# A Disaggregated 5G Testbed for Professional Live Audio Production

Jan Dürre
*Sennheiser electronic GmbH & Co. KG*
Wedemark, Germany
jan.duerre@sennheiser.com

Norbert Werner
*Sennheiser electronic GmbH & Co. KG*
Wedemark, Germany
norbert.werner@sennheiser.com

Panagiotis Matzakos
*Communication Systems Department*
*EURECOM*
Biot, France
panagiotis.matzakos@eurecom.fr

Raymond Knopp
*Communication Systems Department*
*EURECOM*
Biot, France
raymond.knopp@eurecom.fr

Antonio Garcia
*Accelleran NV*
Antwerp, Belgium
antonio.garcia@accelleran.com

Cristina Avellan
*iTEAM Research Institute*
*Universitat Politecnica de Valencia*
Valencia, Spain
criavcar@iteam.upv.es

*Abstract*—**Professional wireless audio equipment such as microphones and in-ear-monitoring systems are popular among live musicians and performers due to their beneficial flexibility and mobility. At the same time, artists have very high technical requirements for these devices e.g., regarding transmission delay, reliability, and battery runtime. In addition, cost, efficient use of scarcely available spectrum, and related scaling factors play a major role in professional audio productions. As of today, the entirety of technical requirements under given constraints can only be met with highly optimized custom RF technology. The new generation of cellular technology 5G is targeted to deliver new ultra-reliable low latency communication (URLLC) use cases similar to forementioned audio scenarios. Within the H2020 project 5G-RECORDS, a consortium of relevant actors from the media and mobile industry are exploring and evaluating 5G technology and related eco systems in the context of professional audio productions. Key element of the consortiums holistic approach of answering the question if requirements can be met and understanding relevant economic trade-offs, is the setup and optimization of a disaggregated 5G testbed. This work describes the testbed in detail and presents first results of the technical evaluation.**

*Keywords—Audio, 5G, URLLC, Latency, Jitter*

## I. INTRODUCTION

Overall, the potential value proposition of 5G technology for professional audio use cases is to be framed in a context where the main driving force is the migration path of content producers to (wireless) IP-workflows. This general trend towards IP-based production is motivated by IP being an established, flexible, and robust interfacing layer to local or wide area networks and software processing systems. The trend originated in the success of IP-based protocols initially designed for local area network deployments (e.g., [1]). Some fundamental differences between local area and wide area IP networks, for example the absence of support for multi-/broadcast IP packets, have created a challenging standardization process in the media industry for potential IP-based interfacing protocols with wide area support. Although this process is still ongoing, some custom deployments have already shown the potential of remote IP productions.

5G natively inherits compatibility to the IP interfacing layer. As such, the use of 5G technology could lower the barriers of implementing native support for IP-based services into media devices. In that way, 5G could act as a catalyst for professional media equipment manufacturers to explore and establish new service-based business opportunities.

The use of cellular technology for live contribution has been, up until now, limited to best-effort, not latency critical use cases and focused on the benefit of the nation- or world-wide availability of mobile internet access. In the future, some concepts introduced with the fifth cellular generation (e.g., URLLC, TSN, Slicing) could potentially allow the new implementation of use cases that rely on quality of service and local deployments, such as professional live audio production.

On the road towards potential productive deployments of 5G technology in professional content production many technical requirements and business aspects are still to be evaluated. A previous work has shown that a hand-optimized 5G Rel. 15 URLLC non-standalone (NSA) implementation might meet a subset of technical requirements for a single professional audio device in a very controlled and simplified lab environment [2]. The European Union's Horizon 2020 project 5G-RECORDS is set to explore a larger subset of technical requirements as well as the scaling of the 5G technology and potential necessary trade-offs for example between latency and spectral efficiency [3].

Following sections describe the considered professional live audio use case, related technical requirements, the deployed and optimized disaggregated 5G testbed, and first technical measurement results.

## II. USE CASE

In a typical live audio production, such as a concert, musical, theatre or studio performance, one or several artists are performing (i.e., acting, dancing, singing, or playing music instruments) live either to create content that can be used later or to entertain an interested audience that can be live on site or follow the content live via stream. Capturing and producing a live event for subsequent use of the cultural and creative content, involves many wireless audio devices. For instance, artists on stage use wireless microphones to capture their voices or instruments' sound while hearing themselves via a wireless in-ear-monitoring (IEM) system.

Consequently, content capture is expected to take place at the highest quality possible, with producers taking steps to ensure the integrity and robustness of content capture and delivery. For these reasons, the quality and reliability of the radio links are fundamental to professional wireless audio device users. For live audio productions especially, the commercial pressure on operators is significant as there is no opportunity for recovery. It is not possible to ask for a repetition during a live concert, so the tolerance for quality of
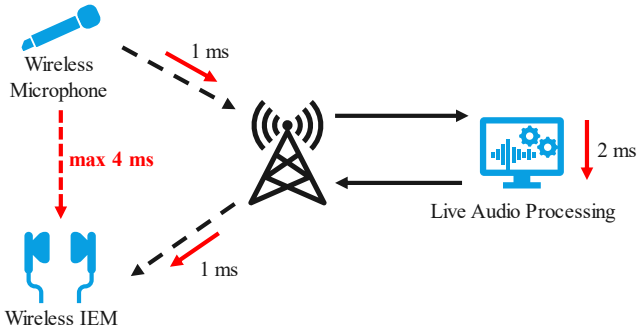
Figure 1: Latency critical self-feedback loop in professional live audio production scenarios

service (QoS) is extremely low. Furthermore, many artists rely on receiving a personalized audio mix of the event streamed back to their IEM device. This self-feedback loop (see Figure 1) has very strict latency requirements from microphone to IEM which is mainly determined as double the transmission latency for up- and downlink (UL/DL) and the audio processing time e.g., mixing and filtering. It is a critical parameter, because above a certain threshold, the artist will no longer be able to perform. Depending on the sort of music, the instrument, and the musician's skills this threshold varies widely, but in most cases, it must not exceed 4 ms. Mixing and filtering of audio signals can take up to 2 ms, leaving only about 1 ms for a wireless transmission in one direction.

## III. REQUIREMENTS AND KEY PERFORMANCE INDICATORS

In addition to forementioned latency requirements, professional content production requires specific QoS and timing parameters. The main requirements for a professional live audio production scenario are listed in Table 1. Most requirements are described in detail in [4]. Some meaningful additions are given in this section.

TABLE I.        REQUIREMENTS AND KPIS FOR A PROFESSIONAL LIVE AUDIO PRODUCTION SCENARIO

| KPI | Use Case Requirement |
| --- | --- |
| Transmission Latency ($d_t$) | < 1 ms |
| Spectral Efficiency (estimated on basis of today's professional wireless narrow-band audio solutions) | > 1.0 audio data bit / s / Hz |
| Synchronicity of Media Clocks | < 500 ns |
| Packet Error Ratio | < $10^{-6}$ |
| Number of Active Devices | 20 mics + 10 IEMs |
| Audio Data Rate per Mono Stream (uncompressed, IEMs are typically stereo) | 1.5 Mbit/s |
| Service Area | 30 m$^2$ |
| Device Speed | 5 km/h |
| Battery duration while transmitting | > 8 hours |

Transfer of audio information through a digital packet-based transmission system such as 5G requires packaging of sampled audio. In professional audio devices sampling is typically done with a fixed rate of 48 kHz at 24-bit resolution. The number of samples bundled together periodically to create IP-packets for transmission is typically not underlying any constraints by media devices and can for example be optimized for the transmission system. At the same time, packaging of multiple samples forces a trade-off between added latency and necessary network overhead. Smaller packages result in less delay for audio sample collection, more frequent IP-packets and therefore more overhead e.g., for

addressing, timestamps and metadata. Larger blocks lead to more latency for sample collection, fewer IP-packets and less overhead. The sample collection delay can also be described as the audio packet periodicity $d_p$, defined as the time between two consecutive audio IP-packets being handed over for wireless transmission. The delay for audio sample collection must be considered in the applications latency budget. Therefore, we define the overall transmission latency $d_t$ as the sum of packet periodicity and 5G end-to-end latency $d_{5G}$:

$$d_t = d_p + d_{5G} \qquad (1)$$

Since the availability of spectrum is today a limiting factor in larger events, the spectral efficiency of audio devices with concurrent and continuous low latency data streams is especially challenging. A potential 5G-based solution should provide similar efficiency as today's custom RF solutions.

## IV. DISAGGREGATED 5G TESTBED

One focus of the live audio production use case within the EU's Horizon 2020 project 5G-RECORDS is the exploration and evaluation of the technical performance of a state-of-the-art 5G system with the goal of assessing the suitability of 5G as a technology for professional live audio productions. Apart from the performance, many other aspects e.g., chip availability or regulatory requirements will play an important role in an overall assessment. This work focusses solely on parts of the technical evaluation.

The experimental 5G testbed used in the context of this work is located at the EURECOM research institute in Sophia Antipolis, France. The testbed incorporates a disaggregated approach and is based around a real-time cloud infrastructure framework. The implementation is 3GPP Rel. 15 compliant and operates in Standalone (SA) mode, with a sub-carrier spacing (SCS) of 30 kHz and a bandwidth of 20 MHz in frequency band n78 at 3.4 GHz. Figure 2 gives an overview on the full setup including the application.

Goal of the setup is to emulate the full self-feedback loop an artist would need to perform on a live stage. While we integrated the microphone and IEM functionalities in this initial setup in a single 5G user equipment (UE) device, splitting the functions into separate UEs would be needed in some real setups to give artists additional degrees of freedom. While one goal of the 5G-RECORDS project is to explore the scaling with multiple audio UEs and its implications e.g., for scheduler, latency, and spectral efficiency, this initial testbed supports only a single UE.

The 5G UE we built consists of two separate units. First, an audio network interface device to connect the analogue audio world and IP-based communication. The implementation of this device is FPGA-based and allows deterministic processing of audio and IP-packets. That way, the interface device allows IP-packet generation with high-precision pacing and timestamping, as well as full logging of received IP-packet latencies, jitter, packet-loss and packet-reordering. A more detailed description of the audio network interface device can be found in [2] and [5], where it was already used for high-precision measurements of IP-network transmissions. The interface device is connected via Ethernet to a 5G Rel. 15 modem. This second unit consists of a small Linux PC combined with a commercially available Quectel RM500Q-GL module.
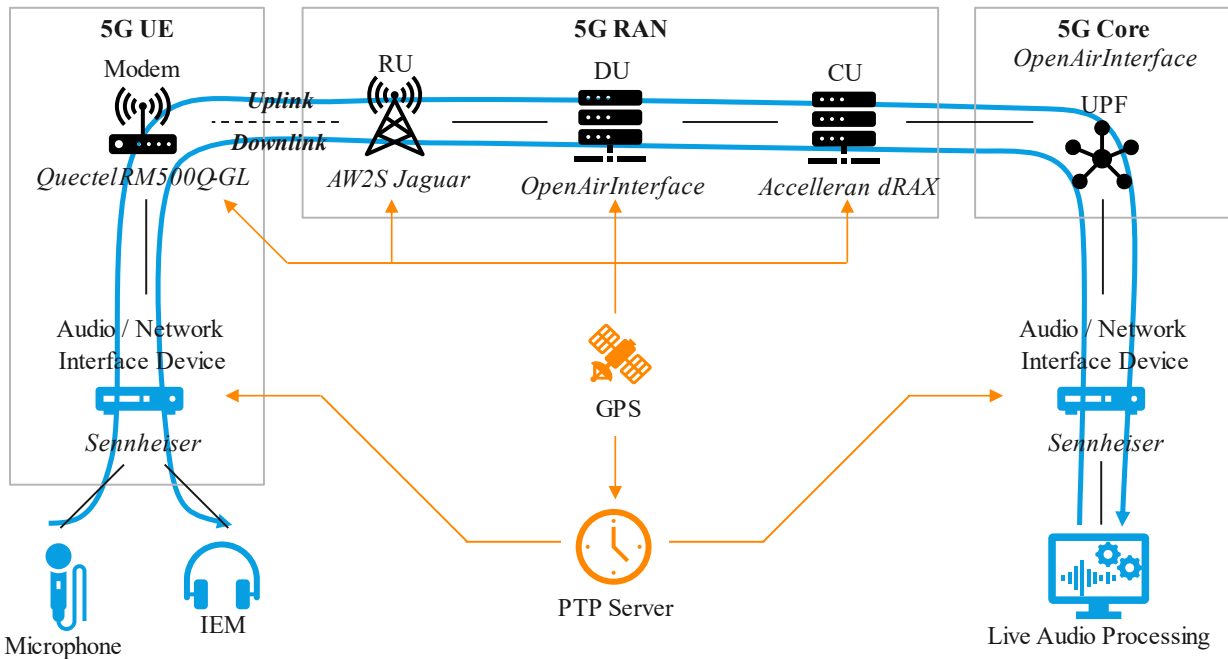
Figure 2: Disaggregated 5G Rel. 15 testbed for evaluation of technical performances in the context of professional live audio production

The modem is connected over the air to the testbeds radio access network (RAN) with a commercially available AW2S Jaguar radio unit (RU), an OpenAirInterface-based distributed unit (DU) and a centralized unit (CU) which is part of the Open RAN compatible [6] dRAX implementation by Accelleran. The 5G Core is also based on the available OpenAirInterface implementation [7]. Connected to the user plane function (UPF) is a second audio network interface device to finally link the live audio processing function to microphone and IEM. This second interface device incorporates the same functionalities as the one connected to the modem. The pair of interface devices allows precise measurement of IP-packets going through the 5G testbed in uplink and downlink direction.

The basis for the high-quality and low-latency processing of audio information is the synchronization of media clocks in all audio devices. Synchronization of the audio devices is achieved in the testbed with a dedicated Ethernet connection of the audio devices to an IEEE Precision Time Protocol (PTP) 1588-2008 server. In a later stage of the project, it is also planned to evaluate the distribution of time information via PTP protocol over the 5G system, to ultimately remove the current constraint of requiring a wired connection in addition to the wireless 5G connection. PTP time is also used in the audio devices to pace, timestamp and log transmitted and received IP-packets.

5G modem and RAN are using GPS as a basis to synchronize radio timing and related packet processing. The PTP server connected to the audio devices in our setup also uses GPS time as a reference. This has some specific benefits when optimizing the interfaces and related timing grids between audio and 5G components. It needs to be pointed out that the UPF is not using synchronized processing in this testbed's implementation, leading to some unfavorable timing characteristics in downlink direction.

## V. APPROACH

This use case has challenging requirements, especially transmission latency and packet error ratio, and many of the KPIs are intertwined and force complex trade-offs making it even more challenging to meet the full set of requirements within the 3GPP standard [8]. In addition to finding the theoretical operation points within the standard, the implementation of 5G components capable of meeting targeted KPIs is equally challenging.

In this work we focus on capturing the state-of-the-art of open-source or commercially available 5G component implementations with the goal of understanding practical challenges, trade-offs and to identify potential needs for further optimizations. Since the current testbed is stationary with radio channel characteristics that are not realistic for our use case and thus not allowing meaningful conclusions with respect to reliability, we focus **only** on the **transmission latency** of a **single audio UE**.

The testbed mandates a limitation due to the forementioned fact that the UPF processing is currently not synchronized to GPS. For the time being, this excludes the latency optimization of the communication between audio device and UPF in downlink direction. Hence, we **only** present measurements and analysis for the optimized **uplink** direction. It is assumed that processing of the UPF can be synchronized. Therefore, we expect, that our results and conclusions also applicable for downlink direction in the future.

## VI. LATENCY ANALYSIS

Media capturing devices such as microphones use media clocks to control sampling of analogue information for digital transport and processing. Playback devices such as IEMs use periodic media clocks to pace retransformation of media data back to analogue signals. In professional systems these media clocks are typically synchronized e.g., to avoid quality-reducing resampling. For network transport multiple media samples are often bundled together. The creation of such packets is typically related to the media clocks in the sense that a fixed number of samples are bundled into one packet. A professional audio system that works with a 48 kHz media clock could for example constantly pack together 48 samples
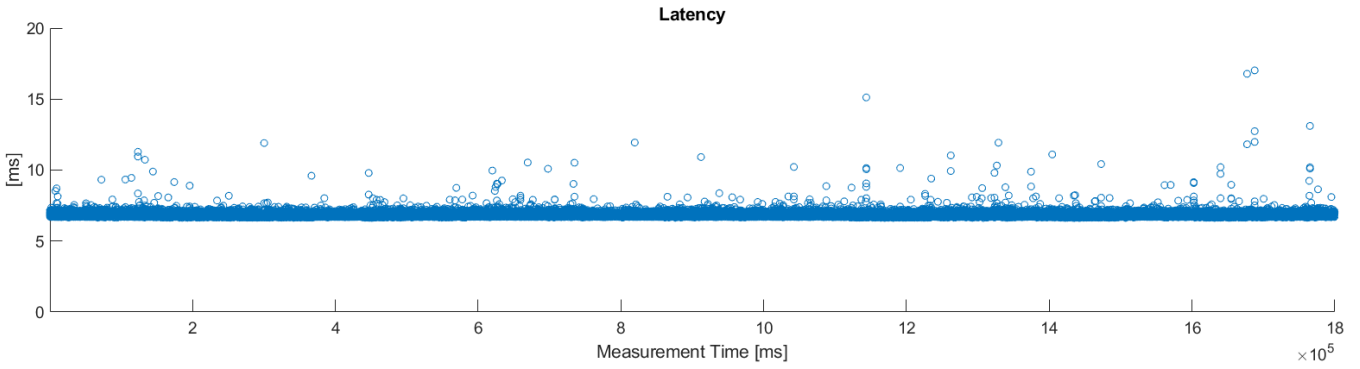
Figure 3: 5G end-to-end uplink latency, 5 ms 5G DL/UL periodicity, 5 ms audio packet periodicity

resulting in a packet ready for transmission every millisecond (see also section III). As such, creation of media packets for transport follows a specific frequency / period and phase.

Digital wireless transmission systems, such as 5G, often work with a fixed synchronized timing grid to control and optimize communication between peers. In 5G this timing grid is dictated by the base-station to which UEs synchronize themselves. The time reference for this grid is typically GPS. Within this timing grid, participating devices have periodic opportunities to transmit or receive data, also following a specific frequency / period and phase.

Having the demanding requirements of professional live audio productions in mind, it stands to reason that understanding the relation between audio and 5G timing grid is of highest importance in order to subsequently parameterize and configure both system in an optimal way.

*A. Initial testbed*

The initial 5G timing grid configuration was based on a repeating 10-slot frame. The ten slots were pre-scheduled in a *DDDDDDDXUU* pattern, where *D* represents an opportunity for a downlink transmission, *U* can be used for uplink transmissions, and *X* can be one or the other. With a 30 kHz SCS each slot has a length of 500 µs, resulting in a downlink / uplink periodicity of 5 ms.

*1) Identical periodicity of audio packet creation and uplink transmission opportunity*

For a first measurement we configured the packing of audio samples in the microphone function to the 5G downlink / uplink periodicity of 5 ms, resulting in IP-packets with 240 audio samples each.

Figure 3 shows the 5G end-to-end latency of every audio IP-packet sent from microphone to live audio processing for

30 minutes. Minimum observed latency is about 7.5 ms. According to (1) this results in a minimum transmission latency of about 12.5 ms. In general, the minimum latency is not relevant for a live media streaming application. Instead, the majority of packets is required to be within the latency budget. Here, the term majority has to be understood in relation to the required reliability as late packets are considered lost. Hence, in professional live audio productions the transmission latency of at least 99.9999% of all packets is relevant. Real packet loss has to be taken into account. Still, reflecting on the theoretical smallest latency helps to understand the structural mechanisms to identify room for optimizations.

Figure 4 depicts the timing grids of the audio and the 5G system. For simplicity, all jitter and processing delays are assumed to be zero. In this theoretical example four audio samples are periodically combined to one packet, which is handed over to the 5G system, transmitted and received. The latency from sampling in the sender to playback in the receiver
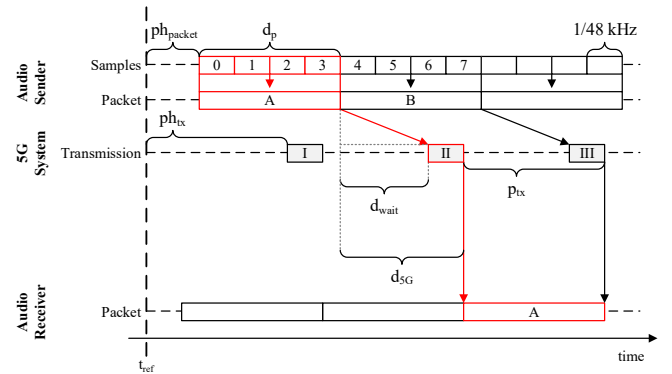


Figure 4: Timing grid with identical periodicity of audio packet creation and transmission opportunity
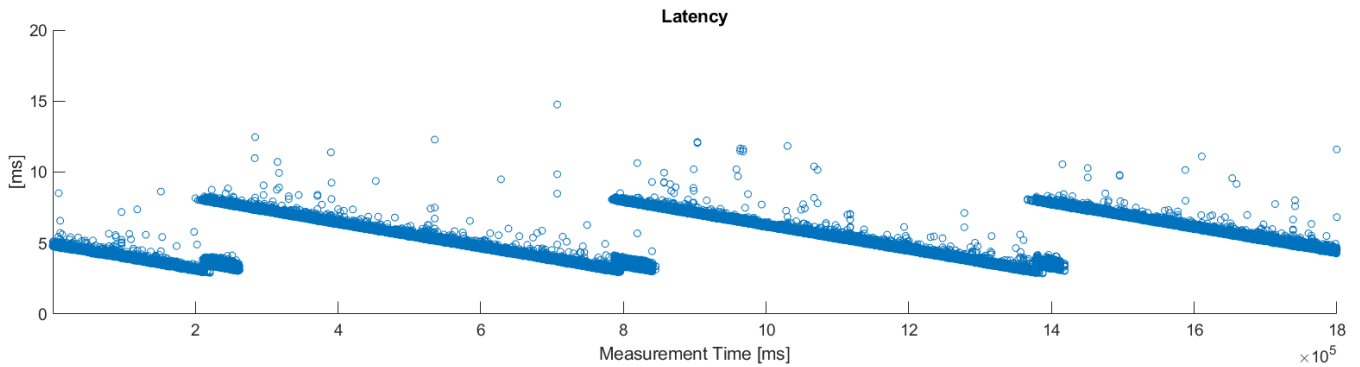


Figure 5: 5G end-to-end uplink latency, 5 ms 5G DL/UL periodicity, 5 ms audio packet periodicity, drifting timing grids
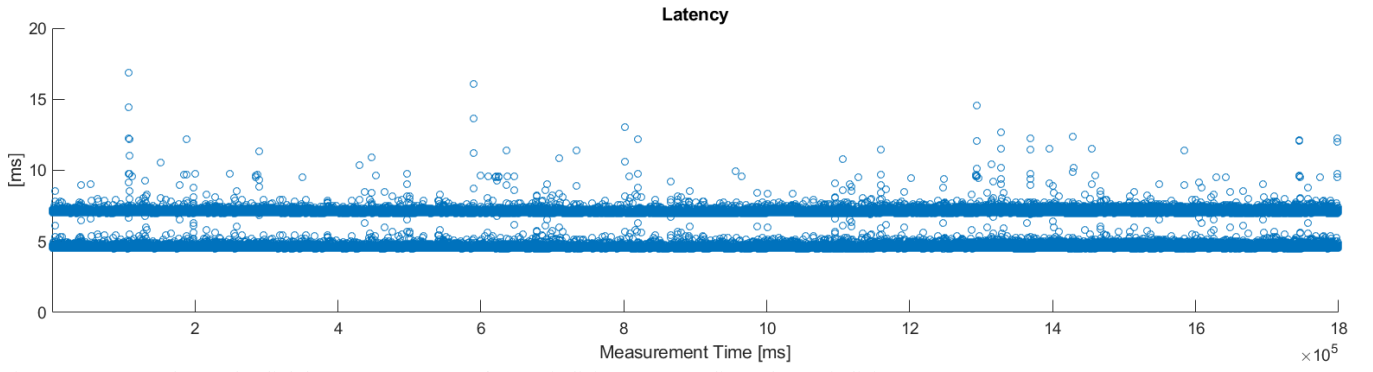
Figure 6: 5G end-to-end uplink latency, 5 ms 5G DL/UL periodicity, 2.5 ms audio packet periodicity

is calculated with (1). The 5G end-to-end latency $d_{5G}$ contains a buffering time $d_{wait}$ that each packet has to wait for the next transmission opportunity that depends on the phase difference between audio and 5G timing grid, and can range between 0 and the periodicity of transmission opportunities $p_{tx}$:

$$d_{wait} = ph_{tx} - ph_{packet} \qquad [0...p_{tx}] \qquad (2)$$

In this theoretical consideration, the phase difference is constant over time if audio system and 5G system use the same time reference, and can be minimized by aligning the packet and transmission grid. In the real measurement shown in Figure 3 the exact phase difference of both timing grids is unknown and significantly influenced by processing delays and jitter in the 5G system.

The effect of grid phase difference can be made visible when removing the time synchronization between both systems. This can be achieved by disconnecting the PTP server from GPS. Media clocks and 5G timing are then based on independent grids that drift past one another. Figure 5 illustrates this effect. The base-line latency is no longer a fixed horizontal line, but changes with the drifting clocks. The minimum latency here ranges from about 3 ms to about 8 ms. From this observation it can be concluded that $d_{5G}$ in this setup could be optimized down to theoretical lower limit of 3 ms by aligning the timing grids. Still, the minimum transmission latency would be 8 ms, including the IP-packet periodicity.

### 2) Reduced audio packet periodicity

In a next step, we reduced the audio packet periodicity to 2.5 ms, half the 5G DL/UL periodicity of 5 ms. As a result, the audio sender generated twice as many IP-packets. 5G end-to-end latency measurement with this configuration is shown in Figure 6.

Here, measured latencies are distributed into two distinct groups, around 4.5 ms and 7.5 ms. To understand the behavior, it is again useful to examine the timing grids of the systems, which are shown in Figure 7. In this theoretical example, each two audio samples a packet is generated and handed over to the 5G system. Since the periodic constant in the 5G system has not changed, half the packets now have to wait for a significantly shorter time for a transmission opportunity due to the grids phase relation. Unfortunately, this is of no benefit for the application. To assure the correct order and pace of sample playback, the faster packets have to be buffered in the receiver now for the full length of a packet. Although, the transmission latency is now significantly reduced for half of the packets, the application latency does not benefit from this circumstance in any way. This illustrates why not the fastest

packets are important in a media streaming application, but the slowest.

Based on this analysis it is evident that the internal 5G system DL/UL periodicity can play a major role in the latency of a live streaming application. Furthermore, it can be concluded that in order to achieve a transmission latency of 1 ms, as the use case requires, the 5G DL/UL periodicity has to be smaller than 1 ms.

### B. Optimization of the testbed

On the road to a transmission latency of 1 ms, we reduced the 5G DL/UL periodicity in the testbed's implementation to a 5-slot frame with a prescheduled *DDXUU* pattern of 2.5 ms length, and configured the audio packet creation periodicity to the same value. Figure 8 show the results of an exemplary measurement with this setup.

Until now our analysis omitted the latency jitter in the 5G system. Looking at the jitter is mandatory for a realistic evaluation of the suitability of 5G for professional live audio productions. Useful for this analysis is considering the cumulative distribution function (CDF), see Figure 8. As explained before, not the fastest packets are of interest in this use case, but at least 99.9999% of all packets. Already a few late packets can shift the operation point significantly up. The CDF shows the respective marker at ~23 ms. With an audio packet periodicity of 2.5 ms we can calculate the transmission latency with (1) to ~25.5 ms.

### VII. CONCLUSIONS

Compared to a previous research demonstrator where a 5G DL/UL periodicity of 500 µs and a transmission latency of ~3 ms was reached [2], this work achieved a periodicity pattern of 2.5 ms and a transmission latency of ~25.5 ms with open-source and commercial components. This shows that available 5G components have not yet caught up with research
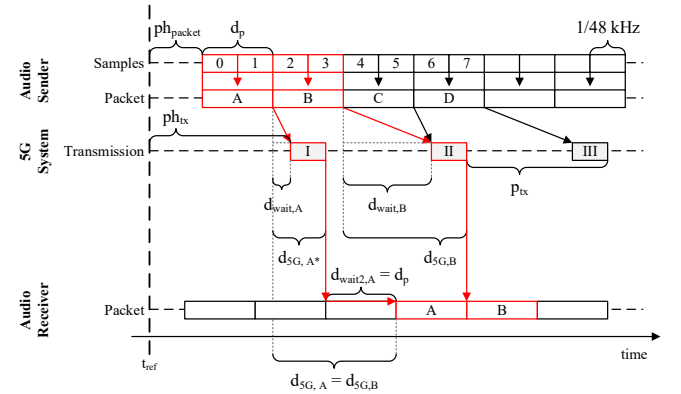


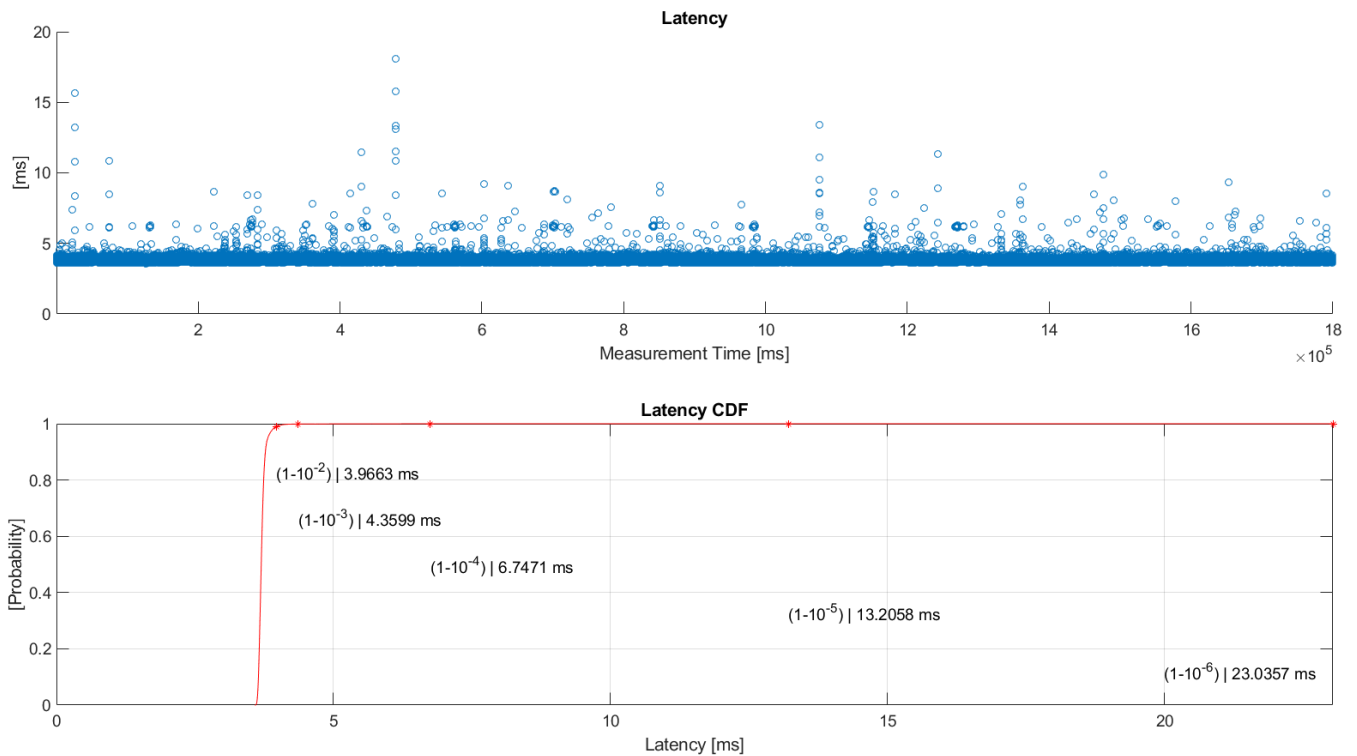Figure 7: Timing grid with half audio packet creation periodicity

Figure 8: 5G end-to-end uplink latency and CDF, 2.5 ms 5G DL/UL periodicity, 2.5 ms audio packet periodicity

in this aspect. Although the DL/UL periodicity pattern is of significant importance in ultimately reaching the goal of promised ultra-low latencies in 5G, the delay in this works testbed is currently mainly defined by general jitter in the 5G system. Consistent use of real-time paradigms in all 5G components is required for the implementation of URLLC use cases.

The transmission latency remains a critical challenge in 5G. Future work also needs to address:

- Concepts to deliver low latency performance to more than one UE

- Reliability, efficiency (spectrum, power), and related trade-offs

- Distribution of time information for media clock synchronization over-the-air

Overall, significant effort is still needed to potentially deliver the full set of requirements of professional live audio production scenarios with 5G technology.

REFERENCES

[1] Audio Engineering Society, "AES67-2018: AES standard for audio applications of networks - High-performance streaming audio-over-IP interoperability," 2018.

[2] S. Hämäläinen and J. Dürre, "Low Latency 5G for Professional Audio Transmission," 2021. [Online]. Available: https://www.bell-labs.com/institute/white-papers/low-latency-5g-professional-audio-transmission/. [Accessed 19 April 2022].

[3] "5G Key Technology Enablers for Emerging Media Content Production Services (5G-RECORDS)," [Online]. Available: https://www.5g-records.eu/. [Accessed 20 April 2022].

[4] 3GPP TR 22.827, "Study on Audio-Visual Service Production," 2019.

[5] R. Hupke, J. Dürre, N. Werner and J. Peissig, "Latency and Quality-of-Experience Analysis of a Networked Music Performance Framework for Realistic Interaction," in *152nd Audio Engineering Society Convention*, Den Haag, 2022.

[6] O-RAN Alliance, "O-RAN: Towards an Open and Smart RAN," [Online]. Available: https://www.o-ran.org/resources. [Accessed 19 April 2020].

[7] F. Kaltenberger, G. Souza, R. Knopp and H. Wang, "The OpenAirInterface 5G New Radio Implementation: Current Status and Roadmap," in *23rd Int. ITG Workshop on Smart Antennas*, 2019.

[8] M. Bennis, M. Debbah and H. V. Poor, "Ultra-Reliable and Low-Latency Wireless Communication: Tail, Risk and Scale," in *Proceedings of the IEEE*, 2018.