Low Latency Wireless Communication System Implemented on a Software-Defined Radio Platform

Yujie Liu[1], Jun Yu[1], Fusheng Zhu[2], Wenru Zhang[2], and Jun Wu[3,4](✉)

[1] College of Electronic and Information Engineering,
Tongji University, Shanghai, China
`1833020@tongji.edu.cn`, `yhbbs7808@163.com`
[2] Guangdong Communications and Network Institute, Guangzhou, China
`{zhufusheng, zhangwenru}@gdcni.cn`
[3] School of Computer Science, Fudan University, Shanghai, China
`wujun@fudan.edu.cn`
[4] Pengcheng Laboratory, Shenzhen, China

Abstract. Low latency communication has attracted much interest in recent years due to the emergence of new types of delay-sensitive applications. Ultra-Reliable and Low-Latency Communication (URLLC) is also considered as one of the important use-cases in 5G cellular system. Traditional wireless communication system is usually optimized for high data throughput, but cannot satisfy the requirement of strict latency threshold. Semi-persistent scheduling and TTI shortening are two methods to address this problem. We implemented these methods by making modification based on OpenAirInterface framework, and build up a realistic cellular network system using software-defined radio technology. Experimental results show a dramatic latency reduction comparing with the baseline LTE scheme. By integrating these effective methods, we implemented a practical wireless communication system that can provide low latency transmission service.

Keywords: Low latency · Software-defined radio · URLLC · OpenAirInterface

1 Introduction

The 5G wireless communication system has been developed rapidly in recent years, and has been gradually used commercially worldwide. One big difference between 5G and 4G is that 5G is expected to support multiple new applications and use-cases rather than traditional communication scenarios [1]. To meet the requirement of new applications, three different service categories have been introduced: enhanced mobile broadband (eMBB), ultrareliable and low-latency communications (URLLC), and massive machine-type communications (mMTC) [2, 3].

The initial phase of 5G deployment is eMBB, which provides higher throughput and data rate with moderate latency. High resolution video streaming is the main application of eMBB, such as AR/VR/MR media, 4 K/8 K video, panoramic video, etc.

H. Gao et al. (Eds.): ChinaCom 2020, LNICST 352, pp. 660–669, 2021.
https://doi.org/10.1007/978-3-030-67720-6_45

With the development of factory automation, autopilot vehicles, smart grid, Intelligent Traffic System (ITS), etc., traditional wireless system has an urgent motivation to evolve and adapt to new applications. These applications usually don't need high data rate transmission, but has a stringent requirement of latency and reliability. URLLC is a typical scenario with such requirements [4, 5]. Figure 1 shows these new applications and their requirements of latency and reliability [6].

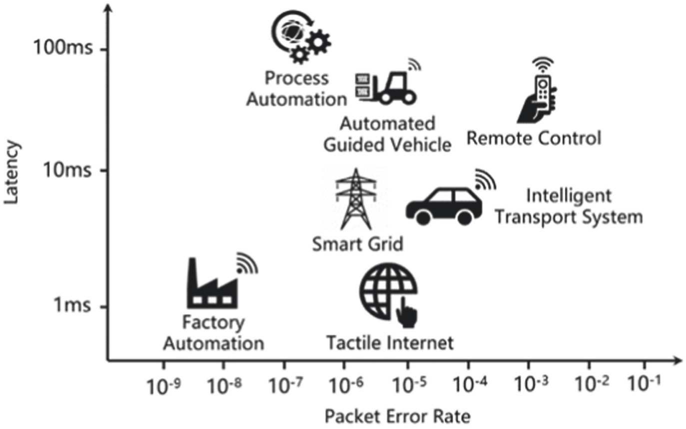

Fig. 1. URLLC use-cases and their requirements of packet error rate and latency

There are many enabling technologies of URLLC, such as short TTI, caching, densification, joint scheduling, grant-free uplink, NOMA, MEC/FOG/MIST and network slicing [7–10]. We would like to survey the related work as follow.

Ashraf et al. proposed a URLLC scheme for factory automation scenario [11]. They used multiple technologies such as shorter TTI, shorter processing time, Semi-Persistent Scheduling (SPS), Instant Uplink Access (IUA) and low complexity channel coding. Simulation results showed good improvement of latency and reliability in a factory deployment model.

Wang et al. gave a comprehensive evaluation of grant-free uplink transmission [12]. Data collision is an unavoidable problem in grant-free transmission, and would thus affect reliability. There are many factors that may affect the performance of grant-free transmission: number of active UEs, HARQ retransmission times, UE detection, SNR, packet arrival rate, etc. Evaluation results showed that a contention-based grant-free transmission scheme could well meet the reliability requirement within the latency bound.

Anand et al. proposed a scheme of joint scheduling of URLLC and eMBB traffic [13]. Superposition, puncturing, overlapping and mini-slot approaches are used to multiplex eMBB and URLLC traffic. Simulation results showed that this joint problem had structural properties, then these properties enabled clean decompositions and corresponding algorithms with theoretical guarantees.

According to our survey and investigation, most of these schemes have been evaluated only by simulations, so we come up with the idea to implement a low latency scheme on a software-defined radio (SDR) system. we implement a realistic wireless system and focus on optimizing the latency to reduce the overall network delay. By doing this, we can evaluate the real-time performance of our scheme, and can use this system as a platform to deploy delay-sensitive applications.

The rest of this paper is organized as follows. In Sect. 2, we analyze the composition of overall radio latency and propose two ways to reduce latency: semi-persistent scheduling and short TTI. In Sect. 3, we described the design and deployment of our system. Section 4 shows the experimental results of our scheme. In Sect. 5, we conclude our work and discuss the direction of future work.

2 Analysis of Latency Composition

A wireless system protocol stack consists of several layers (PHY, MAC, RLC, RRC, PDCP, etc.), and there are many procedures in each layer. In this paper, we focus on MAC and PHY layer, and mainly analyze two factors which contribute to overall transmission latency: scheduling scheme and transmission time interval (TTI).

2.1 Scheduling Scheme

Scheduling is a key feature of MAC layer. In traditional cellular system, the base station acts as the central control entity to allocate time and frequency domain resources, and coordinate all the UEs attaching to this base station by signaling on control channels.

Different scheduling schemes may have different impact on transmission latency. Traditional LTE system uses dynamic scheduling scheme as default, which is fit for mobile broadband traffic with no strict requirement on latency.

In dynamic scheduling scheme, downlink signaling and corresponding downlink data can be sent on the same frame of transmission, so downlink latency is not a critical factor in overall latency. However, in uplink transmission, a “request and grant” procedure is performed when UE has uplink data to send, resulting in a longer latency. Optimization of uplink scheduling scheme is a critical point to reduce the overall latency.

From Fig. 2(a), we can find that the traditional dynamic scheduling has following steps [14]:

1. When an uplink packet arrives at UE’s MAC layer at subframe t, a Schedule Request (SR) will be sent at subframe t + 4 on PUCCH.
2. The base station receives the SR at subframe t + 4, and scheduling algorithm is used to make decisions for radio resource allocation. If UE’s uplink request is granted, base station will send a specific type of Downlink Control Information (DCI) at subframe t + 8 on PDCCH, allocating radio resource for UE’s uplink traffic.
3. UE receives DCI at subframe t + 8, and process the data queueing in buffer. The uplink packet will actually be sent at subframe t + 12.
4. If the uplink packet is correctly received by base station, an ACK would be sent at subframe t + 16, informing UE of a successful uplink transmission.

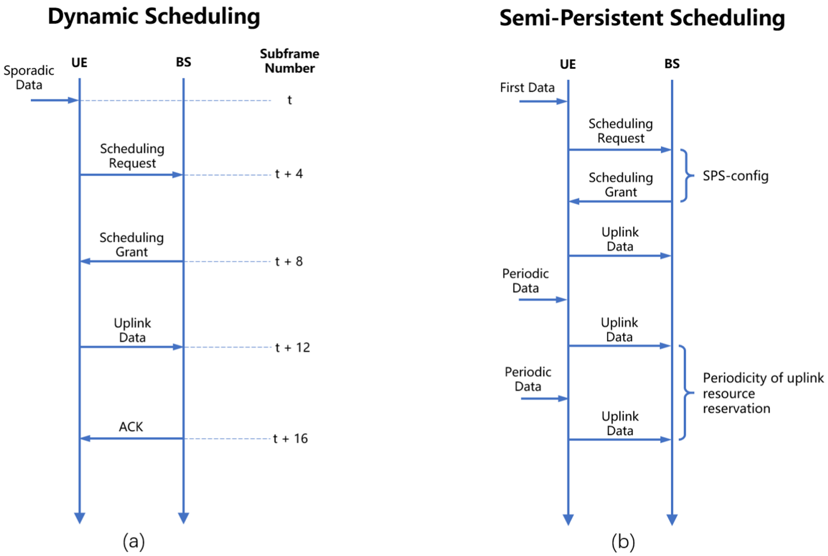

Fig. 2. Procedures of dynamic scheduling and semi-persistent scheduling

Such signaling procedure increases the latency. The overhead would be even bigger when transmitting large amount of small size packet data.

Semi-persistent scheduling (SPS) is a commonly-used scheduling scheme to reduce uplink latency. Figure 2(b) shows that when SPS is applied, the “request and grant” signaling procedure is performed only once at the first attempt of transmission. Afterwards, the uplink data can be sent on pre-allocated radio resource without performing scheduling procedure again. The radio resource is allocated in a periodic manner, and the period of SPS is configurable by the base station. User equipment would wait for its turn to transmit uplink data.

SPS is intended to transmit periodic data, such as VoIP traffic and sensor network data, and the period of SPS could be adjusted to match the production rate of uplink data. For sporadic uplink data, SPS can also be applied to reduce latency, but at the cost of spectral efficiency, because some resource may be over-allocated to user equipment when there is no uplink data to transmit.

2.2 Transmission Time Interval

Transmission Time Interval (TTI) is a very important time parameter in cellular network. TTI is the minimum time unit for many procedures such as channel coding, interleaving and scheduling, so TTI duration would impact the radio latency.

TTI is mainly decided by the design of frame structure. In 4G LTE, TTI is fixed to 1 ms due to the static frame structure, which is equal to the duration of 1 subframe (14 OFDM symbols with normal cyclic prefix). In 5G NR, numerology is introduced to adjust frame structure. Numerologies represent different configurations of Subcarrier Spacing (SCS). According to the theory of OFDM, OFDM symbol duration need to be inversely proportional to SCS to maintain the orthogonality of adjacent subcarriers, which means symbol duration could be shortened with SCS expansion. This new frame structure gives a possibility to shorten the duration of TTI. The baseline configuration of SCS is 15 kHz, which is same with the LTE SCS, and SCS can be scaled with a factor of 2^{μ}, where $\mu = [0, 1, 2, 3, 4, 5]$. Table 1 shows some combinations of numerology, SCS, OFDM symbol duration and TTI.

Table 1. Numerologies and corresponding frame parameters

Numerology (μ)	Subcarrier Spacing (SCS)	OFDM symbol duration	TTI length	TTI duration
0	15 kHz	66.7 μs	14 symbols with normal CP	1 ms
1	30 kHz	33.3 μs	14 symbols with normal CP	0.5 ms
2	60 kHz	16.7 μs	14 symbols with normal CP	0.25 ms
3	120 kHz	8.33 μs	14 symbols with normal CP	0.125 ms
…	…	…	…	…

Apparently, shorter TTI would have better latency performance. TTI consists of several OFDM symbols, so there are two ways to reduce the duration of TTI: reduce the number of OFMD symbols within a TTI, or reduce the duration of an OFDM symbol. For the former case, mini-slot is introduced in 5G NR as a smaller transmission unit. UE can be scheduled either on slot level (duration of 14 OFDM symbols) or on mini-slot level (duration of several OFDM symbols that can be configured). Mini-slot is a good way to shorten the duration of TTI, but requires lots of modifications of frame structure and signal processing procedures. For the latter case, by increasing the numerology, we can shorten the duration of OFDM symbol, and thus shorten the TTI. When numerology increases, the frame structure is expanded in frequency domain, and is compressed in time domain. In our experiment, we choose to increase numerology and keep the number of symbols within TTI unchanged. Figure 3 shows the frame structures of different numerologies.

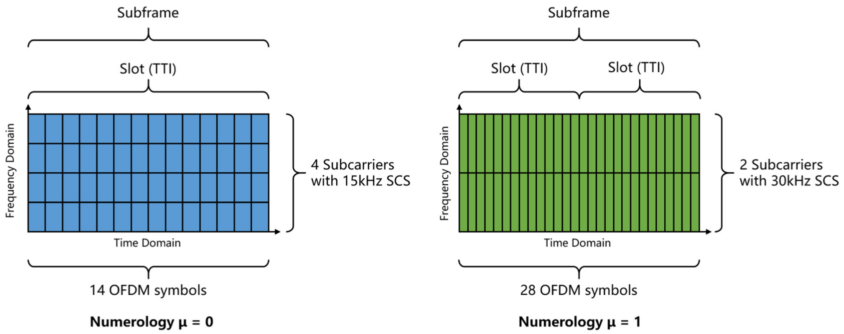

Fig. 3. 5G NR Frame structure of numerology 0 and 1 on time and frequency domain

Shorter TTI have better performance on latency, but TTI can't be reduced infinitely. Shorter TTI would result in higher inter-symbol interference and lower efficiency of channel coding, interleaving, spectrum utilization. A tradeoff should be made to choose a proper TTI to meet the requirement of particular applications.

3 System Design and Deployment

Software-defined radio (SDR) is a kind of radio communication system which implements most of radio signal processing by means of software running on a CPU or DSP, instead of running on specific hardware in traditional wireless system. SDR technology gives more flexibility and convenience on developing radio systems. Researchers can modify the procedures and functions by simply modifying software code, and it's very suitable for implementing prototype system or verifying new schemes. The hardware and software of our SDR system are:

- Host Machine: DELL PowerEdge 8700
- OS: Ubuntu 16.04
- Radio Device: Xilinx Virtex-6 FPGA + Analog Devices AD9361
- SDR Software: Eurocom OpenAirInterface

Our radio frontend device is built up with a Xilinx Virtex-6 FPGA evaluation board and Analog Devices AD9361 radio transceivers, and use PCI-E interface to stream samples to the host machine.

Three host machines are needed by our SDR system. Host Machine 1 runs the components of LTE EPC: HSS, MME and SPGW; Host Machine 2 and 3 are plugged with radio devices and run the softmodem program of eNode and UE respectively. Host Machine 1 and 2 are connected to the same router to provide a backhaul link.

Channel condition is a critical factor in realistic wireless transmission. If channel condition fluctuates a lot, some packets could not transmit successfully at first attempt, and retransmissions would be triggered frequently, causing large fluctuation of latency. To create a relatively stable testing environment, we keep the antennas of UE and eNodeB at a close distance with direct sight to provide a good channel condition, so that most packets can be sent successfully in one transmission.

The SDR software we used in our experiment is OpenAirInterface (OAI). OAI is an open source software radio framework initiate by Eurocom [15], which implements full protocol stack of LTE UE, eNodeB and EPC, forming a complete LTE SDR system. Because the 5G NR version of OAI is under development, we still use the 4G LTE version of OAI as framework and make our modification based on 4G LTE.

Dynamic scheduling is the default option in OAI, and uplink scheduling is triggered by Scheduling Request (SR) or Buffer Status Report (BSR). Following items show our modifications of the scheduling algorithm:

- Disable the uplink scheduling request and grant procedure.
- Use subframe number as index and allocate uplink resource to UE periodically to implement the behavior of SPS.

- Set different period of SPS. For example, if period is set to 1, data can be transmitted on every TTI; if period is set to 3, there is only 1 transmitting TTI in every 3 TTIs.

Next, we try to adjust the transmission time interval to minimize latency. Because we are still using LTE version of OAI as framework, and LTE doesn't support adjustable subcarrier spacing, so we need to set it manually. We set the subcarrier spacing to 15 kHz or 30 kHz by reconfiguring the radio device on both UE and eNodeB side. By doing this, we can make the LTE system running at 1 ms and 0.5 ms TTI respectively.

Figure 4 shows the full deployment of our SDR system.

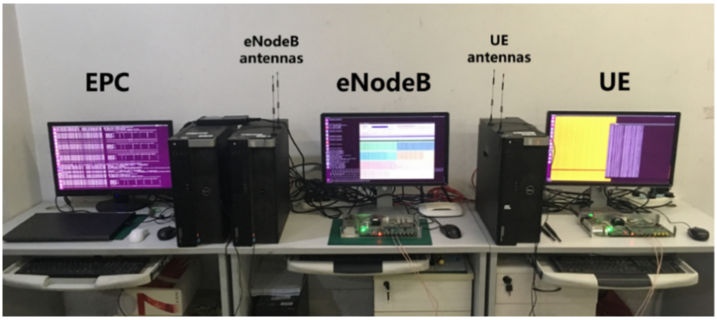

Fig. 4. Full deployment of our SDR system.

4 Experimental Results

First, we give a standard LTE configuration as baseline:

Subcarrier Spacing: 15 kHz
OFDM symbol duration: 66.7 µs
Multiplex Mode: FDD
TTI duration: 1 ms
Scheduling scheme: dynamic scheduling (DS)

In realistic wireless system, it's not easy to measure the latency between UE and base station directly, so we use round-trip time (RTT) of ping as the indicator of latency. Ping is initiate by UE, and the latency includes transmissions of a Scheduling Request, a Scheduling Grant, a ping request and a ping reply. Ping is based on IP packet, and it can better represent the latency that applications may experience in mobile system. The total round-trip time of ping can be calculated by the following equation:

$t_{total} = t_{SR} + t_{SG} + t_{REQ} + t_{REP}$

t_{total}: the total round-trip time of ping;
t_{SR}: time of uplink Scheduling Request transmission;
t_{SG}: time of downlink Scheduling Grant transmission;
t_{REQ}: time of uplink ping request transmission;
t_{REP}: time of downlink ping reply transmission.

When a UE is attached to the base station, a virtual network interface would be created on UE's host machine, and an IP address would be assigned to this UE by the EPC, and this UE is now able to exchange IP traffic with the network. We ping the core network's packet gateway (PGW) from UE for at least 200 times at one second interval, and take the average ping RTT as the value of latency. All the latency in the following parts of this article is calculated in this way.

To verify the impact of different packet size to the network latency, we disabled the function of dynamic radio resource allocation, and manually set some parameters to let each transmission has the same transport block size (TBS). In our experiment, we set the number of uplink resource blocks to 6 RB, and set uplink Modulation and Coding Scheme (MCS) to 8, so that the uplink TBS is fixed to 101 Bytes. When the size of an uplink packet exceeds TBS, multiple transmissions would be needed for one packet (64 Bytes: 1 transmission; 256 Bytes: 3 transmissions; 1024 Bytes: 11 transmissions), and this would have certain impact on latency when using large period SPS scheme.

First of all, we validated the performance of shorter TTI. From Fig. 5, we can find that there is a proportional relation between TTI and ping latency. When TTI is shortened from 1 ms to 0.5 ms, the ping latency also cut in half. We also tried to shorten the TTI to 0.25 ms, but UE failed to complete the random access procedure and couldn't attach to base station. This may be caused by the increasing Inter Symbol Interference (ISI) and the limitation of hardware's capability.

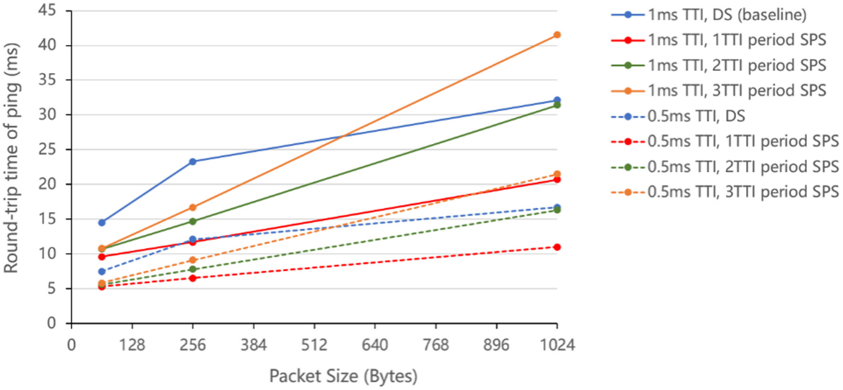

Fig. 5. Latency performance of different TTI duration and scheduling scheme combinations.

Secondly, we validated the performance of optimized uplink scheduling scheme. We applied the modifications mentioned in Sect. 3 to replace the original dynamic scheduling (DS). From Fig. 5, we can find that 1 ms period SPS can reduce the latency by about 5 ms comparing with the baseline when transmitting small packet, and this is contributed by removing the scheduling request and grant procedure, so there is no t_{SR} and t_{SG} components in overall latency.

With Fig. 5,if the uplink packet size is small enough to be sent in one transmission, SPS could have a better latency performance, but if packet size is as large as several times of TBS and the period of SPS is large, SPS performance may be worse than DS. This experiment proved that SPS fitted for periodic small packet traffic. Setting the uplink SPS period to 1 TTI will have the best latency performance because all the data packets could be transmitted immediately upon arrival of MAC layer, but at the cost of reduced spectral efficiency and resource utilization.

By combining shortened-TTI and SPS together, we achieved a minimum 5.3 ms RTT of 64 Bytes ping packet. With this latency level, we have reduced about 65% of network latency comparing with the LTE baseline profile.

5 Conclusion and Discussion

In this paper, we analyzed two major factors contributing to network latency in wireless system: scheduling scheme and transmission time interval. We implemented semi-persistent scheduling and short TTI schemes on a software-defined radio platform and verified the performance. Experimental results show a notable improvement on network latency. The overall delay can be reduced significantly comparing with the baseline LTE system, making it feasible to deploy delay-sensitive applications.

As part of our future work, we also investigate other methods to reduce wireless network latency, such as mini-slot frame structure and HARQ RTT shortening. By exploiting the potential of all these methods, we expect to reduce the latency furthermore, and build up a practical system that can meet the strict latency requirement of URLLC.

Acknowledgement. The authors would like to thank the editors, reviewers and people who provided comments and suggestions for this paper. This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61831018, and 61631017, and Guangdong Province Key Research and Development Program Major Science and Technology Projects under Grant 2018B010115002.

References

1. Li, Z., Shariatmadari, H., Singh, B., et al.: 5G URLLC: design challenges and system concepts. In: 15th International Symposium on Wireless Communication Systems (ISWCS), Lisbon (2018)
2. GPP TR 38.913, Study on Scenarios and Requirements for Next Generation Access Technologies. 3GPP (2017)
3. GPP TR 36.881, Study on latency reduction techniques for LTE. 3GPP (2016)
4. GPP TR 38.802, Study on New Radio (NR) Access Technology, Physical Layer Aspects. 3GPP (2017)
5. International Telecommunication Union (ITU), IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond. Telecommunication Standardization Sector of ITU (2015)

6. Peisa, J.: 5G techniques for ultra-reliable low latency communication. In: IEEE Conference on Standards for Communications & Networking (CSCN). IEEE, Helsinki (2017)
7. Sachs, J., Wikström, G., Dudda, T., et al.: 5G radio network design for ultra-reliable low-latency communication. IEEE Network **32**(2), 24–31 (2018)
8. Fehrenbach, T., Datta, R., Göktepe, B., et al.: URLLC services in 5G low latency enhancements for LTE. In: IEEE 88th Vehicular Technology Conference (VTC-Fall), pp. 1–6. IEEE, Chicago (2018)
9. Shariatmadari, H., Duan, R., Iraji, S., Li, Z., Uusitalo, M.A., Jäntti, R.: Resource allocations for ultra-reliable low-latency communications. Int. J. Wireless Inf. Netwk. **24**(3), 317–327 (2017). https://doi.org/10.1007/s10776-017-0360-5
10. Pocovi, G., Soret, B., Pedersen, K.I., et al.: MAC layer enhancements for ultra-reliable low-latency communications in cellular networks. In: IEEE International Conference on Communications (ICC) Workshop. IEEE, Paris (2017)
11. Ashraf, A., Aktas, I., Eriksson, E., et al.: Ultra-reliable and low-latency communication for wireless factory automation: From LTE to 5G. In: IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1–8. IEEE, Berlin (2016)
12. Wang, C., Chen, Y., Wu, Y., et al.: Performance evaluation of grant-free transmission for uplink URLLC services. In: 85th Vehicular Technology Conference. IEEE, Sydney (2017)
13. Anand, A., De Veciana, G., Shakkottai, S.: Joint scheduling of URLLC and eMBB traffic in 5G wireless networks. In: IEEE International Conference on Computer Communications, pp. 1970–1978. IEEE, Honolulu (2018)
14. GPP TS 36.213, Evolved Universal Terrestrial Radio Access (E-UTRA), Physical layer procedures. 3GPP (2017)
15. OpenAirInterface, https://www.openairinterface.org/