Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 5.258

IJCSMC, Vol. 5, Issue. 6, June 2016, pg.162 – 173

A Study on Benefits and Impact of Cloud Computing on 5G Signal Processing

Dr. I.Lakshmi
Assistant Professor, Stella Maris College, Chennai-600086, Tamil Nadu, INDIA

Abstract: Distributed computing attracts critical consideration the data innovation (IT) people group as it gives pervasive on interest access to a mutual pool of configurable registering assets with least administration exertion. It picks up likewise more effect on the correspondence innovation (CT) people group and is as of now talked about as an empowering agent for adaptable, cost-proficient and all the more capable versatile system executions. Albeit concentrated baseband pools are now researched for the radio access system (RAN) to take into account effective asset utilization and progressed multicell calculations, these advances still require devoted equipment and don't offer the same attributes as distributed computing stages, i.e., on-interest provisioning, virtualization, asset pooling, versatility, administration metering, and multitenancy. In any case, these properties of distributed computing are key empowering agents for future portable correspondence frameworks described by a ultra thick organization of radio access focuses (RAPs) prompting serious multicell obstruction in mix with a noteworthy increment of the quantity of access hubs and colossal changes of the rate prerequisites after some time. In this article, we will investigate the advantages that distributed computing offers for fifth-era (5G) portable systems and examine the suggestions on the sign handling calculations.

INTRODUCTION

The development toward 5G portable systems is portrayed by an exponential development of activity. This development is brought about by an expanded number of client terminals, wealthier Internet content, more successive utilization of Internet-fit gadgets, and by all the more effective gadgets with bigger screens. This infers additionally the requirement for all the more scaling conceivable outcomes in portable systems to handle spatially and transiently fluctuating movement designs, terminals with various quality prerequisites, and more different administrations. Current portable systems are not ready to bolster this assorted qualities proficiently but rather are intended for crest provisioning and run of the mill Internet activity. The utilization of exceptionally thick, low-control, little cell systems with high spatial reuse is a promising approach to take into account taking care of future information rate requests [1], [2]. Little cells misuse two major impacts. To begin with, the separation between the RAP and terminals is lessened, which expands the viewable pathway likelihood and diminishes the way misfortune. Second, the range is utilized all the more productively in light of the fact that every RAP utilizes the same range. Little cells supplement existing macrocellular arrangements that are required to give scope to quick moving clients and in territories with low client thickness. In Third-Generation Partnership Project (3GPP) long haul development (LTE), little cells draw noteworthy consideration on both the physical and higher layer [3], [4], where sways on the RAN convention and framework engineering are talked about. As systems get to be denser, intercell impedance increments and obstruction situations turn out to be more mind boggling because of multitier impedance. Besides, the higher the arrangement thickness, the higher the chance that a RAP will convey no or just low movement load because of spatial and fleeting activity variances. Presently, 15–20% of all destinations convey around half of the aggregate movement [5]. Incorporated preparing licenses to specifically turn RAPs on and off to address the spatiotemporal activity variances. What's more, it takes into consideration proficient obstruction shirking and cancelation calculations over different cells and also joint recognition calculations. Unified RAN (C-RAN) as of late pulled in consideration as one conceivable approach to effectively unify RAN preparing [6]. In C-RAN, remote radio heads (RRHs) are associated through optical fiber connections to a server farm where all baseband preparing is performed [7], [8]. Subsequently, by pooling baseband handling in baseband units (BBUs), centralization increases are accomplished. In any case, BBUs depend on particular equipment stages using advanced sign processors (DSPs) [9]. As a long haul objective, it is valuable to convey distributed computing stages running on broadly useful equipment, prompting a cloud-RAN framework as delineated along these lines in this article. Just fiber connections are equipped for supporting the vital information rates between the RRH and the BBU. This constitutes the fundamental downside of C-RAN, i.e., it requires high information rate connections to the focal BBU. In [8], the creators report a required backhaul (BH) transmission rate of 10 Gbit/s for time-space LTE (TDLTE) with eight get receiving wires and 20-MHz transfer speed. Because of the utilization of optical fiber, C-RAN organizations are less adaptable as just spots with existing fiber access might be picked or fiber access must be conveyed, which is extremely taken a toll extraordinary. Future portable systems will send heterogeneous BH arrangements that are improved for various situations. This blend of BH attributes will likewise suggest a blend of more C-RAN arrangements that require high limit BH and more decentralized arrangements perfect with BH arrangements that present high inactivity and more grounded throughput imperatives [10]. The RAN as a Service (RANaaS) idea is presented in [11]. It addresses the insufficiencies of C-RAN to take into account centralization over heterogeneous BH. The fundamental attributes of RANaaS are the adaptable task of RAN usefulness between the RAPs and the focal processor, the organization of item equipment at the focal processor, and the tight combination of RAN, BH system, and focal processor. In this article, we concentrate on the difficulties and advantages of actualizing sign preparing calculations on a distributed computing stage. Consequently, in the accompanying, we allude to the idea of centralization toward ware distributed computing stages as cloud-RAN. More points of interest on the engineering outline of the basic 5G portable system and also basic ideas from medium access control (MAC) and system layer of the cloud-RAN idea are given in [11]. Further difficulties in 5G versatile systems, which are past the extent of this article, are presented in [2] and [12], among others. Be that as it may, cloud-RAN will encourage approaches as of now under discourse for 5G, for example, enormous various info, different yield (MIMO) and numerous radio access innovations.

FLEXIBLE CENTRALIZATION THROUGH CLOUD-RAN

An adaptable task of RAN usefulness can consider both the cloud-stage asset accessibility and the little cell BH attributes. What's more, distributed computing stages take into account the adaptability that is required to adapt to fleeting and spatial movement variances in versatile systems. This adaptability is a key necessity to enhance the use of portable systems and to take into account a monetarily and biologically feasible operation of versatile systems. Cloud-RAN is a problematic innovation from various perspectives and forces new difficulties on the sign preparing in 5G portable systems. Above all, it will abuse standard processor innovation [general-reason processors (GPPs)] to execute RAN usefulness. By differentiation, right now talked about C-RAN innovation considers a baseband pooling approach where countless are given at a focal element [8], [9]. In spite of the fact that this considers asset sharing, C-RAN still uses particular and costly equipment and programming. Consequently, it is deluding to consider C-RAN as a case of distributed computing as per the IT definition by the U.S. National Institute of Standards and Technology [13]. Cloud-RAN will facilitate foster adaptable calculations that are intended for distributed computing situations and influence gigantic parallelism. This suggests calculations ought not be basically ported to distributed computing stages yet rather updated to pick up from the accessible processing assets. Cloud-RAN takes into consideration the sending of calculations that scale with the requirement for collaboration and coordination among the individual cells, i.e., contingent upon the activity interest and client thickness, RAPs might be distinctively assembled or diverse calculations might be sent. In the accompanying areas, this article gives more definite case to calculations that profit by an application to distributed computing stages. To empower cloud-RAN, it is important to have a framework engineering that gives the required interfaces without troublesome changes to a current arrangement. This design has been presented in [11]. It doesn't suggest changes to existing interfaces yet presents the idea of a virtual eNodeB (eNB). A eNB is made out of one or more RAPs, a distributed computing stage, and the essential BH joins between these hubs. It keeps up the same interfaces as a 3GPP LTE eNodeB (eNB) to amplify in reverse similarity. This framework design requires

- that the usefulness at the eNB can be deteriorated into reassign capable capacities and
- that every capacity can be allocated either to the focal processor or neighbourhood RAPs.

Besides, a tight reconciliation of RAN, BH, and focal processor is required, e.g., through joint coding as presented in [14].

OPPORTUNITIES OF CLOUD-RAN

Distributed computing offers the capacity of computational burden adjusting to RANs. This takes into consideration spending more computational endeavours on basic operations, e.g., on account of obstruction situations or troublesome channel conditions. In these situations, more progressed and computationally extreme calculations might be required and could be executed in a cloud domain. By complexity, customary usage are hard constant frameworks. Thus, a specific assignment, for example, disentangling or booking is constantly executed inside the same time window. An adaptable task of usefulness will likewise consider moulding the flagging burden on the BH association. Case in point, on account of high-limit, low-idleness BH, the focal processor may handle straightforwardly in-stage/quadrature (I/Q) tests. On account of higher idleness and lower data transmission on the BH, the focal processor may just perform upper-layer usefulness. This will oblige changes to the operation of the BH and the sign preparing stage, and it might oblige changes to the RAN convention stack. Cloud-RAN will open the entryway for some new applications in 5G. It offers the likelihood of utilizing sign handling programming devoted to an uncommon reason in light of the genuine administration. It mirrors the assorted qualities of administrations, use cases, and organizations through adaptability and versatility of the sign preparing stage. Also, it might even consider the intricacy and capacities of terminals amid the handling of signs. At long last, cloud-RAN maintains a strategic distance from the run of the mill merchant lock-in as in current organizations that take after a comparative improvement saw in the portable center system, which might be actualized on cloud-stages [15]. The adaptable centralization of RAN usefulness will affect the operation of the 3GPP LTE RAN convention stack and may even be restricted by conditions inside the convention stack. Table 1 gives a diagram of promising elements of the 3GPP LTE radio convention stack, which might be considered for an incomplete centralization. All in all, the lower we put the utilitarian split inside the convention stack, the higher the overhead and the more stringent BH prerequisites are. Unifying usefulness on the physical layer (PHY) takes into account an computational assorted quality that depends straightforwardly on the quantity of clients per RAP. Because of worldly and spatial changes, the computational burden can be adjusted crosswise over cells. Focal preparing additionally takes into consideration actualizing multicell calculations to stay away from or misuse obstruction, e.g., intercell impedance coordination and helpful multipoint handling [16].

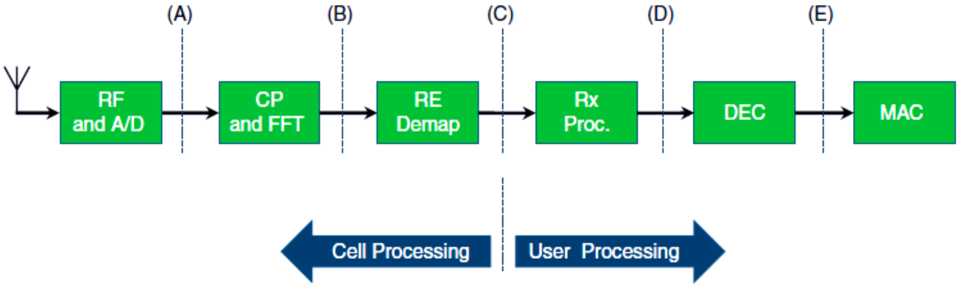

Fig 1 The functional split between RAPs and the cloud-platform for UL transmission.

CENTRALIZED FUNCTIONALITY	CENTRALIZED REQUIREMENTS	CENTRALIZATION BENEFITS	CHALLENGES FOR SIGNAL PROCESSING
DETECTION AND FEC-DECODING	■ DEPENDS ON CONTROL OVERHEAD IN UL ■ LATENCY REQ. DEPENDS ON TIMING REQ. IN DL ■ STRONG RELIABILITY	■ COOPERATIVE RECEIVER (RX) ■ COMPUTATIONAL DIVERSITY	■ PREDETECTION AT RAP TO REDUCE BH OVERHEAD ■ OPTIMAL QUANTIZATION OF SIGNALS AND EXCHANGE OVER BH
FEC-ENCODING AND MODULATION AND PRECODING	■ DEPENDS ON CONTROL OVERHEAD IN DL ■ STRONG RELIABILITY	■ COOPERATIVE TRANSMITTER (TX) ■ ADVANCED PRECODING ■ COMPUTATIONAL DIVERSITY	■ SEPARATE PRECODING DECISION AND EXECUTION AT RAP AND CENTRAL PROCESSOR ■ OPTIMAL QUANTIZATION OF SIGNALS AND EXCHANGE OVER BH
LINK RELIABILITY PROTOCOLS (E.G., HARQ)	■ DEPENDS ON ENTITY THAT PERFORMS RETRANSMISSION DECISION	■ SIMPLIFIED CENTRALIZATION OF SCHEDULING AND DECODING	■ PREDEFINED TIMING OF (N)ACK MESSAGES ■ SEPARATION OF RETRANSMISSION DECISION AND PACKET COMBINING ■ STRONG INTERACTION WITH OTHER FUNCTIONS, E.G., SCHEDULER, EN-/DECODER
SCHEDULING AND INTERCELL RRM	■ FLEXIBLE REQUIREMENTS	■ MULTICELL GAINS ■ COMPUTATIONALLY EXPENSIVE ALGORITHMS ■ GAINS DEPEND ON BH QUALITY	■ SCALABLE LATENCY REQUIREMENTS MUST BE SUPPORTED ■ INTERCELL INTERFERENCE COORDINATION (ICIC) BASED ON CHANGING QUALITY OF CHANNEL STATE INFORMATION ■ CHANGING COMPUTATIONAL COMPLEXITY

Table 1 The benefits and signal processing challenges for the centralization of selected 3GPP LTE radio protocol functionality on the PHY and lower MAC layer.

FUNCTIONAL SPLIT

In this subsection, we introduce several functional split options that determine the execution of processing in the RAP or in the cloud-platform and directly influence the required BH data rate. The discussion is focused on the uplink (UL) since its processing load dominates the downlink (DL) processing. Detailed investigations of such splits have also been conducted in [17], but here we focus more on the opportunities of a flexible split. By relying on GPPs as opposed to dedicated hardware as used in the C-RAN concept, and through extensive use of function virtualization, the envisioned architecture allows us to adapt the functional split flexibly in time (e.g., according to traffic demand) and location (e.g., depending on the density of the deployment). Figure 1 illustrates the principle LTE signal processing chain of an UL receiver and different options of placing a functional split. Notice that similar shifts are also possible for DL processing as considered, e.g., in the context of preceding for massive MIMO systems in [18]. Subsequently, we discuss these split options and give numerical results on the required BH data rates per link between one RAP and the cloud-platform for a simple configuration as specified in Table 2.

PARAMETER	SYMBOL	VALUE
BANDWIDTH	B	20 MHz
SAMPLING FREQUENCY	f_S	30.72 MHz
OVERSAMPLING FACTOR	N_O	2
NUMBER OF USED SUBCARRIERS	N_{SC}	1,200
SYMBOL DURATION	T_S	66.6 µs
QUANTIZATION/SOFT BITS PER I/Q	N_Q	10
RX ANTENNAS	N_R	2
SPECTRAL EFFICIENCY	S	3 bit/cu
ASSUMED RB UTILIZATION	η	50%

Table 2: Exemplary transmission parameter for calculating the impact of functional split choices on the BH data rate

I/Q FORWARDING (A)

By quickly sending the time-area get signals that have been down changed over to the baseband and simple to advanced (AD) changed over (demonstrated by piece RF/AD), the complete get outline including the cyclic prefix (CP) must be transmitted over the BH connection to the cloud-stage. This methodology is normally alluded to as radio-over-fiber (RoF) and is utilized as a part of the basic open radio interface (CPRI) standard [19]. The principle advantage of this split is that no computerized preparing gadgets are required at the RAPs, conceivably making them little and shabby. On the off chance that an adaptable split differing after some time is imagined, the handling gadgets would need to be accessible at the RAPs in any case, invalidating this advantage. Additionally, the required BH information rate for I/Q sending is nearly high and given as

$$D_{BH}^{A}=N_0.f_s.2.N_Q.N_R=2.30.72\text{MHz}.2.10\text{ bit}.2=2.46\text{Gbits/s}. \quad \rightarrow(1)$$

SUB FRAME FORWARDING (B)

By removing the CP and transforming the Rx signal to frequency- domain using fast Fourier transformation (FFT), guard subcarriers can be removed (block CP/FFT). Since the number of guard subcarriers in LTE is ≈ 40%, this decreases the required BH data rate significantly.

$$D_{BH}^{B}=N_{SC}.T_s^{-1}.2.N_Q.N_R=1.200.(66\ \mu s)^{-1}.2.10\text{bit}.2=720\text{ Mbit/s}. \rightarrow(2)$$

As an FFT can be implemented on dedicated hardware very efficiently, the implementation in the RAP is worthwhile compared to the split option I/Q forwarding (A). As the per-cell based processing does not depend on the actual load of the RAP, load balancing gains can be only achieved if RAPs are completely turned off.

DATA FORWARDING (C)

If only a part of the resource elements (REs) are actually utilized by the user equipment (UE) in a cell, only these Res remain after RE remapping (block RE Demap) and have to be forwarded to the cloud-platform. The required BH data rate is directly given by the fraction of utilized RE and thus, the subsequent splits can profit from load balancing gains.

$$D_{BH}^{C} = D_{BH}^{B}.\eta = 720\text{ Mbit/s }.0.5=360\text{ Mbit/s}. \rightarrow (3)$$

To allow for a joint processing of received signals from multiple RAPs, it has to be ensured that only REs of UE not considered for joint processing are removed, even if they are not (primarily) associated with the current RAP.

SOFT-BIT FORWARDING (D)

The receive processing (block Rx Proc) per user consists of equalization in frequency domain, inverse discrete Fourier transformation (IDFT), MIMO receive processing, and remapping. In a MIMO scheme utilizing receiver diversity, the signals of multiple antennas are combined during channel equalization, thus removing the dependency on the number of receive antennas. This results in a reduced BH load of $D_{BH}^{D} = D_{BH}^{C}/N_R$=180 Mbit/s. In contrast, for spatial multiplexing with *N*S layers per UE, the BH would correspond to $D_{BH}^{D}=D_{BH}^{C} \cdot N_S / N_R$ By this split, only joint decoding of soft bits forwarded by several RAPs is possible in the cloud-platform. Also note that usually the number of soft bits per symbol would depend on the modulation scheme (e.g., three soft-bits per information bit), and thus *N*Q and the BH data rate would depend directly on the modulation order, which in turn depends on the access channel quality due to radio resource management (RRM).

MAC (E)

Amid forward blunder amendment (FEC) disentangling (piece DEC), information bits are recouped from the got images and repetitive bits are expelled, bringing about the unadulterated MAC payload at the decoder yield. The subsequent BH information rate depends to a great extent on the utilized balance and coding plan (MCS), which is reflected here by the excellent otherworldly proficiency S = 3 bit/cu. FEC translating is a mind boggling undertaking that is normally performed on devoted equipment and henceforth a brought together disentangling on GPPs has not been considered in C-RAN. In any case, as plot later in this article, late results demonstrate that it can be performed on GPPs. Then again, performing translating in the RAPs as per the split choice MAC (E) ends the likelihood for joint PHY-layer preparing in the cloud-stage and just participation on higher layers, e.g., joint planning, stays conceivable. As PHYlayer participation essentially rotates around impedance relief, this choice is advantageous in situations were RAPs are all around isolated, e.g., for indoor organizations or in restricted road ravines. Clearly, the required BH information rate and the required preparing power in the cloud diminishes essentially when the useful split is moved to the higher PHY handling layers or even to the MAC. Nonetheless, this is exchanged off with lower centralization picks up as far as ghastly effectiveness and computational burden adjusting. The upside of an adaptable split is that we can profit from both extremes: load adjusting for low movement circumstances and high ghostly effectiveness by agreeable handling for high activity. Since current BH guidelines like CPRI just backing a certain utilitarian split, new and more adaptable gauges will must be characterized to empower cloud-RAN designs. The gigantic BH transfer speed necessities of useful movements on the lower PHY layers likewise demonstrates that enhanced and improved BH innovations are required. While advances offering adequate data transfer capacity are as of now accessible [10], a joint outline of radio access and BH connections ought to be likewise considered to utilize the conveyed limit as productively as could be allowed. Also, as far as possible the BH rate between the RAPs and the cloud-stage, helpful preparing systems could be utilized to straightforwardly abuse lower-layer collaboration between RAPs. This would permit the utilization of heterogeneous BH innovations to interconnect the RAPs and execute joint circulated identification procedures as portrayed in Figure 2 and examined in the following segment

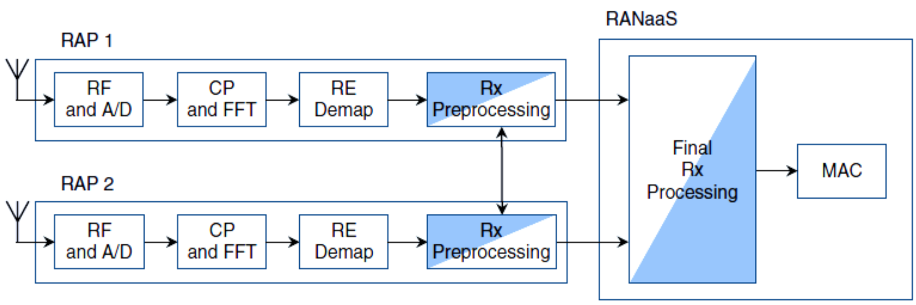

Fig2: The cooperative Rx pre-processing among RAPs with final Rx processing in the central processor.

SIGNAL PROCESSING IN THE CLOUD

The difficulty of implementing RAN functionality in a cloud-platform lies in the tight constraints caused by the 3GPP LTE protocol stack. This implies that individual tasks need to finish within a predefined time window. Figure 3 shows relevant parts of the 3GPP LTE protocol stack and two exemplary functional splits that correspond to options (C) and (D) in

Figure 1. In the following, we discuss the benefits and challenges of a cloud implementation of three representative parts of the signal processing chain.

HYBRID Automatic repeat request

Among all the timers defined in LTE, the one associated to the acknowledgment (ACK) of a UL physical frame at the MAC layer is the most critical one. The reception status of any frame sent through the air interface needs to be fed back to the transmitter, to proceed to the transmission of a new frame ACK or to attempt a retransmission negative ACK (NACK). This hybrid automatic repeat-request (HARQ) operation is performed at the MAC level, after all the physical processing of a codeword is done (detection, demodulation, and FEC decoding). In LTE, each frame sent at sub frame *n* needs to be acknowledged (ACK or NACK) at sub frame $n + 4$ in both UL and DL directions, a sub frame lasting 1 ms [20]. Hence, the overall receive process has to finish in 3 ms to stay compliant with the 3GPP LTE HARQ timing. This timing includes the processing at the RAPs of the physical blocks located before the split (see Figure 3 and both functional split options therein), the processing at the cloud-platform of the physical blocks located after the split and the round-time trip through the BH. However, some algorithms such as turbo-decoders underly a computational jitter which implies that the decoding time may vary. Hence, it may happen that packets are retransmitted even though they would have been decoded with more computational resources, i.e., either more time or more parallel processors. This computational jitter also adds up to the overall delay that needs to be considered. To relax the timing constraint for the receive processing, we may adapt the HARQ process. The authors in [17] suggest for example to suspend the HARQ process until the end of the receive processing. In the case that the receive processing is not finished in time, an ACK is sent after 3ms to meet the timing requirements while receive processing is continued. If, at the end, successful decoding is not possible, a NACK is sent. As the UE does not immediately drop out a package when receiving an ACK to cope with transmission errors on the feedback channel, a retransmission of the particular packet can be scheduled later. However, this approach halves the achievable UE peak rate [17]. This drawback can be avoided by a preliminary HARQ process, where the initial feedback message is determined by estimating the decoding success based on the quality of the received signals (e.g., using models from link level simulations [21]). If correct decoding is likely, a preliminary ACK is sent to the UE, otherwise a preliminary NACK. Again, the standard techniques capturing feedback errors automatically handle erroneous preliminary feedback messages. This approach relaxes the timing constraints for the receive processing chain. It separates the most complex processing parts and the most latency-critical parts but still allows for high data rates depending on the reliability of preliminary ACK/NACK.

FORWARD ERROR CORRECTION

The tight requirement of finishing the overall detection within 3 ms poses a significant challenge for executing FEC decoding within the cloud-platform due to its high complexity. Usually, FEC decoders are implemented in specialized hardware, such as application-specific integrated circuit (ASIC) designs or field-programmable gate array (FPGA) implementations [22]. However, the introduction of many-core architectures opens new perspectives for massively parallel implementations. To meet stringent requirements on data rates, cloud-based FEC decoders will need to fully exploit the available parallelism of a cloud-computing platform. In this context, low-density parity check (LDPC) [23] and turbo codes [24] are two promising candidates because both allow for accommodating various degrees of parallelization. From a high-level perspective, two main approaches can be used to exploit parallelism in multicore platforms. The first approach parallelizes the decoder itself through decomposition of the decoding algorithms into multiple threads that run in parallel. Second, multiple codeword may be decoded in parallel. The first approach decreases the latency per codeword but introduces more synchronization overhead across different threads. By contrast, the second approach uses less synchronization objects and therefore increases the parallelization gain. However, it may introduce a higher latency per codeword compared to the first approach. For very high throughput applications, LDPC codes are known to compare favourably against turbo codes because LDPC decoding allows for a higher degree of parallelism [25], [26]. Hence, LDPC codes are suitable for the first approach of decoder parallelization. However, software-based parallel LDPC decoders barely achieve throughputs of a few tens of Mbit⁄s, as reported in [27] for graphical processing units (GPUs), or in [28] for the signal processing on-demand architecture (SODA). In both cases, the main reason is the need for synchronization across different threads to access shared objects that results in scalability issues [27]. By contrast, parallelizing multiple codeword eliminates the need for synchronizing objects. This results in better scalability properties and the throughput of the multicodeword decoder is known to increase almost linearly with the number of cores [29]. Furthermore, it allows for different codes, algorithms, and configurations running in parallel. Multicodeword LDPC decoders have been reported to achieve throughputs up to 80 Mbit/s on the IBM CELL Broadband Engine [27], [30], with 24–96 codeword decoded in parallel. Recently, central processing unit (CPU) and GPU implementations of multicodeword turbo decoders have also been reported in [31] with a peak throughput from 55 Mbit/s to 122 Mbit/s, as the number of decoding iterations decreases from eight to four. Figure 4 shows experimental results for spectral efficiency and required computational complexity of an 3GPP LTE UL decoder. To obtain these results, the turbo-decoder has been implemented on a default VMWare ESXi server with Ubuntu Linux host operating system, GNU C++ compiler, and codeword multithreading to account for the virtualization overhead. We measured the required CPU time to decode one codeword and determined the average CPU time within the 90% confidence interval. Figure 4(a) shows the achievable spectral efficiency for a given signal-to-noise ratio (SNR) (additive white Gaussian noise, no fading). We illustrate the results for two cases: maximum throughput (high number of iterations possible) and low complexity (number of iterations limited to two). Reducing the complexity of the decoding process results in a performance penalty of 1–2 dB. In Figure 4(b), we show the required computational resources for a 10-MHz 3GPP LTE system. The required complexity strongly depends upon the SNR. First, it increases linearly with the number of information bits, which implies a logarithmic increase of complexity in SNR. Second, the complexity increases with the number of iterations that are necessary to decode a codeword. As shown in [32], the complexity increases super linearly with decreasing SNR (in decibels) for a fixed MCS. In Figure 4(b),

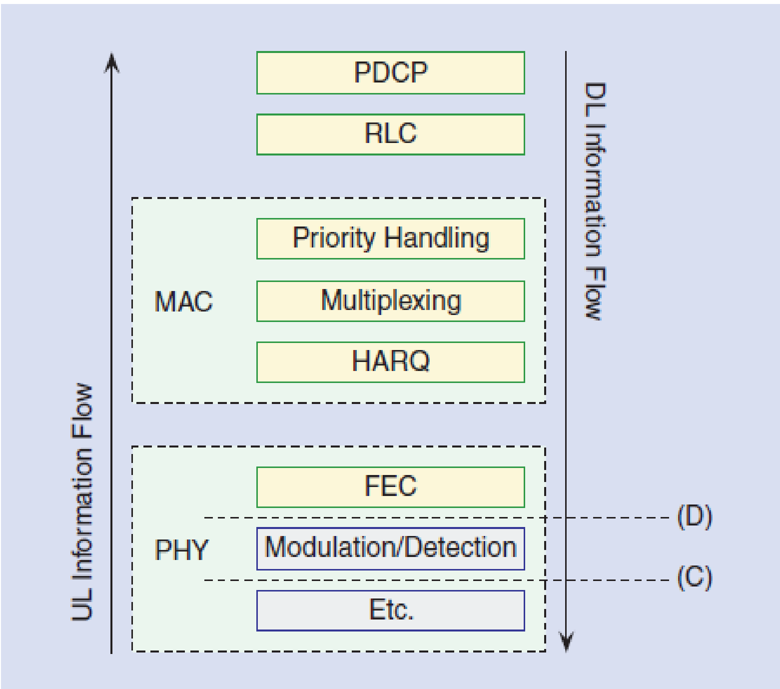

Fig3 The LTE protocol stack and exemplary functional splits.

markers show the SNR where the next higher MCS has been chosen. We notice at each of these markers an increase of the computational demand, which is then quickly decreasing in SNR. Apparently, this strongly varying computational demand allows for the exploitation of multiuser computational diversity at the centralized processor. For instance, the central processor can perform computational load balancing across multiple users to reduce the ratio of peak-to-average computational efforts. Furthermore, the central processor can actively shape the computational demand by selecting MCS to satisfy a computational constraint, e.g., in the case of a traffic burst the computational requirements may significantly increase and may exceed the available resources if MCSs are chosen based on maximum throughput. Finally, the computational load can be actively shaped by adjusting the number of quantization bits *N*Q used for forwarding the Rx signals from the RAP to the cloud-platform over the BH. Figure 5 shows the tradeoffs between number of turbo iterations and quantization bits *N*Q for different modulation schemes at a target bit error rate (BER) of 10-4 . Obviously, the decoding latency can significantly been reduced by increasing the number of quantization bits *N*Q on the cost of a higher BH transmission rate.

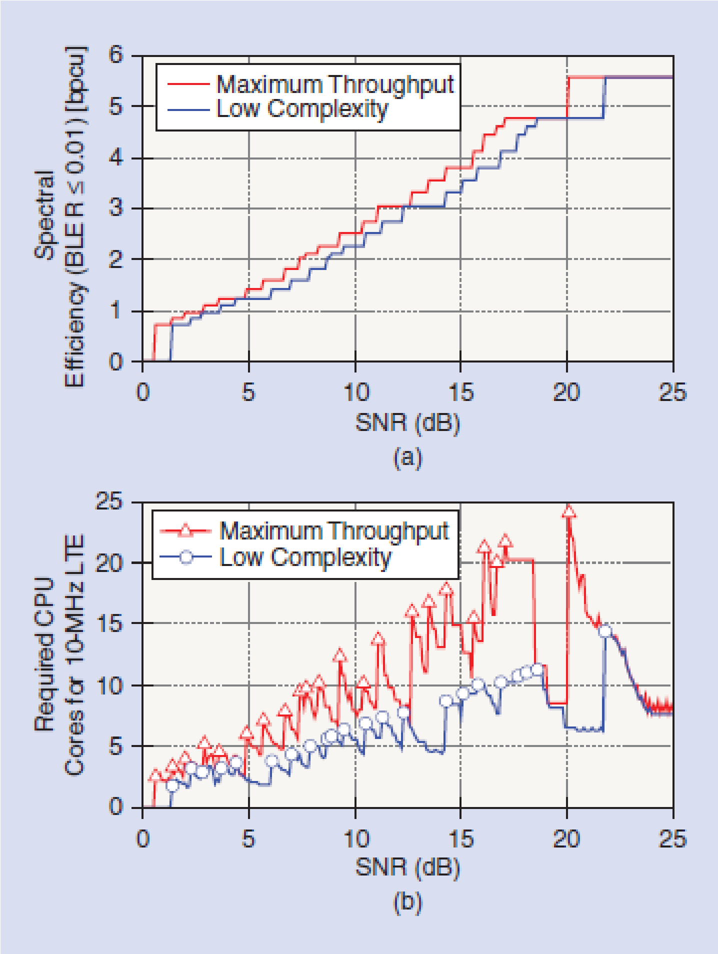

Fig 4 Throughput and computational complexity results for turbo-decoding using an out-of-the-box cloud-computing platform and 3GPP LTE MCSs.(a) Spectral efficiency. (b) Required CPU cores.

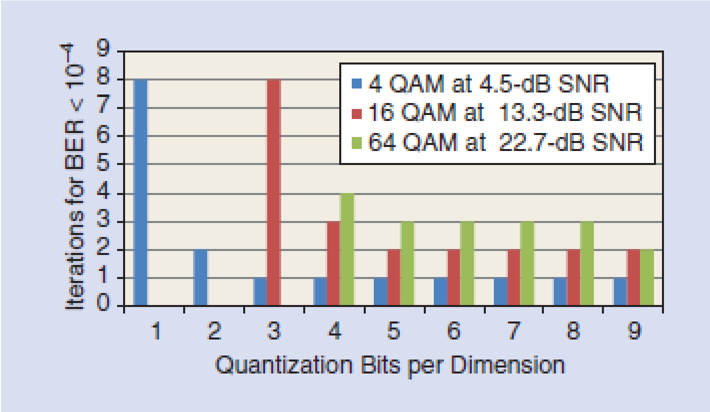

Fig 5: **The number of turbo iterations required for BER 1 10-4 versus number of quantization bits *N*q per I/Q dimension.**

CONCLUSIONS

This article discussed benefits and challenges that may be implied by cloud-computing platforms on signal processing algorithms. The novel RANaaS concept was introduced, which realizes cloud technologies in 5G mobile networks and allows for a flexible functional split between RAPs and the centralized cloud-platform. This allows for centralization benefits, but also introduces challenges due to the strict timing constraints imposed by the 3GPP LTE protocol stack. These challenges were identified and enabling technologies were discussed.

References

[1] M. Dohler, R. Heath, A. Lozano, C. Papadias, and R. A. Valenzuela, "Is the PHY layer dead?" *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 159–165, Apr. 2011.
[2] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: The dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82–89, Feb. 2014.
[3] 3GPP, "Small cell enhancements for E-UTRA and E-UTRAN—Physical layer aspects," Tech. Rep. TR.36.872, Sept. 2013.
[4] 3GPP, "Small cell enhancements for E-UTRA and E-UTRAN—Higher layer aspects," Tech. Rep. TR.36.842, May 2013.
[5] H. Guan, T. Kolding, and P. Merz, "Discovery of Cloud-RAN," in *Proc. Cloud- RAN Workshop*, Beijing, China, Apr. 2010.
[6] NGMN, "Suggestions on Potential Solutions to C-RAN by NGMN Alliance," Tech. Rep., Jan. 2013.
[7] D. Wake, A. Nkansah, and N. J. Gomes, "Radio over fiber link design for next generation wireless systems," *IEEE/OSA J. Lightwave Technol*., vol. 28, no. 16, pp. 2456–2464, Aug. 2010.
[8] K. Chen, C. Cui, Y. Huang, and B. Huang, "C-RAN: A green RAN framework," in *Green Communications: Theoretical Fundamentals, Algorithms and Applications*, J. Wu, S. Rangan, and H. Zhang, Eds. Boca Raton, FL: CRC Press, pp. 279–304, 2013.
[9] G. Li, S. Zhang, X. Yang, F. Liao, T. Ngai, S. Zhang, and K. Chen, "Architecture of GPP based, scalable, large-scale C-RAN BBU pool," in *Proc. Int. Workshop Cloud Base-Station Large-Scale Cooperative Communications, IEEE GLOBECOM 2012 Workshops*, Anaheim, CA, Dec., pp. 267–272.
[10] J. Bartelt, G. Fettweis, D. Wübben, M. Boldi, and B. Melis, "Heterogeneous backhaul for cloud-based mobile networks," in *Proc. 1st Int. Workshop Cloud Technologies Energy Efficiency Mobile Communication Networks (CLEEN 2013), IEEE VTC2013-Fall Workshops*, Las Vegas, NV, Sept.
[11] P. Rost, C. J. Bernardos, A. De Domenico, M. Di Girolamo, M. Lalam, A. Maeder, D. Sabella, and D. Wübben, "Cloud technologies for flexible 5G radio access networks," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 68–76, May 2014.

[12] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.
[13] P. Mell and T. Grance, "The NIST definition of cloud computing," Tech. Rep., National Inst. Standards Technology, Special Publication 800-145, Sept. 2011.
[14] J. Bartelt and G. Fettweis, "Radio-over-radio: I/Q-stream backhauling for cloud-based networks via millimeter wave links," in *Proc. 1st Int. Workshop Cloud Technologies Energy Efficiency Mobile Communication Networks, IEEE GLOBECOM 2013 Workshops (IWCPM 2013)*, Atlanta, GA, Dec., pp. 778–783.
[15] J. Kempf, B. Johansson, S. Pettersson, H. Luning, and T. Nilsson, "Moving the mobile evolved packet core to the cloud," in *Proc. IEEE Int. Conf. Wireless Mobile Computing, Networking, Communications*, Barcelona, Spain, Oct. 2012, pp. 784–791.
[16] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.
[17] U. D.tsch, M. Doll, H. P. Mayer, F. Schaich, J. Segel, and P. Sehier, "Quantitative analysis of split base station processing and determination of advantageous architectures for LTE," *Bell Labs Tech. J.*, vol. 18, no. 1, pp. 105–128, May 2013.
[18] S. Park, C.-B. Chae, and S. Bahk, "Before/after precoded massive MIMO in cloud radio access networks," *J. Commun. Netw.*, vol. 15, no. 4, pp. 398–406, Aug. 2013.
[19] CPRI. (2013, Aug.). Common public radio interface (CPRI); Interface specification (V6.0). Tech. Rep. [Online]. Available: http://www.cpri.info/
[20] 3GPP, "Radio access network; evolved universal terrestrial radio access; (E-UTRA); physical channels and modulation (Release 10)," Tech. Rep. TS.36.211, Dec. 2012.
[21] R. Visoz, A.O. Berthet, and M. Lalam, "Semi-analytical performance prediction methods for iterative MMSE-IC multiuser MIMO joint decoding," *IEEE Trans. Commun.*, vol. 58, no. 9, pp. 2576–2589, Sept. 2011.
[22] F. Kienle, N. Wehn, and H. Meyr, "On complexity, energy- and implementation-efficiency of channel decoders," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3301–3310, Dec. 2011.
[23] R. G. Gallager, *Low-Density Parity Check Codes* (Research Monograph Series). Cambridge, U.K.: MIT Press, 1963.
[24] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, Oct. 1996.
[25] M. M. Mansour and N. R. Shanbhag, "High-throughput LDPC decoders," *IEEE Trans. Very Large Scale Integration (VLSI) Syst.*, vol. 11, no. 6, pp. 976–996, Dec. 2003.
[26] C. Zhang, Z. Wang, J. Sha, L. Li, and J. Lin, "Flexible LDPC decoder design for multigigabit-per-second applications," *IEEE Trans. Circuits Syst. I*, vol. 57, no. 1, pp. 116–124, Jan. 2010.
[27] G. Falcao, L. Sousa, and V. Silva, "Massively LDPC decoding on multicore architectures," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 2, pp. 309–322, Feb. 2011.
[28] S. Seo, T. Mudge, Y. Zhu, and C. Chakrabarti, "Design and analysis of LDPC decoders for software defined radio," in *Proc. IEEE Workshop Signal Processing Systems*, Shanghai, China, Oct. 2007, pp. 201–215.
[29] A. Diavastos, P. Petrides, G. Falcao, and P. Trancoso, "LDPC decoding on the Intel SCC," in *Proc. IEEE 20th Euromicro Int. Conf. Parallel, Distributed Network- Based Processing (PDP)*, Garching, Germany, Feb. 2012, pp. 57–65.
[30] G. Falcao, V. Silva, L. Sousa, and J. Marinho, "High coded data rate and multicodeword WiMAX LDPC decoding on Cell/BE," *Electron. Lett.*, vol. 44, no. 24, pp. 1415–1417, Feb. 2008.
[31] M. Wu, G. Wang, B. Yin, C. Studer, and J. R. Cavallaro, "HSPA+/LTE-A turbo decoder on GPU and multicore CPU," in *Proc. 47th IEEE Asilomar Conf. Signals, Systems, Computers (ASILOMAR)*, Pacific Grove, CA, Nov. 2013, pp. 824–828.

[32] P. Grover, K. A. Woyach, and A. Sahai, “Towards a communication-theoretic understanding of system-level power consumption,” *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1744–1755, Sept. 2011.
[33] iJOIN. (2013, Nov.). State-of-the-art of and promising candidates for PHY layer approaches on access and backhaul network. Tech. Rep. [Online]. Available: www.ict-ijoin.eu/wp-content/uploads/2014/01/D2.1.pdf
[34] G. Caire and R. Müller, “The optimal received power distribution of IC-based iterative multiuser joint decoders,” in *Proc. 39th Annu. Allerton Conf. Communications, Control Computing*, Monticello, IL, Oct. 2001.
[35] R4-132017, “WF on NAICS receiver terminology,” 3GPP TSG-RAN WG4# 66bis, Apr. 2013.
[36] H. Zhu, A. Cano, and G. B. Giannakis, “Distributed consensus-based demodulation: Algorithms and error analysis,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 6, pp. 2044–2054, June 2010.
[37] H. Paul, J. Fliege, and A. Dekorsy, “In-network-processing: Distributed consensus-based linear estimation,” *IEEE Commun. Lett.*, vol. 17, no. 1, pp. 59–62, Jan. 2013.
[38] H. Paul, B.-S. Shin, D. Wübben, and A. Dekorsy, “In-network-processing for small cell cooperation in dense networks,” in *Proc. 1st Int. Workshop Cloud Technologies Energy Efficiency Mobile Communication Networks, IEEE VTC2013-Fall Workshops (CLEEN 2013)*, Las Vegas, NV, Sept.
[39] B.-S. Shin, H. Paul, D. Wübben, and A. Dekorsy, “Reduced overhead distributed consensus-based estimation algorithm,” in *Proc. 1st Int. Workshop Cloud Technologies Energy Efficiency Mobile Communication Networks, IEEE GLOBECOM Workshops 2013 (IWCPM 2013)*, Atlanta, GA, Dec., pp. 784–789.
[40] G. Xu, H. Paul, D. Wübben, and A. Dekorsy, “Fast distributed consensus-based estimation for cooperative wireless sensor networks,” in *Proc. 18th Int. ITG Workshop Smart Antennas (WSA 2014)*, Erlangen, Germany, Mar.

