# 1.4 5G Wireless Communication: An Inflection Point

Vida Ilderem, Intel, Hillsboro, OR

#### Abstract

The 5G era is upon us, ushering in new opportunities for technology innovation across the computing and connectivity landscape. The advent of the Internet of Things has resulted in the generation of massive amount of data, thus creating both challenges and opportunities for industry. Today's networks will not be able to handle the variety and volume of data that is emerging. 5G presents an inflection point where the network will set the stage for data-rich services and sophisticated cloud applications, delivered faster and with lower latency, and where the wireless communication technology is driven by a multitude of applications and expected use cases. Therefore, 5G will optimize not only the network for human-human communication, but it will also deliver a solution for machine-type communications. This paper will review what is 5G and what are its key requirements, and will highlight the disruptive architectures and technology innovations required to make 5G and beyond a reality.

#### 1. Introduction

The advent of the Internet of Things (IoT), where things and devices are becoming more intelligent and connected, requires networks to become faster, smarter, and more agile in order to handle the unprecedented increase in volume and complexity of data traffic. The magnitude of this data challenge is expected to see personal data use reach 1.5GB per day, hospital data use to reach 4000GB per day, and connected factory use to reach an astounding 1 million GB per day [1].

Next-generation wireless communication, 5G, will require intelligence and flexibility across mobile networks and devices; the rollout of 5G from the cloud to the edge will transform the network to handle the data wave we expect in the coming years. From device to data center, Intel is delivering advanced capabilities, intelligent architectures, and powerful solutions that will unlock the future of 5G, see Figure 1.4.1. The wireless communication journey started with the First Generation (1G) analog voice communication. 2G introduced digital voice communication and text messaging. 3G enabled data services resulting in an affordable mobile Internet. The rapid penetration of broadband Internet access due to multimedia traffic workload led to 4G for improved capacity and higher data rates for mobile users. 5G is not simply an evolution of 4G. It presents an inflection point, not only due to major improvements required over 4G/LTE (Long Term Evolution) data rates and capacity, but also 5G will be the first wireless standard to address the inclusion of a massive number of machines/things in the network, some of which will require lower latency and higher reliability. Many pieces have to come together to successfully realize and deploy 5G. The following sections will review the need for network transformation, 5G key requirements for enhanced mobile broadband, and machine-type communication including spectrum requirements, key 5G enabling technologies, and architectures touching on Intel's research in architecture, beamforming, interference cancellation, antennas, RF design, and modems, finally concluding with a summarv.

# 2. 5G Network Transformation

The need for new network topologies is leading to innovations in dense heterogeneous networks, and efficient use of existing spectrum along with introduction of new spectrum.

Ultra-Dense Networks (UDN) is a new frontier and one of the key enablers for 5G which will shift our paradigms towards dynamic networks that are user/machine centric. UDN is comprised of three vectors: air interface, new spectrum, and network efficiency as depicted in Figure 1.4.2 [2]. Air interface covers new coding schemes, full duplex, and full-dimensional multi-input multi-output (MIMO) antennas with 3 to 5× improvement in capacity. New spectrum covers licensed, unlicensed, and shared-license, resulting in 3 to10× capacity increase. Network efficiency covers cell densification, device-device communication, joint multi-cell scheduling, and Wi-Fi offload, to name a few, with a potential of 40 to 50× capacity improvement. In summary, to increase the network capacity by about 1000×, and deliver quality of experience (QoE), innovations in massive MIMO, small cells, new spectrum such as mm-wave, multi-radio access technology (RAT) aggregation, and full duplex are required. Wireless-network densification is, however, not without its challenges such as interference and backhaul capacity, which are active research areas.

5G is designed to address machine-machine communication, referred to as machine-type communication (MTC), in addition to evolving for a better QoE for human communication. The topology for MTC will require agile flexible and high-throughput networks. MTC must address massive number of objects/devices that can impact network traffic, along with mission-critical applications that are sensitive to latency and/or reliability. Interactive virtual reality is an example of an application that requires a few milliseconds (ms) of end-end latency, while autonomous driving will need both reliable and low latency network. A new air interface, 5G NR (New Radio), is being standardized to address MTC use cases.

In IoT applications, the sensed data is mostly generated at the edge of the network, which requires a closer look at the role of centralized cloud computing in the network. Therefore, one of the observed trends is for compute and storage to move closer to where the data is being generated. Multi-access edge computing (MEC) enables distributed computing at the edge where we can intelligently partition complex workloads between edge and cloud to provide the best QoE for the user on various device types. MEC combined with 5G networking will be key to meet the bandwidth, latency, and reliability requirements of a massive number of connected IoT devices [3]. Examples include virtual- or augmented-reality applications, autonomous driving, and so on.

Thus, 5G requires a new network architecture to support a massive number of devices, efficient platforms, and mobile broadband through optimizing cellular network capacity, coverage, throughput, QoE, and spectrum sharing and efficiency. 5G enables devices that will be empowered with anytime/anywhere access to cloud/edge computing resources at 50× the speed, with 10× less latency, and 1000× the capacity of today's network via software-defined virtualized 5G networks. In the next section, 5G use cases and requirements to make this transformation a reality are reviewed.

# 3. 5G and Its Requirements

5G addresses three classes of communication services, see Figure 1.4.3: enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultra-reliable low-latency communication (URLLC). Each class of service has a divergent set of requirements from data rates to latency with different traffic characteristics as depicted in Figure 1.4.4, supporting applications above and beyond the traditional high speed connection. These varied and sometimes conflicting requirements have driven 5G to leverage different frequency spectrum, with <1GHz for mMTC, and <6GHz, in selected mm-wave bands for eMBB and URLLC. Even though eMBB is still an important use case for mobile operators around the world, emergence of other usage scenarios such as those classified under mMTC and URLLC enabled by 5G has opened the door for a myriad of non-traditional applications.

A challenge for 5G is the complexity and variety of its use cases. For instance, URLLC is required for autonomous systems, healthcare, and emergency services. mMTC is an enabler for smart cities, smart agriculture, and manufacturing, whereas eMBB delivers to augmented/virtual reality, mobile office, and entertainment. Each of these market segments have their own unique requirements and needs for bandwidth, response time, energy efficiency, and cost. Enabling these services will require wireless platforms and networks to sense, interpret, and act in real-time. Standardization efforts are underway to provide guidance on the essential functional support needed to cope with the wide collection of use cases 5G will make possible.

# 3.1 Spectrum – Requirements and Status

Spectrum is the center piece of any generation of wireless communication. The diverse 5G requirements mandate the streamlining of spectrum utilization, and exploring spectrum segments that were not previously available. Therefore, the emerging network capitalizes on a variety of radio interfaces across licensed, licensed shared, and unlicensed spectrum in low-, mid-, and high-frequency bands. Different physical characteristics of spectrum leads to different applications deployed in a suitable spectrum [4]. Low frequencies (those below 1GHz) provide obstacle penetration and non-line of sight (NLOS) operation in cluttered environments, whereas high frequencies (those in mm-wave) are needed for applications requiring very large bandwidths within a relatively short range. For applications where low frequencies do not provide sufficient bandwidth, for example eMBB in mixed line-of-sight (LOS)/NLOS situations, a spectrum range in between (for instance those in the 3 to 6GHz band) would be optimum.

A new air interface, 5G NR, is required to take full advantage of the advanced capabilities of 5G networks. One key feature that differentiates 5GNR from 4G is

self-contained frame structure as well as dynamic time division duplex (TDD) with instant cross-link measurement, which is critical to support the upcoming diverse requirements of connecting both people and machines/things, especially for low latency applications. Channel sizes, and other building blocks of 5GNR specifications are customized to the environment where these different frequency ranges would be used. The 3GPP Release 15 standard is the first specification of 5G NR, which defines the fundamental air interface and network architecture of 5G.

# 4. 5G Key Technology Enablers

5G use cases and real-time-data analytics are driving system, architecture, and technology requirements, while leveraging process nodes for density, energy efficiency, and performance. Due to the diverse set of applications, no single frequency band (low, mid, high) can meet all 5G requirements. It is widely acknowledged that Moore's Law enables new devices with higher functionality and complexity while controlling power, cost, and size. 5G enabling technologies can significantly benefit from not only process scaling, but also from advanced packaging for passive/antenna integration. In the following subsections, Intel's research solutions to address these 5G challenges are highlighted.

# 4.1 Low-Power Architectures

Several wireless protocols address the needs of IoT devices, including low-power long-range (LoRa), NB-IoT, Wi-Fi, Bluetooth, and so on. The key metrics for this market segment are energy efficiency and latency, depending on the application. One way to address energy efficiency is through re-architecting the radio at the system level. In both Wi-Fi and cellular protocols, the main radio is always listening or pinging the base station to check for data communication. Wake-up radio (WuR) is a new concept to optimize platform radio power consumption. WuR is a companion radio in Wi-Fi that serves as a wake-up for the main radio, and it is not meant to transfer or access data. Figure 1.4.5 depicts simulated savings for a Wi-Fi wake-up receiver with two different latencies. We can expect power savings of about 7× at 5 sec latency and over 220× at 100 msec latency [5]. Figure 1.4.6 shows the differences in architecture between a Wi-Fi and a cellular WuR (C-WuR). In the case of C-WuR the main radio is reused: however. the baseband is modified to implement the wake-up protocol. C-WuR power savings can be a combination of low-complexity C-WuR digital block, faster sync, and maintaining baseband receiver in deep sleep mode.

As the carrier frequency increases, the effective antenna aperture must be maintained to preserve acceptable link budgets. At mm-wave frequencies this often requires arrays of antennas, which also introduce directionality and hence the need for phased array systems to electronically steer the beam and preserve spatial coverage. Reducing the area and power consumption of each individual transceiver in the phased array is critical to allow scalability and system integration. For the same total isotropic radiated power, the power of each power amplifier (PA) decreases with the square of total number of elements in the array, potentially resulting in better overall efficiency and linearity, while enabling integration in low-cost technologies, such as CMOS. Direct-conversion (zero-IF) architectures can also lead to low power and low area transceiver designs, since no IF mixing stages with large resonant coils are required. Intel has designed a scalable E-band phased-array radio transceivers [6] in our 22FFL process [7]. The proposed architecture and design provides a low-cost solution for fullyintegrated phased-array systems with substantial scalability at silicon, package, and board level. This solution paves the way for fully-integrated digital beamforming phased-array systems, which will further boost network capacity.

# 4.2 Beam-Management Architecture

There are several "Gigabit"-class wireless solutions which manifest very substantial differences in terms of how they perform in practical deployments. The mm-wave spectrum from 24 to 86GHz can serve for both backhaul and network access scenarios, and paves the way for higher capacity and throughput usages in part due to a wider available bandwidth (800MHz to 14GHz). However, these frequency ranges are not without challenges, such as propagation particular characteristics due to atmospheric effects on link availability [8], and LOS/NLOS and blockage issues. Therefore, propagation and channel characteristics, beam management, MIMO/smart antennas, and system architecture must be studied and optimized. To take advantage of the higher available bandwidth, there is a need for wideband high-performance circuits and antennas. Beam management, including channel acquisition, beamforming, beam steering, and beam tracking is an important area of research to allow for spatial diversity and managing of interference, specifically in dense areas of use.

Figure 1.4.7 depicts three architectures for beamforming: analog, hybrid, and digital. Analog beamforming is commonly implemented in mm-wave phased array systems because of its simplicity and low power consumption. However, it may limit the beam-tracking capability of the mm-wave system since it is implemented at RF to avoid array-wide local-oscillator (LO) distribution. Furthermore, analog beamforming is not well suited to MIMO applications, where the array should steer, with full-array gain, in multiple directions simultaneously. In the hybrid architecture, the array is divided into subarrays, each with its own baseband and RF chain. This architecture permits each subarray to produce a unique beam, although the smaller subarrays may limit the performance relative to that of a full array. Finally, full-digital beamforming provides low latency beam management and optimal beamforming using all available antenna information for digital processing and storage. However, power consumption of this architecture is large compared to its analog beamforming counterpart due to the need for analog-digital converters (ADCs) and digitalinterface power dissipation. Therefore, we propose a digital beam management architecture that combines a direct-conversion frontend, high performance low power ADCs, and blind spatial compression to reduce interface throughput requirements. Antenna-element information is preserved for digital beam tracking algorithms to provide near-optimum low-latency beamforming. RF circuitry is implemented in scaled CMOS, permitting low-power digital processing. The solution results in an overall power consumption that is competitive with analog beamforming while offering improved performance.

### 4.3 Self-Interference Cancellation

Simultaneous transmit (Tx) and receive (Rx) signals in the same carrier frequency band, also known as full duplex (FD), is an active area of research aimed at doubling spectral efficiency. Figure 1.4.8(a) depicts the challenge for in-band FD and demonstrates how the Tx signal can impact the Rx signal. To address this issue, we have developed an echo-cancellation technique that is effective in cancelling this in-band self-interference [9]. Self-interference cancellation (SIC) needs to optimize the antenna or circulator isolated Rx/Tx antennas, along with the analog/RF SIC to address the Tx echo, and the digital SIC to address the remaining echo cancellation, see Figure 1.4.8(b). Our solution has resulted in a maximum of ~150dB SIC with an average SIC of 130dB using a 2×2 MIMO. Our closed-loop technique continuously updates the system parameters without requiring a special-training signal and synchronizations, such as a orthogonal frequency division multiplexing (OFDM) boundary. This results in fast and continuous tracking even during random data transmission; hence, it is agnostic to any air interface.

# 4.4 Antenna/Frontend Module

Development of multi-band/multi-mode mobile terminals in combination with implementation of MIMO operation for communication, present considerable challenges for efficient antenna-system integration in a wireless platform. MIMO antennas are a key technology ingredient to improve spectral efficiency, and thus network capacity and coverage. On the infrastructure side, MIMO dimensions will be large (sometimes referred to as massive MIMO (mMIMO)) to serve multiple users. On the terminal/edge side, the size of the MIMO is limited by the device/terminal form factor. One advantage of operation at mm-wave frequencies is a reduction in the antenna size due to the shorter wavelengths. Various beamforming networks (multi user-MIMO (mu-MIMO), mMIMO), as well as antenna-array antenna-element topologies at sub-6GHz, and mm-waves are being developed for 5G advancement [10,11]. Researchers are also exploring various approaches to embed sub-6GHz, as well as mm-wave bands, within the tightlyintegrated small-form-factor and challenging material environments of a smartphone. For instance, we have designed a quad-feed differential broadside antenna topology that exhibits >35dB of cross-polarization discrimination (XPD), which is capable of achieving more than 14GHz of bandwidth, demonstrating an antenna architecture integrated with CMOS RFIC [12].

Design and integration of the antenna system, RF frontend module (FEM), and CMOS RFIC are, as well, becoming more complex and challenging. Traditionally, antennas, FEM, and RFICs are designed independently, and assembled with interconnects and matching networks to create a wireless-communication system. Figure 1.4.9 presents a conventional mm-wave phased-array module cross-section with passive networks, power combiners, and impedance transformers. In this design, lossy interface networks and large impedance-transformation ratios degrade the total power and impact the system efficiency. Co-design of these components is desirable to optimize the effects of FEM on beam patterns and to avoid any performance degradation [13]. For example, the frontend integration and transceiver-performance-fluctuation effects on beam

patterns and steering antenna arrays can be identified with such a co-design approach. A dual-polarized (DP) MIMO Tx/Rx CMOS chip integrated with impedance-matched DP-broadside stacked patch antenna structure is shown in Figure 1.4.10 [14]. A proximity-coupled backside-excitation scheme is used to achieve an 8GHz antenna bandwidth on both vertical and horizontal polarizations. The simulated DP radiation patterns including wire bonds, matching circuitry, and antenna feeds show a 90° spatial 3dB bandwidth per polarization.

### 4.5 RF Design

4G/LTE evolution with more bands and tighter specifications presented challenges to form factor, power, and cost. 5G exacerbates these challenges through introduction of more diverse frequency bands and wider bandwidths. Figure 1.4.11 depicts a generic model of a radio. For instance, the wide-bandwidth and high-throughput mm-wave systems must not only provide power-optimized RF and baseband, but must also address the heavy burden on the power dissipation at the input/output (I/O) data interface between the ADC and the baseband processor at the receiver. Therefore, the fabrication process must simultaneously provide high-performance analog and RF/mm-wave transistors, combined with high-density and low-leakage digital transistors. While Moore's Law continues to benefit digital baseband performance, new processes must also provide RF specific optimizations in order to preserve wideband high-linearity and highly tunable transceiver design.

An excellent balance between integration density and RF/mm-wave performance has been demonstrated by FinFET technologies. The tri-gate transistor addresses short-channel effects, such as drain-induced barrier lowering (DIBL) and increased device leakage. The benefits are immediately apparent when comparing the I-V curves of planar devices, Figure 1.4.12 [15]. However, these deeply scaled CMOS can exhibit worse thermal dissipation compared to bulk CMOS due to higher local junction temperatures, referred to as self-heating [16], which must be taken into consideration during the design. In the absence of self-heating, energy-efficient designs must co-optimize transconductor efficiency and RF performance, resulting in device operations below peak F<sub>1</sub>/F<sub>max</sub> values of the process. In addition to transistor performance requirements such as gain, leakage, Ft, and Fmax, the quality factor for the passives and overall system parasitic due to interconnect and layout must also be addressed. We have leveraged Intel's 22FFL process to design and optimize an state-of-the-art CMOS E-Band PA with peak gain of 16.7dB, and 26.3% power-added efficiency, and a 11mW low noise amplifier (LNA) with 20dB gain and 4dB noise figure [17].

In summary, the feasibility of designing mm-wave frontend blocks in a deeply scaled CMOS process is demonstrated. In combination with faster and denser digital circuits, this technology will provide a platform for implementing low-cost, fully-integrated mm-wave phased-array system-on-chips (SoCs) for next-generation wireless systems.

#### 4.6 Modems

5G requires new modem technologies that can accommodate a wide range of use cases, air interfaces, and device form factors. It needs modems that can connect big/small things, fast/slow things, and consumer/industrial/machine things. Additionally, these modems will need to accommodate the previous wireless generation air-interface protocols, such as LTE and 3G.

The modem architecture must address the peak data-rate requirements, while the power requirements are a strong function of the use case. For instance, Internet browsing or voice calls require lower power envelopes, whereas a video download needs a higher data rate. Therefore, the modem design mandates co-optimization of power-performance and of course cost. The modem SoC must focus on dynamic voltage and frequency scaling (DVFS), power gating, network-on-chip-SoC subsystem state management, and so on. New circuit techniques, such as near-threshold-voltage (NTV) operation to maximize active energy efficiency becomes very important [18]. Advanced baseband algorithms such as error correction codes (ECC) such as polar coding and low density parity check (LDPC) could become complex to implement. Therefore, both architecture and circuit design must be co-optimized along with silicon process technology. One of the key blocks of a modem is memory, which makes density, power, and performance important parameters to balance. Logic power/performance must be optimized during the operation, whereas memory-power optimization becomes more important during sleep modes. Various design techniques can be used to optimize the memory size/power [19]. Consequently, application of Moore's Law for the modem is crucial where density/power/performance needs to be co-optimized. Integration of cellular and connectivity protocols will require an extensive amount of validation and testing to ensure that when we integrate all these radios, they will operate to deliver the desired QoE.

### 5. Summary

5G will not just reshape the wireless, computing, and cloud ecosystem, it will impact nearly every market segment with its ability to render everything smart and connected. Wireless networks must transform to become more powerful, agile, and intelligent, to realize the potential for IoT, and to enable richer experiences throughout daily life. Smart systems and future edge devices need to have capabilities that will require energy-efficient solutions at all levels, including architecture, computation, communication, sense/actuation, power management, and security, to name a few.

5G will encompass a broad variety of cellular and connectivity options including new spectrum such as mm-wave. A wide variety of technologies are required to enable 5G from the infrastructure to the terminal/edge devices. Moore's Law scaling is very critical to provide the power, performance, and cost requirements for this network transformation.

#### Acknowledgement

I would like to thank my Wireless Communication team and the Next Generation & Standards team for their support and contribution to this paper. I would also like to acknowledge and thank Dr. Vivek De for his support and input.

#### References

[1] Amalgamation of analyst data and Intel analysis, and VNI (Visual Networking Index) Global Traffic Forecast.

[2] Qian (Clara) li, *et al.*, "5G Network Capacity – Key Elements and Technologies," IEEE Vehicular Technology Magazine, vol 9, no. 1, p71, March 2014.

[3] ETSI White Paper on "MEC on 5G Networks", June 2018.

[4] 5G Americas White Paper on 5G Spectrum Recommendation, April 2017.

[5] E. Alpman *et al.*, "802.11g/n Compliant Fully Integrated Wake-Up Receiver With –72-dBm Sensitivity in 14-nm FinFET CMOS," in *IEEE Journal of Solid-State Circuits*, vol. 53, no. 5, pp. 1411-1422, May 2018.

[6] S. Pellerano, et al., "A Scalable 71-76GHz 64-Element Phased-Array Transceiver Module with 2x2 Direct-Conversion IC in 22nm FinFET CMOS Technology," *ISSCC Dig. Tech. Papers*, pp. 174-175, Feb. 2019.

[7] B. Sell, *et al.* "22FFL: A High Performance and Ultra Low Power FinFET Technology for Mobile and RF Applications," IEDM, p. 685, December 2017.

[8] T. S. Rappaport, *et al.*, "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!" IEEE Access, pp. 335-349, 2013(1).

[9] Y.S Choi, et al." Simultaneous Transmission and Reception: Algorithm, Design and System Level Performance" IEEE Transactions on Wireless Communications, Dec 2013.

[10] L-P. Shen *et al.*, "Dual-Polarized Wideband Multi-Beam Arrays in Wireless Communications," 2015 IEEE International Symposium on Antennas and Propagation (AP-S), pp.1544-1545, 2015.

[11] N. Ojaroudiparchin *et al.*, "Multi-Layer 5G Mobile Phone Antenna for Multi-User MIMO Communications," 2015 23rd Telecommunications Forum Telfor (TELFOR), pp. 559-562, 2015.

[12] Chintan Thakkar, *et al.* "A 42Gb/s 4.8pJ/b 60GHz Digital Tx with 12b/symbol Polarization MIMO," *ISSCC Dig. Tech. Papers*, pp. 172-173, Feb. 2019.

[13] D. Choudhury, "mmW Antenna Integrated Front-End IC-Module Co-Design and Testing for 5G Applications," IEEE IMS2017 workshop WSC presentation, Honolulu-HI; June 4, 2017.

[14] S. Daneshgar *et al.*, "A 27.8Gb/s 11.5pJ/b 60GHz Transceiver in 28nm CMOS with Polarization MIMO," *ISSCC Dig. Tech. Papers*, pp. 166-167, Feb. 2018.

[15] M. Bohr. "Intel 22nm Tri-Gate Announcement, Apr. 2011, https://www.intel.com/content/www/us/en/silicon-innovations/standards-22nm-3d-tri-gate-transistors-presentation.html,

[16] M. Shrivastava *et al.*, "Physical Insight Toward Heat Transport and an Improved Electrothermal Modeling Framework for FinFET Architectures," in IEEE Transactions on Electron Devices, vol. 59, no. 5, pp. 1353-1363, May 2012.

[17] W. Shin *et al.*, "A Compact 75 GHz LNA with 20 dB Gain and 4 dB Noise Figure in 22nm FinFET CMOS Technology," RFIC 2018.

[18] Jain, S. *et al*, "A 280mV-to-1.2V wide-operating-range IA-32 processor in 32nm CMOS", *ISSCC Dig. Tech. Papers*, pp. 66-68, Feb. 2012.

[19] Khellah, M *et al.*, "A 256-Kb Dual-VCC SRAM Building Block in 65-nm CMOS Process With Actively Clamped Sleep Transistor," IEEE Journal of Solid-State Circuits, Volume 42, Issue 1, Page(s):233 – 242, Jan. 2007.



2019 IEEE International Solid-State Circuits Conference



