

Energy footprint reduction in 5G reconfigurable hotspots via function partitioning and bandwidth adaptation

Nikolaos Bartzoudis*, Oriol Font-Bach*, Marco Miozzo*, Carlos Donato^{†‡}, Pavel Harbanau*, Manuel Requena*, David López*, Iñaki Ucar[†], Arturo Azcorra Saloña^{†‡}, Pablo Serrano[‡], Josep Manges* and Miquel Payaró*

*Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)-CERCA, Spain

Email: {nbartzoudis, ofont, mmiozzo, pharbanau, mrequena, dlopez, mpayaro, jmanges}@cttc.es

[†]IMDEA Networks Institute, Spain

Email: {carlos.donato, arturo.azcorra}@imdea.org

[‡]Universidad Carlos III de Madrid, Spain

Email: inaki.ucar@uc3m.es, pablo@it.uc3m.es

Abstract—Cloud-based radio access networks (C-RAN) are expected to face important challenges in the forthcoming fifth generation (5G) communication systems. For this reason, more flexible C-RAN architectures have recently been proposed in the literature, where the radio communication stack is partitioned and placed across different RAN nodes to tackle the 5G capacity and latency requirements. In this paper, we show that this functional split also supports energy efficiency, especially when it is combined with bandwidth adaptation. To this aim, we have built a dynamic hotspot prototype, where the hardware-accelerated physical-layer is placed in the remote radio head, and higher software-based layers are placed in a server (either directly connected or remotely accessible). This setup allowed us to experimentally evaluate the power consumption of key hardware modules when adapting the bandwidth and the modulation and coding scheme. The real-time operation of the testbed allows further experimentation with different 5G use cases and the evaluation of other key performance indicators.

Keywords: HW-SW partitioning, 5G networking, 5G dynamic hotspots, energy efficiency, reconfigurability, power measurements.

I. INTRODUCTION

Fifth generation (5G) wireless communication systems will be required to support a much wider range of devices and applications than the currently deployed technologies. The new service requirements (e.g., wider bandwidth and shorter latency) together with key 5G technologies such as Cloud radio access networking (C-RAN) [1], network function virtualization (NFV), software defined networking (SDN) and mobile edge computing (MEC) [2] manifest the emerging need for flexibility and network programmability in 5G systems [3]. Through them, network operators would leverage traffic dynamics and optimize the utilization of resources across different segments of the network [4].

The classic C-RAN architecture, where the Baseband Units (BBU) are located in the Cloud and connected with the remote radio heads (RRHs) using common public radio interface (CPRI) links, would have to be revised to serve specific 5G use

cases. In more detail, considering the current industry proposals and standardization efforts [5], this C-RAN scheme will no longer be able to handle the increased capacity requirements in a 5G network architecture [6]. Such limitation could be overcome by employing a flexible functional split between the BBU and RRH, applied either at communication stack-level or algorithm-level (e.g., MAC-PHY, RLC, PDCP); this will allow to place more baseband processing functions at the RRHs end [7]. The goal of this split is to maintain the benefits of C-RAN, while relaxing at the same time the stringent latency and bandwidth requirements of CPRI. Moreover, in 5G networks this function split might even be applied dynamically according to a specific network reconfiguration which will rely on SDN, NFV and, of course, on programmable and configurable devices that will be used to serve a number of logical networks operating over a single physical network infrastructure.

Albeit very important, the capacity and low latency are not the only key performance indicators (KPIs) that could be addressed by applying a flexible function split of the 5G communication stack. In this paper, we focus on the energy efficiency (NRG) KPI and how this can be achieved when applying a given functional split or bandwidth adaptation. The main contribution of this work is the development of all the necessary software (SW) and hardware-accelerated (HWA) building blocks and their integration in a testbed to experimentally measure and evaluate the power that is consumed by key signal processing components under different conditions and system parameterizations. The KPI-driven and multi-objective functional split of the communication stack is a quite young research area with very scarce experimental implementations. A significant difference introduced by our work compared to other all-software experimental evaluations [8] is the use of HWA functions (i.e., the entire PHY-layer in this respect), which guarantee that computationally-intensive real-time functions can run without performance degradation, allowing in this way to realistically assess the chosen KPIs.

II. SYSTEM DESCRIPTION

We consider a number of reconfigurable hotspots deployed in places where the coverage, number of subscribers and overall capacity requirements cannot be sufficiently served by the macro base stations (BSs) such as in shopping malls, large venues, train stations and public service buildings. The reconfigurability of the hotspots is based on their bandwidth adaptation and communication stack partitioning features. The NRG KPI is particularly relevant in this 5G use case, because, unlike in conventional macro BSs where the baseband processing consumes a small fragment of their total energy budget (i.e., the 13%), in small cells this does not hold true anymore, reaching up to the 42% [9].

In the absence of a standardized and fully defined 5G air interface and protocol stack, the 4G long term evolution (LTE) has been considered. The LTE stack is based on the following SW and HWA building blocks:

- SW: The LENA LTE open source simulator and emulator, born and maintained at CTTC [10]; LENA is able to simulate and emulate Evolved Node Bs (eNBs), user equipment (UE) and the evolved packet core (EPC).
- HWA: The LTE-based downlink (DL) physical-layer (L1) developed at CTTC targeting field programmable gate array (FPGA) devices [11].

The HWA and SW communication stack functions defined above could be partitioned, placed and executed in different processing elements distributed across the RAN. In this paper we define the following three functional splits, denoted hereafter as network configurations (NETCFG):

- In NETCFG1 the L1 is configured locally at the eNB, whereas the higher layers together with the EPC are placed in the Cloud in order to offload the processing of the eNB.
- In NETCF2 the L1 is also configured locally at the eNB, whereas the higher layers are placed in a proximity MEC-like server encountered at the Cell's edge.
- In NETCFG3 the eNB is configured as a typical RRH, while the entire eNB communication stack is placed in the Cloud together with the EPC, implementing likewise a centralized C-RAN setup.

In the developed prototype, apart from switching between NETCFGs, a number of wireless communication parameters (WCP) can be flexibly modified, including among others the signal bandwidth, the traffic load, the modulation and coding scheme (MCS) and the output power of the radio frequency (RF) transceiver integrated circuit (IC). Adapting the WCPs enables system-wide optimizations, provides a quality of experience (QoE) that adapts to specific scenario conditions (e.g., data traffic) and conforms with energy saving policies or operational expenditure (OPEX) demands.

In order to leverage energy saving benefits, a NETCFG transition could take place or a bandwidth adaptation could be applied in response to actual traffic conditions. The relation of traffic load with the three selected NETCFGs is shown in

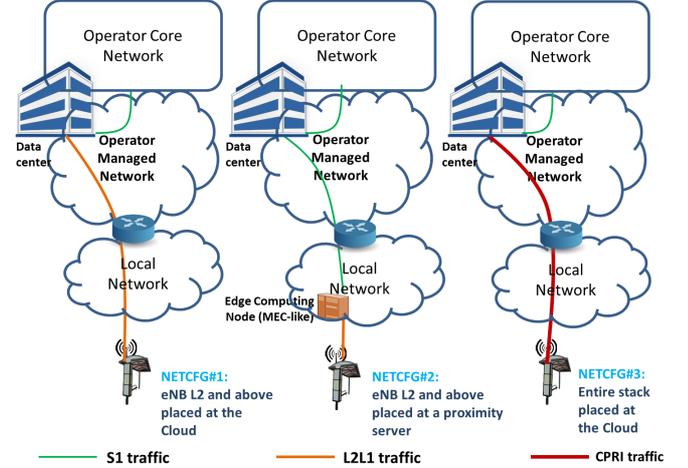


Fig. 1. The relation of traffic load with the three selected NETCFGs.

Figure 1. More specifically, while the common public radio interface (CPRI) traffic has stringent requirements that can be satisfied mainly by fiber (ideal transport channel with 250 μ s of one way latency and 2.5 Gbps for supporting a standardized LTE 2x2 20 MHz), the medium access (MAC)-L1 (L2-L1) traffic can be served with less stringent requirements, such as a sub-ideal transport channel with 6 ms of latency and 150 Mbps of total capacity [14]. The S1 traffic between the eNB and the EPC has even less demanding needs, since it requires a non-ideal transport channel, which implies one-way latency of up to 30 ms and limited and variable bandwidth.

Originally, the SW and HWA functions were not designed to inter-operate with each other. For this reason they had to be modified, adapted, significantly extended and suitably interfaced to serve the goals of the flexible function split. As a result, LENA was converted from an LTE simulator to a real-time experimentation environment able to transmit and receive LTE DL signals over-the-air. The low-level insights of this development and integration effort will not be detailed herein, as they fall beyond the scope of this paper. It is important however to give a brief overview of the L2-L1 interface, since it is a critical part of the functional split and in charge of providing reliable connectivity between the SW and HWA building blocks.

LENA's L2 provides specific messages on a subframe basis containing data control indication (DCI) and transfer blocks (TB). DCIs are filled with control information such as MCS, TB size and allocation bitmap. A dedicated L2-L1 software interface runs in real-time Linux at the processing system (PS) of a Xilinx System-on-Chip (SoC) device, while HWA blocks are implemented in the programmable logic (PL) area of the same device (Figure 2). On the one end, the L2-L1 interface application is connected with LENA's L2 using a gigabit Ethernet (GigE) link and on the other end it is connected with the PL using a proprietary advanced extensible interface (AXI4) by Xilinx. The L2-L1 interface application receives UDP packets containing L2-L1 interfacing frames assembled according to an agreed format. The received frames are then parsed and forwarded to the PL (including control information and TBs) using an appropriately designed ring buffer solution. LENA uses the ns-3 internal scheduler in hard real-time mode

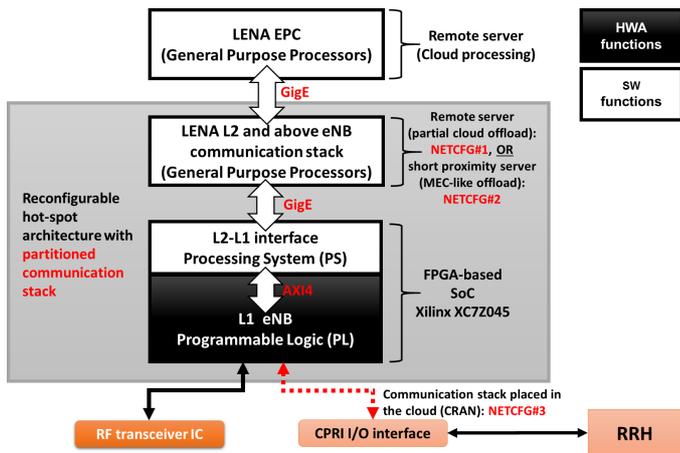


Fig. 2. Overview of the SW and HWA building blocks.

for assuring transmission of UDP packets each millisecond with minimal jitter. A number of performance optimizations were applied towards this end. At the same time, the Linux operating system running at the PS features a fully pre-emptive kernel in order to achieve real-time deterministic behaviour, minimal latencies and reduce the application and kernel-level jitters.

III. EXPERIMENTAL SET-UP

This section includes information of the different HW equipment comprising the assembled testbed and also the measurement solution that was employed to evaluate the NRG KPI at the partitioned eNB (reconfigurable hotspot).

A. Testbed description

1) *Hardware boards, components and equipment*: The L1 HWA processing blocks and the L2-L1 SW interface of the eNB prototype are hosted in the Xilinx ZC706 board which features the SoC device mentioned in section II (Xilinx Zynq XC7Z045). The Analog Devices AD-FMCOMMS3 RF transceiver board is plugged to the Xilinx ZC706 board. A power amplifier (PA) unit, RF band filter and an antenna were also interfaced with the AD-FMCOMMS3 board. A Linux kernel space application that runs at the PS side of the Zynq XC7Z045 device was used to tune and program the AD9361 radio frequency (RF) transceiver IC (RFIC) at the AD-FMCOMMS3 board. The Xilinx ZC706 board is interfaced with CTTC's EXTREME Testbed [12] using a GigE connection. EXTREME comprises generic purpose servers configured either as a datacenter or distributed throughout the network. For the case of the currently described testbed EXTREME hosts LENA's SW-based eNB and UE stack (i.e., L2 and above), as well as the EPC. LENA generates DL traffic which is forwarded from the L2 to the Xilinx ZC706 board, whereas the uplink (UL) traffic maintains the original SW functionality and connectivity of the LENA simulator assuming an ideal channel: a software process at the UE (UL transmitter) communicates via GigE to another software process in the eNB (UL receiver). A simplified representation of the eNB prototype hardware components is shown in figure 3.

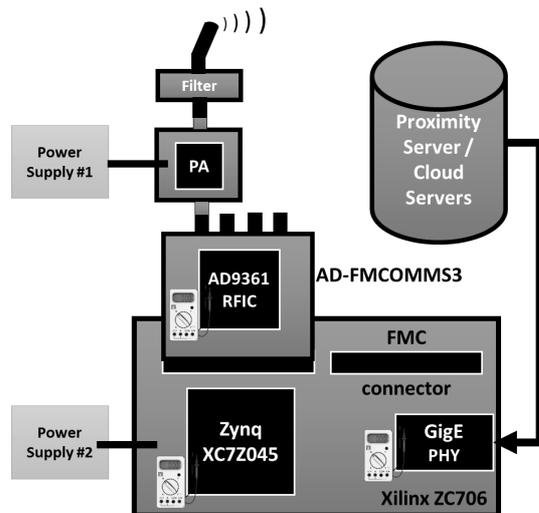


Fig. 3. The components comprising the eNB prototype and the ICs where power measurement validation was made.

2) *Power measurements framework*: The NRG KPI was assessed by measuring the energy footprint of the following devices under test (DUT) of the eNB prototype: the Zynq XC7Z045 baseband processor, the AD9361 RFIC and the Ethernet PHY IC (however results of the latter are not included in this paper). The voltage and current measurements for the Zynq XC7Z045 device were conducted using the Texas Instruments (TI) power controllers of the Xilinx ZC706 board. The controllers were reached via the power management bus (PMBus) connector of the board, using the TI Fusion USB cable and the TI Fusion Digital Power Designer graphical user interface (GUI).

On the other side, the power measurements for the AD9361 and Ethernet PHY ICs were made feasible by using the experimental setup presented in Figure 4. Three main blocks can be distinguished in this setup: the DUTs, the system to acquire and quantify the power measurements and, finally, the controller where the software framework runs. In the case of the AD-FMCOMMS3 board, the measurements were taken through the 3.3V power rail of the AD9361 RFIC.

The DUTs were connected with the measurement framework through two custom circuits, denoted hereafter as *ad-hoc measurement circuits*, which were specially designed to operate over these voltage ranges. The *ad-hoc measurement circuits* extract the voltage and convert the current with a high-precision sensing resistor and amplifier. The second block of the power measurement framework is based on a shielded connector block (SCB), implemented by National Instrument's (NI) SCB-68A, that allows to interconnect several inputs and multiplex them in the data acquisition (DAQ) card. Both outputs of the *ad-hoc measurement circuits* are connected to SCB rails. This block sends all the signals to the NI PCI-6289 multifunction DAQ device that quantifies the measurements to post-process the results. Finally, the third block is the software controller running in a general purpose computer, where the peripheral component interconnect express (PCIe) Ethernet and DAQ cards are attached. The software application tool *DAQ-Acquire* [13] demultiplexes the measurements of the DAQ card

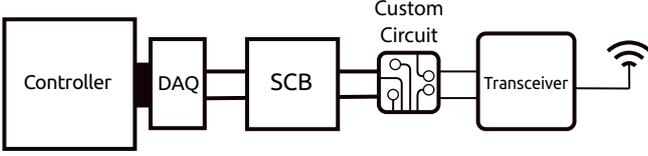


Fig. 4. The measurement setup for the AD9361 RF transceiver IC device.

and dumps all the values to a file. Finally, the captured results were analyzed using the R tool.

B. Methodology

A set of scenarios were initially defined and preliminary experiments were conducted with the intention to characterize the hardware setup features and limitations and also to identify the hardware ICs that contribute most to the power consumption. The results of these experiments were obtained as a function of the MCS index, the carrier bandwidth, the utilized resource block groups (RBGs) and in the case of the AD9361 RFIC also as function of the output power of the RF signal. The time resolution of the DAQ card responsible for gathering the measurements, is up to $0.1 \mu\text{s}$ (10 MHz sampling frequency). This feature provides a wide margin of values to work, which in our case was fixed to $1 \mu\text{s}$. Using this resolution and 30s for each data capture (some of which were repeated), we could sufficiently characterize the power consumption of the DUTs. The power measurements at the Zynq XC7Z045 baseband processor were conducted using NETCFG2 and NETCFG3 with four different bandwidth configurations (1.4, 5, 10 and 20 MHz), three RBG values for each bandwidth (lowest, half and maximum) and three MCS indexes. The functional split applied to the NETCFG1 and NETCFG2 provide the same results for the Zynq XC7Z045 baseband processor in this respect. The power measurements for the AD9361 RFIC apply to all three NETCFGs and apart from the mentioned parameters (bandwidth, RBGs and MCS index), three values for the RFIC output power were also used (i.e., -19, -29, -39 dBm). The results were obtained by generating DL traffic using the eNB prototype in each experiment, while fixing in each iteration a given bandwidth, MCS index, RBG load and output power (for the case of the AD9361 RFIC). Once each experiment was finalized, the results were analysed using an R script to assess their validity. As the number of samples obtained in each experimental iteration was over 30 million, we only used a fraction of these datasets to alleviate the data post-processing. Finally, in order to validate the experimental results, we compared them with vendor's datasheets.

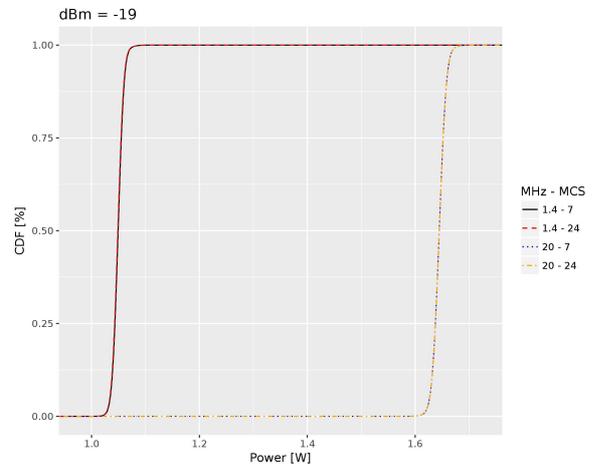
IV. PERFORMANCE EVALUATION

This section includes the results per DUT as obtained after post-processing the captured data per each experiment and iteration.

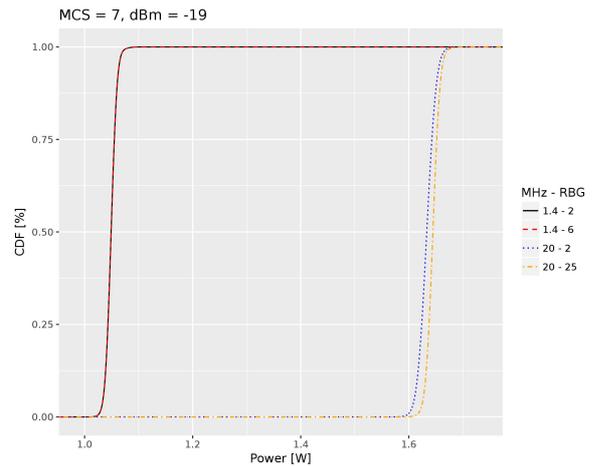
A. Power measurements of the AD9361 RFIC

As mentioned before, the power measurements of the AD9361 RFIC apply to all three NETCFGs. The 3.3 Voltage rail of the AD9361 RFIC was measured considering that it is the main power consumption contributor. Following the methodology presented in Section III-B, we first analyzed the

effect of the MCS index and the RBG load in the power consumption of the RFIC for two bandwidth configurations. As it is shown in Figure 5, neither the MCS nor the number of utilized RBGs have a notable impact on the power consumption. On the other hand, it can be clearly observed a significant difference in the power consumption between the lowest and highest bandwidth configuration. Similarly, Figure 6 highlights the impact of the output power on the consumption of the AD9361 device. Although the difference observed in power consumption when the RFIC output power is attenuated is not as impressive as in the case of bandwidths adaptation, both WCPs can serve energy efficiency goals in specific end use scenarios. It is worth mentioning that the bandwidth adaptation would require a rapid reprogramming of the RFIC to satisfy the latency requirements related with the reconfiguration procedure.



(a) CDF of two MCSs and bandwidth configurations.

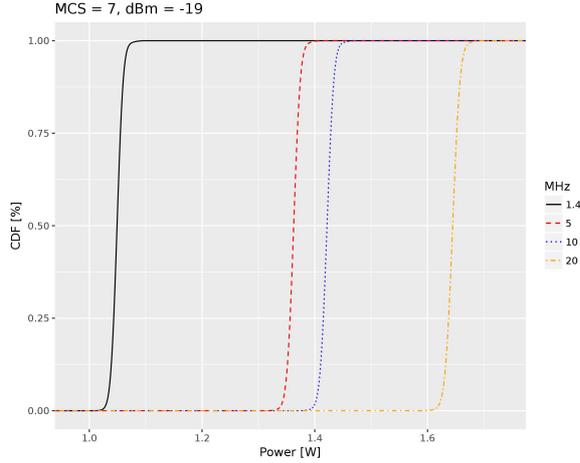


(b) CDF of two RBGs and bandwidth configurations.

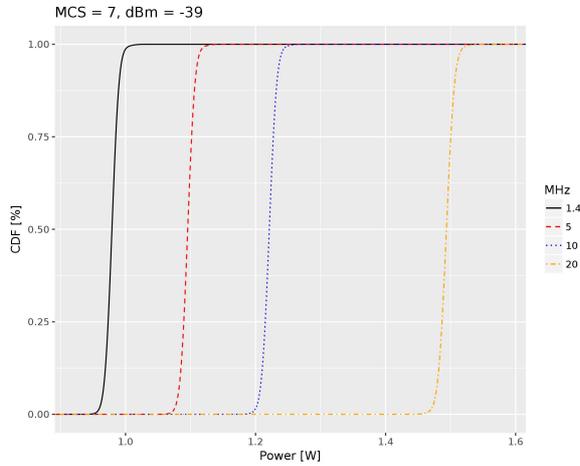
Fig. 5. CDF curves showing the negligible impact of the MCS index and RBG load on the power consumption of the AD9361 RFIC.

B. NETCFG1/2: power measurements of the Xilinx Zynq L1 processor

As mentioned before, considering that the Xilinx ZC706 board is configured with the same HWA functions in NETCFG1 and 2, the power measurements of the Xilinx Zynq



(a) CDF of captured data at -19 dBm.



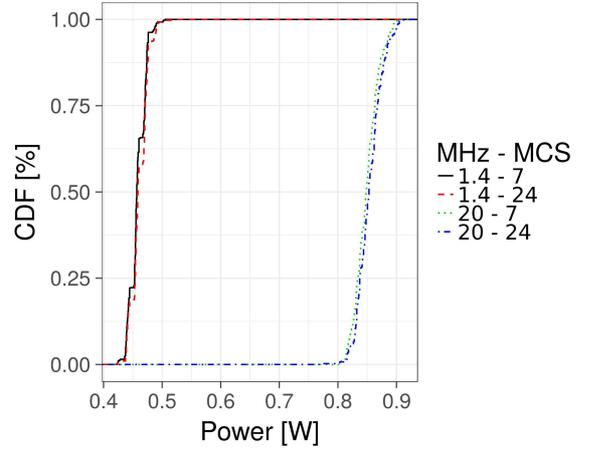
(b) CDF of captured data at -39 dBm.

Fig. 6. CDF curves showing the impact of the RFIC output power on its power consumption.

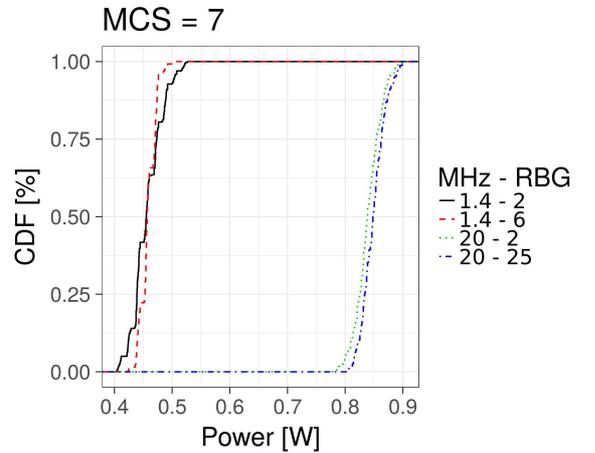
XC7Z045 SoC device apply to both of them. The VCCINT rail of the SoC device was selected to be measured because it is the one driving both the internal logic of the PL and the PS. As in the case of the results of Subsection IV-A, while the modification of the MCS index and RBG load is not affecting the power consumption of the device (Figure 7), the bandwidth adaptation indeed has a great impact on the power it drains (Figure 8). The data captured in the Xilinx Zynq baseband processor with the help of the TI measurement solution is not having the same precision compared to the measurement solution employed for the AD9361 RFIC. The bandwidth adaptation would require reconfiguration of the baseband processor and exchange of specific signalling messages between the eNB and the UE (coming from higher layer of the stack), which will have to abide with system-wide latency restrictions.

C. NETCFG3: power measurements of the Xilinx Zynq baseband processor

In NETCFG3 the Xilinx Zynq XC7Z045 SoC device acts a RRH that hosts a CPRI slave interface. The implementation occupies a small fraction of the PL and as expected the consumption does not vary significantly when different



(a) CDF of two bandwidth configurations featuring two MCSs each.



(b) CDF of two bandwidth configurations featuring two RBG loads each.

Fig. 7. CDF curves showing the negligible impact of the MCS index and RBG load on the power consumption of the Xilinx Zynq XC7Z045 SoC device.

bandwidth configurations are used. This is mainly due to the fact that the CPRI traffic is practically the same and there are only some minor differences in the datapath according to the configured bandwidth. For the MCS index and RBGs apply the same conclusions as those mentioned in the previous subsection. When compared to NETCFG1 or 2 (Figure 8), the different bandwidth configurations of NETCFG3 consume less power and this is made more clear as the bandwidth increases (Figure 9). This means for instance that there is a notable decrease in the power consumption of the 20 MHz configuration in NETCFG3 when compared to NETCFG1 or 2.

V. SUMMARY AND FUTURE WORK

In this paper we presented power consumption measurement results of two key ICs encountered in a 5G reconfigurable hotspot prototype [15]. The latter was built to serve 5G C-RAN architectures that feature HW-SW function split at communication stack level. Reconfigurable hotspots with such characteristics will be able to serve the high capacity and low latency requirements of C-RAN. On top of that, we

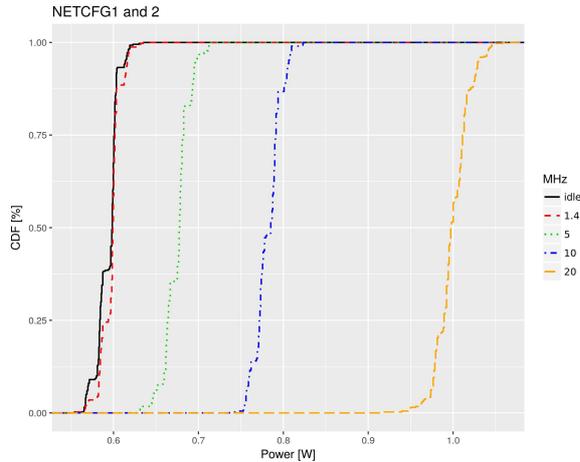


Fig. 8. NETCFG1/2: Comparison of the CDF curves for different bandwidth configurations using a 64 QAM modulation.

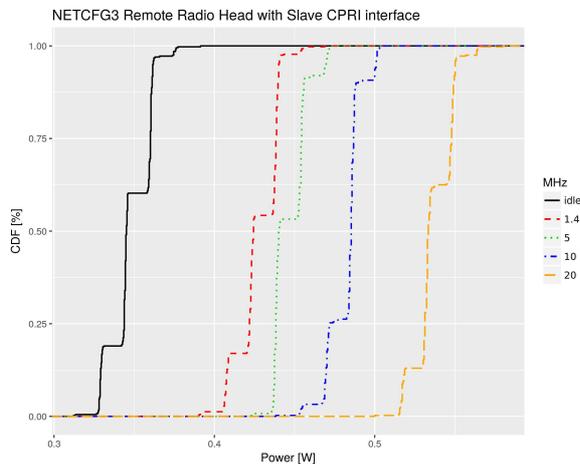


Fig. 9. NETCFG3: Comparison of the CDF curves for different bandwidth configurations using a 64 QAM modulation.

demonstrated that significant energy efficiency benefits can be achieved by adapting the bandwidth and by switching to a different function split (NETCFG). These flexible modifications can be envisioned in 5G C-RAN architectures where traffic load conditions allow or justify such changes to be applied.

An immediate follow-up step would be to elaborate and present also the results of the GigE PHY layer IC. The PA is also another IC in the existing setup that could be characterized adapting its output power according to different UE traffic scenarios and peak-to-average power ratio (PAPR) profiles. The ultimate future goal that goes beyond current standardized efforts, is the inclusion of the required signalling and message exchange in the partitioned communication stack that will allow to apply both the NETCFGs and the bandwidth adaptation in a dynamic fashion, according to different KPIs and use case objectives.

ACKNOWLEDGMENT

This work was supported by the European Commission in the framework of the H2020-ICT-2014-2 project Flex5Gware (Grant agreement no. 671563). The work of CTTC was also partially supported by the Generalitat de Catalunya (2014 SGR 1551) and by the Spanish Government under project TEC2014-58341-C4-4-R.

REFERENCES

- [1] Peter Rost et al., "Benefits and challenges of virtualization in 5G radio access networks," in *IEEE Communications Magazine*, vol. 53, no. 12, pp. 75-82, Dec. 2015.
- [2] Tarik Taleb, Sunny Dutta, Adlen Ksentini, Muddesar Iqbal, and Hannu Flinck, "Mobile Edge Computing Potential in Making Cities Smarter," in *IEEE Communications Magazine*, vol. 55, no. 3, pp. 38-43, March 2017.
- [3] Andreas Maeder et al., "A Scalable and Flexible Radio Access Network Architecture for Fifth Generation Mobile Networks," in *IEEE Communications Magazine*, vol. 54, no. 11, pp. 16-23, November 2016.
- [4] Mugen Peng, Yaohua Sun, Xuelong Li, Zhendong Mao and Chonggang Wang, "Recent Advances in Cloud Radio Access Networks: System Architectures, Key Techniques, and Open Issues," in *IEEE Communications Surveys and Tutorials*, vol. 18, no. 3, pp. 2282-2308, thirdquarter 2016.
- [5] 3GPP, Technical Specification Group Radio Access Network, Study on New Radio Access Technology, Radio Access Architecture and Interfaces (Release 14), 3GPP TR 38.801 V1.0.0, December 2016.
- [6] Chia-Yu Chang, Ruggero Schiavi, Navid Nikaein, Thrasyvoulos Spyropoulos and Christian Bonnet, "Impact of packetization and functional split on C-RAN fronthaul performance," 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, May 2016, pp. 1-7.
- [7] "Small Cell Virtualization Functional Splits and Use Cases", Small Cell Forum Release 5.1 (159.05.1.01), June 2015.
- [8] Ghizlane Mountaser, Maria Lema Rosas, Toktam Mahmoodi and Mischa Dohler, "On the feasibility of MAC and PHY split in Cloud RAN", in the Proceeding of the 2017 IEEE Wireless Communications and Networking Conference (WCNC), 19-22 March 2017, San Francisco, USA.
- [9] Zeller D. et al. (2013) Sustainable Wireless Broadband Access to the Future Internet - The EARTH Project. In: Galis A., Gavras A. (eds) The Future Internet. FIA 2013. Lecture Notes in Computer Science, vol 7858. Springer, Berlin, Heidelberg.
- [10] Nicola Baldo, Marco Miozzo, Manuel Requena-Esteso, and Jaime Nin-Guerrero, "An open source product-oriented LTE network simulator based on ns-3", In Proceedings of the 14th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems (MSWiM '11), ACM, New York, NY, USA, pp. 293-298.
- [11] O. Font-Bach, N. Bartzoudis, A. Pascual-Iserte, M. Payaró, L. Blanco, D. López, M. Molina, "Interference Management in LTE-based HetNets: a Practical Approach", *Transactions on Emerging Telecommunications Technologies (Wiley)*, Vol. 26, Issue 2, Pages 195-215, February 2015.
- [12] R. Muñoz et al., "The CTTC 5G End-to-End Experimental Platform: Integrating Heterogeneous Wireless/Optical Networks, Distributed Cloud, and IoT Devices", in *IEEE Vehicular Technology Magazine*, vol. PP, no. 99, pp.1-1, January 2016.
- [13] DAQ-Acquire tool GitHub repository <https://github.com/Enchufa2/daq-acquire>.
- [14] Small Cell Forum White Paper, Document 106.06.01, "Virtualization for small cells: Overview", June 2015.
- [15] Fredrik Tillman et al., "Deliverable 1.1 5G system use cases, scenarios, and requirements break-down", H2020 Flex5Gware project, Dec. 2015. Available online: http://www.flex5gware.eu/images/documents/Flex5Gware_D11.pdf