5G Base-Station with Hardware Acceleration for Non-Terrestrial Networks on a Space-Grade System-on-Chip

Michael Petry [1,2] MICHAEL.PETRY@TUM.DE
Raphael Mayr [2] RAPHAEL.R.MAYR@AIRBUS.COM
[1] Airbus Defence and Space GmbH, Willy-Messerschmitt-Straße 1, Taufkirchen, D-82024 Germany
[2] Technical University of Munich, Arcisstraße 21, Munich, D-80333 Germany

Michael Hennerich MICHAEL.HENNERICH@ANALOG.COM
Analog Devices GmbH, Otl-Aicher Strasse 60-64, Munich, D-80807 Germany

Abstract

Delivering 5G connectivity from space to consumer hardware via Non-Terrestrial Networks serves a variety of safety and convenience use-cases for consumers. This transformation of the cellular network incurs major technical changes, with space-based base stations posing the main disruption. Although a healthy competition between space companies has been sparked, technical details specifically from the satellite payload perspective, i.e., satellite processing technology and 5G stack integration, remain mostly proprietary and undisclosed, hindering system understanding, standards development, and overall research. We present a heterogeneous payload design for a completely on-board 5G base station (gNodeB) using the next-generation space-qualified VERSAL hardware, a multi-core CPU platform, and a tightly integrated RF front-end. We propose a "processing split" by distributing 5G layers between these components and show its benefits, particularly on the example of a hardware-accelerated 5G PHY layer. A brief technology readiness assessment and a discussion on open points concludes this paper.

1. Introduction

The concept of Non-Terrestrial Networks (NTNs) is not new and has been around since the mid-20th century. One of the earliest and most popular examples of a global-scale space-based communication system was satellite television, which, while transformative, offered only simplex communication—broadcasting signals in one direction. Full-duplex NTN systems, capable of both sending and receiving signals, became operational with the launch of systems like Iridium and Globalstar in the late 1990s. However, the initial promise of these systems often fell short due to technical limitations, high operational costs, and limited consumer adoption.

Today, NTNs are experiencing a resurgence, and this time, the chances for success seem guaranteed. Driven by advancements in satellite technology and radio frequency (RF) systems (specifically compact phased-array antennas), NTNs are on the brink of enabling truly global, high-speed communication, with first operational networks, such as Starlink, setting new standards. However, for most NTNs, the main disadvantage of proprietary systems of requiring vendor-specific hardware, applies.

On the contrary, 5G NTN distinguishes itself with technological openness and a standardized radio access mechanism, characteristics that have been the game changer before, such as for the Global System for Mobile Communications (GSM). For the first time, satellite connectivity is made available to anyone owning a 5G-capable device, removing the restriction of owning expensive satellite-compatible transmitter hardware. Furthermore, technological openness fosters an open development ecosystem, both on the space and consumer hardware side, delivering the benefits of competition (Azari, 2022).

While still in their infancy, multiple competitors have entered the 5G NTN market. At the heart of this new wave of NTNs are two primary approaches to NTN realization, as depicted in Fig. 1: (a) Relaying the raw 5G New Radio (NR) waveform between terrestrial 5G base stations and users using a simple "transparent" satellite, or (b) hosting the 5G base stations (including 5G processing) and hence handling the 5G waveform in orbit on a powerful "regenerative" satellite.

In this paper, we present a fully regenerative, heterogeneous payload design for a 5G gNodeB onboard a satellite.

Proceedings of the 14^{th} GNU Radio Conference, Copyright 2024 by the author(s).

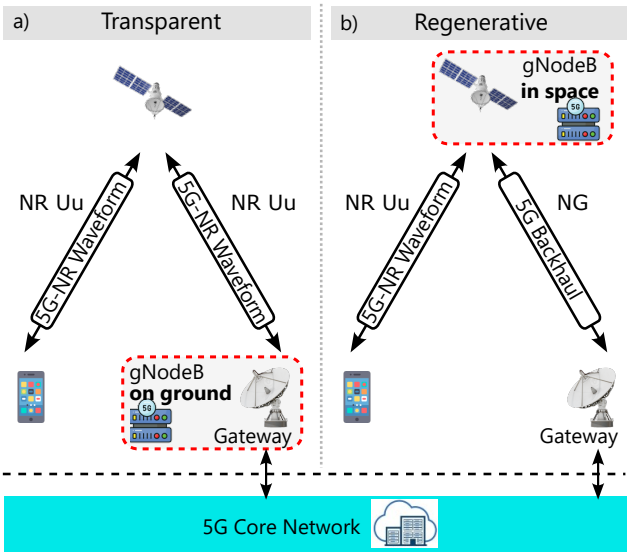

Figure 1. Illustration of the two different satellite system architecture for a link between a 5G user equipment and a 5G base station via a (a) transparent or (b) regenerative satellite.

Our approach leverages cutting-edge VERSAL hardware, a multi-core CPU platform, and an integrated RF front-end to achieve an efficient "processing split" across system components. By offloading key 5G stack layers to specialized hardware, including hardware-accelerated 5G PHY layer processing, we optimize performance while minimizing power consumption and address the unique challenges of space-based operation. The design offers a path forward for scalable and cost-effective NTNs and sparks future research directions needed to achieve full commercial deployment.

The paper is structured as follows: Sec. 2 provides an industry overview of NTNs that are already operational or in roll-out and systems in planning or evaluation stage. Sec. 3 dives into the presented system design of the full-fledged on board gNodeB, starting with an introduction of the concept and hardware platform, followed by a look at key design decisions, such as PHY acceleration, higher 5G layer distribution, and RF front-end integration. Section 4 concludes with a brief technological readiness assessment and an outlook on future work to make this system fully operation.

2. State-of-the-Art

While still in their infancy, multiple competitors have entered the 5G NTN market. At the heart of this new wave of NTNs are two primary approaches to satellite communication architectures, as denoted in Fig. 1: Transparent (a) and regenerative (b). The transparent architecture, where satellites act as relays, passing signals between user devices and terrestrial base stations, is simpler and requires less onboard processing power. However, its reliance on continuous high-bandwidth communication with ground stations can introduce latency and coverage gaps, particularly in maritime or polar regions.

The regenerative architecture, by contrast, places more processing power onboard the satellite, enabling the satellite to function as a full or partial 5G base station (gNodeB) in space. This approach significantly reduces dependence on ground stations through inter-satellite links, offers lower latency, and can provide coverage to more isolated regions. However, the increased complexity of onboard processing in space introduces new challenges related to power efficiency, radiation resistance, and thermal management, which must be overcome for widespread adoption.

In the following we provide an overview of selected key players in the industry.

2.1. Available Systems / In Roll-out

The most prominent example is SpaceX's Starlink Direct-to-Cell (D2C) constellation, also called Generation 2, which is currently in the roll-out and early system testing phase. Although not directly supporting the 5G NTN extension, but instead based on a server-side modified implementation of Long Term Evolution (LTE, 4G), which has a very similar waveform, it targets to serve unmodified smartphones with text, voice, and data. By offering two bandwidth options, the target data rates for Downlink are 4.4/18.3 Mbps with 1.4/5.0 MHz bandwidth, and 3.0/7.2 Mbps with 1.4/5.0 MHz bandwidth for Uplink operation (Space Exploration Holdings, LLC, 2024b). This service is achieved by means of 29988 (Space Exploration Holdings, LLC, 2024a) satellites in a multi-shell multi-inclination arrangement between 340 - 614 km, providing complete surface coverage.

Another company that is developing into a serious competitor is AST SpaceMobile. Created in 2017, their solely technological focus lies on designing a satellite constellation that is optimized for satellite to smartphone connectivity (AST SpaceMobile, LLC, 2024). With a planned constellation of 243 satellites with each carrying an extremely large antenna, their approach is to generate very narrow beams to close a high-datarate link budget to the smartphone. AST SpaceMobile has also secured partnerships with cellular providers that account for 70% of US mobile users and shared access to 850 MHz low-band spectrum (AST SpaceMobile, 2024). It is not yet confirmed, whether LTE or 5G is used.

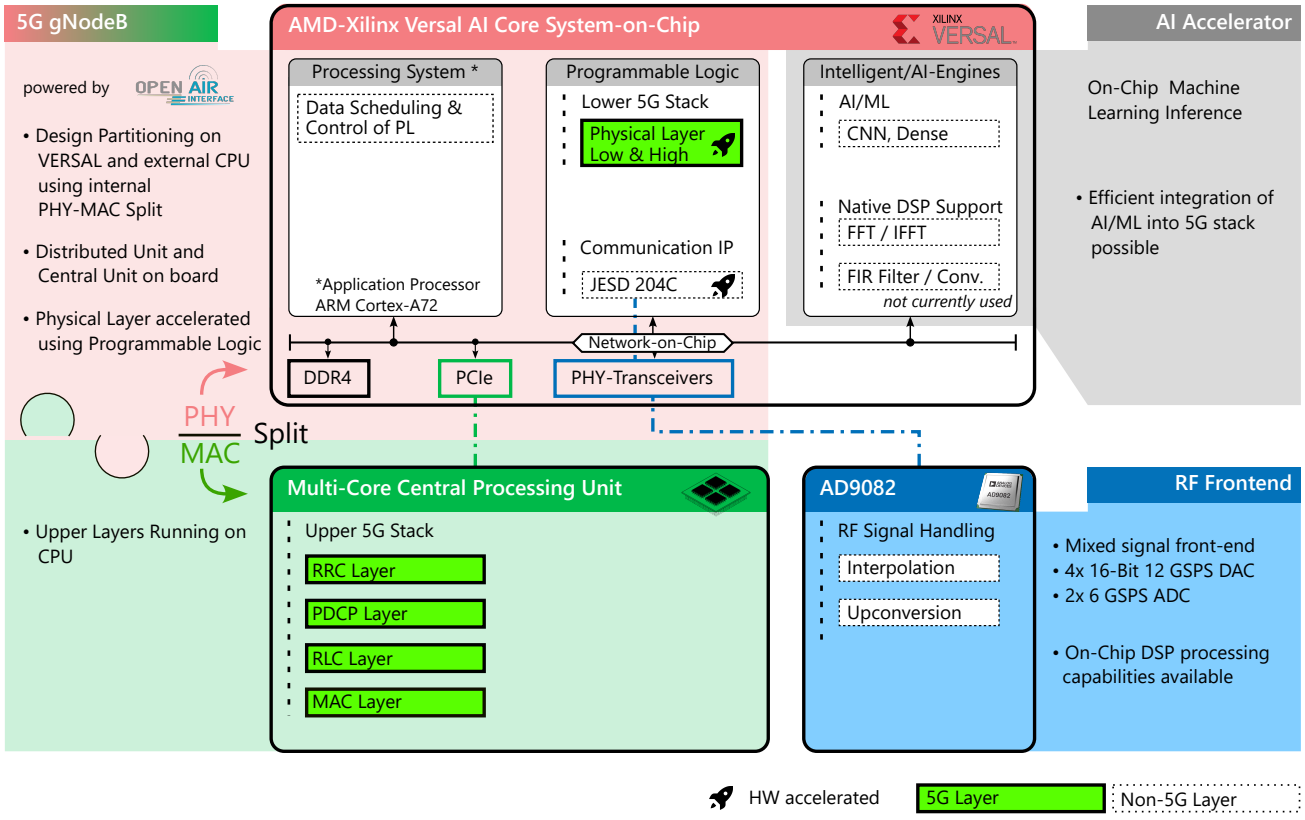

Figure 2. System design of a full 5G base station (gNodeB) including central and distributed unit realized on a heterogeneous processing platform. 5G layer processing (green-shaded boxes) is distributed between VERSAL (physical layer) and multi-core CPU (higher 5G layers), realizing a PHY-MAC processing spit. The AD9082 mixed signal front-end is connected to the VERSAL via the JESD204B/C procotol. AI inference capacity remains available trough the VERSAL's AI-Engines.

2.2. Planned Systems / In Evaluation

One of the pioneers in direct-to-smartphone communication is the company Lynk, who demonstrated the first smartphone-based satellite communication for text messages in 2020 and launched first 5G payloads in 2023. Although initially planning with ten satellites, they state that they require 1000 satellites for full continuous broadband coverage with plans for 5000 satellites at the completion (newspace.im, 2024).

Iridium, an established satellite communications provider, also recently announced their plans to re-purpose parts of their constellation to support 4G Narrow-Band Internet-of-Things and 5G NTN connectivity (Iridium Communications Inc., 2024). However, the achievable quality of service is likely very limited due to restrictions in RF satellite capabilities.

Some ventures even expand beyond the boundaries of Earth, such as Nokia's LTE/4G network on the moon, funded by NASA's Tipping Point Program. As more initiatives are being launched, the necessity of a space-grade cellular base station becomes obvious. However, no key player, including the above, is sharing technical details of their system. The only available information is within filings to the Federal Communications Commission.

In the next section, we present our implementation of a space-grade 5G base station.

3. 5G Base Station Concept

This section begins by outlining the general system concept and the target hardware platform. Afterwards, we focus on the hardware acceleration of the physical layer, followed by higher 5G layers. We conclude this section by details on the utilized RF front-end and its integration into the system.

3.1. Concept and Target Hardware Platform

The goal of this design is to realize a full gNodeB, i.e., distributed and central unit, on a space-grade platform. The system concept is designed around the fact that process-

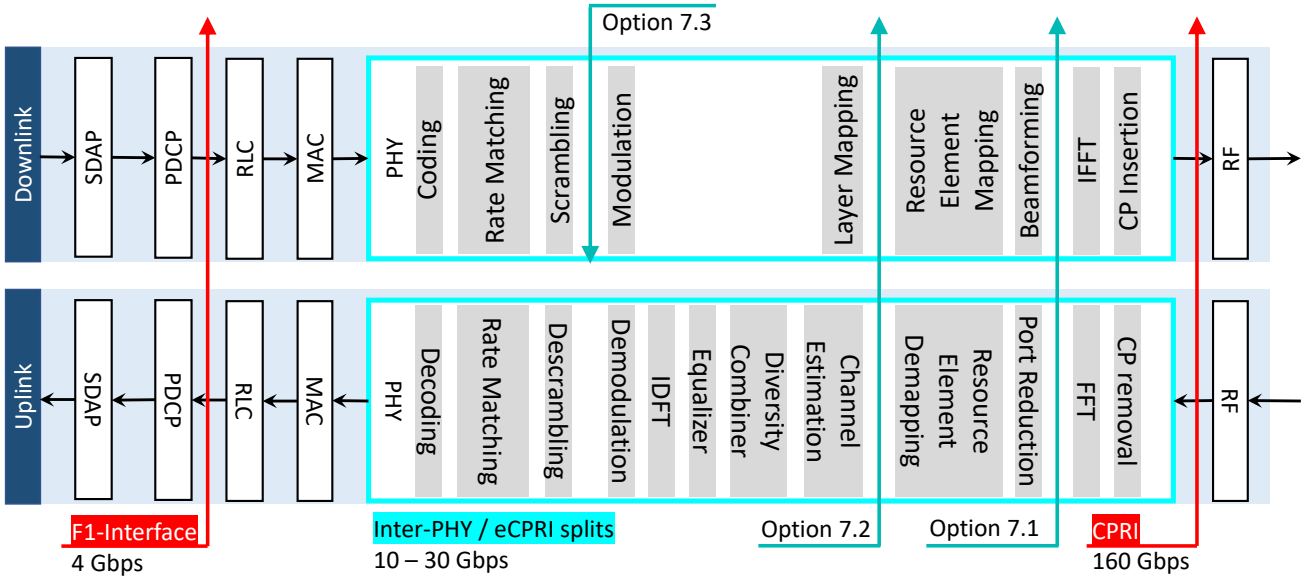

Figure 3. Illustration of the 5G layers within the 5G gNodeB for downlink and uplink. The physical layer (light-blue region) is presented in a more detailed view showing its main signal processing operations. Various central/distributed unit (CU/DU) splits are depicted with their associated interface bandwidth.

ing inside a 5G base station can be grouped into two categories. Category 1, which features highly parallelized and computationally-demanding algorithms, can be mapped to the physical layer. Those include, among others, coding, rate matching, scrambling and modulation, resource element mapping, beamforming, fast-fourier transform, and the respective inverse operations. On the other side, category 2 features algorithms that require dynamic decision-making based on real-time network conditions, which is why they benefit from a flexible general purpose processing architecture. We map the higher 5G layers, such as Medium Access Control (MAC), Radio Link Control (RLC), Packet Data Convergence Protocol (PDCP), and Radio Resource Control (RRC) to this category. This includes tasks such as scheduling, higher-level error correction, packet reassembly, encryption, integrity protection, and signaling control.

To best serve the requirements of both categories, a heterogeneous processing architecture, consisting of AMD-Xilinx's VERSAL AI core System-on-Chip (SoC) and a multi-core CPU is utilized. The VERSAL SoC is chosen due to its powerful Field Programmable Gate Array (FPGA) in combination with a CPU while being available in a radiation-tolerant / space-qualified version. Because the VERSAL's dual-core ARM A72 CPU is not powerful enough to handle the full load of the upper 5G layers, an additional multi-core CPU is added to serve this task.

The system architecture is depicted in Fig. 2.1. It can be seen that 5G related processing (green shaded boxes) is distributed between VERSAL and multi-core CPU, with the physical layer (both high and low part) being accelerated on the FPGA, while all upper 5G layers are outsourced to the multi-core CPU. The gNodeB processing is, hence, distributed within both components, realizing a design partitioning using a PHY-MAC layer processing split. Both components are interconnected using PCIe to provide sufficient bandwidth for the PHY-MAC interface.

In the following section, we discuss physical layer implementation.

3.2. Hardware Acceleration of the Physical Layer

Offloading the signal processing operations in the physical layer from the CPU to optimized hardware is a key requirement towards enabling a scalable and power efficient gNodeB implementation. Fig. 3 gives an overview of the 5G layers with a detailed look into the physical layer.

The VERSAL's FPGA is chosen as processing platform due to it being perfectly suitable to accelerate communications related algorithms within a rich ecosystem of available Intellectual Property (IP) cores. Another benefit is that the input and output RF data of the physical layer can be directly sent to the RF front-end via the I/O-transceivers attached to the FPGA, ensuring a seamless data transfer, as detailed in Sub-sec. 3.4. It is crucial to accelerate all components within the physical layer. 3GPP additionally defines a set of inter-PHY "splits", also called eCPRI splits,

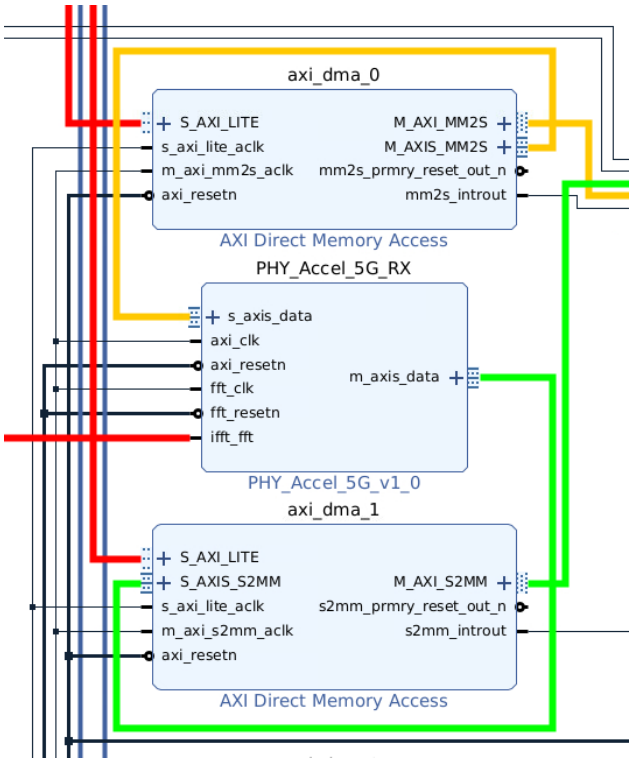

Figure 4. Screenshot of the Vivado software showing the block diagram of the RX-Path for 5G PHY acceleration. Input data stream lines are marked in yellow, output lines in green, and control lines in red.

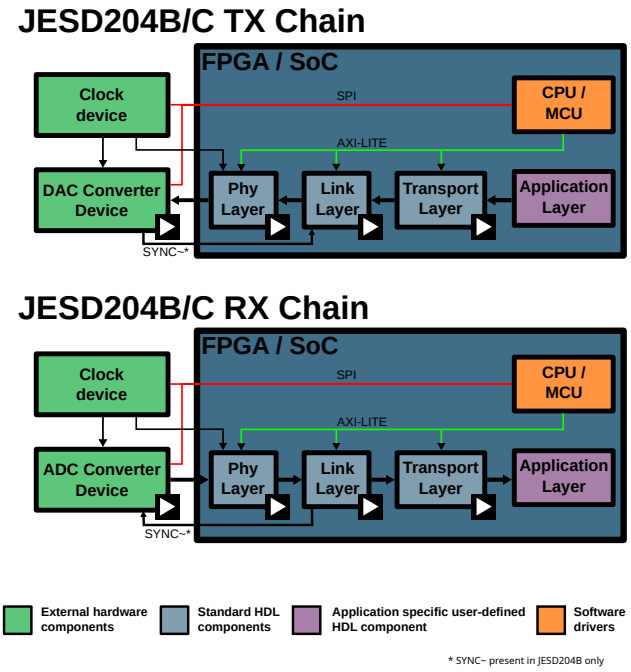

Figure 5. Illustration of the transmit and receive JESD204B/C datapath from application to RF front-end hardware.

that allows for grouping various operations together, which can be useful when combining cross-vendor IPs due to their standardized interfaces.

Up to now, we accelerate the Fourier Transforms in receive and transmit direction including scaling as well as cyclic prefix insertion and removal on the FPGA. See an extract of the FPGA's block design implementation in Figure 4. Hereby, the receive path is depicted with the input data stream lines in yellow, the output lines in green, and the control signals to and from the processing system in red. Running at 500 MHz, the Direct-Memory-Access (DMA) core zero fetches received data from memory, passes it to the PHY layer acceleration core running at the same frequency, and computes previously mentioned functions.

Following this, the acceleration core feeds the data to DMA core one to write the resulting data back to memory. For the transmit path, the inverse operations are accelerated with separate DMA cores for reading and writing. This second processing path runs at 500 MHz as well. Although, currently the processing chains have to fetch data from memory, it is planned to directly connect the acceleration cores to the RF front-end's I/O transceivers to diminish latency and lower the strain on the Network-on-Chip.

In the following subsection the higher 5G layer implementation is discussed.

3.3. Implementation of Higher 5G Layers

The higher 5G Layers, as shown to the left of the physical layer in Fig. 3, are deployed on the multi-core CPU. While the physical layer is implemented by synthesizing a combination of manually assembled IPs for the FPGA, the higher layers are directly used with little change as provided by the OpenAirInterface 5G gNodeB framework.

Although the 5G software stack has to be compiled for the AARCH64 instruction set architecture (ISA) to be executable on the multi-core CPU, which is not natively supported by OAI, we enable this by identifying x86 specific passages in the software stack and porting these to AARCH64, thereby remaining generic to be still able to compile for x86. On the one hand, these x86 specific code pieces are intrinsics, which are mostly found in the PHY and MAC layers. As we aim to accelerate most of the PHY layer, the PHY intrinsics do not have to be ported. For the MAC layer, the remaining instrinsics are mapped to AARCH64 leveraging SIMDE. Other x86 specific functions include, for instance, timing functions. Furthermore, due to the restricted resources on the VERSAL itself, the software stack is cross-compiled on a powerful Desktop-

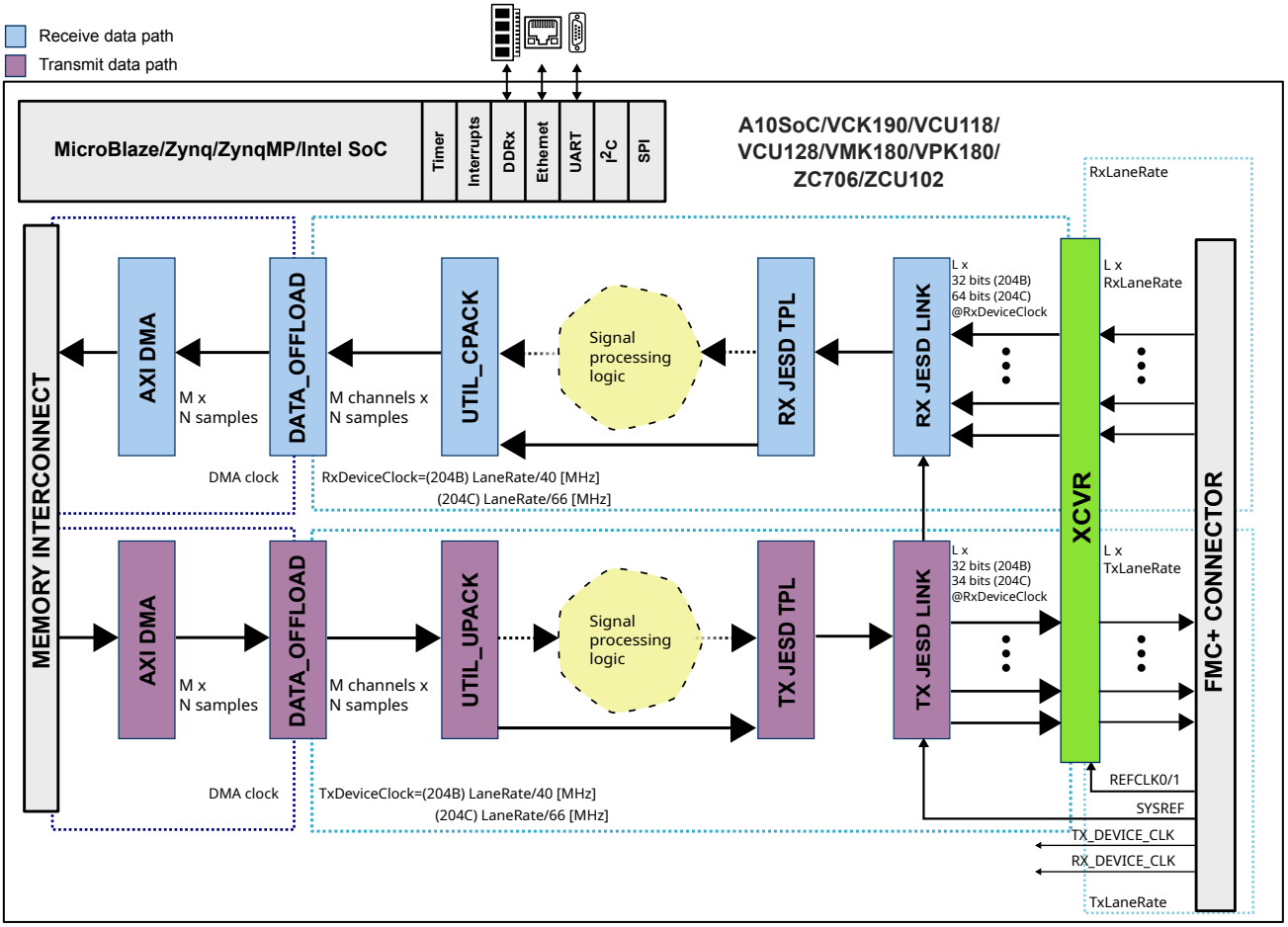

Figure 6. In-depth view into the JESD204B/C driver showing the dataflow from the I/Q buffer in memory to the FMC connector to which the AD9082 is connected.

PC to build the ARM executables. Hereby, containerization is employed to be able to distribute the 5G gNB software part in an efficient manner independent of the underlying operating system's features.

3.4. Radio Front-end Integration

After performing 5G processing eventually a waveform signal is generated that is to be transmitted through means of an RF front-end. For this task, Analog Devices AD9082 mixed signal front-end (MxFE®) is used in this design. We begin by introducing this highly specialized chip and its capabilities and then detail the related firmware design and technical integration of it into the system architecture.

3.4.1. Chip Capabilities

The AD9082 is a highly integrated device with a 16-bit, 12 GSPS maximum sample rate RF digital-to-analog converter (DAC) core, and 12-bit, 6 GSPS maximum sample rate RF analog-to-digital converter (ADC) cores. The AD9082 is well-suited for applications requiring both wideband ADCs and DACs to process signals with wide instantaneous bandwidth. With typical applications including wireless communications infrastructure, broadband communications systems, microwave point-to-point, phased array radar and electronic warfare, and electronic test and measurement systems, it is perfectly suited for the bandwidth hungry demand generated by 5G base station operation.

To interface with the VERSAL's FPGA, the device features eight transmit lanes and eight receive lanes that support 24.75 Gbps/lane JESD204C or 15.5 Gbps/lane JESD204B standards. It also has an on-chip clock multiplier and digital signal processing (DSP) capability that includes configurable digital upconverters (DUCs) and downconverters (DDCs) with 48-bit numerically controlled oscillators (NCOs), fast frequency hopping support, and a pro-

grammable 192-tap finite-impulse response (FIR) filter for receive equalization. Since in this implementation the 5G PHY is implemented on the VERSAL itself, the DSP datapath is bypassed to allow a direct connection between the converter cores and the JESD204B/C data transceiver port.

3.4.2. Firmware Design

Accessing the AD9082 from the application side and handling these extremely high data rates between VERSAL and AD9082 requires a hardware-software co-design. Analog Devices provides publicly available FPGA reference designs; here, the reference design for the AD9082 and AD9081 MxFE is used. It contains a JESD204B/C HDL solution that defines four layers: physical layer, link layer, transport layer, and application layer, as shown in Fig. 5. For the first three layers, Analog Devices provides standard components that can be linked to provide a full JESD204B/C protocol processing chain. Each Verilog HDL IP core is supported by Linux device drivers and fully integrated with a JESD204 interface framework. Additionally, Linux Industrial I/O Subsystem (IIO) device drivers are utilized that allow full configuration and control over the MxFE data converter and transport layer IP and DMA cores. This setup can either provide IIO DMA streams (IQ samples or payload data buffers) or allow for the insertion of custom application logic and signal processing. Fig. 6 provides an in-depth look into the data flow on the FPGA side. Since 5G physical layer processing is achieved with external custom logic (as described in Sub-sec. 3.2), the DMA streams are directly utilized as I/Q data interface. The 5G physical layer IPs are integrated into Analog Device's reference design.

4. Conclusion

In this paper we motivated the development of 5G base stations for the deployment on board satellites in space. We presented a system design that realizes a complete gNodeB (distributed and central unit) on a heterogeneous processing system utilizing the VERSAL, a separate multi-core CPU, and a specialized RF front-end chip. While the VERSAL accelerates physical layer signal processing on its FPGA, the higher 5G layers are deployed on a tightly interconnected multi-core CPU. The RF front-end is connected to the VERSAL's FPGA using their GTY transceivers operating on the JESD204B/C protocol.

Although the concept's feasibility has been confirmed on a hardware setup consisting of evaluation boards of the related chips, key system parameters, such as capacity of the complete system with respect to total throughput, number of users handled simultaneously, maximum speed per user link, etc., must still be determined. Furthermore, a custom platform design of the flight-ready system is still in the development stage and requires further validation for reaching a sufficient Technology Readiness Level (TRL).

Future Work Follow-up work can be roughly grouped into three categories: Continuously updating the 5G stack to the newest released version of Open Air Interface (currently LEO), connecting the gNodeB to a 5G core network and implementing a corresponding back-haul via satellite radio interface (SRI), and extending the test campaign of the gNodeB to multiple virtual and physical smartphones. Furthermore, with the outlook to the deployment of inter-satellite links, the Xn interface, which allows to directly connect two gNodeBs without the necessity to loop over the core network used for faster handover procedures, is to be implemented.

References

AST SpaceMobile. BlueBird 1-5 - AST SpaceMobile. https://ast-science.com/spacemobile-network/bluebird-1-5/, 2024. [Online; accessed 15-September-2024].

AST SpaceMobile, LLC. Our Technology - AST SpaceMobile. https://ast-science.com/spacemobile-network/our-technology/, 2024. [Online; accessed 15-September-2024].

Azari, M. et al. Evolution of non-terrestrial networks from 5g to 6g: A survey. *IEEE Communications Surveys Tutorials*, 24(4):2633–2672, 2022. doi: 10.1109/COMST.2022.3199901.

Iridium Communications Inc. Iridium Unveils Project Stardust; Developing the Only Truly Global, Standards-Based IoT and Direct-to-Device Service. https://www.iridium.com/ntn-direct/, 2024. [Online; accessed 15-September-2024].

newspace.im. Lynk (Tower). https://www.newspace.im/constellations/lynk, 2024. [Online; accessed 15-September-2024].

Space Exploration Holdings, LLC. Application for approval for orbital deployment and operating authority for the SpaceX Gen 2 NGSO satellite system. https://fcc.report/IBFS/SAT-LOA-20200526-00055/2378671, 2024a. [Online; accessed 15-September-2024].

Space Exploration Holdings, LLC. SpaceX Gen2 Direct-to-Cellular System - Attachement A: Technical Information to supplement schedule S. https://cdn.arstechnica.net/wp-content/uploads/2023/05/SpaceX-T-Mobile-Technical-Narrative.pdf, 2024b. [Online; accessed 15-September-2024].