WP476 (v1.0) June 13, 2016

Toward 5G
Xilinx Solutions and Enablers for Next-Generation Wireless Systems

By: Sassan Ahmadi, PhD

Xilinx envisions a virtualized, programmable, flexible 5G network with edge computing capability, based on open and standard interfaces. Xilinx solutions and products play a significant role in the implementation, trials, and commercialization of pre-5G and 5G technologies, enabling and promoting flexibility in network design and deployment.

ABSTRACT

The 5th generation of wireless systems will usher in tremendously revolutionized mobile communication systems that can provide ubiquitous, extremely high-throughput, and low-latency user experiences anytime, anywhere. Significantly increased system capacity and real-time responsiveness of 5G systems will enable new services, providing the users with a truly immersive and rich experience. The realization of the mission-critical machine communication or IoT requires reliable connectivity among massive number of devices, which can only be fulfilled via 5G systems. Xilinx is playing an important role by providing state-of-the-art all-programmable FPGA platforms and SoCs that can accelerate development of 5G proof of concepts, technology trials, and early implementation of large-scale commercial wireless products for 5G networks.

Introduction

The 5th generation (5G) of mobile broadband wireless networks are expected to meet the challenging system and service requirements of the existing and the new applications in 2020 and beyond. 5G is envisioned to be a fully-mobile and widely, securely, and reliably connected ecosystem comprising humans and machines, enabling more efficient connectivity, increased industrial productivity, automation, sustainability and improved health care. The demand for a mobile and connected society is characterized by substantial growth in connectivity and traffic density, heterogeneous network densification, and a broad range of new use cases and business models. As a result, there is a need to push the envelope of the 5G system/network performance to the limits to guarantee higher network capacity, higher user throughput, higher spectrum utilization, wider bandwidths, lower latency, lower power consumption, more reliability, more connection density, and higher mobility through virtualized and software-defined networks.

While extending the performance envelope of mobile networks, 5G is expected to include, intrinsic flexibility and configurability to optimize the network usage and services, accommodating a wide range of use cases and business and partnership models. The 5G architecture will include modular network functions that can be deployed, configured, and scaled on demand to accommodate various use cases in a smart and cost efficient manner.

Xilinx has been a leading provider of wireless/wireline infrastructure all-programmable semiconductor components for digital radio front end, connectivity, baseband acceleration, and fronthaul/backhaul and packet processing functions. Programmability, performance, flexibility, and scalability of Xilinx silicon products are the key value propositions that make these devices perfect solutions for these applications with the fastest time to market. As a result, Xilinx is playing an important role by providing state-of-the-art all-programmable components that can accelerate the development of 5G proof of concepts, technology trials, and large-scale commercial deployment of wireless networks and products.

5G Vision

The scope of 5G is best understood by scrutinizing three usage models targeted by this megatrend: high throughput mobile broadband (eMBB), ultra-reliable low-latency communication (URLLC), and low-energy massive machine type communication (mMTC) [Ref 1]. The high throughput mobile broadband use case will be based on the evolution and enhancement of LTE technology. The evolutionary nature of the mobile broadband use case implies inherent backward compatibility but this consideration is limited to the sub-6GHz RF spectrum. Any mobile broadband deployments above 6GHz are expected to use new baseband and radio technologies that improve spectral efficiency, throughput, and latency. The mission-critical machine type communication and massive machine type communication use cases are revolutionary in nature. Technology and design choices for the latter use cases are expected to be unencumbered by the burden of legacy or backward compatibility. See Figure 1.

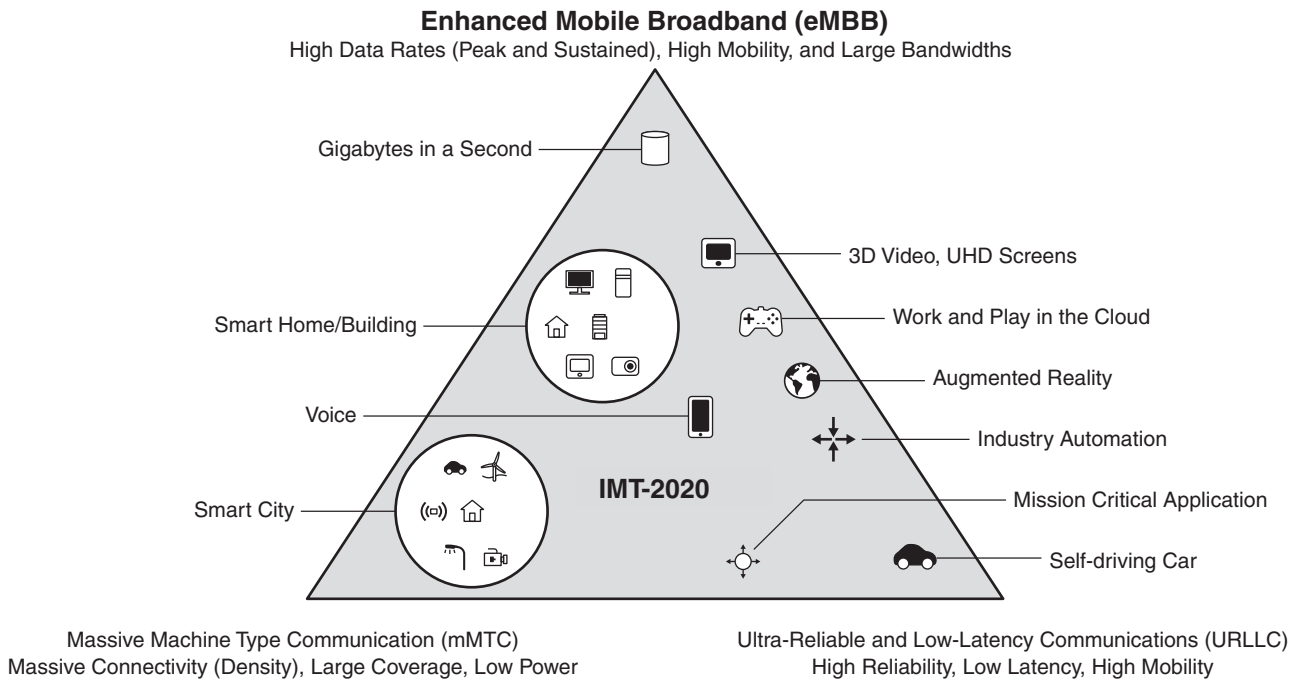**Enhanced Mobile Broadband (eMBB)**
High Data Rates (Peak and Sustained), High Mobility, and Large Bandwidths

Gigabytes in a Second

3D Video, UHD Screens

Smart Home/Building

Work and Play in the Cloud

Augmented Reality

Voice

Industry Automation

IMT-2020

Mission Critical Application

Smart City

Self-driving Car

Massive Machine Type Communication (mMTC)
Massive Connectivity (Density), Large Coverage, Low Power

Ultra-Reliable and Low-Latency Communications (URLLC)
High Reliability, Low Latency, High Mobility

WP476_01_050916

Figure 1: **5G Use Cases and Requirement Categories (Source: ITU-R [Ref 2])**

As shown in Figure 1, peak and average data rates as well as transport delay are the leading performance indicators of the high throughput mobile broadband experience. Interactive gaming, augmented/virtual reality, and immersive entertainment are notable consumer mobile broadband services in the nascent state and hold significant promise of new business opportunities and experience. 5G is expected to deliver over 10Gb/s peak data rates with average data rates of approximately 100Mb/s—anywhere, anytime. This will require more than a 1,000X increase in network capacity by 2020. Network operators are gearing up toward meeting such a dramatic increase in capacity by a combination of new spectrum utilization, spectral efficiency improvements, and ultra-dense network deployments.

The Industrial Internet of Things (IoT), automotive, smart grid, traffic safety, and emergency response services are some examples of low latency machine type communications. Today, some of these applications are using proprietary communication technologies that are expensive and difficult to maintain and scale. The goal of 5G is to introduce reliable, resilient, low latency, and highly secure communication to accelerate the much needed transition to standard solutions. Tactile Internet is an interesting aspect of this use case that is expected to introduce new services.

Utilizing the RF spectrum above 6GHz for mobile broadband, low latency mission critical machine type communication, and low energy massive machine communication requires development of new radio interfaces and access technologies. However, 5G is not only about the development of a new radio interface. Rather, 5G is expected to operate in a highly heterogeneous environment characterized by the existence of multiple types of radio access technologies, multilayer networks, devices, user interactions, etc. In such an environment, there is a fundamental need for 5G to achieve a seamless and consistent user experience across time and space. New business models and economic incentives with fundamental shifts in cost, energy, and operational efficiency are expected to make 5G feasible and sustainable. 5G also is going to enable value creation towards

customers and partners through the definition and exposure of capabilities that enhance today's overall service delivery. Enabling 5G use cases and business models requires the allocation of additional spectrum for mobile broadband and flexible spectrum management capabilities.

5G networks will employ a set of new frequency bands to augment existing wireless frequencies, enabling larger bandwidth and substantially improved capacity. Such tremendous connectivity and capacity will demand many more base stations or access nodes. These will be small cells and high-capacity C-RAN or V-RAN architectures serving a large number of remote radio heads. Content caching and processing at the cloud edge in 5G networks will overcome bottlenecks in the network transport for delay-sensitive applications.

5G will inevitably require some of the sub-6GHz spectrum to be repurposed for use with the new technologies. Existing cellular bands will be augmented with new spectrum allocations above 6GHz to create wider operating bandwidths. To further improve peak data rates, carrier aggregation techniques will be used to combine segments of spectrum that are not co-located within the same band. The core bands will provide up to 100 or 200MHz of contiguous bandwidth and the new extended bands will provide a total bandwidth of up to 2GHz.

In addition to supporting the evolution of the established mobile broadband usage models, 5G will support numerous emerging use cases for a wide-range of applications with versatile performance attributes, e.g., from delay-tolerant applications to ultra-low-latency applications, from high-speed scenarios in trains/airplanes to nomadic/stationary scenarios in home or office and from best-effort applications to ultra-reliable applications such as in healthcare and public safety. Furthermore, content and information will be delivered across a wide range of devices, e.g., smartphones and wearables, and through heterogeneous environments.

Advanced driver assistance systems (ADAS) and autonomous vehicles are emerging trends in the automotive space. Together, they bring a number of benefits, including improved safety, reducing collision risks and road congestions, improved fuel economy, and higher productivity for the drivers. 5G wireless technologies that support high-speed, low-latency vehicle-to-vehicle and vehicle-to-infrastructure communications are key enablers of ADAS and autonomous vehicles. In addition, the automotive industry is demanding richer infotainment options, which are adding more traffic bottlenecks to wireless networks.

As in past generations, improvements in radio access technologies, evolutionary and revolutionary changes, will continue to be the core focus. However, network architecture, functionality, programmability, and APIs will become equally important to 5G in order to meet the goal of an integrated communication platform that can serve multiple vertical markets and diverse consumer categories. Network function virtualization will be the core component to allow carriers to create network slices per end-user application or business vertical with guaranteed performance and service-level agreements. Both cloud and edge computing components will be needed in these network slices to address varying performance and latency requirements. Each network slice will rely on enhanced LTE and/or new radio access technologies.

The ongoing 5G research and standards development is revealing a number of key technologies that are expected to help achieve the ambitious system requirements, including:

- Millimeter wave spectrum: Using a higher RF spectrum (up to 100GHz) will allow the use of wider bandwidths (up to 2GHz), which will lead to higher peak data rates and system capacity.

- New waveforms and multiple access schemes: An OFDM-based LTE air-interface might not be suitable for some use cases, and therefore, a number of new air-interface candidates are being studied. OFDM with adaptive numerology would allow configurable frame structures and radio resource planning based on various use cases, spectrum and bandwidth without affecting backward compatibility with legacy LTE systems.

- Massive MIMO: Extension of multi-user MIMO concept to hundreds of antennas at the base station is a promising solution to significantly increase user throughput and network capacity by allowing beamformed data transmission and interference management. The significantly increased pathloss in very high-frequencies has to be compensated by higher antenna gains, which is made possible by increasing number of antennas at the base station. Additionally, adaptive beamforming algorithms are required and can be implemented using active antenna technology.

- Device-to-device (D2D) communication: While being an existing use case for LTE to support public safety requirements, D2D communications would also allow low-latency for new proximity-based applications and ad hoc mesh networks.

- Network virtualization cloud-based radio access and core network: The aim is to run today's network functions on the dedicated hardware as virtualized software functions on general-purpose hardware in the core network. This is extended to the radio network by separating base stations into remote radio units and baseband processing units (connected via optical fiber links, high-speed Ethernet cables and wireless front-hauling) and pooling baseband functions in the virtualized environment to handle a high number of radio units and to form different network slices along with core network functions.

- Splitting control-plane and user-plane and/or decoupling downlink and uplink control/data paths: Making it possible to control all user devices on a macro layer, whereas user data is independently sent/received via one or several small cells in a heterogeneous network.

- Optimized radio resource management: Considering the high number of potentially very small cells and dense distribution of users, radio resource management needs to be optimized, requiring more efficient scheduling strategies that can also be deployed in uncoordinated scenarios.

See Figure 2.

Key Performance Indicators (KPIs)   **Candidate Technologies**
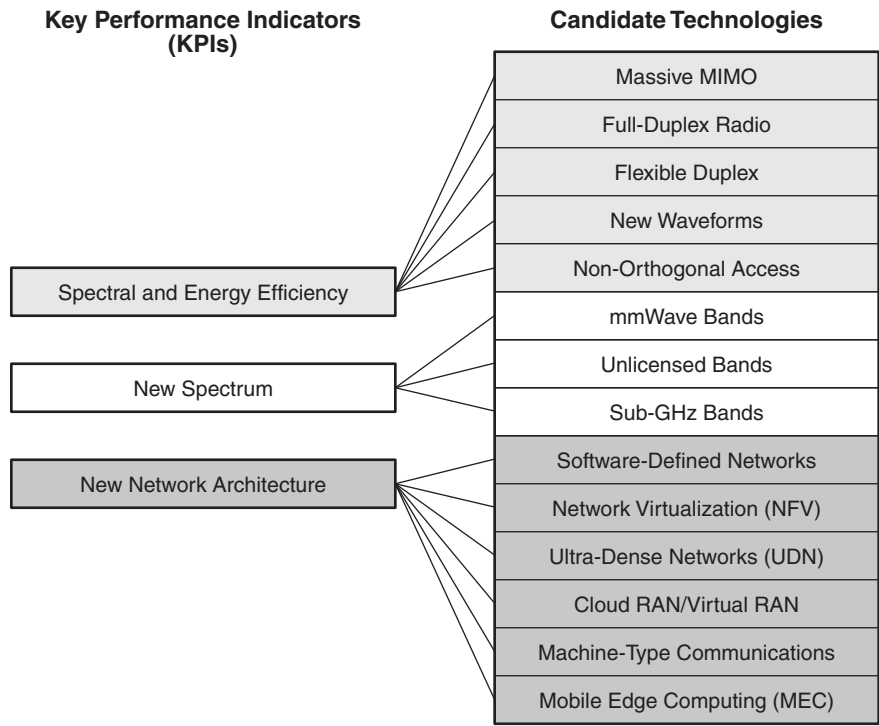
WP476_02_042616

Figure 2: **Mapping 5G Candidate Technologies to KPIs [Ref 3]**

5G System and Service Requirements

The system and service requirements along with the key performance indicators (KPIs) for 5G systems will define the capabilities of the future networks and will provide a framework for 5G technology development. The user experience depends on the type of applications, network capabilities, and coverage quality, which will translate into various sets of KPIs with specific system requirements. As shown in Figure 3, peak data rates in excess of 10Gb/s is a basic KPI for a 5G network, but a minimum guaranteed user throughput of 50–100Mb/s might be more important when it comes to accessing the Internet ubiquitously. Moreover, the 5G ecosystem must ensure high data rates are sustained in various mobility scenarios in the range of zero (stationary) to 500km/h (high mobility). Besides the data rates and mobility, the traffic density and connection density must be included in capacity requirements. End-to-end latency at the millisecond level and radio link latency of less than 1ms are required to ensure support of tactile Internet, mission critical, and interactive applications.
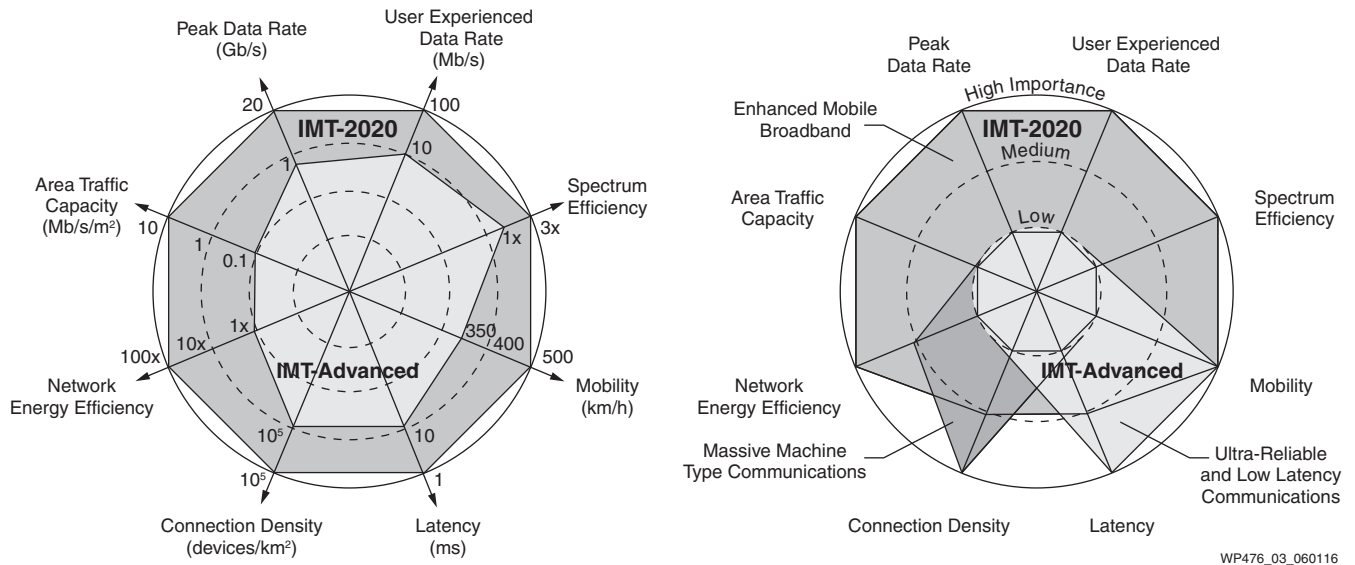
Figure 3: **Enhancement of Key Capabilities from IMT-Advanced to IMT-2020 (Source: ITU-R [Ref 2])**

The 5G use cases demand very diverse and often challenging requirements, where a single solution to simultaneously satisfy all requirements can lead to over-engineering, complexity, and higher cost of implementation and deployment. Nevertheless, some applications and use cases might need to concurrently be provided by the same network, requiring a high degree of flexibility and scalability for the 5G networks. To reflect their use case dependency, the 5G requirements are specified according to the use case scenarios. For each category, one set of requirements represent the extreme use cases in the category. As a result, satisfying the requirements of a category leads to satisfying the requirements of all the use cases in that category. The following sections describe various aspects of 5G radio access network requirements and the candidate technologies under consideration.

Network Architecture

The network architecture not only plays an important role in enabling new 5G services and accommodating exponential traffic growth, but also in supporting multiple radio access technologies (RATs) in a flexible manner. Current mobile core networks are implemented in the form of tightly-coupled software and hardware platforms. Many EPC network functionalities such as network monitoring, IP address allocation, QoS, security, and charging are performed in dedicated entities. The cost of upscaling the network capacity is expensive because it often requires expensive equipment installations. To resolve the high cost structure, SDN-based virtualized mobile core network infrastructure must be used.

5G mobile network architecture based on C-RAN/V-RAN, SDN and NFV concepts as well as Xilinx focus areas are shown in Figure 4. In this framework, the control and user planes are decoupled and the control plane is (logically) centralized in the network controller whereas the user-plane is positioned closer to the users (mobile edge computing). The network functions are decoupled from dedicated hardware and are implemented in software that can run on virtual machines instantiated on generic computing platforms. The network operators have the flexibility to choose the network resources in different locations, which means that network services and the resources can be configured on demand. On the core network side, network functions are virtualized and

modularized, thus new services can be easily developed and tested. On the radio access side, some of the control functions can be logically centralized in the radio controller to perform joint optimization to improve user experience. Furthermore, the use of open network interfaces allows a simplified representation of the network to the applications via APIs, streamlining conception of new services and applications.
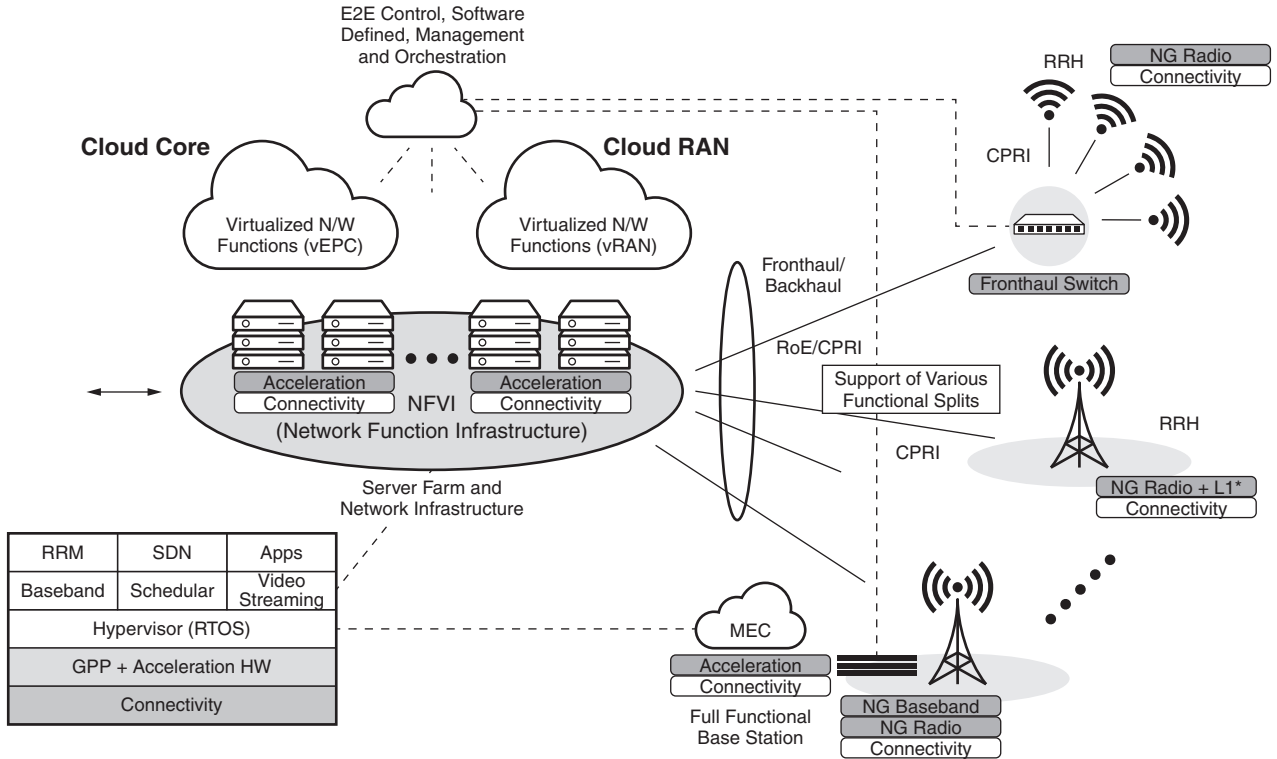
*Figure 4:* **5G Network Architecture and Xilinx Value Proposition**

Despite numerous attractive advantages of SDN and NFV technologies in 5G network architecture, the true realization of this concept is quite challenging because it is not proven whether the standard hardware can meet the performance and reliability requirements of cellular networks. The standardization of the interfaces and protocols of the controller and forwarding hardware and ensuring interoperability require substantial efforts. Furthermore, the potential security threats and solutions brought about SDN and NFV need to be comprehensively studied.

The network function virtualization along with software-defined networking concepts will significantly simplify the implementation of the core and access network entities, enabling cost-efficient network deployment, operation and management. In a software-defined network, the computing, storage and radio resources are virtualized and centralized to achieve dynamic and user-centric resource management compatible with service requirements. Software-defined networks are expected to build upon a carrier-grade cloud platform in order to enable network-as-a-service featuring open-interface network capability and network sharing. This makes it possible to achieve network flexibility and scalability and provide users with a variety of services and consistent quality of experience. In the access network, a variety of network architectures ranging from high-capacity macro base stations to low-power low-cost femto-cells or from an ultra-dense urban network to rural area coverage are realizable.

Xilinx All Programmable technologies enable a wide range of SDN/NFV applications, including intrusion detection, virtual switching/routing, and traffic engineering, as well as on-the-fly provisioning of services. Xilinx supports optimized workload acceleration, efficient management and routing of data flows, a wide range of communication protocols, and programmable data plane acceleration on demand. Xilinx provides a flexible, standards-based solution that combines a software programmable data plane, VNF workload optimization, and high-performance connectivity with the security needed for SDN/NFV networks. To support SDN, Xilinx's SDNet™ software-defined environment is used to build programmable data plane solutions to meet the network's unique performance and latency requirements. To support virtualized and software-defined networking, Xilinx's SDAccel™ software-defined environment is used to accelerate the performance of virtualized network functions.

As shown in Figure 5, when migrating from the traditional Cloud Radio Access Network (C-RAN) model, where all baseband functions are performed in the baseband units running on customized hardware, to virtualized radio access networks where RAN functions become virtual machines running on top of general-purpose hardware, various forms of virtual machines can be defined and RAN functions can be split across the baseband processing unit (BBU) and remote radio head (RRH) depending on the deployment scenario. Hardware accelerators are used to ensure computationally-intensive real-time functions can be performed without performance degradation. One of the areas where hardware acceleration plays a critical role is the network interface controllers (NICs) and virtual switches (VS). While both ASIC-based and FPGA-based network interface controllers can be used in practice, the former has to be pre-designed to support desirable features whereas FPGA-based NICs do not have to support the full feature set. With partial reconfiguration, Xilinx FPGA-based NICs are able to support different features by loading different bit images at any time.
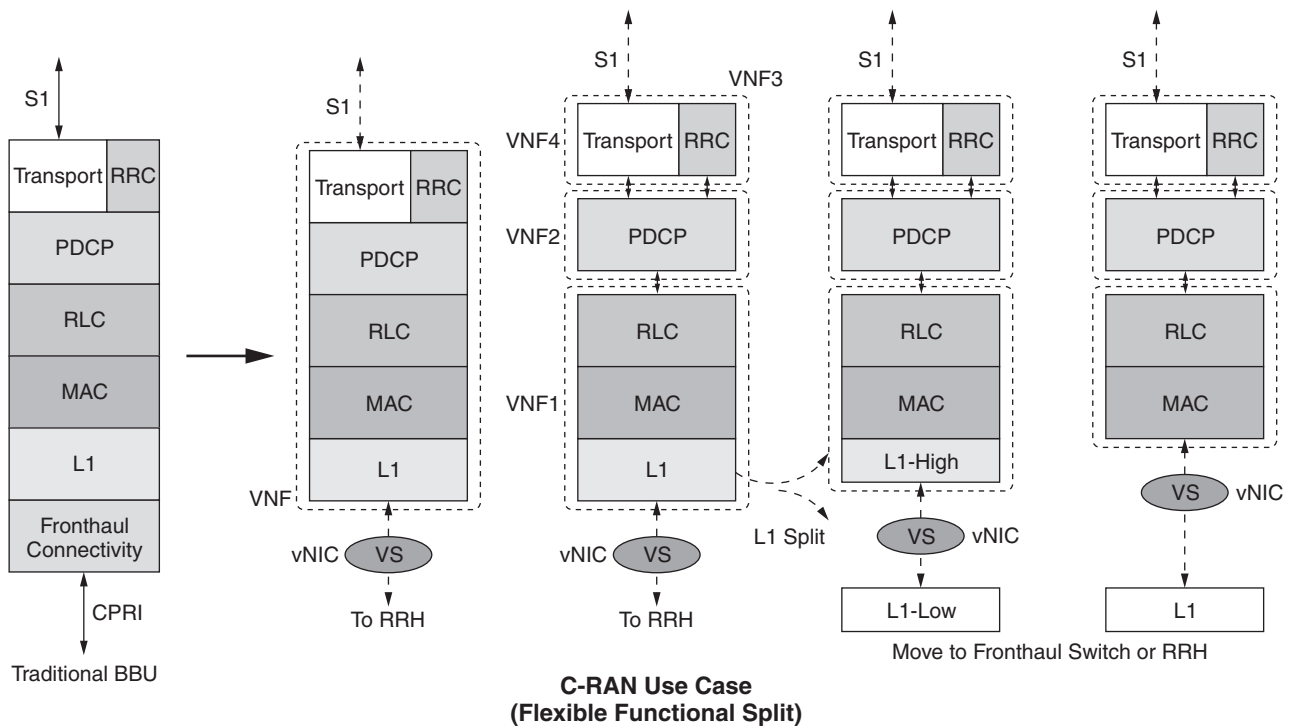
WP476_05_051616

Figure 5: **C-RAN and Baseband Network Function Virtualization**

Figure 6 illustrates two alternative data center (and C-RAN) server models. Traditional data centers have a relatively static computing infrastructure with certain number of servers, each having a predefined number of CPUs, and a fixed amount of memory, storage, and hardware acceleration resources. However, due to the dynamic nature of network loading and user traffic and in order to ensure availability of sufficient resources to meet the peak demand, more servers are typically provisioned—some of which might become underutilized during non-peak loading and traffic conditions. Disaggregating the data center (cloud) resource components would provide the operators with flexibility, reduced CAPEX and OPEX, and on-demand resource provisioning and availability while ensuring optimal utilization. Xilinx FPGAs and MPSoCs provide on-demand hardware acceleration and fast connectivity for NFV services while reducing total cost and power consumption. The run-time programmable solutions create service differentiation for the network operators and accelerate services with substantially improved performance and deterministic QoS. The FPGA size can be further reduced by using Xilinx partial reconfiguration technology. The SmartNIC solution from Xilinx provides acceleration on demand and fast network connectivity for virtual switches in the virtual C-RAN architectures, achieving maximum throughput and portability for the virtual machines.
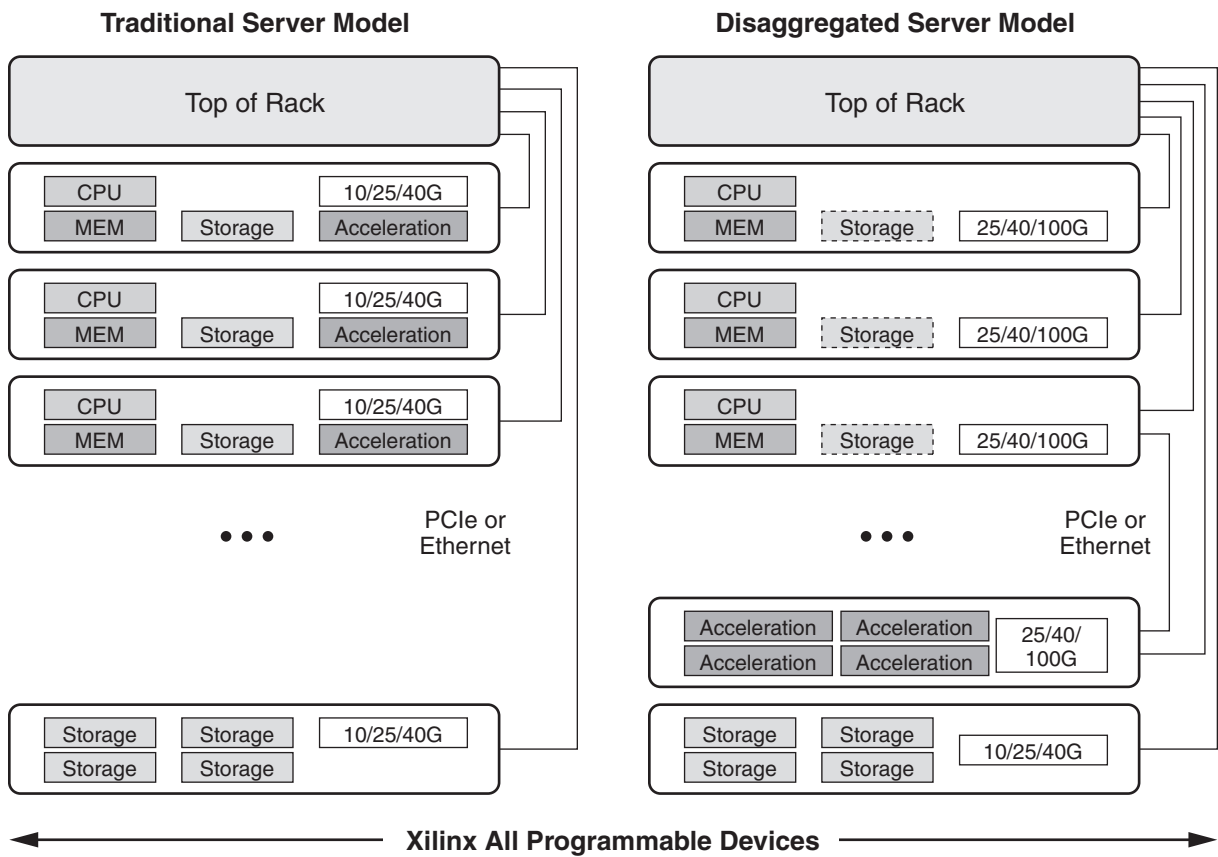
Figure 6: **Xilinx SmartNIC and Programmable Acceleration Solution for Mobile Telco Cloud and Data Center**

As a promising architecture, C-RAN exploits a combination of virtualization, centralization, and coordination techniques with separation of control and user planes. In C-RAN architecture, network functions in core network and access network can be close, even on the same platform, which helps reduce latency and improves QoE. An alternative approach known as mobile edge computing (MEC) architecture places computing resource at the edge of the access network near the base

stations and mobile users. MEC has the potential to offer an ultra-low latency, high-bandwidth environment that provides real-time access to radio networks at the edge of the mobile network. There are challenges in the realization of MEC architecture that need to be addressed in different deployment scenarios, including mobility support, security, resource management, etc. As shown in Figure 4, hardware acceleration is one of the Xilinx competencies in mobile edge computing. Furthermore, Xilinx offers extremely efficient solutions for C-RAN backhaul/fronthaul connectivity, digital radio and baseband processing in the remote radio heads, and the baseband units. RAN virtualization enables multiple tenants with many relatively small VMs to share compute, storage, and network resources in an efficient manner.

The network aspects of 5G requirements must include the following:

1. Programmability and scalability of the network architecture based on SDN and NFV

2. Ability to provision the right network slices depending on the service requirements

3. Coexistence and integration with non-3GPP technologies (e.g., Wi-Fi)

4. New business enablers, including eMBB, mMTC, and URLLC use case categories

5. Improvement of spectrum utilization efficiency, energy, and cost of implementation and deployment

The 5G core network can use virtualized functions and support efficient and low-cost reconfiguration, upgrade, and orchestration of the network entities with minimal or even no requirements for changing network hardware. The control and user planes must be decoupled with open interfaces between them. 5G core network functions are expected to be highly modularized and virtualized. The functions and capabilities are called upon by relevant APIs and form different network slices on demand. The different types of network functions performing similar tasks should be minimized and a trade-off between flexibility and complexity must be made. A group of dedicated and virtualized network entities implement a network slice for selected use cases where network slices share the infrastructure resources.

To support the diverse use cases and requirements in a cost-efficient manner, the system design methodology should move away from the traditional monolithic design optimized for mobile broadband. Therefore, reconsideration of design principles in the areas of control/data bearers, APNs, gateways, tunneling, and mobile IP are needed. In addition, the UE state-machine and entities that store UE context should be revisited and redesigned and mandatory functions should be minimized.

To enable massive connectivity for a wide range of new applications and use cases, the capabilities of 5G radio access networks must be extended beyond previous generations of cellular networks. Examples of these capabilities include very high peak data rates and user throughputs, very low latency, ultra-high reliability, and the ability to admit and serve massive number of connected devices. These objectives will be realized by the continued evolution of 3GPP LTE RAN in conjunction with development of new radio access technologies. Key technology components include extension to higher frequency bands, advanced multi-antenna transmission schemes, very efficient and low-overhead radio access design, user plane and control plane separation, flexible spectrum usage, complementary device-to-device communication, and backhaul/access integration.

Flexibility, intelligence, and programmability can be leveraged in radio access technologies in a cost-efficient manner if all or part of RAN functions are performed in the cloud environment. This also enables resource sharing and radio service on demand to form different network slices, which lowers the latency for the interconnection between radio access and core networks while both are running in a centralized data center environment. The C-RAN concept is about centralization, virtualization, and coordination of RAN functions to provide a better user experience and to accelerate time to market of new services. In 5G networks, distributed and C-RAN will coexist in heterogeneous access network deployment scenarios depending on the transport network availability and performance.

Increasing the data rate and the operating bandwidth would constrain a fronthaul transport network where CPRI is the most common interface over dedicated fiber, requiring tighter constraints on latency and jitter on connections to the RF modules. More affordable transport solutions, e.g., radio over Ethernet and next-generation fronthaul interface (NGFI), can be exploited with different baseband functional split options, reducing the fronthaul requirements and increasing transport efficiency. For example, L1 or part of L1 can be integrated into an RF module close to antennas, which can improve the fronthaul bandwidth efficiency up to 10X.

While LTE will continue to evolve in a backward compatible manner and will be an important part of the 5G wireless access solution for frequency bands under 6GHz, parallel development of a new radio access technology with no backward compatibility constraints in the new spectrum is inevitable. For network operators with limited spectrum assets, the possibility to introduce 5G capabilities in a backward compatible manner is highly advantageous. In the longer-term, the new non-backward-compatible solutions might also be extended to the spectrum below 6GHz. Although the overall 5G radio access networks will consist of different components, including the evolution of LTE as well as new RATs, different components must be highly integrated with the possibility of tight interworking for faster transitions/mobility. This includes dual-connectivity between LTE operating in lower frequencies and new technologies in higher frequency bands. The possibility of user-plane aggregation or joint delivery of data via both LTE and a new RAT to the mobile devices should also be considered.

Design of radio access networks with minimal overhead is important to achieve high efficiency in future wireless networks. The underlying principle of ultra-efficient air-interface design is to minimize any downlink/uplink transmission that is not directly related to the delivery of user data. Such transmissions include signals for synchronization, system acquisition, and channel estimation, as well as the broadcast of different types of system and control information. Ultra-efficient design is especially important for dense deployments with a large number of network nodes and highly variable traffic conditions by enabling network nodes to rapidly enter low-energy states when there is no user data transmission. Ultra-efficient design will also enable higher achievable data rates by limiting resource wastage and reducing interference from unnecessary transmissions.

Capacity and User Throughput

The 5G eMBB scenario will be a viable substitute to fixed access in terms of capacity and user throughput. To efficiently support the exponential growth in traffic demands, 5G networks must be able to deliver user data with lower cost per bit compared to the existing networks. Furthermore, to operate with relatively lower energy consumption, 5G systems must enable drastically lower energy per bit. Another aspect of 5G system capacity is the capability to support a very large number of devices compared with today's systems. Some 5G use cases encompass the deployment of massive

number of wirelessly-connected sensors, actuators, and similar devices. Each device will typically be associated with very small payloads, implying that cumulatively they will have a limited impact on the overall traffic volume. However, one of the challenges presented by the large number of devices is how to design efficient signaling protocols. Over-the-air signaling efficiency should be enhanced so that the radio resource and energy consumption are minimized. Defining the number of active connections per unit area as a performance metric will help objectively measure the density of devices that can be simultaneously connected to a particular network in any deployment scenario.

User throughput and latency attributes for a network define the typical average data rate and round-trip delay, respectively, which users experience in a network. Ultimately the achievable values for these attributes determine the type of applications that can be supported on a network. It is estimated that by 2020, there will be new classes of service with high data rate and low latency requirements. The studies show that new applications such as augmented reality, interactive gaming, and Tactile Internet will require a 1,000X increase in achievable data rates compared to today's figures and up to a 10X reduction in latency.

Small access points with low transmit power and no planning requirements can be densely deployed, resulting in ultra-dense networks. This approach will improve spectral efficiency by reducing the distance between transmitters and receivers, and improve macro-cell service by offloading user traffic, thus releasing radio resources in the access network. Network densification is a way to increase the capacity and data rate in 2020 and beyond.

Energy Efficiency

Classic design principles for wireless communication systems tend to maximize data rate, capacity, and coverage, typically leading to solutions that are not necessarily energy efficient. As a result, the energy consumed in the network can increase, implying high operational costs for the operator that will ultimately translate into more expensive services for the end users. Furthermore, some radio communication strategies involve intensive computational complexity at the terminal side with negative impact on the battery life. Therefore, the efficient use of energy must become a new consideration in addition to the classic design criteria. 5G radio access technology design must target better energy efficiency while the number of devices/networks and their cumulative energy consumption is increasing. A Joules/Bits or Bits/Joules metric must be defined to benchmark the energy efficiency of the 5G wireless underlying radio access technologies.

There are currently two approaches for reducing energy consumption over the radio link. The small-cell architecture reduces the transmit power by reducing the distance between the access point and the terminal. The main challenges of this approach are related to providing an economic backhaul solution and to minimize the additional deployment cost. The second approach is massive MIMO, where energy is concentrated toward the user by means of increased antenna directivity and beamforming. In this way, the energy is directed toward the desired user and less energy is wasted through inter-user interference. The challenge of massive MIMO includes the dissipation of energy due to scattering in NLOS scenarios, limiting the achievable directivity and the complexity of spatial multiplexing of users. The goal of minimizing the energy consumption per bit requires a paradigm shift in wireless system design to dramatically improve efficiency in terms of power and spectrum usage. In terms of device implementation, the power efficiency implies more focus on low-power hardware design and energy efficient signal processing.

Future network architectures and deployment scenarios must take advantage of traffic characteristic to improve both spectral efficiency and energy efficiency. Various types of base stations and cell sizes can be deployed in an energy-efficient manner, taking into account their respective power consumption models and different capabilities, to provide coverage and capacity. The variation of traffic volume over time should be exploited to manage and allocate radio resources such as energy-efficient and spectrally-efficient configuration of RF transceivers and antennas in massive MIMO.

Availability and Reliability

The level of resilience and robustness required by ultra-reliable IoT applications tends to be much higher than that required for typical subscriber services. High-grade network slices, which are capable of providing the desired level of availability, robustness and resilience to malicious attacks and radio link failures, can be offered to public safety and emergency services with the additional option for extra capacity or coverage on demand in the event of an emergency. It is essential to maintain network emergency services in case of a disaster. The network availability is typically objectively characterized by an availability rate that is defined as the rate at which the network services are available for the targeted communication in the locations where the network is deployed and over certain period of time. Resilience is the capability of the network to recover from failures and is an important feature to maintain high availability rates which includes self-healing feature.

Latency

The access network latency can be viewed from different perspectives namely end-to-end latency and transport delay. The end-to-end latency measures the duration of transmission of a small data packet from the application layer at the source node to successful reception of the packet at the application layer of the destination node plus the equivalent time needed to send the acknowledgement. The transport delay is the one-way air-link transmission time between the availability of a packet at the IP layer in the source node and the availability of this packet at IP layer of the destination node including the processing delay introduced by the associated protocols. The 5G systems must be able to achieve a transport latency of less than 1ms in lightly-loaded conditions.

The end-to-end latency is the latency perceived by the end user. It accounts for the time needed for a data packet to traverse the network from the source node (e.g., application server) to the destination node (e.g., mobile station), which includes intermediate nodes inside and outside the 5G system. In contrast, the user-plane latency is limited to the air-link transport over a 5G system. Both latency metrics approximately coincide when the application server is located within the 5G system. In the latter case, the latency is minimized when the application server is co-located with the radio node; e.g., the radio base station or another user terminal for the case of device-to-device communication.

Depending on the use case, the 5G systems should be able to provide less than 10ms end-to-end latency in general and less than 1ms latency for ultra-reliable machine-type communication and Tactile Internet applications. Note these latency targets assume the application layer processing time is negligible compared to the delay introduced by transport and switching. The 5G system must further provide the end-users an always-on perception. The establishment of the initial access

to the network (or status change from idle state to connected) should then be virtually instantaneous from the user perspective.

Mobility

Mobility refers to the system's ability to provide seamless service experience to users that are moving across network. In addition to mobile users, the identified 5G use cases show that 5G networks will have to support an increasingly large segment of static and nomadic users/devices. 5G solutions therefore should not assume mobility support for all devices and services rather provide mobility on demand only to those devices and services that need it. In other words, mobility on-demand should be supported, ranging from very high mobility scenarios such as high-speed trains to low mobility or stationary devices such as smart meters. The mobility requirements are expressed in terms of the relative speed between the user and the network node (base station or access node), at which consistent user experience should be ensured.

5G networks will likely comprise a collection of layers or slices, technologies, and frequency bands that should seamlessly interwork when moving across networks, layers, and/or frequencies. Interruption times in the order of a few milliseconds for both inter-RAT and intra-RAT handovers can be expected in these scenarios. 5G networks might require up to 500km/h mobility support for certain applications.

Coverage

Conventional perception of cellular coverage is that there are areas of strong signal near the base stations and weak signals in the cell edges; however, 5G systems are required to provide a consistent user experience at any time and any location. This means that the future networks will have virtually uniform coverage with sufficiently high performance for all use cases at any time and location. As the future networks are expected to have flexibility in how resources are configured and deployed, the dynamic selection of different radio resources will be used to provide coverage based on service requirements.

Traditionally, devices associate with a single cell, thus the link performance degrades as a device moves away from the cell center. In a device-centric network, the network determines which access point(s) should be associated with the device based on the measurement reports and location information that is collected from the device. In this case, each device is served by a preferred set of access points typically surrounding the device. The actual serving set for a device can contain one or multiple access points and the device's data is partially or fully available at some or a small set of potential serving access points. The access point controller will accommodate each device with its preferred set and transmission mode at every communication instance while considering load and channel state information knowledge associated with the access points.

The concept of heterogeneous networks, initially conceived during LTE-Advanced development, is expected to grow in importance in 5G systems. Ultra-dense heterogeneous networks in the form of numerous macro-overlaid small cells will operate with a single or multiple RF carriers serving small areas. Multi-antenna systems based on diversity schemes, spatial multiplexing or beamforming techniques will be used for significant improvement of transmission quality as well as increasing capacity and reducing inter-user/inter-cell interference. Massive MIMO or large number of one or two-dimensional antenna arrays will be used to facilitate multi-user scheduling and interference reduction. Cooperative non-collocated antenna techniques, i.e., transmission and reception from

multiple nodes in the network, are also considered a promising technology to improve capacity, user throughput, and interference management in the future networks. In addition to D2D use cases and proximity services, user devices can also act as relays to help the communication between the network and other devices in their vicinity.

Spectrum

5G carrier frequencies will include bands above 30GHz, known as the mmWave spectrum, with the potential of ten times as much spectrum as is currently available in IMT bands under 6GHz. Radio channels with aggregated or contiguous bandwidth of 1GHz or more will enable multi-Gb/s peak data rate requirements of 5G systems. As one of the most innovative and effective solutions to realize the 5G vision and requirements, the use of large segments of underutilized spectrum in the very high frequencies such as the mmWave bands has recently gained significant interest. Traditionally, due to the high propagation loss and lack of cost-effective components, among other reasons, these frequencies have mostly been utilized for outdoor LOS backhaul links or for carrying high resolution multimedia streams for indoor applications, but not for outdoor wireless access links. To repurpose these underutilized spectra for future outdoor cellular applications, two major obstacles must be overcome: sufficiently large geographical coverage, and support for mobility even in NLOS environments where the direct communications path between the transmitter and the receiver is obstructed.

In addition to identification of the new spectrum, more spectrum resources can be virtually obtained through more efficient spectrum utilization policies. Authorized shared access (ASA)/licensed shared access (LSA) can be adopted under well-defined conditions where the incumbent usage of spectrum is low. Co-primary shared access models may be explored to share a spectrum block among multiple network operators, rather than divide it between them, through cognitive radio access procedures.

5G mobile communication systems target extremely high data rates and as such one of the key technologies for achieving this goal is the carrier aggregation, which has been used in LTE since Release 10 and it is expected to be further extended in 5G. The future network is expected to be focused on heterogeneous carrier aggregation where a large number of small cells are overlaid by a macro cell. It is also expected that terminals with dual cellular/Wi-Fi support will proliferate, and more wireless network interfaces will be added. Although Wi-Fi offload is being used at the present time to relieve the overloaded cellular networks, it is expected that 5G systems will further utilize heterogeneous carrier aggregations and tightly-integrated cellular and Wi-Fi as well as fast and seamless inter-RAT handover to achieve stringent QoS/QoE service requirements.

A large amount of unlicensed spectrum is available in 2.4 and 5GHz frequency bands. The LTE has support for unlicensed spectrum under LTE unlicensed (LTE-U), licensed assisted access (LAA), and LTE Wi-Fi link aggregation (LWA) initiatives since Release 12, which are expected to evolve toward to 5G. Furthermore, the unlicensed spectrum is mostly used in small coverage areas, resulting in high-frequency reuse and much higher throughput per square meter of coverage versus typical cellular deployments. New spectrum below 6GHz has been allocated for mobile communication in the world radio conference 2015 (WRC-15) and new bands above 6GHz are expected to be allocated in WRC-19. Spectrum sharing and flexible spectrum usage strategies can be exploited to optimize spectrum utilization and more importantly to provide opportunities for operators to access additional spectrum typically allocated to other radio services. Depending on the specific

regulatory framework, a different spectrum usage strategy can be selected individually or jointly to satisfy diverse requirements.

5G Candidate Technologies and Xilinx Competencies

In addition to extending system operation to higher frequencies, several other key technology components are relevant for the migration to 5G wireless access. In this section, some of the prominent schemes that have the potential to be part of the 5G wireless standards are described. The role of programmable SoCs in efficient implementation and commercialization of 5G candidate technologies are also reviewed.

Duplexing and Multiple Access Schemes

Existing cellular systems rely on FDD or TDD duplexing schemes for downlink/uplink transmissions. In an FDD system, the uplink and downlink signals are separated in the frequency domain by an offset in order to protect the receiver from self-interference, whereas in a TDD system, the downlink and uplink transmissions are time-multiplexed on the same RF carrier. The TDD duplex scheme can theoretically support dynamic downlink/uplink resource partitioning, adapting to the instantaneous user traffic in each direction. In traditional macro-cell deployments, the TDD downlink/uplink ratio was fixed across the network to ensure no strong inter-cell interference was introduced. However, in small-cell deployments, due to limited coverage and low output power of the access node, this constraint can be relaxed and each access node operating in TDD mode can use a traffic adaptive downlink/uplink ratio. Furthermore, to optimize the spectrum utilization based on downlink/uplink traffic volume, a flexible duplexing scheme can be used.

Full-duplex schemes are exhibiting a great potential to overcome the constraints of FDD (e.g., use of guard bands) and TDD (e.g., use of guard time and switching point synchronization) systems. It is not generally possible to use the same frequency channel for simultaneous transmission of downlink and uplink signals. The main challenge is the large power difference between the strong self-interference due to the node's own transmissions and the weak signal of interest coming from the distant transmitter. Experiments have shown that reduction of self-interference by 110dB or more is possible, which has generally been achieved through a combination of analog, digital, and hardware cancellation techniques. The interference-reduction levels achieved so far are sufficient for low transmit-power systems over very short distances, where the received signal is strong enough to make the difference between the base station's transmit power and the received signal. These levels of interference reduction are clearly not sufficient for current cellular systems, which operate with much higher transmit powers and in greater path loss environments. In a typical LTE device, the transmitter output power can be up to 23dBm and the requirement for the receiver sensitivity is about -113dBm, thus a total of 136dB isolation would be required. Further investigation is required to achieve a practical transmit/receive isolation in the time frame of 5G radio access technology development.

Multiple access schemes are the most fundamental aspect of the physical layer design and operation. The multiple access schemes have evolved in each cellular generation from frequency division multiple access (FDMA) and time division multiple access (TDMA) in 1G and 2G to code division multiple access (CDMA) in 3G and Orthogonal Frequency Division Multiple Access (OFDMA) and single carrier FDMA (SC-FDMA) in 4G. When encountering the stringent requirements and diverse use cases in 5G, the classic orthogonal multiple access approach might

not be adequate. As a result, some alternative non-orthogonal multiple access schemes that support multiple users over the same frequency/time resources via code domain and/or power domain multiplexing have been investigated. These include multi-user shared access (MUSA), power-domain non-orthogonal multiple access (NOMA), sparse code multiple access (SCMA), and various forms of (filtered) OFDMA with fixed or adaptive numerology. A non-orthogonal multiple access scheme that can be combined with OFDMA would be advantageous, enabling the continued benefits of OFDMA while maintaining backward compatibility—all while enhancing the data rate and accommodating scalability for massive IoT connectivity and reduced access latency. The advanced low-complexity receiver, multi-user scheduling, power allocation, reference signal, codebook, feedback mechanism designs, and the combination with multi-antenna technologies are the key issues that must be further investigated before non-orthogonal multiple access schemes are specified in the standards.
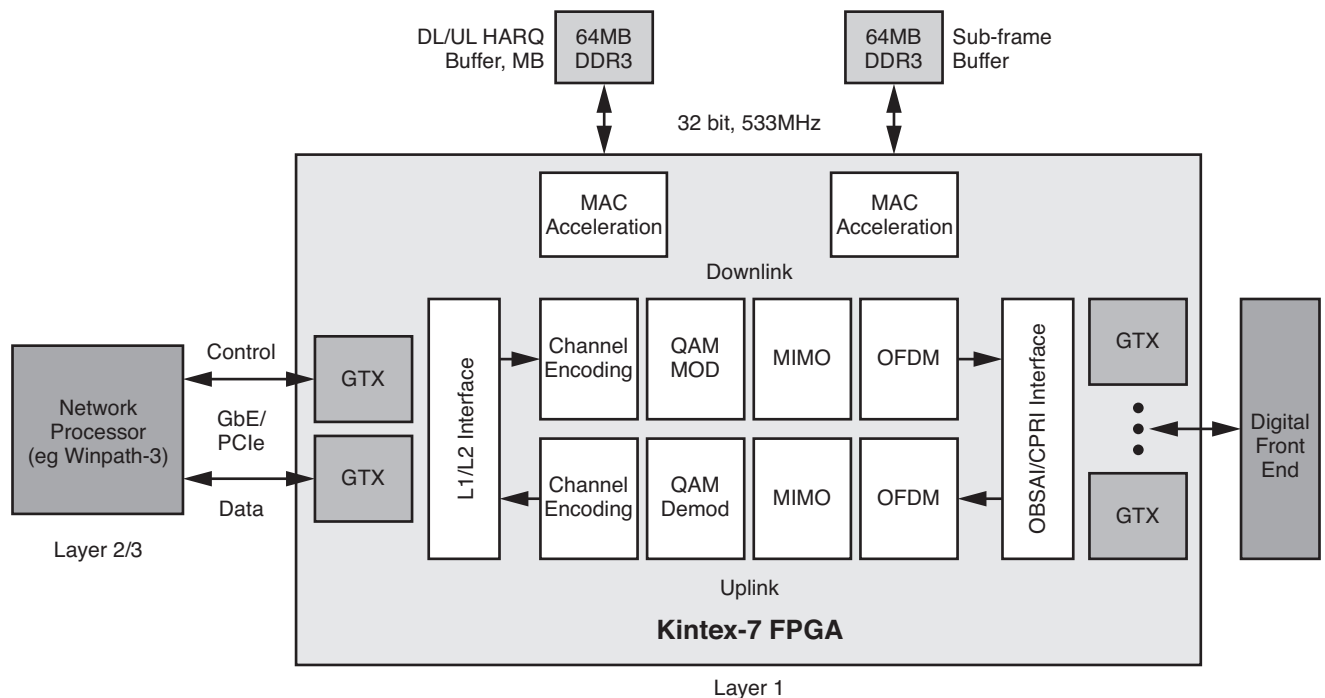
WP476_07_041316

Figure 7: **Example Single-Chip LTE Baseband Solution from Xilinx [Ref 4]**

Xilinx devices have enabled single-chip LTE baseband and radio solutions for small base stations achieving minimal power consumption, implementation complexity, and cost. Xilinx devices can be used to efficiently implement and accelerate 5G baseband and radio transceivers irrespective of the wide range of possible solutions under consideration. Examples include LTE L1 baseband design (4×4 multi-user MIMO) on a single Kintex®-7 XC7K410T device, replacing multiple DSP cores (shown in Figure 7) or the use of SoC/MPSoC all-programmable devices for designing small base stations ranging from femto-cell to macro-cells. In the latter case, the on-chip processors perform the real-time processing tasks including but not limited to L1 (+L2/L3) baseband processing, digital radio front end (DUC/DDC, CFR, DPD), Gigabit Ethernet backhaul management, timing and synchronization, hardware acceleration for higher layer functions such as IPsec, packet encryption, integrity protection, and RoHC header compression protocol implementation. The FPGA logic on these devices can be used to accelerate L2/L3 functions, offloading the processors to provide designers with more flexible acceleration options. On-chip logic resources enable the integration

of DFE functions (DUC/DDC, CFR, DPD), which contributes to further reduction in power consumption with minimal processing delay. The integrated PCIe® connectivity ports enable seamless integration of Wi-Fi for data offloading.

New Waveforms

OFDM was the key feature of the 4th generation of cellular systems that enabled the use of wider channel bandwidths and various MIMO techniques due to its capability to cope with frequency selectivity of the wireless channels as well as sub-carrier orthogonality and ease of implementation using FFT processing. Despite OFDM's need for precise frequency and timing synchronization and overhead due to the use of cyclic prefix, this multi-carrier scheme is still considered the most promising waveform for the future radio access design. However, a pure OFDM scheme might be insufficient to address the challenges of diverse applications and various deployment scenarios of 5G. To accommodate the new services such as mission-critical machine type communication or Tactile Internet and to overcome the disadvantages of OFDM modulation, poly-phase filtered forms of OFDM such as filter-bank multi-carrier (FBMC), filtered OFDM, universal filtered multi-carrier (UFMC), and generalized frequency division multiplexing (GFDM) are being studied. The common property of these transmission schemes is that they, at least in principle, can provide more localized spectrum compared to conventional OFDM. This is relevant to spectrum-sharing scenarios.

It should be noted that localized spectrum is a desirable property of the waveform and that transmitter non-linearity can cause additional spectral spreading that might reduce the benefits offered by the localized waveforms. To generate the localized waveforms, the underlying time-domain/frequency-domain basis functions are shaped by applying different digital filters to reduce the out-of-band emissions and to reduce or remove the guard bands/guard times that can improve the spectral coexistence and efficiency. Some of these waveforms can relax the tight requirements for time/frequency synchronization due to good spectral/temporal localization of the (filtered) basis functions. Filtered OFDM is a promising technique that preserves OFDM advantages through the use of digital filters to shape the spectrum of the signals in each sub-band in order to reduce the out-of-band emissions, making the waveform more localized, thus more suitable for use in flexible spectrum sharing and over radio access links. Moreover, by introducing the sub-band filters, different OFDM numerologies such as FFT size, CP length, and sub-carrier spacing can be applied for different users with different service requirements within the same unified frame structure. Filtered OFDM further supports asynchronous uplink transmissions from multiple users whose signals can be separated by appropriate filters with minimal inter-user/inter-cell interference with a performance comparable to synchronous OFDM.

Fast Fourier transform (FFT) is a fundamental building block in OFDM-based baseband physical layer processing. Xilinx provides four different highly optimized FFT architectures (solutions) along with system-level fixed-point tools. FFT provides the users with the ability to make all the necessary algorithmic and implementation specific trade-offs demanded by DSP algorithms and hardware design. These trade-offs enable the users to select the most resource and power efficient solutions for a specific FFT size and processing time required for various applications. Xilinx FFT solutions support run-time configurable forward and inverse complex-valued FFT with FFT sizes ranging from 8 to 65536 in ascending order of integer powers of two. Xilinx FFT solutions expand the focus on increased dynamic range by increasing data and phase factor width support up to 34-bit precision to further allow IEEE single-precision floating-point data type processing. The floating-point option is implemented by utilizing a higher precision fixed-point FFT internally to

achieve similar noise performance to a full floating-point implementation, with significantly fewer resources and thus reduced power consumption.

Coding and Modulation Schemes

To meet the stringent link budget and throughput requirements of 5G systems under various conditions, the modulation and coding schemes need to be redesigned to further close the performance gap with the maximum achievable channel capacity. Turbo codes and various forms of quadrature amplitude modulation (QAM) were widely employed in the previous generations of cellular systems due to their good error resilience and performance; however, some of the 5G use cases require very low error rates that cannot be achieved with turbo codes. The interest of the wireless industry in low-density parity check (LDPC) codes has increased over the past few years. The LDPC codes have proven to be efficient coding schemes to improve transmission robustness and reliability with manageable implementation complexity and maximal throughput per (chip) area. The LDPC codes belong to the class of linear block codes with a sparse parity check matrix. The purpose of using sparsely distributed parity check architecture is to apply probabilistically independent XOR calculations, whereby the estimates of parity check values are aggregated across each row of the parity check matrix and every subsequent iteration improves the estimates. The exact number of iterations is determined by the code design based on the application and specifications of the system. Numerous trade-offs are possible before determining the minimum number of iterations required to meet the requirements of a particular system. Modern LDPC codes use soft-decision algorithms to further enhance the coding gain. With larger block length, the confidence level in the error correcting capability increases. However, it increases the computational complexity by increasing the size of the parity check matrix and thereby increases the number of calculations required to estimate each bit. New constellation arrangements (baseband modulation schemes) are being investigated to shape the signal statistical distribution in order to achieve additional capacity gain relative to traditional Gray-coded QAM schemes. Considering the effect of inter-carrier interference, a combination of FSK and QAM modulation, has been found to improve the cell-edge performance.

Polar code is a new scheme in coding theory, which is claimed to achieve near Shannon capacity with a simple encoder and a successive cancellation decoder when the code block size is sufficiently large. Polar codes have attracted a significant amount of interest, and research has been focused on code design and decoding algorithms. One of the most important decoding algorithms for polar codes is successive cancellation list decoding whose performance is as good as the optimum maximum likelihood decoding with a list size of 32 for moderate code block sizes. Some recent study has claimed that polar codes concatenated with cyclic redundancy checks and an adaptive successive cancellation list decoder can outperform turbo/LDPC codes for short and medium block sizes.

Xilinx has been providing high-throughput (up to 4Gb/s) low-power (less than 350mW/Gb/s) LDPC encoding/decoding IP cores for wired/wireless backhaul applications, which can be particularly suitable and applicable to mMTC and URLLC scenarios in 5G where small payloads and extremely low error rates are targeted. Xilinx LDPC core based on quasi-cyclic LDPC (QC-LDPC) codes uses advanced code construction methodology and a proprietary optimization technique to achieve best code performance and very low error floor. It supports various frame lengths (64 to 32,768 bits per frame) and various modulation schemes (QPSK, 16QAM, 64QAM, 256QAM). It employs a layered decoding algorithm and log-likelihood ratio (LLR) generation logic combined with the decoder to increase the iteration convergence rate and to achieve lower error floor.

Multi-Antenna Schemes

The use of multi-antenna schemes at the transmitter and receiver can increase the spectral efficiency of the wireless link. When the number of user device antennas is much smaller than base station's antennas, multiple users can be served simultaneously using the same time-frequency resources (multi-user MIMO). This requires knowledge of the channel between the base station antennas and the receiver antennas so that appropriate precoding methods can be used to eliminate inter-user interference. Single-user MIMO and multi-user MIMO were both part of the LTE standards.

Massive MIMO extends the multi-user MIMO concept by dramatically increasing the number of antennas at the base station. The number of base station antennas in this case is significantly larger than the number of users being served simultaneously. With hundreds of antennas serving tens of users at the same time, spectral efficiency can increase five to ten-fold, while cell-edge users would experience higher throughputs. Massive MIMO performance scales up by orders of magnitude compared to the current multi-antenna schemes and promises energy efficiency and exploitation of excess degrees of freedom, relaxing the TDD RF calibration requirement, while requiring new techniques to mitigate pilot contamination and to estimate channel state information.

The massive MIMO paradigm is an evolution of multi-user MIMO. It should be noted that MIMO, in current wireless standards such as LTE and Wi-Fi, is typically implemented in its single-user form, thus on a given channel and in a given time slot, all the base station antennas are used to communicate with a single user terminal where the multiplicity of antennas at both ends allows the creation of multiple data streams in spatial domain, resulting in the link capacity increase by a significant factor. Since it is more feasible to have a large array of antennas (for size, implementation, power consumption, and cost reasons) at the base station compared to the terminal, these additional degrees of freedom can be used to communicate with multiple users at the same time. However, this is a more difficult problem given that multiple users addressed simultaneously cannot easily perform joint processing to eliminate the inter-user interference created by this method. Massive MIMO is based on the fact that if there is a sufficiently large number of antennas at the base station, efficient linear processing in the form of eigen-beamforming and maximal-ratio combining become optimal. This is attractive because such processing is not only extremely simple, but also it scales linearly with the size of the array and requires very limited inter-processor communication. However, while the antennas are inexpensive, such very large arrays with hundreds of RF front-ends and ADCs/DACs are complex and power consuming. It is therefore of interest to explore the quasi-massive cases, where the number of antennas is not sufficient to achieve the massive effect, nevertheless large enough to utilize linear MIMO receivers such as minimum mean-squared error (MMSE) and zero forcing (ZF). MMSE and ZF are typically undesirable because their complexity scales according to the power of three of the number of antennas.

Xilinx offers a wide range of optimized IP blocks for fast and efficient processing of MIMO encoding/decoding algorithms based on its FPGAs and MPSoCs. In general, the type of hardware acceleration that can be envisaged based on FPGAs includes baseband signal processing, higher-layer and transport acceleration, and preprocessing for radio signals. Baseband signal processing acceleration can be envisioned for all radio air-interface types and is partially applicable to advanced system configurations and functionalities that require significant computational complexity.

Fronthaul and Backhaul Solutions

In a C-RAN architecture, BBUs and RRHs are separated from each other, where BBUs are typically located at a central office or in the cloud while RRHs are distributed across the cell sites. These physically separated radio units are typically connected via CRPI interface, with one CPRI port connecting each RRH to the corresponding BBU. If each RRH has two antennas with an LTE channel bandwidth of 20MHz, the capacity of the CPRI interface should be 2.45Gb/s for transport of IQ samples. When massive MIMO is employed, the point-to-point CPRI capacity between BBU and RRH is not sufficient and has to drastically increase to allow the transport of IQ samples. For example, if the channel bandwidth is 20MHz and the RRH has 16 antennas, then 19.66Gb/s of CPRI capacity is required between the BBU and the RRH. The problem will be aggravated if the channel bandwidth is extended beyond 100MHz. This means that today's C-RAN fronthaul architecture will no longer be able to handle such increased capacity requirements in a 5G network architecture. Currently, the maximum transmission capacity per CPRI port is 10–25Gb/s. To overcome this limitation, a flexible functional split between the BBU and RRH would be required to allow not only the radio functions but L1 or L1+L2 functions to be performed in the RRHs. In addition, the interface between BBU and RRHs should be changed from circuit-switched CPRI to packet-switched Ethernet to make fronthaul data transport more efficient (see Figure 8).
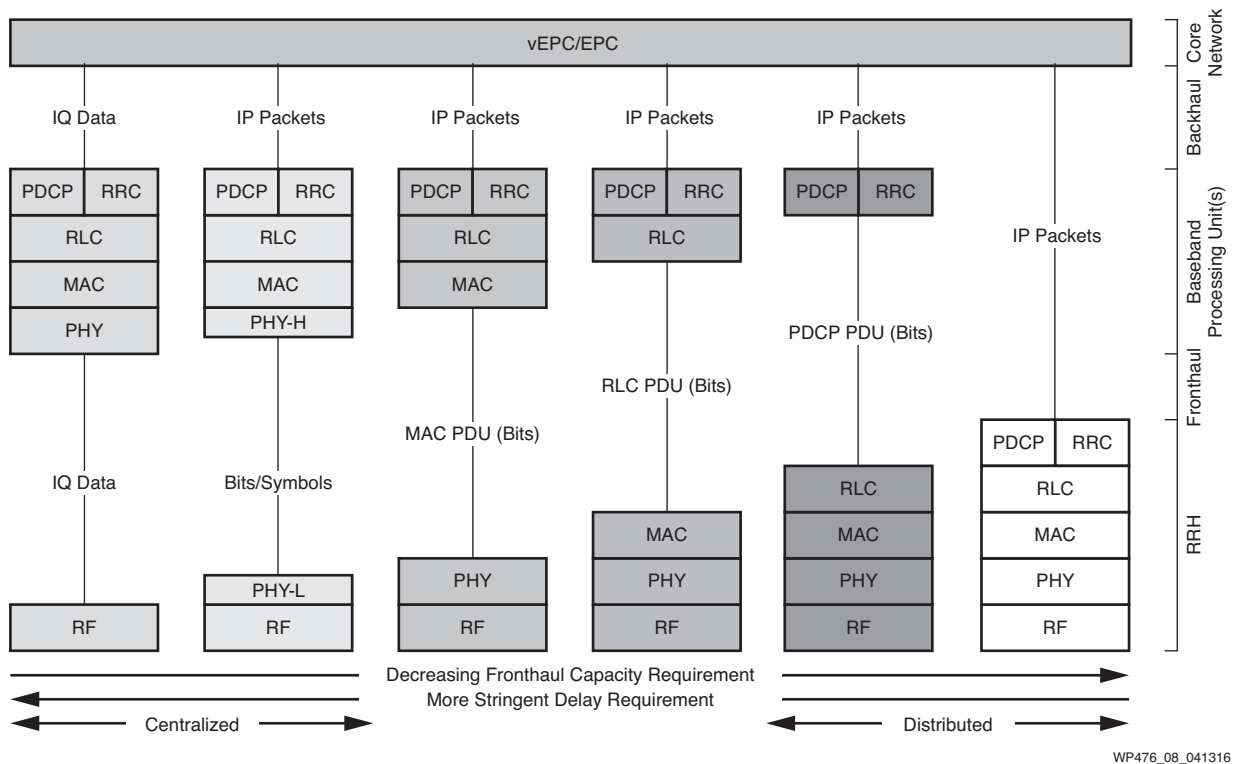

WP476_08_041316

Figure 8: **Various Functional Split between RRH and BBU [Ref 3]**

To implement these changes, various functional split options have been studied, each offering trade-offs such as reduced fronthaul capacity, CoMP effect, RAN virtualization gain, etc. To achieve minimal end-to-end latency in terms of network architecture, which is critical for some 5G applications, it would be imperative to eliminate the backhaul delay by distributing 5G core and application servers closest to the mobile devices (mobile edge computing). The requirements on the fronthaul can be changed depending on how the digital processing is split between data

centers and remote radio units. Depending on the available fronthaul capacity and the latency requirements, it might be necessary to utilize a functional split that allows more baseband processing to be done at the RRH to relax the fronthaul requirements. In 5G networks, this might even be done dynamically and adapted to specific application requirements, relying on software-defined networking and network function virtualization as well as programmable and configurable devices.

With a wireless backhaul or fronthaul connecting many small base stations, many solutions for the access link are also of interest for the backhaul link. For example, higher frequency bands that have so far been used for point-to-point backhaul radio links are being considered for the access link. To allow more flexible topology and simpler installation, beamforming based on large scale MIMO is also being considered for the backhaul links. Since the RRHs are stationary in general, the challenges of tracking a mobile user and frequently estimating the channel are relaxed compared to the access link, where these are major problems for millimeter wave and massive MIMO. Because the technology between access links and backhaul links will be similar, it might be possible for much of the interface to be the same. It is already possible in LTE to use relays. In future systems, the efficiency could be further improved for example if the relay uses full duplex transmission within the same frequency band. A more limited approach to integrated backhaul and access is to take the backhaul limitations into consideration in the radio resource management algorithms, even if the backhaul uses separate resources.

As shown in Figure 9, the programmability and the computational capabilities of the current and the planned Xilinx FPGA and MPSoC devices tremendously facilitate the efficient design and flexible C-RAN functional splits between the RRHs and BBUs. The single-chip RRH designs which encompass L1 and radio functions (DUC/DDC, CFR, DPD, and ADC/DAC), CPRI gateways, and NGFI interfaces, BBU baseband functions combined with high-speed connectivity for backhaul interfaces provide great flexibility and ease of upgrade for 5G network deployments.
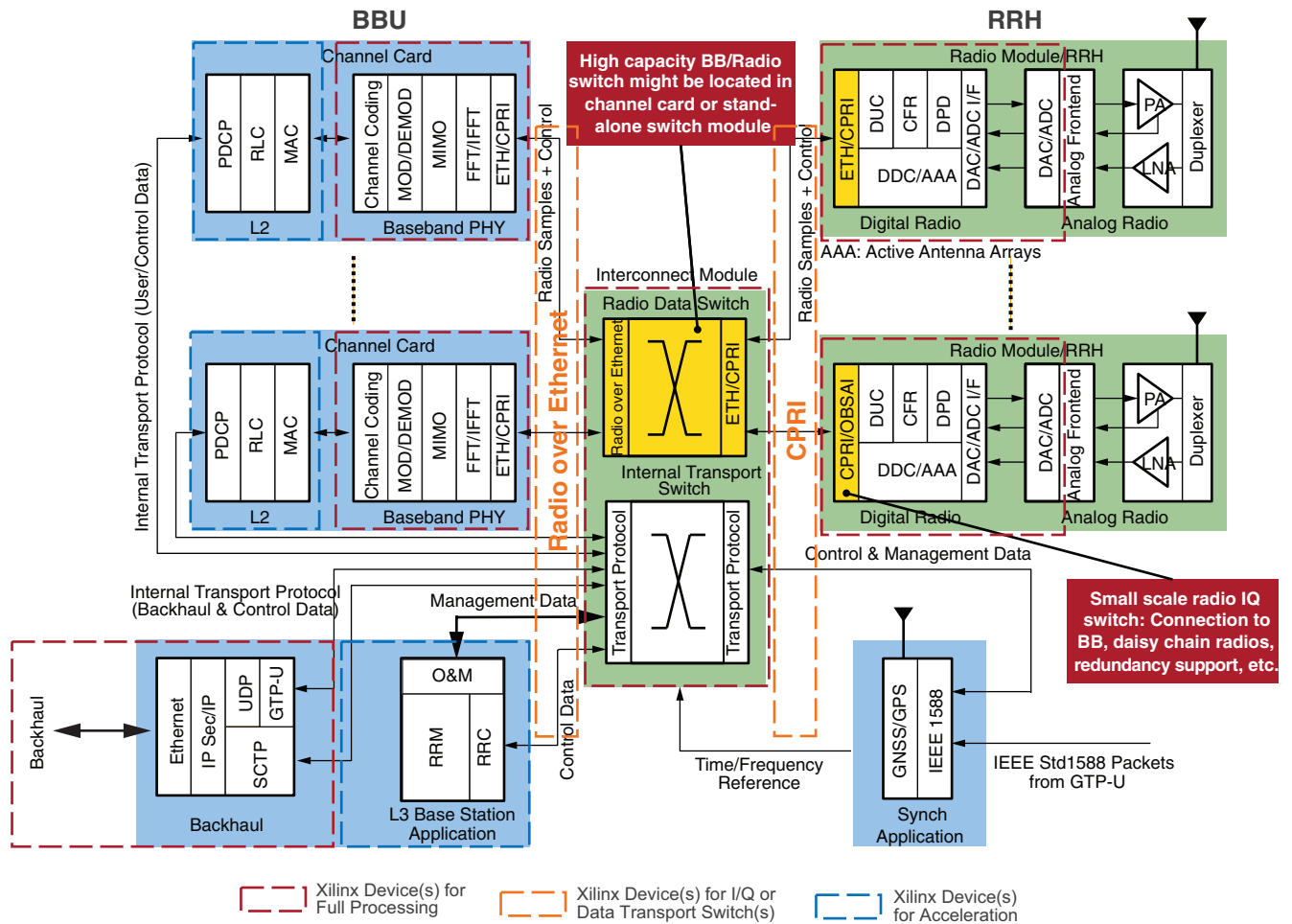
Figure 9: **Example C-RAN Solutions (RRH, Fronthaul/Backhaul Interface, BBU) for 5G Ultra-Dense Networks from Xilinx**

Modern RF Architectures

Direct RF conversion using high-speed ADCs/DACs is a technology disruption for modern RF transceivers. The ultra-high speed DACs use direct digital synthesis to move the quadrature modulator, agile local oscillator, and signal-filtering analog functions into the digital domain. The ultra-high speed DAC with direct digital synthesis partition capitalizes on the fact that digital processes scale better than analog in terms of lower power consumption, faster speed, smaller die area, and lower cost. However, these benefits can only be realized if direct signal conversion from the digital domain to analog domain is achievable. The ultra-high-speed DAC is the enabling technology that makes this possible because it bridges the digital-to-analog domain. An ultra-high- speed DAC is generally characterized as a mixed-signal device that operates in multiple Nyquist zones with conversion rates above 1.5GSPS to perform direct RF signal synthesis. An ultra-high-speed DAC synthesizes output signals of at least 500MHz analog bandwidth at carrier frequencies of 2.0GHz or higher. See Figure 10.
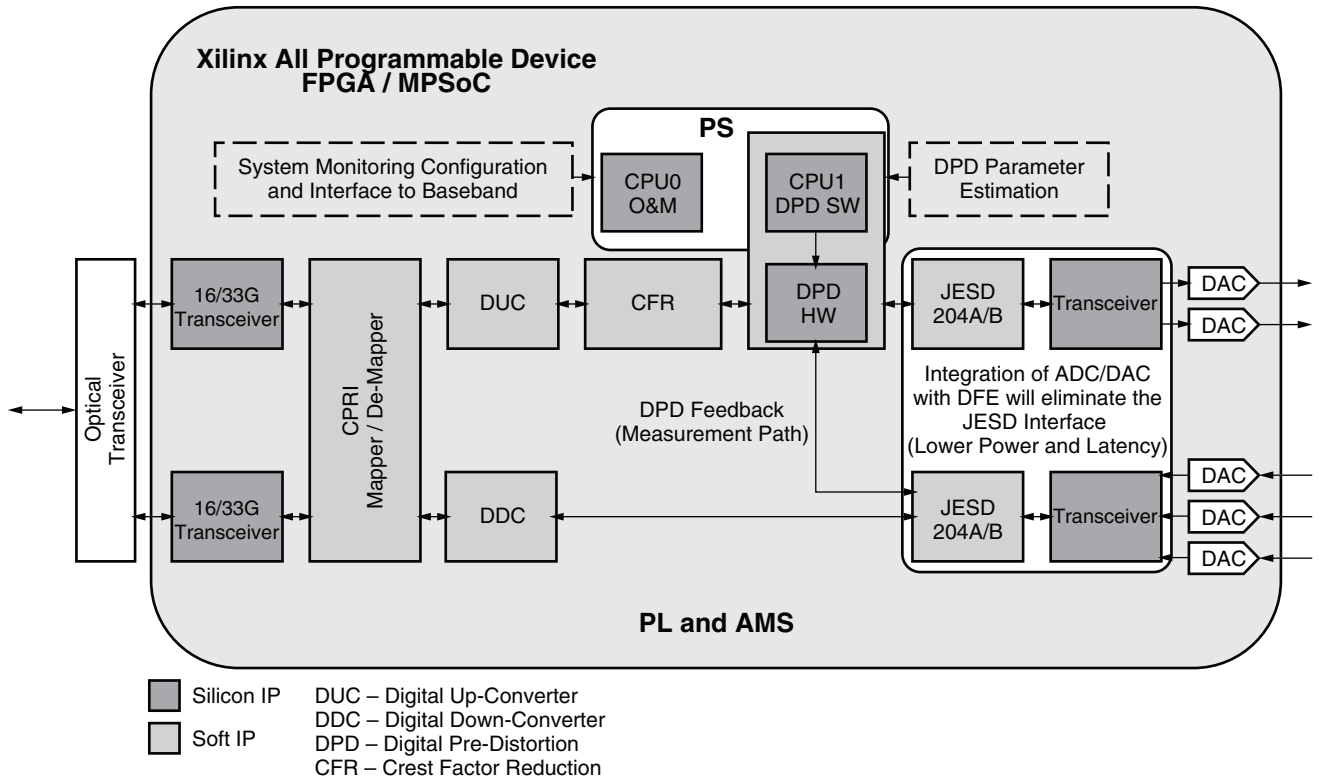
Figure 10: **Single-Chip Radio Solution by Xilinx [Ref 4]**

Compared to conventional RF transmitter architectures (zero-IF, complex-IF, and real-IF), a direct RF sampling architecture occupies less die area with fewer components. It operates at lower power and can deliver excellent dynamic performance. In terms of RF performance, an ultra-high speed DAC has significant benefits relative to other topologies. The digital up-conversion with digital filtering implemented in direct digital synthesis eliminates gain-phase errors and achieves perfect carrier suppression with minimal local oscillator leakage. The result is excellent EVM performance when transmitting high-order modulations like 64QAM. The embedded quadrature modulator and numerically-controlled oscillator make the direct RF sampling to be an agile transceiver capable of tuning across the entire spectrum below 6GHz. For example, a single device can synthesize multi-carrier, multi-band, and multi-standard signals including GSM, WDMA, LTE, and future 5G air-interfaces. Another benefit of direct RF conversion is that it allows lower-cost and more efficient implementation of digital pre-distortion functions. The base stations use DPD techniques for power amplifier linearization. The use of DPD techniques requires monitoring the PA output and to detect the PA distortion products (measurement path in Figure 10) to compensate for intermodulation and adjacent-channel leakage power. Typically, DPD bandwidth expansion requires the DPD feedback bandwidth to be five times the signal bandwidth. In 100MHz carrier aggregation applications, this means that the DPD bandwidth must be at least 500MHz. Also, the DPD feedback cannot add impairments to the observed signal because it cannot be discerned from the main transmit path impairments. Consequently, the DPD feedback path must have excellent linearity (which adds cost and circuit complexity). The single-chip digital radio and integrated ultra-high-speed ADC/DACs further simplify design and implementation of active antenna arrays and massive MIMO architectures because they reduce the number of active components on the board. Therefore, direct RF sampling transceivers are the ideal RF architectures for 5G wireless

applications in sub-6GHz (without external mixers) and above 6GHz (with external mixers) frequency bands.

Role of Programmability in 5G

5G systems are required to support much wider range of devices and applications than the existing networks. Flexibility through programmability is a significant characteristic that will enable 5G networks to support fast time to market for new services and efficient scaling. Programmability allows network operators to adopt innovations rapidly and adapt to continuously changing network requirements. The network planners need to determine the required degree of flexibility or ability to reconfigure as well as the layer or layers that need to be programmable in the network. The new service requirements, connection capabilities (e.g., wider bandwidth and shorter latency), network function virtualization, etc., underline the inevitable need for network programmability and flexibility. Transport network programmability enables network operators to exploit traffic dynamics and to optimize the utilization of resources across different segments of the network. A programmable transport network facilitates the division of transport resources into multiple slices. These slices can be allocated to different clients (e.g., enterprises or service providers) enabling efficient sharing of resources. A programmable transport network further enables the resources allocated to a service to be scaled up or down, when and where it is needed across the network. Network control can be centralized or distributed, and networks are operated differently depending on the usage scenarios.
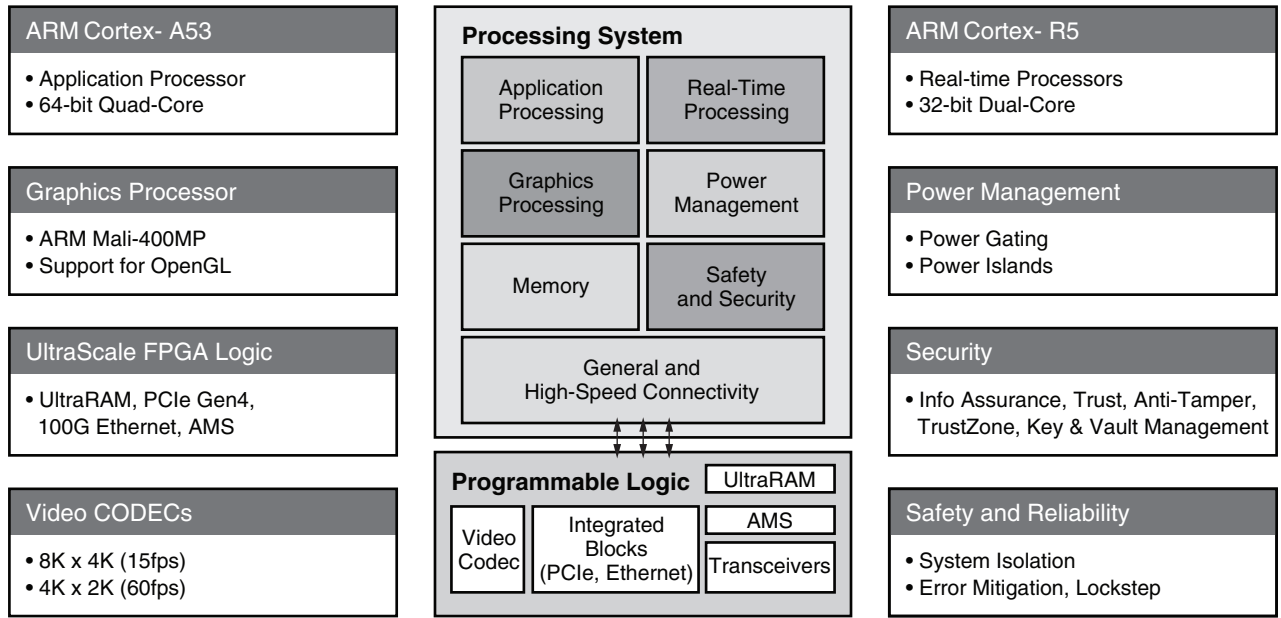
Network slices can be configured to meet the needs of broad range of applications supported by various device types, whether they require low latency, minimal power consumption or wide area coverage. Slicing can therefore provide the necessary support for differentiated service offerings. The use cases envisioned for the 2020 time frame will need new types of wireless connectivity that are highly scalable and programmable in terms of speed, capacity, security, reliability, availability, latency, and power consumption. SDN, NFV, MEC, and general cloud technologies enable networks to be detached from their underlying physical infrastructure so that they can programmatically provide network and connectivity as a service. The advantage of SDN lies in its ability to provide an abstraction of the physical network infrastructure. Through network-wide programmability, the capability to change the behavior of the network as a whole, SDN greatly simplifies the management of networks. The level of network programmability provided by SDN allows several network slices, customized and optimized for different service deployments, to be configured using the same physical and logical network infrastructure. One physical network can therefore support a wide range of services and deliver these services in an optimal manner.

By separating network functions from the hardware, NFV allows a network function to be implemented programmatically on a general-purpose processor instead of a dedicated hardware. This capability enables instant scalability, which supports the delivery of on-demand services. The most significant benefit offered by NFV is the flexibility to execute network functions independent of their location. By virtualizing a network function, it is no longer bound to a specific location or node. The same network function can be executed in different places for different network slices. Depending on the use case, a network function could either be placed in a centralized data center or close to a base station. By placing network functions accordingly, the same physical infrastructure can provide connectivity with different latencies. Even in non-NFV environments, programmability is greatly required to support various deployment scenarios and interaction with NFV-based platforms.

Considering the strong desire for programmability and configurability in 5G networks, FPGAs, general-purpose processors, and DSPs are the candidates that can offer different design options and flexibilities and can be used in various phases of design, including proof of concept through production release. Compared to ASICs, FPGAs can provide significant values to designers including flexibility, a faster product development cycle, and a rapid time to market. In contrast to GPPs and DSPs, FPGAs can utilize real-time clocking, acceleration, and parallelism in processing of computationally intensive algorithms. While FPGAs used to be selected for lower speed, lower volume designs in the past, today's FPGAs easily push the 500MHz performance barrier. With unprecedented increase in logic density and integration of other features such as embedded processors, DSP blocks, zone clocking, and high-speed serial interfaces at lower price points, new generations of FPGAs are a compelling value proposition for almost any type of design. The enhanced FPGA design flow eliminates the complex and time-consuming floor planning, place and route, timing analysis, and mask/re-spin stages of the project since the design logic is already synthesized to be mapped onto an already verified, characterized FPGA device.

New generations of FPGAs are also supporting block-level partial reconfiguration in run time without disturbing networking flows and services. This enables additional flexibility to radio, baseband, and packet core design in virtualized and non-virtualized environments. While 5G intends to cover a greater number of use cases, spectrum bands, and bandwidth combinations, new generations of FPGAs reduce total cost of ownership and extend the product life cycle, and more importantly, they can be customized to meet specific use case requirements in the 5G era.

Building on the success of Xilinx's 20nm UltraScale™ family, Xilinx has introduced the new 16nm UltraScale+ family of FPGAs, 3D integrated circuits, and MPSoCs, combining new memory and multiprocessing SoC technologies, delivering a generation ahead of value. In addition, to enable an even higher level of performance and integration, the UltraScale+ family also includes a new interconnect optimization technology known as SmartConnect. These devices extend Xilinx's UltraScale portfolio, now spanning 20nm and 16nm FPGAs, 3D integrated circuits, and MPSoCs, and leverage a significant boost in performance-per-watt based on 16nm FinFET technology. Optimized at the system level, the UltraScale+ portfolio delivers value far beyond a traditional process node migration—providing 2–5X greater system-level performance-per-watt over 28nm devices, far more systems integration and intelligence, and the highest level of security. Xilinx UltraScale+ products resolve the number one bottleneck for processing-intensive functions, which is the memory interface. These new memory enhanced programmable devices include UltraRAM, delivering up to 360Mb of capacity. UltraRAM provides optimal system power, flexibility, and predictable performance while replacing external memory in some applications, allowing for total cost reduction. Xilinx has developed advanced tool-driven interconnect optimization technology for FPGAs called SmartConnect. SmartConnect intelligently bridges different interface types to match the interconnection to specific application requirements. The high end of the Xilinx UltraScale+ portfolio leverages the combined power of 3D transistors and the third generation of Xilinx 3D integrated circuits (see Figure 11).

ARM Cortex- A53
• Application Processor
• 64-bit Quad-Core

Graphics Processor
• ARM Mali-400MP
• Support for OpenGL

UltraScale FPGA Logic
• UltraRAM, PCIe Gen4, 100G Ethernet, AMS

Video CODECs
• 8K x 4K (15fps)
• 4K x 2K (60fps)

Processing System

Application Processing	Real-Time Processing
Graphics Processing	Power Management
Memory	Safety and Security
General and High-Speed Connectivity	

Programmable Logic — UltraRAM
Video Codec — Integrated Blocks (PCIe, Ethernet) — AMS — Transceivers

ARM Cortex- R5
• Real-time Processors
• 32-bit Dual-Core

Power Management
• Power Gating
• Power Islands

Security
• Info Assurance, Trust, Anti-Tamper, TrustZone, Key & Vault Management

Safety and Reliability
• System Isolation
• Error Mitigation, Lockstep

WP476_11_050316

Figure 11: **Xilinx UltraScale+ Architecture and Features [Ref 4]**

The Zynq® UltraScale+™ MPSoCs utilize advanced FPGA technologies in addition to providing unprecedented level of heterogeneous multiprocessing. These new devices deliver approximately 5X system-level performance-per-watt relative to previous alternatives. At the center of the processing subsystem is the 64-bit Quad-core ARM Cortex-A53 processors, which are capable of hardware virtualization and asymmetric processing. The processing subsystem also includes a Dual-core ARM Cortex-R5 real-time processor for deterministic operation, ensuring responsiveness, high throughput, and low latency for the highest levels of safety and reliability. A separate security unit enables military-class security solutions such as secure boot, key and vault management, and anti-tamper capabilities, standard requirements for machine-to-machine communication and industrial IoT applications. For complete graphics acceleration and video compression/decompression, the MPSoCs incorporate an ARM Mali-400MP dedicated graphics processor as well as an H.265 video codec unit. A dedicated platform and power management unit has been added that supports system monitoring, system management, and dynamic power gating of each of the processing engines.

Xilinx Value Proposition in 5G and Product Roadmap

Programmable logic devices play a pivotal role in creating platforms for validating algorithms and proposals, building proof-of-concept systems to investigate, measure, demonstrate feasibility of technology proposals, to create consensus within research groups ahead of standardization, early commercialization and ultimate productization.

Xilinx devices are preferred components for system vendors, carrier ecosystem and academia to develop 5G proof of concepts. The migration to higher performance lower-cost 16nm Zynq UltraScale+ MPSoC enables path to field trial systems and early commercialization of 5G networks. Technical collaboration with the industry leaders in enabling 5G standardization, technology development, and proof of concept programs has allowed Xilinx to gain in-depth

understanding of the system-level network planning, deployment, and the associated challenges. The partnerships and valuable experience obtained through these industry collaborations help Xilinx improve the existing designs, innovate new technologies for the next generation of its products; i.e., 7nm silicon family, optimize design tools, software and hardware libraries, IP blocks, and reference designs to address 5G commercialization and wide-scale deployment of networks at optimal power, performance, and cost while providing the customers with fast time to market advantage and much needed flexibility and scalability.

The most notable and invaluable tool to mention in this context is the Vivado® high-level synthesis (HLS) tool and libraries. Vivado HLS is extremely useful in rapid exploration and fine-tuning of multitude of algorithms that are under consideration for 5G. Translating algorithmic models to hardware using RTL takes months of effort, leaving no room for modification and iteration. Vivado HLS shrinks design timelines from months to weeks using translation of C models to verified RTL and test benches. Vivado HLS ties well with algorithmic modeling and simulation tools such as MATLAB™ software. This tool is much friendlier to algorithm and system designers by allowing them to import C models and C/C++ implementations without having to learn about RTL or block-level details and clocking constraints within FPGAs.
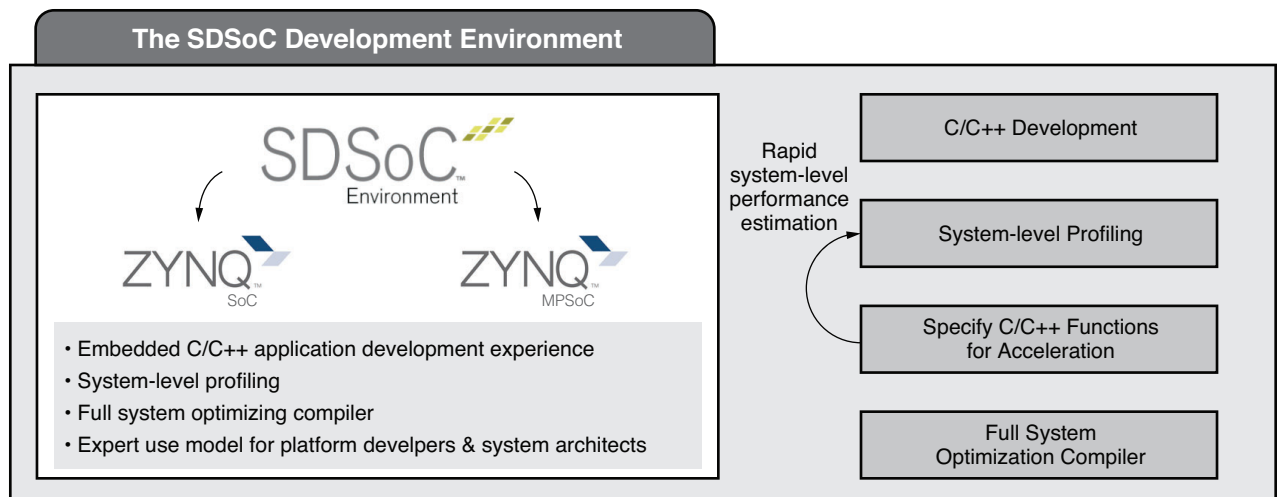

WP476_12_041316

Figure 12: **Xilinx SDSoC Software Development Environment [5]**

Xilinx SDSoC™ software development environment provides an embedded C/C++ application development tool including an easy to use comprehensive design environment for heterogeneous all programmable SoC and MPSoC implementations. Complete with the industry's first C/C++ full-system optimizing compiler, SDSoC delivers system level profiling, automated software acceleration in programmable logic, automated system connectivity generation, and libraries to speed up programming. It also enables end-user and third-party platform developers to rapidly define, integrate, and verify system-level solutions and enable their end customers with a customized programming environment.

The Xilinx SDAccel development environment for OpenCL, C, and C++, enables up to 25X better performance/watt for data center application acceleration leveraging FPGAs. The SDAccel development environment combines the industry's first architecturally optimizing compiler supporting any combination of OpenCL, C, and C++ kernels, along with libraries, development

boards, and the first complete CPU/GPU like development and run-time experience for FPGAs. See Figure 13.
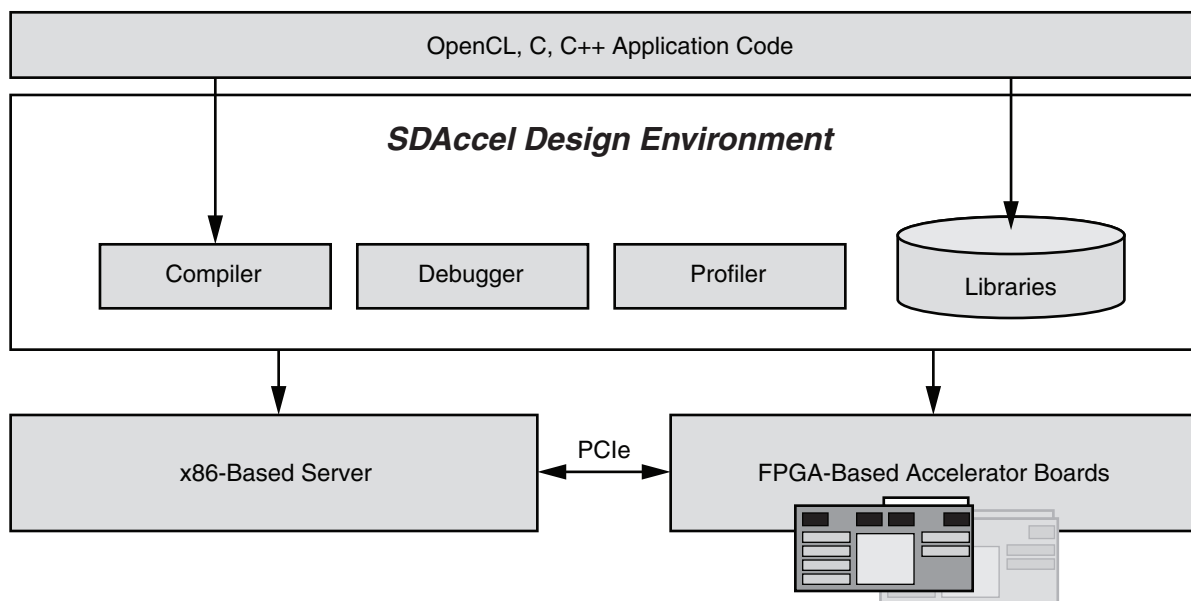
Figure 13: **Xilinx SDAccel Software Development Environment for OpenCL, C, and C++ [Ref 5]**

Xilinx is the leading provider for the wireless infrastructure for digital radio front end, connectivity, baseband acceleration and fronthaul/backhaul model and packet processing functions. Programmability, flexibility, and scalability of Xilinx silicon platforms are key value propositions that make the Xilinx portfolio of products an extremely good fit for these applications, particularly for radio where variety of form factors, frequencies, bandwidths and radio access technologies need to be supported. Xilinx introduced targeted devices for wireless infrastructure at 28nm that were well aligned with LTE rollout. Xilinx is the industry's leading provider of programmable devices for cellular radio, baseband CPRI switching, and baseband offload functions. Xilinx 20nm silicon platform is a preferred platform for 5G proof of concepts. Xilinx 16nm silicon technology is focused to address LTE-Advanced Pro product requirements and early commercialization of 5G proof of concepts. Xilinx is using deep market presence, internal expertise and experience, partnerships with system vendors to actively invest in technologies at 7nm to enable productization and broad rollout of 5G equipment in 2020 and beyond. Xilinx is committed to being the leading semiconductor supplier for wireless/wired infrastructure and C-RAN data centers.
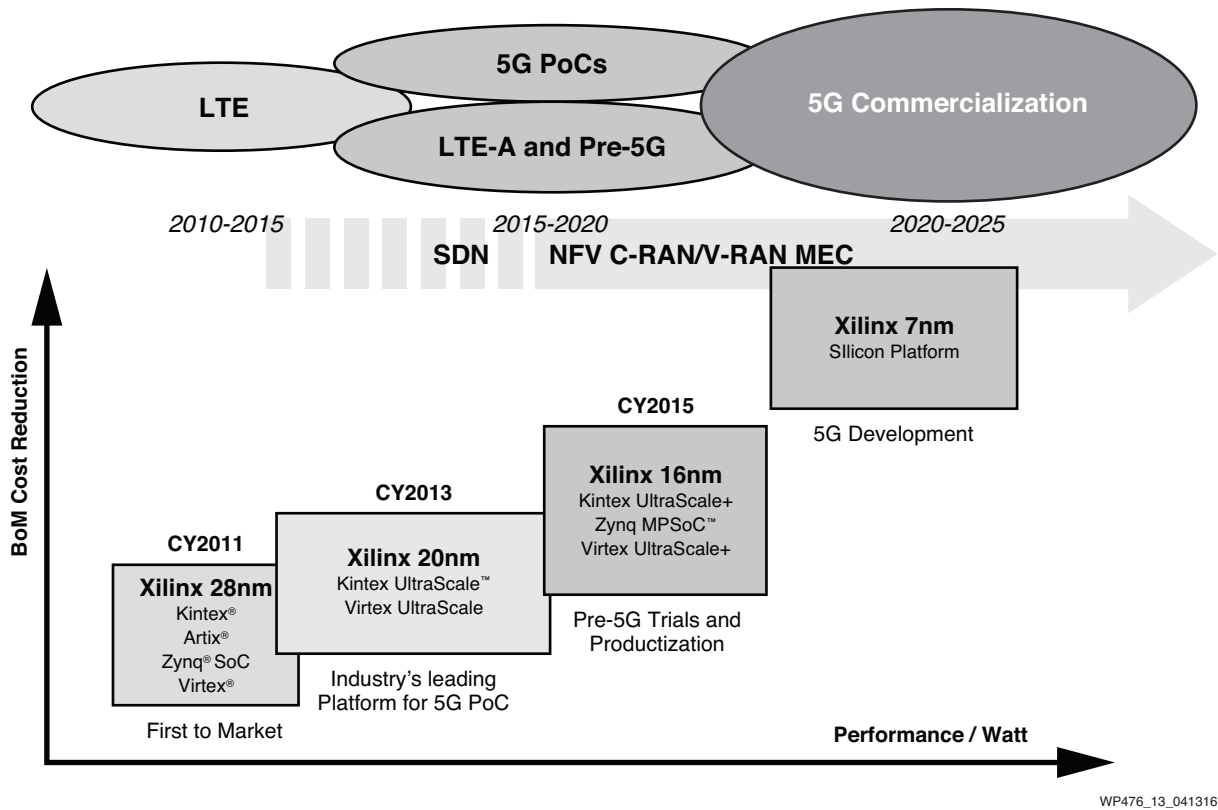
Figure 14: **Xilinx Product Roadmap is Aligned with 5G Development and Deployment Schedule**

Summary

The 5th generation of wireless systems will usher in tremendously revolutionized mobile communication systems that can provide ubiquitous, extremely high-throughput, and low-latency user experience anytime anywhere. Significantly increased system capacity and real-time responsiveness of 5G systems will enable new services providing the users with a truly immersive and rich experience. The realization of the mission-critical machine communication or IoT requires reliable connectivity among massive number of devices which can only be fulfilled via 5G systems.

Xilinx is playing an important role by providing the state-of-the-art all-programmable FPGA platforms and SoCs that can accelerate development of 5G proof of concepts, technology trials, and early implementation of large-scale commercial wireless products for 5G networks.

References

1. 3GPP RAN 5G Workshop Summary (RWS-150073), September 2015

2. Recommendation ITU-R M.2083-0, IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond, September 2015

3. NGMN 5G whitepaper, March 2015

4. Xilinx UltraScale+ Architecture and Product Overview

5. Xilinx Software Development Environments

Revision History

The following table shows the revision history for this document:

Date	Version	Description of Revisions
06/13/2016	v1.0	Initial Xilinx release.

Disclaimer

The information disclosed to you hereunder (the "Materials") is provided solely for the selection and use of Xilinx products. To the maximum extent permitted by applicable law: (1) Materials are made available "AS IS" and with all faults, Xilinx hereby DISCLAIMS ALL WARRANTIES AND CONDITIONS, EXPRESS, IMPLIED, OR STATUTORY, INCLUDING BUT NOT LIMITED TO WARRANTIES OF MERCHANTABILITY, NON-INFRINGEMENT, OR FITNESS FOR ANY PARTICULAR PURPOSE; and (2) Xilinx shall not be liable (whether in contract or tort, including negligence, or under any other theory of liability) for any loss or damage of any kind or nature related to, arising under, or in connection with, the Materials (including your use of the Materials), including for any direct, indirect, special, incidental, or consequential loss or damage (including loss of data, profits, goodwill, or any type of loss or damage suffered as a result of any action brought by a third party) even if such damage or loss was reasonably foreseeable or Xilinx had been advised of the possibility of the same. Xilinx assumes no obligation to correct any errors contained in the Materials or to notify you of updates to the Materials or to product specifications. You may not reproduce, modify, distribute, or publicly display the Materials without prior written consent. Certain products are subject to the terms and conditions of Xilinx's limited warranty, please refer to Xilinx's Terms of Sale which can be viewed at http://www.xilinx.com/legal.htm#tos; IP cores may be subject to warranty and support terms contained in a license issued to you by Xilinx. Xilinx products are not designed or intended to be fail-safe or for use in any application requiring fail-safe performance; you assume sole risk and liability for use of Xilinx products in such critical applications, please refer to Xilinx's Terms of Sale which can be viewed at http://www.xilinx.com/legal.htm#tos.

Automotive Applications Disclaimer

XILINX PRODUCTS ARE NOT DESIGNED OR INTENDED TO BE FAIL-SAFE, OR FOR USE IN ANY APPLICATION REQUIRING FAIL-SAFE PERFORMANCE, SUCH AS APPLICATIONS RELATED TO: (I) THE DEPLOYMENT OF AIRBAGS, (II) CONTROL OF A VEHICLE, UNLESS THERE IS A FAIL-SAFE OR REDUNDANCY FEATURE (WHICH DOES NOT INCLUDE USE OF SOFTWARE IN THE XILINX DEVICE TO IMPLEMENT THE REDUNDANCY) AND A WARNING SIGNAL UPON FAILURE TO THE OPERATOR, OR (III) USES THAT COULD LEAD TO DEATH OR PERSONAL INJURY. CUSTOMER ASSUMES THE SOLE RISK AND LIABILITY OF ANY USE OF XILINX PRODUCTS IN SUCH APPLICATIONS.