Flexible Fine-Grained Baseband Processing with Network Functions Virtualization: Benefits and Impacts

Maicon Kist[a,b], Juliano Araújo Wickboldt[a], Lisandro Zambenedetti Granville[a], Juergen Rochol[a], Luiz A. DaSilva[b], Cristiano Bonato Both[c]

[a]*Institute of Informatics (INF)*
Federal University of Rio Grande do Sul
Av. Bento Gonçalves, 9500 – Porto Alegre, Brazil
[b]*Trinity College Dublin, Ireland*
[c]*Department of Applied Mathematical and Social Sciences*
Federal University of Health Sciences of Porto Alegre
R. Sarmento Leite, 245 – Porto Alegre, Brazil
Email: {mkist, jwickboldt, granville, juergen}@inf.ufrgs.br,
dasilval@tcd.ie,
cbboth@ufcspa.edu.br

Abstract

The increasing demand for wireless broadband connectivity is leading mobile network operators towards new means to expand their infrastructures efficiently and without increasing the cost of operation. Network Functions Virtualization (NFV) is a step towards virtualization-based, low-cost flexible and adaptable networking services. In the context of centralized baseband architectures, virtualization is already employed to run baseband processing units as software on top of conventional data center hardware. However, current virtualization solutions consider atomic virtualization, *i.e.*, single virtual machines implementing all baseband functionalities. In this article, we propose the fine-grained virtualization of baseband processing to achieve a more flexible distribution of the processing workload in centralized architectures. We also evaluate the benefits of our approach in terms of (*i*) the bandwidth requirements for each fine-grained distribution option, (*ii*) the latency experienced by mobile users for each fine-grained distribution option, and (*iii*) the total CPU usage of each fine-grained baseband processing function.

1. Introduction

The fifth generation of mobile networks (5G) is expected to offer substantially higher data rates, as well as support for massive machine-type communications and ultra-low latency applications. In addition, it is likely that 5G will have to coexist with multiple Radio Access Technologies (multi-RAT), *e.g.,* 4G and NarrowBand-IoT (NB-IoT) [1]. To achieve such aggressive targets, the 5G architecture needs to focus its design objectives around efficiency, adaptability, and versatility [2] [3].

The Network Functions Virtualization (NFV) paradigm can provide 5G networks with the required efficiency, adaptability, and versatility [4]. NFV decouples the physical network equipment from the function it provides and moves its functionality to software, namely Virtual Network Functions (VNFs) [4]. VNFs can be consolidated on top of standard commercial hardware located in data centers, network nodes, or in end-user devices.

Although NFV is commonly associated with the softwarization and virtualization of core network elements, substantive research focuses on the radio access network and enabling baseband centralization architectures through NFV, *i.e.,* executing Virtual Baseband Processing Units (vBBUs) in VNFs on top of standard data center hardware [5] [6] [7] [8] [9]. The common approach is to split the LTE access network into three major components: (*i*) the Remote Radio Head (RRH), responsible for signal digitization, forwarding the digital signal to a central data center in the uplink or transmitting the digital signal received from the data center in the downlink, (*ii*) the data center, which provides the processing resources, and (*iii*) atomic VNFs implementing full-blown vBBUs. This functionality split allows mobile operators to expand the mobile network coverage and capacity quickly, simply by deploying RRHs, connecting them to the data center through a fronthaul network, and allocating the vBBUs required by the RRH [10].

The adoption of atomic and centralized vBBUs may be infeasible in low-bandwidth or high-latency fronthaul links due to the high throughput required to transport raw digital signals in both uplink and downlink directions and the strict latency constraints of the LTE standard [10]. These limitations restrict the fronthaul network to be composed only of optical links (to satisfy the bandwidth constraint) and data centers to be located close to RRHs (to satisfy the latency constraints). Moreover, execution of atomic VNFs is restricted to high volume servers due to the large amounts of processing and memory resources required, limiting the adoption of more

energy- and processing-capable hardware, such as Field Programmable Gate Arrays (FPGA) and Digital Signal Processors (DSP) [7].

In this article, we propose the employment of NFV to enable fine-grained vBBUs, *i.e.,* a flexible and adaptable solution in which the baseband functions are distributed in multiple independent VNFs. We also present a solution, following the European Telecommunications Standards Institute (ETSI) NFV Management and Orchestration (MANO) reference architecture, designed for fine-grained vBBU execution that achieves three core properties: (*i*) programmability, *i.e.,* baseband functions VNFs can be reprogrammed on-demand according to mobile network needs, (*ii*) adaptability, *i.e.,* resources allocated for baseband function VNFs can be changed on-demand to satisfy data center requirements, (*iii*) flexibility, *i.e.,* baseband VNFs can run on top of data center hardware or in specialized hardware for better performance and energy efficiency.

Our proposal can be illustrated with an example: the uplink of a LTE BBU encompasses operations such as Fast Fourier Transform (FFT), a channel estimator, and a Resource Element (RE) de-mapper [11], which are sequentially applied to transform the received analog signal into user data.

Based on fine-grained vBBUs, each function is moved to an independent VNF; they are then aggregated to compose a complete LTE vBBU. We evaluate the feasibility of this approach in terms of (*i*) the bandwidth requirements for each fine-grained distribution option, (*ii*) the latency experienced by mobile users for each fine-grained distribution option, and (*iii*) the total CPU usage of each fine-grained baseband processing function. The results obtained show that fine-grained vBBU can: reduce the fronthaul bandwidth demand by moving part of baseband functions closer to RRHs; reduce the processing delay by allocating computation-heavy VNFs to specialized processors; and distribute the processing workload by enabling VNFs to be distributed in different processing hardware and data centers.

The remainder of this article is organized as follows. We first present the state-of-the-art in BBU virtualization for centralized baseband architectures, followed by technologies required in 5G, and a brief overview of the ETSI NFV MANO architecture. Then, we discuss our fine-grained vBBU virtualization concept. Next, we conduct a performance evaluation to demonstrate the benefits and impact of the proposed architecture. Finally, we present our concluding remarks.

2. Next Generation 5G: current architectures and technologies

In this Section, we discuss the current state-of-the-art in BBU virtualization for centralized architectures, followed by an overview of essential technologies for 5G deployment. We close this section with a discussion of the ETSI NFV MANO architecture, which serves as a basis for our proposed solution.

2.1. BBU Virtualization in 5G

Recent pioneering developments, such as EASE [6], CONCERT [7], and MobileFlow [12], explore atomic vBBUs in centralized architectures. In this virtualization, an atomic vBBU VNF is responsible for all the baseband and MAC layer functions of a particular technology, such as LTE, while the RRH only implements the signal digitization. Signal samples must be transported between the RRH and the atomic vBBU, which demands extremely high data rates on the fronthaul network.

The architecture for vBBUs introduced in SoftAir [8] considers keeping baseband functions that handle raw signal samples, *e.g.,* modulation and demodulation, in the RRH, while the remaining functions are executed on vBBUs in the data center. The idea of moving only part of the baseband processing to a vBBU while keeping low-level physical layer functions in the RRH gained traction recently as a viable solution to reduce the fronthaul bandwidth demand while preserving the benefits of centralized architectures [11] [13]. Although there are dozens of functions that can be moved to the RRH, given the type of data that is transferred between the RRH and the vBBU, it is possible to summarize the split options into: IQ Forwarding, SubFrame Forwarding, RX Data Forwarding, SoftBit Forwarding, and MAC Forwarding (we consider each of these split options in Section 4). However, a static split of functionalities means that it is not possible to dynamically adapt fronthaul or processing demands at the data center.

2.2. Enabling densification, multi-RAT, and fronthaul technologies

5G networks may need to handle a 1000x increase in current traffic volumes, provide a 100x increase in the edge data rate, support 1ms latency, provide ultra-high reliability and availability while reducing or at least maintaining current energy consumption and costs [3]. To achieve such goals, centralized baseband architectures must consider the joint exploitation of three

aspects: (*i*) dense deployment of RRHs [14], (*ii*) multi-RAT support [7], and (*iii*) heterogeneous fronthaul links [10].

Dense deployments of RRHs is considered a key aspect to meet the increasing demand for higher data rates and lower latencies expected for 5G. The high inter-cell interference resulting from densification requires advanced interference mitigation techniques, *e.g.,* cooperative scheduling and CoMP. Such mechanisms need fast synchronization of transmission and reception parameters between multiple vBBUs to adapt to the fast-paced changes of wireless channels. The distance between where the atomic vBBU resides and the RRH can make the adoption of these advanced interference mitigation techniques infeasible. Fine-grained vBBUs provides the framework to move a subset of baseband functions closer to the RRH, significantly reducing the communication latency between vBBUs performing CoMP [15].

Multi-RAT involves supporting the coexistence and joint operation of different radio technologies, *e.g.,* LTE and NB-IoT, for the vast range of applications and scenarios expected in future 5G deployments. Deploying atomic vBBUs with multi-RAT capabilities is challenging because of the large footprint regarding high memory and processing usage. Fine-grained vBBUs provide the programmability and adaptability to allow a low-cost multi-RAT network. Moreover, it enables real-time adaptation of the air-interface to the signal conditions, controlling the signaling overhead according to application-specific requirements, and supporting a broad range of data rates, going from low-rate applications up to ultra-high-rate multimedia services.

Heterogeneous fronthaul links refer to the adoption of different types of physical links connecting the RRH with the data center. Traditional centralized baseband architectures with atomic vBBUs require the fronthaul network to transport raw signal samples from the RRH to the data center, usually in the form of IQ samples, demanding data rates that only optical links can provide [16] (approximately 4.92 Gbps for the downlink and uplink of one RRH). Unfortunately, in practice, the fronthaul network is often composed of a plethora of link types with considerably less capacity than optical ones, thus limiting the adoption of atomic vBBUs. By enabling fine-grained vBBUs, mobile operators can adopt fronthaul networks consisting of a blend of high and low capacity links by moving baseband VNFs to the RRH to reduce the data rate required in the fronthaul.

Considering the technologies required by 5G deployments, the current atomic virtualization approach hinders some of the benefits of baseband centralization because of the low flexibility, limited programmability and adapt-

ability, and high fronthaul requirements.

2.3. NFV Management & Orchestration Architecture

The NFV Industry Specification Group, created under ETSI, aims to establish a consensus regarding the virtualization of network functions to provide integration among NFV solutions [9]. ETSI defined an architecture where the central elements are VNFs, with corresponding functional blocks for Management and Orchestration (MANO) [17]. We provide a brief description of the most important components in the remaining of this subsection. All elements described also are part of the fine-grained vBBU architecture presented in the next section.

The central elements of the architecture are the VNFs, which are software implementations of functions deployed in the NFV Infrastructure (NFVI) Layer. A VNF can be composed of a set of interconnected components, named VNFCs, which are the smallest elements of a VNF. The different arrangements of the set of VNFCs into one or more processing resources give rise to different composition options. Each composition option has its processing, storage, and networking requirements. Moreover, a Virtualization Layer maps processing resources to dedicated virtualization containers (*e.g.,* virtual machine or Linux container) or even physical resources that do not support virtualization (*e.g.,* FPGAs or DSPs). In the next section, we present our solution for a flexible fine-grained BBU virtualization.

3. Fine-Grained BaseBand Processing Unit Virtualization

We provide an overview of the proposed fine-grained BBU virtualization architecture in Subsection 3.1. Next, we present an example of the main interactions of each component of the architecture in Subsection 3.2.

3.1. Architecture Design

Now we discuss our flexible fine-grained vBBU architecture, as depicted in Figure 1. As an example, consider the physical uplink functions of an LTE vBBU, which is composed of the FFT, RE de-mapper, receive processing, Forward Error Correction (FEC), and MAC. In our proposal, instead of moving all baseband function operations to an atomic vBBU, we encapsulate each of them in independent VNFCs. VNFCs are distributed across multiple processing resources (which can be located in different data center sites)
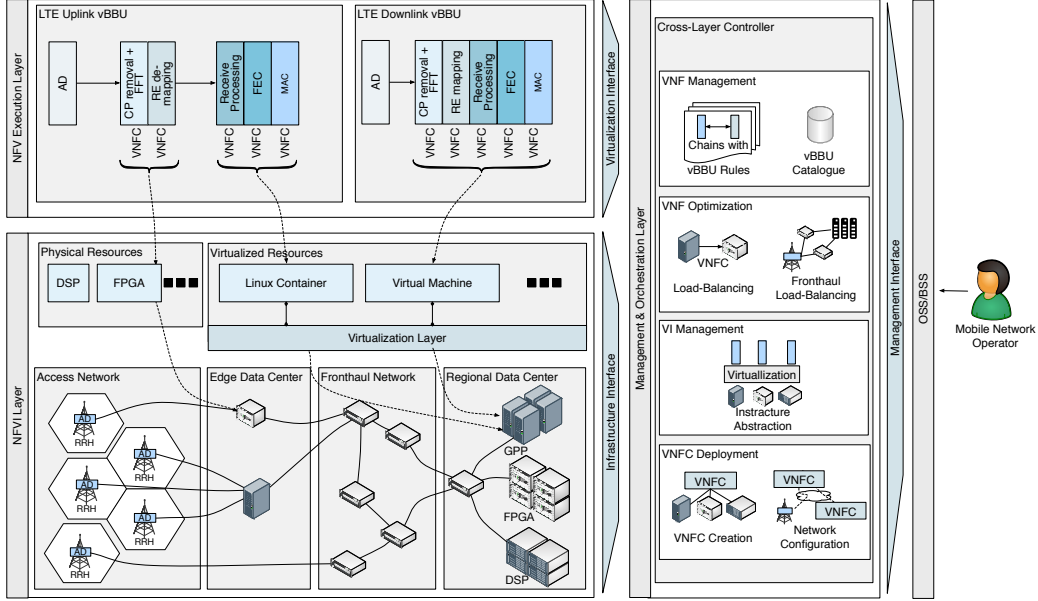
6

Figure 1: Centralized baseband architecture enhanced with fine-grained vBBU

to build a full-blown vBBU. We describe each layer of the architecture as follows.

The NFVI Layer encompasses three architectural elements found in centralized baseband architectures [16]: (*i*) the regional data center contains the processing hardware to execute all computational tasks. Different from traditional architectures, which consider only commercial servers as the available processing hardware, we also consider specialized hardware accelerators, *e.g.,* FPGAs and DSPs; (*ii*) the fronthaul network connects the regional data center with edge data centers or directly to RRHs; (*iii*) the edge data center is a set of processing resources in proximity to RRHs. Thus, specific baseband functions, *i.e.,* VNFCs, can be moved closer to the RRH to reduce bandwidth and latency requirements over the fronthaul network.

The NFV Execution Layer provides the framework for the execution of fine-grained vBBUs. Each fine-grained vBBU is composed of a set of interconnected baseband VNFCs. The main benefit of this approach is the small virtualization footprint, while simultaneously enabling the execution of baseband functions in heterogeneous hardware devices. Each baseband VNFC is a black box that receives data, which range from IQ samples, in the case of the FFT, to user and control data, in the case of FEC. vBBUs are composed

in such a way that the result of data processing of one baseband VNFC is forwarded to the next one in the chain. Moreover, computationally intensive baseband VNFCs, such as FFT and RE de-mapper, can be migrated to high-performance and non-virtualizable FPGAs and DSPs to achieve higher data throughput, at the cost of less flexibility.

The Cross-Layer Controller is a logical entity (which can be distributed into different physical boxes to improve adaptability and performance) responsible for the MANO. The main functionalities of the Cross-Layer Controller are: (*i*) VNF management, (*ii*) VNF optimization, (*iii*) Virtualized Infrastructure (VI) management, and (*iv*) VNFC deployment. VNF management covers all aspects related to the life-cycle of vBBUs, *e.g.,* creation, installation, and migration according to the available physical resources, and ensuring that processing and latency requirements are fulfilled during the vBBU operation. To this end, the Cross-Layer Controller must be aware of the resources required by each baseband VNFC composing the vBBU and manage its scheduling and decisions such as migration, resource scaling, and failure recovery. We highlight that a single centralized controller might not be able to cope with the real-time monitoring in ultra-dense deployments. In this case, adopting multiple controllers, each responsible for a cluster of vBBUs, is a more appropriate approach.

Optimizing VNFs, *i.e.,* vBBUs, includes a broad range of adjustments, such as distributing baseband VNFCs among edge and regional data centers, selection of hardware to execute VNFCs aiming to increase the overall vBBU performance, and adjustments in the fronthaul network forwarding to reduce bottlenecks. Such optimizations should be located in the Cross-Layer Controller, as it has accurate and instantaneous information regarding baseband functions. Due to the time-sharing nature of processing resources, different bandwidth and latency constraints, and heterogeneous processing capabilities in the data centers, such joint optimizations come with challenges because of their non-convex nature, making it a research topic on its own.

VI Management encompasses the abstraction and presentation of physical processing resources as virtual resources to higher layers of the architecture. This increases vBBU portability and ensures that different physical hardware resources are viewed uniformly by the Cross-Layer Controller. Virtual resources can be created in diverse ways, each one with its advantages and drawbacks. For example, FPGAs and DSPs are characterized by predictable performance, which is suitable for physical layer processing operations. VNFC deployment encompasses the steps to instantiate baseband

8

processing functions in the chosen processing resource.

The Cross-Layer Controller interacts with other components of our architecture through three well-defined interfaces:

- The Management Interface with external applications allows gathering information regarding any aspect of the network from the Cross-Layer Controller. It enables the direct management of fine-grained vBBUs, such as instantiation and distribution of vBBUs in the data center, and configuration of specific parameters of a particular baseband function.

- The Virtualization Interface with the NFV Execution Layer provides the means to instantiate full-blown vBBUs, and migrate and monitor baseband VNFCs.

- The Infrastructure Interface is used to configure the physical resources to match the expected behavior of vBBUs, *e.g.,* the forwarding path connecting all baseband VNFCs.

The Operations/Business Support System (OSS/BSS) interacts with the Cross-Layer Controller to gather information from the network, *e.g.,* interference map, and channel and processing hardware utilization, and provide visualizations in terms of charts and reports of all layers of the architecture. Based on this, an operator can manage the RAN with a global view of the vBBUs and their associated resources. In the next subsection, we detail the main interaction from the OSS/BSS to the physical infrastructure when instantiating a fine-grained vBBU.

3.2. Fine-Grained vBBU Instantiation

The main interactions during the instantiation of a fine-grained vBBU VNF are illustrated in Figure 2. The network operator utilizes the OSS/BSS to specify the characteristics of the fine-grained vBBU, *e.g.,* RAT, center frequency, and channel bandwidth (1). The description of the vBBU is sent to the Cross-Layer Controller through the Management Interface (2). After that, the Cross-Layer Controller must select processing resources that can fulfill processing and latency demands (3) and start the instantiation of the vBBU through the Virtualization Interface with the NFV Execution Layer (4). Next, the baseband processing VNFCs are instantiated (5) and their data forwarding configured to compose the fine-grained vBBU (6). When the process is finished, the Cross-Layer Controller sends a notification message
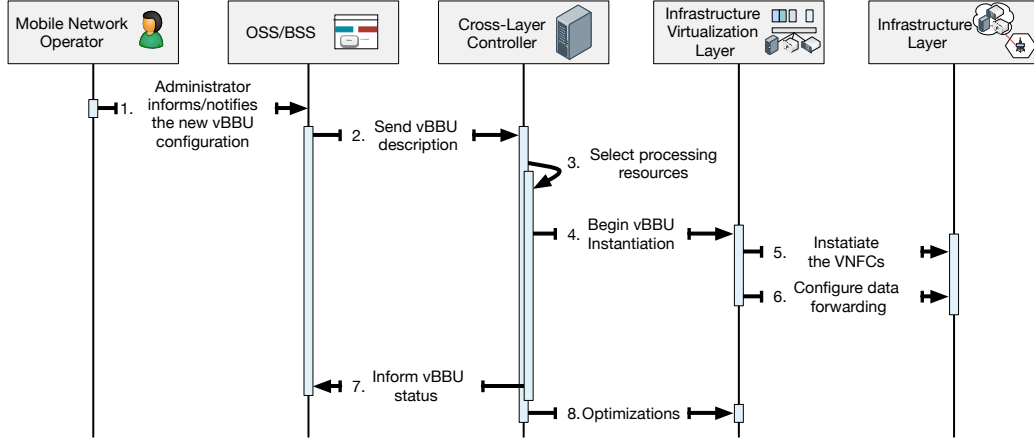
Figure 2: High-level interactions of the fine-grained vBBU architecture

to the original application through the Management Interface (7). After that, the Cross-Layer Controller or the network operator can optimize the distribution of VNFCs (8).

Many challenges appear in this example. For example, step (2) requires a comprehensive VNF description language that can express rules to deploy VNFs, and instructions to handle failures or to solve conflicting configurations. Selecting the appropriate set of processing resources to be used in step (3) requires up-to-date information on the network resources, which is hard to obtain because of fast fluctuations caused by user mobility, wireless channel characteristics, and even other VNFs. Step (5) requires the development of a cross-platform system that abstracts the underlying processing resource, while at the same time making efficient use of different hardware capabilities. Finally, step (8) requires the development of algorithms to solve a non-convex optimization problem.

Fine-grained vBBUs enable RRH densification, multi-RAT, and heterogeneous fronthauls, three aspects essential for future 5G deployments. First, edge data centers facilitate densification of RRHs by moving interference mitigation techniques closer to them, allowing fast adaptation of transmission and reception parameters according to changes in the wireless environment. Second, our approach also presents unique opportunities to deploy multi-RAT, as the low memory and processing footprint enables the execution of multiple RATs without significant overhead. Moreover, RATs can be reconfigured by changing specific vBBU parameters, similarly to what is done in

current baseband processing functions implemented in Software-Defined Radio platforms [18], or by adding or removing functionality, *e.g.,* by adding a baseband VNFC that perform carrier aggregation to the chain of baseband VNFCs that build the vBBU. Third, heterogeneous fronthaul links are enabled with the dynamic and flexible distribution of baseband VNFCs to reduce the fronthaul demand. For example, all baseband functions can be moved to the regional data center if high bandwidth and low latency fronthaul links are connecting it to the RRHs. In the case of low bandwidth or high latency links, baseband functions implementing the physical layer operations can be moved to the edge data center, while the regional data center takes responsibility only for the MAC layer functions. This flexibility is also necessary to deal with stringent latency requirements of mobile standards.

4. Evaluation of Benefits and Impacts

In this section, we illustrate the benefits and impact of the fine-grained vBBU architecture. First, we conduct a mathematical analysis to obtain the bandwidth required for each VNFC distribution option. Second, we run a simulation to measure the impact of different distribution options on the latency experienced by mobile users when considering a bandwidth constrained fronthaul link. Finally, we create an experimental scenario to measure the CPU usage of each baseband VNFC and the total processing power required in the edge and regional data center to execute a fine-grained vBBU.

4.1. Fine-grained vBBU Distribution Options and Bandwidth Requirements

Different from centralized baseband architectures, fine-grained vBBU allows the flexible distribution of VNFCs according to fronthaul link constraints or data center requirements. We exploit the bandwidth requirements of the five most common split options between the regional and edge data centers considering possible baseband VNFCs in a LTE vBBU, as shown in Figure 3.

We assume that the edge data center is connected to one RRH using a sampling rate (f_S) of 30.72 MHz (for a 20 MHz channel) with a MIMO configuration (N_R) of 2x2 and oversampling factor (N_O) of 2. In this configuration, the vBBU uses 1200 subcarriers (N_SC) and a symbol duration (T_S) of 66μs. We considered RRHs with 10%, 50%, and 100% utilization rate of REs (η), representing different levels of mobile user load. The analog signal to IQ samples conversion (N_Q) is set to 10 bits/sample. Table 1 summarizes the parameters and values used for this evaluation as well as for the simulation
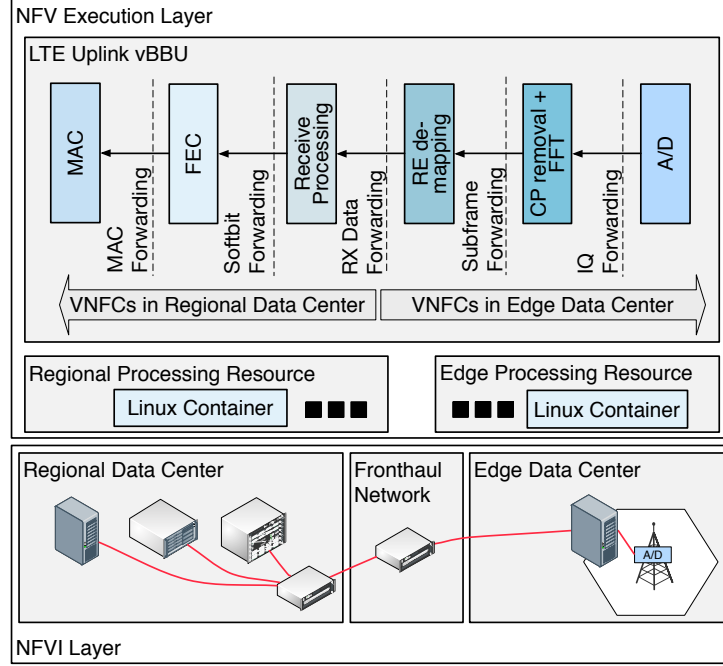
Figure 3: Fine-grained vBBU distribution options

in the next subsection. The graph in Figure 4 shows the required fronthaul bandwidth for each fine-grained vBBU distribution option.

- In **IQ Forwarding**, samples are transported over the fronthaul to the regional data center, which centralizes all baseband VNFCs. The fronthaul data rate required in this distribution is fixed. Thus, mobile operators can determine beforehand whether it can handle the traffic

Symbol	Description	Value Used	Impacts on
f_S	Sampling Rate	30.72 MHz	Bandwidth
N_R	Number of Antennas	2	Bandwidth
N_O	Oversampling Factor	2	Bandwidth
N_{SC}	Number of Used Subcarriers	1200	Bandwidth
T_S	Symbol Duration	66.6 μs	Bandwidth
η	Fraction of RE used	[0.1, 0.5, 1.0]	Bandwidth
N_Q	Quantization Bits per IQ	10	Bandwidth
F_{BW}	Fronthaul Capacity	10 Gbps	Latency
F_L	Distance Regional-Edge	15 Km	Latency

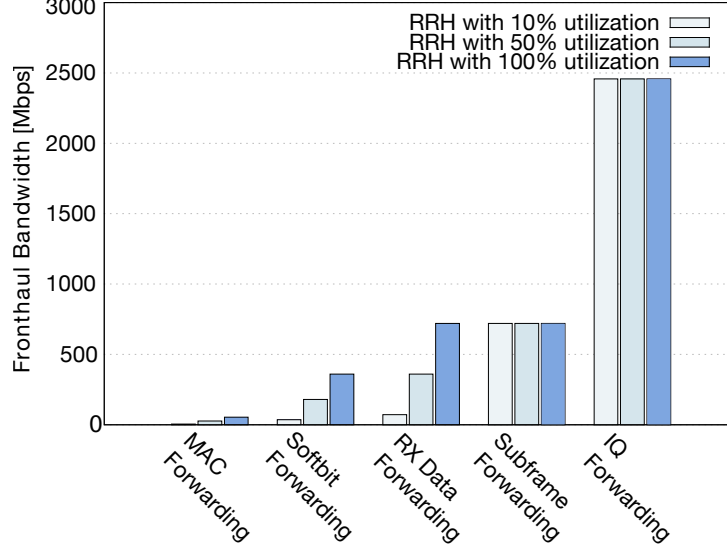Table 1: Parameters used in the analytical and simulated scenarios

Figure 4: Fronthaul bandwidth for each vBBU distribution option

of an RRH. The main benefit of this distribution is that almost no digital processing is required at the edge data center. Moreover, this distribution eases the adoption of Large-Scale Cooperative Multiple Antenna Processing (LS-CMA) because of the centralization of all IQ samples [19]. This option is interesting only in the cases where the edge data center is already overloaded with other baseband processing tasks or if the cost of fronthaul transport is low. The fronthaul demand when using this split option is given by:

$$B_{FH}^{IQ} = N_O \cdot f_S \cdot 2 \cdot N_Q \cdot N_R = 2 \cdot 30.72\text{MHz} \cdot 2 \cdot 10\text{bits} \cdot 2 = 2.46\text{Gbit/s}$$

- In **Subframe Forwarding**, the VNFC implementing the CP Removal and FFT is moved to the edge data center. In this case, only the IQ samples of useful subcarriers are transported over the fronthaul, representing roughly 60% of the total subcarriers in our configuration. Eliminating this overhead reduces the bandwidth required to 720 Mbps. Subframe forwarding is attractive when 100% of the wireless resources are being utilized because the fronthaul data rate required is always the same, while at the same time enabling LS-CMA mechanisms. Also, the processing workload does not depend on the actual load of the RRH. The fronthaul demand when using this distribution option is given by:

$$B_{FH}^{SF} = N_{SC} \cdot T_S^{-1} \cdot 2 \cdot N_Q \cdot N_R = 1.200 \cdot (66\mu s)^{-1} \cdot 2 \cdot 10\text{bits} \cdot 2 = 720\text{Mbit/s}$$

- In **RX Data Forwarding**, the VNFC implementing the RE de-mapper is moved closer to the RRH. In this distribution option, the regional data center receives the IQ samples of REs allocated to mobile users, *i.e.,* 10% of 720 Mbps if 10% of REs are allocated (which is something that can change in each LTE frame). Because of this, the fronthaul data rate required is not constant. Based on fine-grained vBBU this distribution option can be selected on-the-fly when less than 50% of the wireless resources of an RRH are being allocated, significantly reducing the overhead in the fronthaul network. This fronthaul demand when using this distribution can be calculated using the factor of REs allocated and B_{FH}^{SF}:

$$B_{FH}^{RX} = B_{FH}^{SF} \cdot \eta = 720\text{Mbit/s} \cdot [0.1, 0.5, 1.0] = [72, 360, 720]\text{Mbit/s}$$

- In **SoftBit Forwarding**, the edge data center executes all VNFCs required to recover bits from the radio signal, which includes both user data and higher layer control data, such as MAC headers. This distribution option reduces the fronthaul data rate required to a fraction of the standard IQ forwarding adopted in atomic vBBUs, but at the cost of disabling LS-CMA at the regional data center. However, fine-grained vBBUs allows LS-CMA to be performed between all RRHs connected to the same edge data center. The fronthaul required is given by:

$$B_{FH}^{SB} = B_{FH}^{RX}/N_R = [72, 360, 720]\text{Mbit/s}/2 = [36, 180, 360]\text{Mbit/s}$$

- **MAC Forwarding** is the approach of current mobile networks, in which MAC packet data units are transported over the fronthaul. Although the burden of the fronthaul is significantly reduced, the processing demand at the edge data center becomes the major bottleneck, as the VNFC implementing the FEC requires considerable computational capacity. This bandwidth required for this split is given by:

$$\begin{aligned} B_{FH}^{MAC} =& N_{SC} \cdot T_S^{-1} \cdot \eta \cdot S \\ & 1.200 \cdot (66\mu s)^{-1} \cdot [0.1, 0.5, 1] \cdot 3\text{bit/cu} = [5.4, 27, 54]\text{Mbit/s} \end{aligned}$$

The fine-grained vBBU distribution options discussed in this subsection represent huge opportunities to enable the dynamic adaptation of the fronthaul bandwidth. In addition, fine-grained vBBU still preserves the flexibility and scalability required in 5G networks. By adapting the bandwidth, mobile network operations can adopt more diverse and cost-effective links. Finally, different distribution options lead to different numbers of vBBUs that can have their traffic transported over a given fronthaul link with a fixed bandwidth constraint. We analyze the number of vBBUs that can be transported over a constrained fronthaul in the next subsection.

4.2. Latency of fine-grained vBBU distributions

We also sought to understand how distributing VNFCs between the regional and edge data centers affects the latency experienced by mobile users. Therefore, we simulated the infrastructure shown in Figure 5 in the Mininet network emulator. We varied the number of vBBUs from 1 to N and measured the latency given a fronthaul link with capacity limited to 10 Gbps (F_{BW}) and length (F_L) of 15 km. As the number of vBBUs increases, it is expected that the competition for the shared and limited capacity of the fronthaul link will increase the overall latency. The traffic generated at each RRH was according to the results presented in the previous section, considering a 50% utilization rate, *i.e.,* 2.46 Gbps for IQ, 720 Mbps for Subframe, 360 Mbps for RX Data, 180 Mbps for Softbit, and, finally, 27 Mbps for MAC Forwarding. The latencies obtained are shown in Figure 6.

Considering that vBBUs need to generate an ACK/NACK response in 3 ms to stay compliant with the 3GPP LTE Hybrid Automatic Repeat-reQuest (HARQ) timing, we have an estimate of the maximum number of fine-grained vBBUs that can be executed simultaneously in each distribution
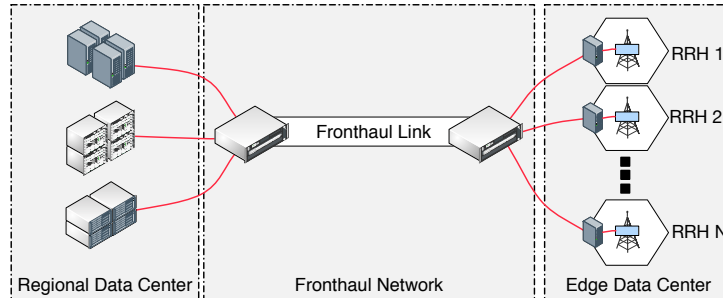

Figure 5: Fine-grained vBBU infrastructure with constrained fronthaul
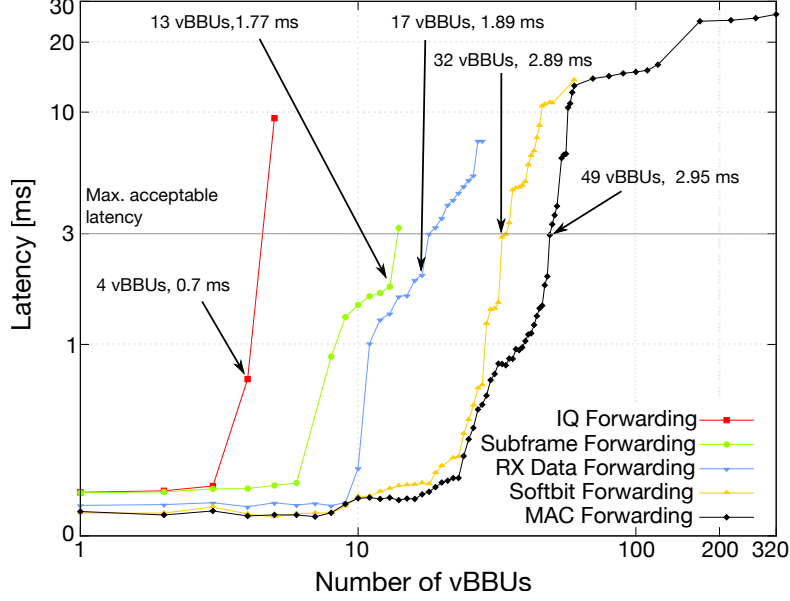
15

Figure 6: Average latency considering the number of vBBUs

option. This number was as follows in our emulated network: 4 for IQ, 13 SubFrame, 17 for RX Data, 33 for SoftBit, and 49 for MAC Forwarding. We highlight that the processing at the regional data center must compensate for higher latencies in the fronthaul network. For example, 13 vBBUs in the SubFrame Forwarding distribution lead to an average latency of 2.77 ms, leaving the regional data center only 0.23 ms to perform the processing to generate the HARQ message.

From the evaluations conducted so far we can observe that offloading the physical layer processing to edge data centers presents two major advantages: *(i)* enabling the adoption of heterogeneous fronthaul links, and *(ii)* enabling an ultra-low latency network due to the proximity with mobile users. Clearly, one has to be careful not to sacrifice the performance of fine-grained vBBUs due to the lack of LS-CMA; hence, a fundamental research question still to be tackled is to find the CPU usage of each VNFCs, which we evaluate next.

4.3. CPU usage of fine-grained vBBU distributions

Finally, we evaluate the computational overhead of each baseband processing VNFC of a LTE vBBU. For this, we run a fine-grained LTE vBBU and measure the CPU usage in an Intel Core i5-4250U2 1.3 GHz in the w-

16

iLab-t testbed. Table 2 show the obtained results for each VNFC baseband function and for all six valid LTE channel bandwidths. First, we can note that the same VNFC, *e.g.,* CP removal + FFT, requires more CPU time as the channel bandwidth increases. The CPU usage increase is a side effect of higher channel bandwidths due to the high number of digitized IQ samples being processed.

		LTE Channel Bandwidth					
		1.4 MHZ	3 MHZ	5 MHZ	10 MHz	15 MHz	20 MHz
VNFC	**CP Removal + FFT**	8.9	10.7	10.7	10.9	13.4	16.1
	RE demapping	3.4	4.7	8.7	13.1	15.4	16.1
	Receive Processing	10.2	12.0	18	32.8	40.3	46.9
	FEC	10.2	11.4	16.7	32.1	41.6	47.6
	MAC	6.8	6.7	8.0	8.5	10.2	12.8

Table 2: CPU usage for each vBBU function

We highlight that increasing the channel bandwidth does not correlate to the same increase in the CPU usage, *e.g.,* doubling the channel bandwidth does not incur in double CPU usage. This happens because several baseband processing operations take advantage of modern processor instructions, such as Single Instruction Multiple Data (SIMD). Moreover, we can see that the FEC and RX processing are by far the most CPU intensive functions (with the first being slightly more intensive than the latter). Both functions together require up to half of the total CPU consumed by the vBBU. Finally, the CPU usage is reduced drastically as the vBBU functions shift from the low physical operations, which encompass all operations except the MAC, to the higher MAC layer.

We further explore the fine-grained vBBU possibilities by measuring the CPU usage in the edge and regional data centers, as shown in Figure 7. We highlight the fact that usage below 100% uses only one processing core of the CPU, whereas above it uses two CPU cores. As expected, all CPU usage is concentrated in the regional data center when adopting the IQ Forwarding option, but at the cost of huge fronthaul bandwidth, as we mentioned earlier. For comparison, adopting the Subframe Forwarding option moves only a fraction of the total processing required to the edge, while at the same time reducing the fronthaul demand from 2.46 Gbps to 720 Mbps.

5. Concluding Remarks

In this article, we presented a fine-grained baseband processing virtualization for centralized baseband architectures. We exploited moving baseband
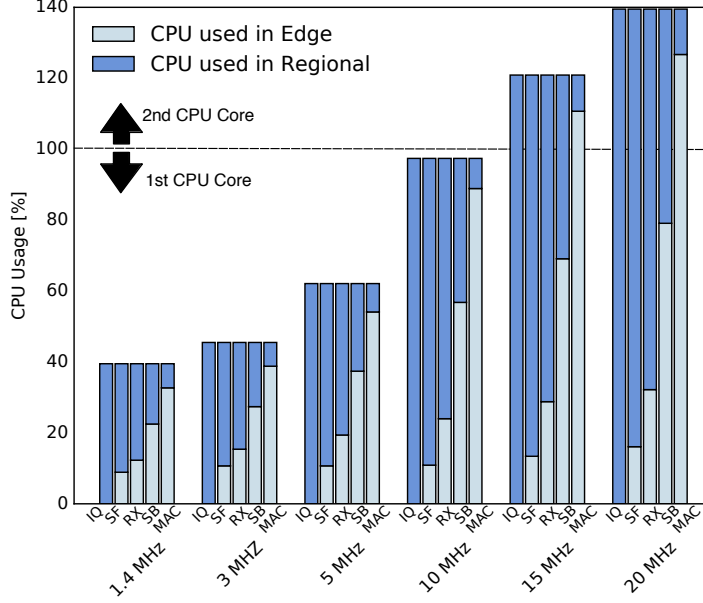
Figure 7: Total CPU usage in edge and regional data center for each distribution options

processing functions to VNFs, which can be chained to build a fine-grained vBBU and enable essential features for future 5G deployments, in particular, RRH densification, multi-RAT, and heterogeneous fronthauls. Fine-grained vBBUs avoid the drawbacks of centralized baseband architectures by allowing baseband functions to be distributed according to the fronthaul and data center constraints. We also assessed the benefits and impact of our architecture by demonstrating that: (*i*) different baseband distribution options required different fronthaul network bandwidth, which can enable a mix of cost-effective fronthaul link technologies; (*ii*) the latency reduces by moving baseband VNFCs from the regional to edge data centers, which can enable the adoption of ultra-low latency radio access technologies; and (*iii*) the CPU workload of the regional and edge data center can be balanced according to the network requirements by distributing fine-grained vBBU VNFCs accordingly.

We expect that fine-grained vBBUs can catalyze mobile networks innovations in a range of areas, from the introduction of new air-interfaces specialized in specific services, to management of data center and fronthaul resources. Fine-grained vBBUs solves the challenges of current centralized

18

baseband architectures while enabling an unprecedented control over any aspect of the RAN.

Acknowledgements

References

[1] Y. D. Beyene, et al., NB-IoT Technology Overview and Experience from Cloud-RAN Implementation, IEEE Wireless Communications 24 (2017) 26–32.

[2] J. G. Andrews, et al., What Will 5G Be?, IEEE Journal on Selected Areas in Communications 32 (2014) 1065–1082.

[3] C. Sexton, et al., 5G: Adaptable Networks Enabled by Versatile Radio Access Technologies, IEEE Communications Surveys & Tutorials 19 (2017) 688–720.

[4] R. Mijumbi, et al., Network Function Virtualization: State-of-the-Art and Research Challenges, IEEE Communications Surveys & Tutorials 18 (2016) 236–262.

[5] I. Gomez, V. Marojevic, A. Gelonch, Resource Management for Software-Defined Radio Clouds, IEEE Micro 32 (2012) 44–53.

[6] T. Taleb, et al., EASE: EPC As a Service to Ease Mobile Core Network Deployment Over Cloud, IEEE Network 29 (2015) 78–88.

[7] J. Liu, et al., CONCERT: A Cloud-Based Architecture For Next-Generation Cellular Systems, IEEE Wireless Communications (2014) 14–22.

[8] I. F. Akyildiz, P. Wang, S.-C. Lin, SoftAir: A Software Defined Networking Architecture for 5G Wireless Systems, Computer Networks 85 (2015) 1–18.

[9] S. Abdelwahab, et al., Network function virtualization in 5G, IEEE Communications Magazine 54 (2016) 84–91.

[10] M. Peng, et al., Fronthaul-constrained Cloud Radio Access Networks: Insights and Challenges, IEEE Wireless Communications 22 (2015) 152–160.

[11] Wubben, et al., Benefits and Impact of Cloud Computing on 5G Signal Processing, IEEE Signal Processing Magazine 31 (2014) 35–44.

[12] K. Pentikousis, Y. Wang, W. Hu, MobileFlow: Toward Software-Defined Mobile Networks, IEEE Communications Magazine 51 (2013) 44–53.

[13] J. Bartelt, P. Rost, D. Wubben, J. Lessmann, B. Melis, G. Fettweis, Fronthaul and Backhaul Requirements of Flexibly Centralized Radio Access Networks, IEEE Wireless Communications 22 (2015) 105–111.

[14] D. Boviz, C. S. Chen, S. Yang, Effective Design of Multi-User Reception and Fronthaul Rate Allocation in 5G Cloud RAN, IEEE Journal on Selected Areas in Communications 35 (2017) 1825–1836.

[15] X. Wang, et al., Virtualized Cloud Radio Access Network for 5G Transport, IEEE Communications Magazine 55 (2017) 202–209.

[16] A. Checko, et al., Cloud RAN for Mobile Networks: A Technology Overview, IEEE Communications Surveys & Tutorials 17 (2015) 405–426.

[17] Van Lingen, et al., The Unavoidable Convergence of NFV, 5G, and Fog: A Model-Driven Approach to Bridge Cloud and Edge, IEEE Communications Magazine 55 (2017) 28–35.

[18] O. Font-Bach, et al., When SDR meets a 5G candidate waveform: Agile use of fragmented spectrum and interference protection in PMR networks, IEEE Wireless Communications 22 (2015) 56–66.

[19] M. Peng, et al., Heterogeneous Cloud Radio Access Networks: A New Perspective for Enhancing Spectral and Energy Efficiencies, IEEE Wireless Communications 21 (2014) 126–135.