall data center. Wired schemes tend to further reduce and limit the layout design of new solutions, architectures, and even topologies. Wireless technologies, by its turn, are becoming mature and fully capable of replacing the wired infrastructure, and therefore they can not only be seen as an overlay solution anymore. Indeed, wireless mobile systems can provide reliability, performance, and security requirements of complex and critical processing data systems. Wireless data center networking has been introduced to improve data center connectivity and circumvent current bottlenecks or even to enable fully wireless data centers.\textsuperscript{10–15}

Innovative wireless terminals, air interfaces, and network transmission technologies are accelerating the evolution not only of devices, but also of architectures. Mobile data-centers can replace hundreds of meters of cable with millimeter wave (mm-wave) wireless connections along with servers addressing both intra- and inter-rack connections. Accordingly, mobile data centers will potentially attain higher aggregate bandwidth, lower latency, and substantially higher fault tolerance than a conventional wired data center while improving ease of construction and maintenance. Hence, radio technology based on 5G mobile and beyond has become a prospective hybrid solution for data center infrastructure. Several technologies such as mm-wave transmission and beamforming with phase antenna arrays have been developed within the scope of 5G mobile and beyond, i.e., converged optical and wireless networking with radio frequency (RF) carriers of tens of GHz enabling a new paradigm of wireless mobile network conception (here referred to as beyond 5G) so as to provide high capacity, density, flexibility, and reconfigurability data centers.\textsuperscript{11}

In this paper, wireless technology beyond 5G such as spectrum unification, in which frequency bands from 5G mobile up to the visible light window, covering mm-wave and THz regions, are addressed, as well as beamforming and self-configuring networks are discussed as a potential solution for wireless data center connectivity. The paper further discusses the implications on the network architecture and data center design coming from the introduction of wireless connectivity and highlights the potential gain from the resulting increased flexibility and reconfigurability. Discussing the relation of major advances in the mobile and wireless world to data center networking, this paper introduces the concept of a unified technological framework for different types of network based on beyond 5G technology, including an extension of the useable spectrum, adaptive beamforming and steering, as well as AI assisted self-configuring networks.

2. WIRELESS TECHNOLOGIES IN 5G AND BEYOND

The evolution of mobile and wireless technologies has been rapid between 3G, 4G/long term evolution (LTE) and 5G, enabling a rapid advance of the supported link and network performance in terms of data rates, connection density, overall throughput, achieved latencies and the range of employed spectrum. With the level of performance offered by 5G technologies, especially with the introduction of mm-wave transmission, the capacities and
Figure 1. Evolution of mobile telecommunications standards: achievable KPIs and underlying technologies.

capabilities of technologies developed for mobile communications start to be possible candidates for wireless data center deployments.

Extrapolating from the range of technologies introduced in 4G and 5G, to beyond 5G technologies discussed as candidates for 6G\textsuperscript{16,17} – as shown in Fig. 1 –, highlights a few technologies that are key to continue the growth of capacities and capabilities of mobile networks. The resulting link and network capacities – also shown in Fig. 1 – firmly puts beyond 5G technologies into the range of technologies meeting the requirements for wireless data centers.

Of the range of technologies discussed as the next steps beyond 5G and as key technology candidates for 6G, some are in direct continuation of the technology advances made in 5G, e.g., with respect to using ever higher carrier frequencies supporting ever larger bandwidths and in shaping RF beams to concentrate energy in a smaller region. Others, on the other hand, such as the introduction not only of software defined networks and radio, but self-configuring networks based on AI, as well as satellite roaming or even direct hum-to-human communication via implants, add new elements to the scope of technologies to be considered and open entirely new directions.

For wireless data center networks, it is mainly the technologies in the former group that are of interest, with the addition of self-configuring networks and the introduction of AI. The following will thus discuss the continued unification and extension of the usable spectrum, the use of spatial control via beamforming and -steering, and the introduction of AI and their benefit to wireless data centers in more detail.

2.1 Spectrum unification and extension of the usable spectrum

The evolution of the RF spectrum used in mobile networks has seen ever higher frequencies being introduced, extending the frequency ranges employed well beyond the traditional mobile bands and into the mm-wave, with the THz region or even optical frequencies considered beyond 5G or for introduction in 6G\textsuperscript{16–18}. The unification of the available spectrum and the concurrent use of spectrum from entirely different frequency bands in a single connection are key to increase efficiency in spectrum usage and achieve the ever larger capacities required from wireless links – in the mobile world, just as in wireless data centers.

With the first 5G specifications not only extending the available spectrum in the sub-6 GHz range, but more importantly opening up the mm-wave frequency range and introducing spectrum up to 40 GHz\textsuperscript{19}, they set an important trend, bringing higher frequencies into range and accelerating the required technology development. As frequencies up to 100 GHz are already firmly in sight for future use in subsequent updates of the 5G standards, the continuation of this trend towards frequencies in the higher mm-wave THz, and even optical frequencies is a clear strategy to augment the capacity of wireless networks beyond the few Gbit/s, into the range of tens and hundreds of Gbit/s\textsuperscript{20,21}. Figure 2 shows the evolution of the employed spectrum, from current networks existing in the crowded and congested sub-6 GHz range with its narrow channel assignments, over current 5G assignments...
for the first time providing contiguous spectrum of up to 3 GHz to the potentially available frequency bands for 6G and the even larger and very wide, contiguous range of interest for wireless data centers.

The unification of spectrum assignments in the mm-wave range with frequencies in the THz region – which are rapidly being explored by wireless communication research\textsuperscript{22, 23} – allows the use of large communications bandwidths and/or tailoring the employed technology to the specific link requirements. The THz region here is a key field of development, combining many of the properties of radio communications with the possibility to employ techniques for controlling and directing the signal in free space more commonly found in optics, such as collimators, lenses and mirrors.\textsuperscript{13}

Beyond the THz range, at much higher frequencies, optical wireless communications, where optical signal are transmitted through open space with collimated beams, rather than guided via optical fiber, can offer significant capacities and has been suggested as an alternative for data center networking.\textsuperscript{13, 24} While the typical telecommunications and fiber optics wavelengths – at 850 nm, 1310 nm and 1550 nm – are obvious candidate for such free space optics (FSO) implementations, the available spectrum is much larger and extends across the entire infrared region and beyond to visible light. As such the available spectrum is multiple times larger than in RF mm-wave and THz, making up for the potentially lower spectral efficiency, while benefiting from potentially improved transmission characteristics, immunity to RF interference, higher transmitter density and the chance for increased spectrum reuse. Based on these and the large range of available technologies especially in the telecommunications wavelengths, FSO is a candidate for wireless data center networks that is already under heavy investigation and may see usage even before the lower mm-wave and THz ranges are exhausted.

In wireless data center networks spectrum regulations may be somewhat relaxed compared to outdoor applications, especially for higher mm-wave, THz or optical frequencies where penetration through walls is very low or non-existent, spectrum unification offers more than three decades or ten octaves of contiguous spectrum, as also seen in Fig. 2. Mobile standards such as the currently developed 5G and those beyond have shown to drive a great acceleration of technology development in the related areas and with high frequencies now in focus, technology, components and equipment availability can be expected to improve rapidly. In such a development trajectory from 5G to 6G and beyond, spectrum unification will be a major cornerstone and the introduction of wireless data center networks as a viable and compelling usecase for beyond 5G technologies appears as a logical step.

2.2 Spatial control and distribution with beamforming and -steering

Spatial distribution and control of signals and the focus of emitted energy onto a confined area around the receiver plays a key role in wireless communications both in order to increase the received signal power – and as a result thereof signal to noise ratio (SNR) – and in order to minimize emission in unwanted directions, allowing increased spectrum reuse and avoiding interference. While traditionally achieved with high gain and highly directive reflector antennas, the move to phased antenna arrays is a key advancement where technology development was greatly fuelled by the emergence of 5G standards.

In such systems, an densely packed array of antenna elements – typically spaced at half the wavelength $\lambda$ – is fed with copies of the same signal, where progressive temporal differences between the signal copies across each dimension of the array allow directing the beam by changing the emission angle at which the different copies will
interfere constructively. Equivalently, a constructive addition of the signals received on the elements of the array can be achieved by matching the introduced delays introduced to the temporal delays resulting from the signal's angle of arrival in both horizontal and vertical directions. As a result, phased array antennas not only achieve similar confinement of the transmitted energy, but also allow changing of beam direction without mechanical movement of the antenna.

In the development of 5G, but also for the use in data center networks, not only the generation of a single, highly focused beam is of interest, but also the simultaneous transmission of multiple, independently steerable beams from a single array.\textsuperscript{12,13} The latter is key as it greatly increases spatial reuse of spectrum and allows concurrent transmission of many beams – the larger the array the larger the number of non-overlapping beams and the larger overall antenna gain --, which in combination enable the levels of connection density and required capacities for data center networks. The development of the required beamforming networks or arrays that adapt a number of incoming signals targeted at different beams into a number of outgoing signals to feed the antenna array – where outgoing signal each carries all incoming signals but at different differential delays -- has greatly progressed with the development of 5G,\textsuperscript{25,26} but further technological advances can be expected, reducing their cost, footprint and energy consumption.

While beamforming based on phased arrays of emitters is the dominant approach in the mm-wave domain, THz and especially free space optics (FSO) signals may allow for other strategies and other technologies are being developed. For instance, two-dimensional steering can be achieved using a micro-mechanical mirror or with fully integrated optical chips based on tunable lasers and waveguide gratings.\textsuperscript{13,24}

The potential to concurrently emit tens or even hundreds of high-bandwidth wireless signals from a single array while maintaining spatial separation and avoiding interference by beamforming is key whenever a high density of large capacity connections is required – i.e., in crowded areas with many mobile users and in wireless data centers alike. Similarly, the flexibility afforded by being able steer the beam is of great value in both mobile networks and wireless data centers, allowing dynamic following of users in the former and providing dynamic reconfigurability to meet changing load and requirements in the latter.

2.3 Adaptive and self-configuring networks and artificial intelligence

Artificial intelligence (AI) has been discussed for data center optimization and to reduce power consumption, but with continued development of AI and the ever growing complexity of data center networks and connectivity the potential gains from AI assisted self-configuring networks become more and more clear.\textsuperscript{27–29} Adding additional flexibility through wireless data center networking to achieve either full or partial reconfigurability (see section 3) greatly amplifies this gain, as not only the degree and configuration of connectivity can be optimized, but also the capacity per link tuned to the requirements.

The use of AI has been discussed in the radio domain under the term cognitive radio for the optimization of the radio signals to current link conditions or for local network optimization, however with the increased reconfigurability expected from fully software-defined networking (SDN) and network function virtualization (NFV) enabled networks in 5G and beyond its applicability is extended to a larger scale.\textsuperscript{27,30} For networks beyond 5G and in the next generations, AI may move from the role of local optimization towards a more global – and with sufficient virtualization potentially cross-operator – vision and optimization of the network as well as intelligent prediction of changing user locations and traffic demands.\textsuperscript{28,31}

In a similar fashion, it is the connection of knowledge about dynamic server loads, link availability and loads with that of network topology which suggests AI to have a bright future in dynamically reconfigurable data center networks. The combination of wireless data center networks and AI controlled or assisted networks thus appears as an ideal combination, where the inherent reconfigurability and more meshed connectivity of wireless networks augments the potential of optimization through AI and the latter alleviates the challenge of handling additional complexity coming from a more complex and variable network topology and multiple available technologies for any potential link.
3. WIRELESS DATA CENTER NETWORKS

The majority of the global research effort on the topic of data center networks has and is focused on wired networks and their respective components, topologies and performance. While architectures different to the well established and understood fat-tree architecture – such as butterfly, hypercube, torus or even mesh networks – can offer performance, cost or connectivity advantages, their adoption is limited and those that are practical to implement typically share one problem that is common to all wired data center networks: a lack of reconfigurability. The latter is of increasing importance, the more dynamic and/or modular data centers become and typically scales opposite to data center size, resulting in a challenging trade-off.

In today’s data centers it is common for thousands of devices such as computing servers, data storage and switches are arranged in racks, which in turn are arranged in rows or clusters a number of which form a section of a data center. Despite physical distances usually being introduced between rows or groups of racks, the vast majority of links in typical data centers and with common hierarchical data center network architectures are within the rack or its close vicinity and only a small fraction of links approaches or exceeds cable lengths of 100 m.\textsuperscript{13}

Wireless connectivity can be introduced at any level of the hierarchy of a data center network and at any physical level within the arrangement of devices, i.e., between clusters of racks, between racks within a cluster, between the devices within a rack or even within a single device. However, contrary to wired networks, network connectivity no longer needs to follow the logical and hierarchical physical layout of the data center and – provided spatial arrangement allows for a suitable link – any point to point connection within the same level of the hierarchy or across hierarchy levels can be established without regard for whether the involved nodes belong to the same group in the higher layers of the hierarchy.

Depending on the intended role of wireless connectivity, i.e., whether wireless technology is considered to alleviate local hotspots and bottlenecks or if a fully wireless data center is desired, the required physical modifications or re-designs of data centers may vary. While such modifications at first may appear like major changes, many alternative physical and networking architectures have been suggested.\textsuperscript{10,13} The use of ad-hoc wireless connectivity to alleviate local bottlenecks and hotspots, as shown in Fig. 3, requires less reconfiguration and can already provide significant gain in usage efficiency of data centers as situations where whole clusters of servers are slowed or even stalled by a single bottleneck link are avoided.

A more challenging scenario is the deployment of fully wireless data center networks where all connectivity must be achieved wirelessly and a significant amount of reconfiguration will be required. While the use of spatial control, i.e., beamforming and steering, as discussed in section 2.2, certainly helps in maximizing connectivity via line of sight (LOS) paths and with achieving the required link density, it will be impossible to ensure a LOS exists between any two arbitrary points in the data center. Initially the arrangement of servers within a rack into circular or hexagonal structures will alleviate the problem for intra-rack connections and can significantly increase the number of available connections between servers within adjacent or closely located racks,\textsuperscript{11,13} but for large data centers and such with densely packed clusters of racks additional measures will be required. Among the suggested solutions is the introduction of ceiling mirrors in specific locations to reflect high-frequency (i.e.,
high mm-wave or THz) radio or fso signals and thus multiply the number of reachable other racks and nodes. While often expected to be sufficient to achieve near complete connectivity among the racks in a cluster and in many cases even directly among a large number of the servers in the racks of a cluster, this solution is unlikely to be practicable to achieve connectivity across an entire large data center.

To achieve connectivity for the entire data center, it may be necessary to re-introduce a certain degree of hierarchy, i.e., while maintaining practically mesh-type connectivity within the rack and among many of the devices in the racks of the same cluster, it may be impossible avoid aggregation stages such as a top of rack (TOR) node or a node central to the cluster. Contrary to traditional data centers however, where such nodes handle both vertical traffic and horizontal traffic and their connectivity is mainly with devices of other hierarchy levels, in the case of a wireless data center networks only traffic crossing the boundaries of the local hierarchical cell must be handled and any node will additionally have mesh-type connectivity with closely located nodes of the same hierarchical level. Figure 4 shows such a scenario, where ceiling mirrors establish connectivity within a cluster, but cross-cluster connectivity will require a ceiling node, which may be either an active relay node or – where physically feasible – a multi-faceted changeable mirror where independent control of the facets may allow flexibly establishing many links between different clusters.

While the combination of introducing a whole multitude of new technologies, i.e., a whole range of wireless technologies from mm-wave to FSO and from antenna arrays, beamforming and steering to ceiling mirrors or similar devices for establishing LOS connections, and changing to a radically different connectivity paradigm in the data center network may appear daunting, a great deal of effort and re-design will be required for any technology introducing full flexibility and reconfigurability with the potential to scale to large data center sizes. Indeed the introduction of new data center network topologies with increased intelligence on the ingress and egress nodes of the network as well as in the network control offers the possibility to maximize the gain from a more meshed network and from the availability of different connection types – wired or wireless and in the latter case anything from mm-wave and THz to FSO – and may present the most viable option to scale to ever larger data centers, while maintaining the dynamicity and efficiency.

4. CONCLUSIONS

In this paper, we have introduced an innovative concept, based on beyond 5G technologies, which accounts for the unification of the complete frequency spectra range towards new solutions based on different technologies such as beamforming, AI, adaptive systems, and free-space optics for a full wireless data center connectivity solution. We further discussed networking architectures and topologies for future wireless data centers based on beyond 5G concepts that can profit from diverse technologies. Beamforming and steering technologies that contribute to further data center connectivity and to higher density. The proposed spectrum unification, covering from millimeter wave over THz to visible light allows wireless data center networks to support higher capacities, paving the way towards a single, interoperable, and flexible solution for wireless data centers in a cost-efficient...
manner. The main technology developments behind 4G, 5G, and beyond 5G have been addressed. The core performance indicators for 5G and beyond communication systems have been introduced, as have adaptive schemes and self-configuring networks along with AI. The technologies and techniques behind this potential solution have been briefly discussed. The introduction for the first time of AI for network optimization as well as novel network architectures are new open topics to be exploited within future data centers. Therefore, the 5G beyond concept can be regarded as an imminent solution in the coming years, where unpredictable massive data storage capacity and real-time flexibility capabilities in data centers will be essential.

ACKNOWLEDGMENTS

This work was partially funded by the blueSPACE, 5G-PHOS, Metro-Haul, and SILIKA projects with funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement numbers 762055, 761989, 761727, and 721732, respectively.

REFERENCES


