2021 17th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)

Performance Modelling and Assessment for Social VR Conference Applications in 5G Radio Networks

João Morais
Radio Systems Group
Instituto Superior Técnico
Lisbon, Portugal
joao.de.morais@tecnico.pt

Sjors Braam
Department of Networks
TNO
The Hague, The Netherlands
sjors.braam@tno.nl

Remco Litjens
Department of Networks
TNO
The Hague, The Netherlands
Faculty of EEMCS
TU Delft
Delft, The Netherlands
remco.litjens@tno.nl

Sandra Kizhakkekundil
Faculty of EEMCS
TU Delft
Delft, The Netherlands
s.kizhakkekundil@student.tudelft.nl

Hans van den Berg
Department of Networks
TNO
The Hague, The Netherlands
Faculty of EEMCS
University of Twente
Enschede, The Netherlands
j.l.vandenberg@tno.nl

***Abstract*— One of the most challenging applications targeted by evolving (beyond-)5G technology is virtual reality (VR). Particularly, 'Social VR' applications provide a fully immersive experience and sense of togetherness to users residing at different locations. To support such applications the network must deal with huge traffic demands, while keeping end-to-end latencies low. Moreover, the radio access network must deal with the volatility and vulnerability of mmWave radio channels, where even small movements of the users may have substantial effects on the Quality of Experience. We present an integral modelling framework for feasibility assessment and performance optimization of the radio access network for Social VR applications in indoor office scenarios. Using the presented modelling approach, we conduct an extensive simulation-based assessment to determine the performance impact of head motion, the frequency band (3.5 GHz, 26 GHz) and radio network configurations, and derive the required carrier bandwidth for a range of 'Social VR' scenarios. Insights into these issues are a prerequisite for setting up guidelines for network deployment and configuration as well as for the development of (AI/ML-based) methods for dynamic resource management to optimally support Social VR applications.**

Keywords—Social XR, VR, 5G, modelling, performance assessment

I. Introduction

It is widely recognized that emerging 5G network technology [1] will boost the development of new, highly innovative applications in virtually all domains. One of the most challenging application classes targeted by (beyond-)5G networks is augmented/virtual reality (AR/VR), in particular scenarios with multiple distant users who are able to interact fluently with each other through all human senses [2][3]. Transporting one's social and functional self to any place on earth is an exciting idea that will save travelling time and costs (and reduce carbon footprint) by e.g. enabling virtual meetings, exploitation of remote expertise or skills in smart industry or other contexts in the form of collaborative working, remote inspection or maintenance, and remote education and training.

To support such so-called 'Social VR' applications with compelling visual, haptic, audio and possibly even olfactory experiences for remote users, the networking infrastructure must be able to handle extremely high-bandwidth streams while keeping the end-to-end latency low [3]-[5]. Furthermore, the infrastructure should provide powerful (in-network) processing capabilities, reducing the need for heavy computational capabilities at the client devices, thus increasing the clients' flexibility and mobility. Actually, natural Social VR experiences can only be realized when all components in the end-to-end capture, transmission and rendering pipeline optimally work together [5]-[8]. Optimization of the Quality of Experience (QoE) requires dynamic orchestration of these components, allocation of resources in the networking and compute infrastructure, and an optimized exploitation of the connect-compute trade-off, noting that the end-to-end latency budget can be flexibly distributed over computational and transport tasks. In particular, dynamic orchestration and resource management complemented with e.g. multi-connectivity should also be able to cope with the intermittent nature of high-frequency (mmWave) radio channels used in 5G to satisfy the huge capacity and throughput requirements [5][9][10].

Other (beyond-)5G Radio Access Network (RAN) features and capabilities to support extremely demanding wireless applications like Social VR, are the use of massive MIMO-based beamforming to overcome the associated attenuation challenges, the flexible OFDM numerology, the use of scheduling mini-slots and self-contained subframes, and the use of channel-adaptive latency-based packet schedulers to efficiently achieve high data rates while satisfying latency requirements.

This paper focuses on radio access challenges for supporting Social VR applications. As a first key contribution, we present an *integral modelling framework* covering all relevant aspects needed for an adequate feasibility assessment and performance optimization of 'Social VR' applications in indoor office scenarios. Studies presented in literature mostly focus on individual aspects, particularly blockage of mmWave channels and possible ways to mitigate such blockage effects [10][11]. Comprehensive studies of VR performance in the RAN covering multiple relevant aspects (and their dependencies) simultaneously are lacking, while there is a clear need for such an integral view. The modelling aspects covered in our study, besides the specifics of the use case scenario (office layout), include: *(i)* user behavior; *(ii)* application's traffic characteristics and service requirements; (*iii*) the radio network deployment, incl. the deployed antennas, the carrier assignment and the configuration of the Grid of Beams (GoB); and (*iv*) key 5G radio resource management mechanisms, incl. MU-MIMO/latency-based packet scheduling. Exploiting the integrated implementation of the proposed models in a simulation tool, our second key contribution is an extensive *scenario-based performance assessment*, conducted to determine the impact of several use case/application-level aspects, the employed frequency band (3.5 GHz, 26 GHz) and radio network configurations on the

978-1-6654-2854-5/21/$31.00 ©2021 IEEE

attainable performance and the resources required to support a given Social VR scenario with adequate service quality.

Such insights are particularly useful, a prerequisite even, for setting up guidelines for local network deployment as well as for the development of (potentially AI/ML-based) methods for dynamic resource management and tuning of radio access parameters to best support Social VR applications in indoor office scenarios. The envisioned ultimate goal is then to use these results for the design of a fully automated 'Social VR network slice', i.e. a self-configuring and autonomously managed 5G (radio) network slice to support this type of Social VR applications.

The remainder of the paper is organized as follows. First, in Section II an overview of related work is given. Next, in Section III, we give a high-level description of the addressed 'Social VR' use case. Section IV then presents our modelling framework for the performance assessment and optimization of Social VR scenarios, covering user behavior, application aspects, network deployment and radio resource management. Next, in Section V we outline the different assessment scenarios and define the Key Performance Indicators (KPIs). Section VI then presents and analyses the simulation results obtained for the specified scenarios. Finally, in Section VII we summarize our key contributions and insights, and give an outlook for further research.

II. Related work

There is a large body of research addressing virtual reality systems in a wireless setting. In [12], a framework is presented that analyses the performance of VR services over wireless networks. The framework captures the tracking accuracy, transmission delay, and processing delay, but most radio characteristics such as frequency-selective fading, antenna configurations and blockage effects are not considered. In [13], the authors study the impact of blockage by hand, head and body on wireless mmWave links and suggest an algorithm to overcome the corresponding challenges. The proposed solution uses a fixed relay to increase robustness against blocking and is assessed in an experimental setup. The attainable gains strongly depend on numerous assumptions and deployment configurations which are not described in any detail. In [10], the challenges and enablers for ultra-reliable and low-latency VR are discussed. The authors state that for a truly immersive VR user experience, three aspects should be designed and orchestrated together: communication, computing and caching. A case study has been worked out where multi-connectivity is used to mitigate the blockage and disturbance caused by several impulse actions. Although the results of this study are quite extensive, a lot of modelling aspects and choices are not covered. The authors of [9] propose a new latency-based MAC scheduling approach to optimize VR applications over 5G radio networks.

Recently, 3GPP also showed interest in supporting Extended Reality (XR) over 5G networks. An XR 360° conference meeting has been defined in [4]. However, QoS requirements for such use cases are not clearly defined, likely because the QoE assumptions vary so tremendously. 3GPP considers 50-100 Mbps [4], while other sources mention bit rates in the order of Gbps [5][10]. Despite lacking concreteness, there is more of an agreement on latency requirements, in the sense that latency should be lower than for current real-time applications.

To the best of our knowledge, there is no study available in literature providing a comprehensive and integral framework of all relevant modelling aspects for the considered type of scenario. This paper aims to provide a modelling framework and an extensive simulation study based on this framework, enabling a variety of sensitivity analyses, derivations of deployment guidelines and the development and assessment of radio network management solutions.

III. Use case

We consider a use case of running a Social VR application in an indoor office scenario. More specifically, we consider a square meeting room with a large round conference table placed in the center of the room (see Figure 1). Uniformly spread around the conference table are $N = N_P + N_V$ chairs, seating in this example $N_P = 4$ participants that are physically present in the room, and $N_V = 4$ participants located elsewhere and virtually attending the conference meeting. The physical/virtual participants are assumed to be seated in an alternate fashion (P – V – P – …).

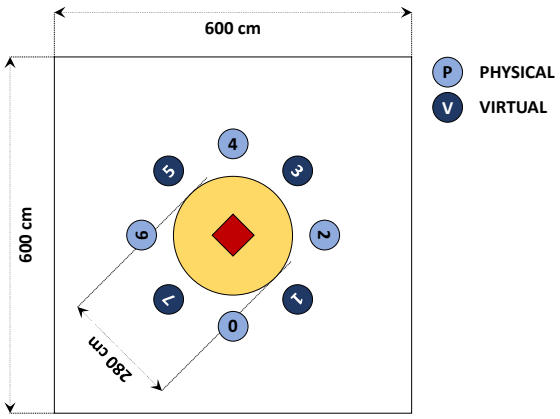

Figure 1: Visualization of the 'Social VR' application scenario with eight physically or virtually present meeting participants.

Each physically present participant wears a VR headset or Head-Mounted Display (HMD) towards which video images are wirelessly transferred via an integrated User Equipment (UE). These tailored images are transferred by the Multi-point Control Unit (MCU) [7] residing in the media processing cloud. Two on-table video capture devices are installed and pointed to the participant to capture images for wireline upstream transfer (see Figure 2).

IV. Modelling

In this section we present the integrated modelling framework, covering user behavior, the VR application, the radio network deployment and key radio resource management aspects.

A. User behavior

Each physically present participant is and remains seated in his chair, but the participant's head is modelled to exhibit some degree of randomized *head motion* in terms of changes in position and orientation [15][16]. In general, the relevance of modelling head motion lies therein that the antennas used for downlink reception are fixed to the VR headset and hence any realistic motion of a participant's head directly impacts the position and orientation of the receive antennas. We purposely differentiate modelling the position and orientation due to their inherently different natures, although, when put together, they should realistically emulate the head motion of

a seated participant. More precisely, the proposed model consists of:

- The change of a head's *position* can be modelled according to a modified random waypoint model in 3D. More specifically, a participant's default (or average) head position has coordinates (x_0,y_0,z_0), with (x_0,y_0) given by its appointed chair and z_0 fixed to, e.g. 1.4 m. The initial head position is sampled according to a three-dimensional Gaussian distribution with mean vector (x_0,y_0,z_0), standard deviations σ_x, σ_y, σ_z, and all covariances set to 0 m^2. From there, the head is modelled to move linearly to its next position, sampled from the same distribution, at a speed of $0.1 \times \mu$ m/s, where μ is the head motion index in the range [0,10]. Once there, the process repeats itself, ad infinitum.
- The head's *orientation* is given by a randomly sampled 3D vector of angular offsets around a straight line aimed at the current speaker (the speaker itself is assumed to aim his view on the previous speaker). Offsets are independently sampled w.r.t. the pitch, yaw and roll axes perpendicularly cutting through the head, assuming uniform angular distributions on the ranges $[-\alpha_P, \alpha_P]$, $[-\alpha_Y, \alpha_Y]$ and $[-\alpha_R, \alpha_R]$, respectively, with e.g. $\alpha_{P/Y/R} = 5^o$. Whenever the speaker changes, such offsets are sampled and the corresponding change in orientation is modelled to take $1.5 - 0.2 \times \mu$ seconds, with μ as defined above. Once the targeted orientation is reached, a new offset is sampled and pursued, and so on, until the speaker changes once again.

Aside from substantial changes in the head orientation that are due to speaker changes, the randomizations of the head position and its orientation effectively constitute some degree of 'head wobbling' in 3D, with the intensity configured by head motion index μ.

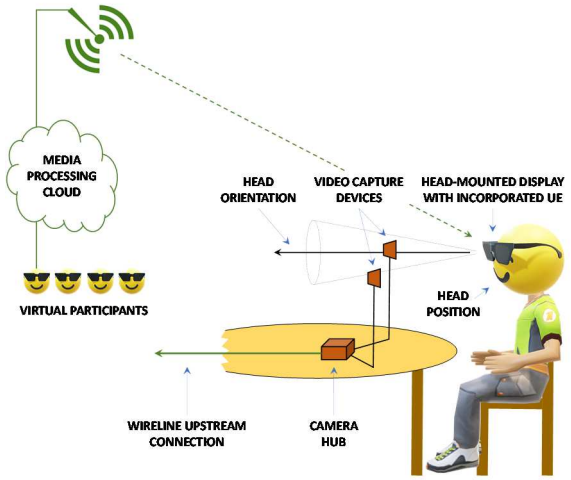

Figure 2: Illustration of user setting and end-to-end architecture.

The *meeting dynamics* are modelled by traversing through a pre-generated speaker list, letting each designated participant speak for a deterministic time of τ_S seconds, for instance 4 seconds, and then switch to the next speaker on the list. For the above-mentioned scenario with N = 8 participants, the speaker list could e.g. be given by '0 – 5 – 2 – 7 – 4 – 1 – 6 – 3', where we choose to alternate between physically and virtually present users.

B. VR application

The VR application is modelled as a persistent real-time UDP-based video streaming application [17], where each up- or downstream is a sequence of GoPs (Groups of Pictures) and each GoP comprises a single I ('intra-coded') and $M_{GoP} - 1$ P ('predicted') video frames. The size of the relatively large fresh I and the differential P frames is denoted S_I and $S_P < S_I$, respectively. A typical frame generation rate of R_F FPS (frames per second) is assumed, which is noted to apply at the source. The application-level bit rate R_B is given by

$$R_B = \frac{8\,R_F(S_I + (M_{GoP} - 1)\,S_P)}{10^6\,M_{GoP}}.$$

For an example case inspired by [18] and visualized in Figure 3, with $M_{GoP} = 6$, $S_I = 1250$ kB, $S_P = 250$ kB and $R_F = 30$ FPS, the application bit rate is 100 Mbps, which is labelled as an 'entry-level VR' quality, providing a 2048 × 2048-pixel visual field. In the downlink, the assumed fixed-rate (non-adaptive) video stream is noted to be a (potentially processed) aggregation of the distinct uplink video streams captured for the different meeting participants.

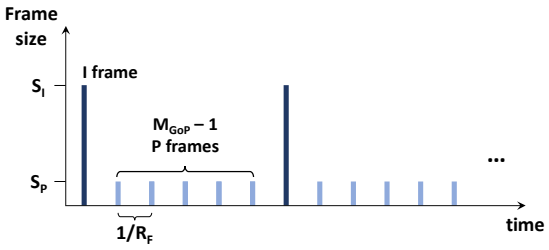

Figure 3: VR application traffic model.

For the considered indoor office scenario, the source of the downlink transfers is the above-mentioned MCU, which receives, aggregates and potentially processes all captured video streams (incl. e.g. HMD removal) before generating a downlink video stream to a given UE. Generated as application-level frames at the source (see Figure 3), the video information arrives in the form of 1500-byte IP packets at the Base Station (BS) transmission buffer. For the downlink streams, the packets are dispersed in time due to the variabilities of packet latencies on the path from source to transmit buffer. The time dispersion for the downlink streams is modelled as follows. Upon generation of a video frame, the frame is segmented into IP packets at a 'packet extraction bit rate' given by $R_B / (1 - \beta)$, with R_B the application-level bit rate as introduced above and $\beta \in [0,1]$ the configured burstiness parameter. Note that for the extreme case of $\beta = 0$ the IP packets are maximally dispersed (effectively yielding a constant bit rate flow), while for the case of $\beta \to 1$ each video frame is upon generation instantaneously segmented into IP packets, reflecting the maximum degree of burstiness. In our simulations, we assume a β of 0.5. Figure 4 shows how β influences the traffic profile. Note that β can make packets from distinct frames overlap.

A frame is generated each time the scene is fully rendered in the MCU. The most likely case for an entry-level setup is for the MCU to render the complete scene, which includes all participants, and use that information to generate the frames for each. As opposed to rendering a user-specific scene, this means the IP packets are sent simultaneously, and thus we

assume the packet arrival traffic profile represented in Figure 4 is the same for each participant. Despite the most realistic, this is a worst-case scenario for RAN performance because each participant will require resources concurrently.

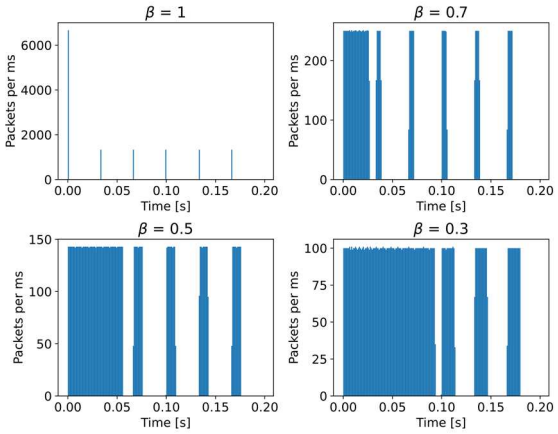

Figure 4: Packet arrival traces for scenarios with M_{GoP} = 6, R_B = 100 Mbps and different burstiness values β.

The QoS requirement of the VR application is given by a maximum tolerable end-to-end one-way frame-level latency of e.g. Δ_{E2E} = 150 ms, considering the application as 'conversational video (live streaming)' [4], where any VR frame delivered with a latency exceeding Δ_{E2E} is considered useless. In our study, we assume that the external network consumes part of this end-to-end budget, and assume the remainder as a packet-level budget for the RAN segment of $\Delta_{RAN} < \Delta_{E2E}$ ms, acknowledging that the RAN handles IP packets and is unaware of application-level video frames [18]. Distinct values Δ_{RAN} are considered, effectively addressing scenarios with distinct Δ_{E2E}. Δ_{RAN} is used by the packet scheduler to give due priority to packets approaching the associated transmission deadline and potentially even drop packets that (are estimated to) exceed the deadline.

C. Radio access network

The meeting room is adequately equipped to support social VR meetings, including, besides video capturing and playback devices, also advanced wireless networking hardware. More specifically, we assume that a single massive MIMO antenna array is mounted at the ceiling center (at a height of 3 meter) and pointed downwards (see Figure 1). Considering both 3.5 GHz and 26 GHz deployment scenarios, we assume an antenna array comprising 4 × 4 or 8 × 8, respectively, of cross-polarized AEs (Antenna Elements) described by 3GPP [19], with a half-wavelength inter-AE spacing in both dimensions (see Figure 5). No subarray structure is assumed for these antenna arrays, hence all AEs of the 32T32R and 128T128R antennas, respectively, are fed with independent RF (Radio Frequency) chains and fully digital beamforming is applied. The maximum BS transmit power is 20 dBm [20].

The BS antenna is assigned a single TDD carrier of distinct bandwidths in the 3.5 or 26 GHz band. The carrier is configured with *numerology* 2, implying a subcarrier spacing of 60 kHz and a slot duration of 0.25 ms.

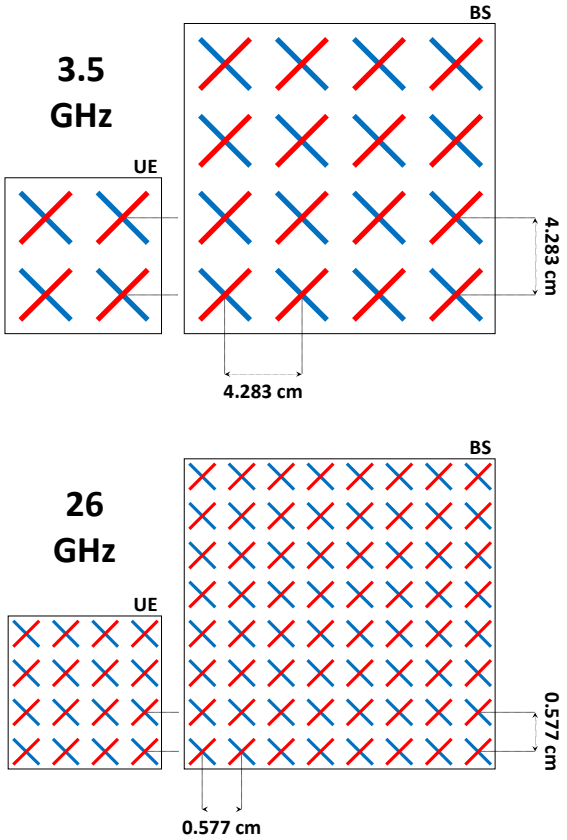

Figure 5: Assumed UE (HMD; left) and BS (right) antenna arrays for the 3.5 GHz (top) and 26 GHz (bottom) scenarios.

At the BS antenna a *Grid of Beams* (GoB) is configured to comprise a set of pre-defined transmission beams (wideband precoders) in a three-dimensional angular grid covering an angular span of 120° in both the azimuth and elevation planes. Since we assume a persistently on-going VR-aided meeting, we are not concerned with the more coarsely defined access-oriented SS/PBCH-based (Synchronisation Signal / Physical Broadcast CHannel) GoB, but rather model the traffic-oriented CSI-RS (Channel State Information Reference Signals) beams in a more finely granular GoB. Specifically, the proposed grid of CSI-RS beams has an approximate angular resolution of 12° for both the 3.5 GHz and 26 GHz scenarios. Given the angular span and the applied granularity, the GoB comprises 11 × 11 = 121 beams.

The UEs incorporated into the HMD are equipped with an antenna array comprising 2 × 2 or 4 × 4 cross-polarized AEs with half-wavelength inter-AE spacings, for the 3.5 GHz and 26 GHz scenarios, respectively (see Figure 5). A receiver noise figure of 8 dB is assumed at the UE.

The characteristics of the *propagation environment* in terms of path loss, shadowing and multipath fading, are taken from the line-of-sight 'indoor office' scenario specified by 3GPP [19] with implementations provided by Quadriga [21].

D. Radio resource management

Connected to the ceiling-mounted BS antenna, the UE selects the strongest CSI-RS beam, indicated via the CSI-RS Resource Indicator (CRI) feedback. Receive beamforming is used at the UE side in the form of Maximum Ratio Combining (MRC). The combination of a selected CSI-RS beam and a selected MRC configuration at the UE receiver, constitutes a

so-called downlink 'beam pair'. The best CSI-beam used by the BS $\boldsymbol{w}_{BS}$ is chosen from the GoB using to the expression below, which simplifies when the BS knows the UE beamformer $\boldsymbol{w}_{UE}$ is derived with MRC, i.e. $\boldsymbol{w}_{UE} = (\boldsymbol{Hw})^{\mathrm{H}}$,

$$\boldsymbol{w}_{BS} = \underset{\boldsymbol{w} \in \boldsymbol{W}_{GoB}}{\operatorname{argmax}} |\boldsymbol{w}_{UE}\boldsymbol{Hw}| = \underset{\boldsymbol{w} \in \boldsymbol{W}_{GoB}}{\operatorname{argmax}} |\boldsymbol{Hw}|$$

where H represents the Hermitian matrix, $\boldsymbol{W}_{\boldsymbol{GoB}}$ are the beamformers in our GoB, and $\boldsymbol{H}$ is the channel matrix that connects the AEs in the UE's antenna array to the BS's AEs.

The matrix $\boldsymbol{H}$ used for beam choice is outdated by a small delay of 5 TTIs due to the time past between transmission and measurement of the reference signal. And the beam pairs do not change until a new CSI update. CSI is acquired based on a configurable CSI periodicity τ_{CSI} parameter that can vary from one TTI to infinity. Other mechanisms, like scheduling, rely on the available CSI to operate.

The latency-oriented Modified-Largest Weighted Delay First (M-LWDF) packet scheduler is modelled to govern wideband-based downlink transmissions [22][23], given the resources configured by an applied TDD duplexing scheme, configured by a frame size of five time slots and a fixed 1:4 UL/DL resource split. In the DL, the slot is DL-heavy, i.e. all symbols are used for DL transmission, however we consider control signaling overhead given by 3GPP in [24] of 14% and 18% when using a 3.5 GHz and 26 GHz carrier, respectively. Head of line packet dropping is applied in the transmit buffers in case the packet's corresponding VR frame is determined not to meet the aforementioned latency budget Δ_{RAN}. At each scheduling opportunity, hence at each TTI (Transmission Time Interval; taken equal to a time slot), the M-LWDF scheduler ranks the UEs based on priority levels given by

$$P_i(t) = -\frac{\log_{10}\delta_i}{\Delta_{\mathrm{RAN}}} W_i(t) \frac{R_i(t)}{\hat{R}_i(t-1)}$$

for UE i at TTI t, where $\delta_i = 0.05$ is the maximum allowed packet drop rate of flow i, $W_i(t)$ denotes the flow's experienced head-of-line packet latency, $R_i(t)$ denotes its instantaneously attainable bit rate and $\hat{R}_i(t-1)$ is the exponentially smoothed bit rate it experienced so far. After TTI t, the $\hat{R}_i(t)$ is updated according to

$$\hat{R}_i(t) = \begin{cases} (1-\alpha)\hat{R}_i(t-1) + \alpha R_i(t) & \text{if scheduled} \\ (1-\alpha)\hat{R}_i(t-1) & \text{otherwise} \end{cases}$$

MU-MIMO-based co-scheduling of multiple UEs in the same time-frequency resources is applied. The applied criterion thereby is that two UEs can be co-scheduled if their selected CSI-RS beams are at least κ beams apart in the GoB, with $\kappa \in \{0, 1, 2, \ldots\}$ a configurable scheduling parameter. The smaller κ, the higher the applied degree of MU-MIMO, while for very large κ the MU-MIMO is effectively turned off and only single-user scheduling is used. Note that in the extreme case of $\kappa = 0$, spatial diversity or, more specifically, polarization diversity, must be applied to avoid excessive interference. The system currently operates with a dual-polarized single-layer per UE, therefore users cannot be multiplexed effectively within the same CSI-beam. In line with this, a unit value for κ was found to yield the best performance, i.e. users are always co-schedule as long as they reported a different optimum beam.

The adaptive modulation and coding scheme estimates the received Signal-to-Interference-plus-Noise-Ratio (SINR) and maps this to a selected and fed back MCS (Modulation and Coding Scheme) based on a 10% BLER (BLock Error Rate) target and link-level results available from [26].

V. Scenarios & KPIs

We have implemented all above-mentioned modelling considerations in a system-level simulator. Using this simulator we investigate how the Social VR performance and the feasibility of use case scenarios are impacted by different scenario aspects. We organize a scenario-based assessment around a baseline scenario, studying the impact of the number of meeting participants N (with $N_P = N_V$), the motion index μ, the application bit rate R_B, the RAN latency budget Δ_{RAN} and the CSI periodicity τ_{CSI}. In the default scenario we assume N = 8, $\mu = 3$, R_B = 100 Mbps, Δ_{RAN} = 10 ms and τ_{CSI} = 5 ms. Both scenarios with 3.5 GHz and 26 GHz carriers are considered with a default carrier bandwidth of 50 MHz. In different experiments, distinct model parameters are unilaterally varied around their default settings. Regarding the number of meeting participants N, we note that the room and table are dimensioned to seat up to 16 participants, hence when varying N, the average distance between users will vary accordingly, which is expected to influence the experienced interference. In two of the conducted experiments we derive for a range of use case scenarios the amount of bandwidth needed in the 3.5 GHz or 26 GHz band to support the scenario with sufficient QoS. For all other parameters, we used the examples provided in Section IV.

The considered key performance indicator is the packet loss ratio (PLR), i.e. the ratio between the number of lost packets and the total number of generated packets. For the experience in video streaming to be tolerably affected, the PLR should not exceed 5% [26]. As described in Section VI.D, packets are dropped when they cannot be successfully transmitted within the RAN latency budget. Factors that can cause packet losses are:

- Incidental overestimation of the achievable SINR may cause the selection of too high an MCS and hence transport block errors, where multiple such block errors belonging to a given packet may cause the packet delay to exceed the scheduler's dropping threshold.
- The rate at which packets are successfully transferred is momentarily lower than the rate at which packets are generated, e.g. due to poor channel conditions aforementioned block errors or simply a scarcity of transmission resources. This may lead to an accumulation of packets in the BS buffer and consequently to excessive delays and packet dropping.

In all experiments, to attain statistical significance, for each scenario we simulate twenty different meetings, by generating the channels for twenty random initializations. Each meeting is simulated for a duration of sixteen seconds, which proved sufficient for the PLRs to converge.

VI. Numerical results

A. Performance impact from user behavior

Figure 6 shows how the PLR varies for motion indices μ = 1, 3 and 5, and for 3.5 GHz and 26 GHz. The observation

that the latter (mmWave) band yields lower PLRs can be related to beamwidth. Since our mmWave antenna has more elements, its beams are narrower than the sub-6 GHz counterpart, and that can result in less interference each participant experiences due to other nearby participants.

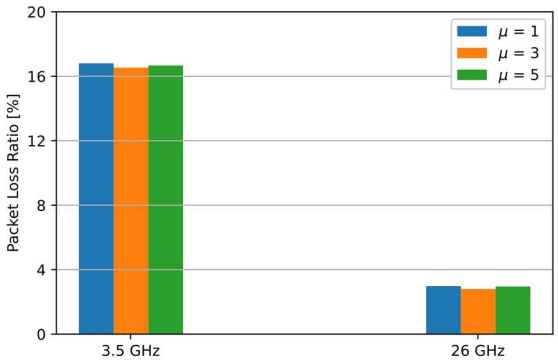

Figure 6: PLR for different motion indices μ.

The second noticeable thing is the similarity across different μ. One would expect that if users move more, that causes a higher variability of the channel, resulting in degraded performance, but it seems that head motion barely impacts performance. However, recall that every 5 ms (default value for τ_{CSI}) CSI is acquired. Now, GoB-based operation permits less frequent CSI acquisition: the optimal beam changes less often than the optimal weights to make AEs coherently interfere, as is the case with reciprocity-based beamforming. Therefore, it can simply be the case the channel is measured so often that we cannot see the effect of different motion indices, because the channel variability caused by them is within what can be accurately estimated with our rate of measurements. But further testing shows that performance for μ = 1, 3 or 5 stays practically unchanged as we increase τ_{CSI}, so that hypothesis is invalid.

Alternatively, the similarity can be justified by an insufficient variability in movement rate between the considered μ, thus hiding the impact of user behavior on performance behind rare random-nature phenomena inherent of simulations. However, recall that for μ = 5 the rate of head position translation is 5× bigger than for μ = 1. Therefore, there certainly is enough variability between motion indices already, regarding the head positions, but when we look at the orientations, we see that for μ = 5 users rotate their heads only 1.3/0.5 = 2.6 times faster than with μ = 1. On this basis, maybe head rotation has more impact than head position, and we simply need to test sufficiently different rates of rotation.

To assess that hypothesis, we compare μ = 7 to μ = 5, since the former has 5× the rotation rate of the latter, and less difference than translation speed. Table 1 shows the results. In all cases, the highest μ resulted in the highest PLR, effectively proving that change in head orientation affects performance more than change in head position. This can be explained since the channel varies little between positions close to one another, because large-scale fading is spatially correlated, a product of using a realistic channel generator. Antenna orientation, on the other hand, is directly associated with its radiation pattern, and so, aggressively changing orientations can cause immense signal oscillations.

Table 1: PLR for a given motion index μ, frequency and three distinct CSI periodicities.

CSI periodicity	3.5 GHz		26 GHz	
	μ = 5	μ = 7	μ = 5	μ = 7
5 ms	16.7%	17.3%	3.0%	4.2%
10 ms	16.9%	18.1%	3.3%	4.7%
20 ms	17.4%	19.9%	4.1%	6.9%

Regarding the CSI periodicity, note that as τ_{CSI} increases, the difference between speeds becomes more pronounced, thus supporting the theory that we could not accurately measure the difference between motion indices at first because the rate of CSI acquisition is much higher than what is required for this scenario. Overall, this also shows us that we can use the CSI periodicity to reduce the impact of head motion. Let us see in more detail other ways the CSI periodicity impacts performance.

B. Performance impact from RAN parameters

Among all RAN parameters, the period between two CSI reports is of major relevance. Figure 7 compares the performance across many CSI periodicities. Firstly, it is generally true that a higher τ_{CSI} corresponds to higher PLR, which makes sense since sampling the channel less often increases the likelihood of block errors, and eventually of packet errors. Secondly, the performance in higher frequencies is more dependent on τ_{CSI}. Because of the increased volatility of higher spectrum [27], the channel estimation gets outdated quicker, requiring more frequent CSI updates. These two patterns are also present in Table 1.

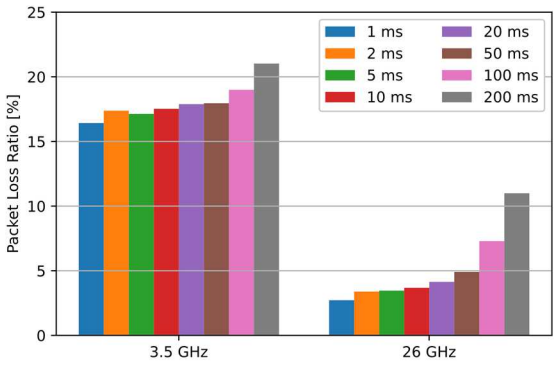

Figure 7: PLR for different CSI periodicities.

C. Performance impact of service requirements

Let us assess now how service requirements such as the application bit rate or a smaller RAN latency budget impact performance. In Figure 8 we vary application bit rates. We see a threshold, which seems to be around 75 Mbps for 3.5 GHz and 100 Mbps for 26 GHz, after which the increase in PLR is very similar for subsequent increases in application bit rate, independently of the frequency. This suggests that past the threshold the system gets overloaded, and all additional packets are lost.

Additionally, we observe that the 3.5 GHz carrier can on average support for each participant a throughput of 75 Mbps with a similar PLR as the 26 GHz carrier supports 100 Mbps.

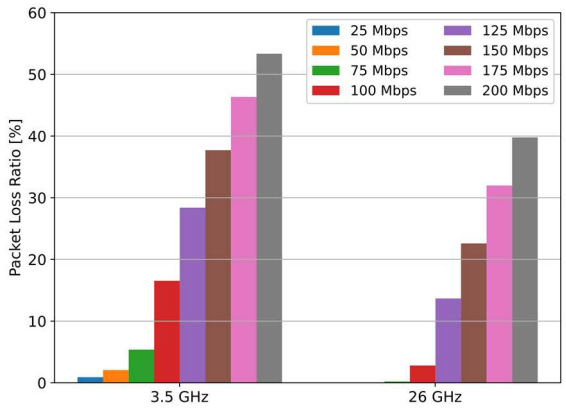

Figure 8: PLR for different required application bit rates.

With respect to latencies, Figure 9 compares PLR for distinct RAN latency budgets Δ_{RAN}. The lower the latency budget, the more difficult it is to cope with this latency requirement and PLR performance gets worse, as expected. More interesting is to see the convergence of PLR to zero slows down in 3.5 GHz and not in 26 GHz. 3.5 GHz has inherently lower average SINR, and so, naturally, there are periods (that may last longer than the latency budgets we considered) where the SINR is so low that practically no packets are transmitted, leading to an accumulation and eventual loss. In mmWave, such reduction in convergence speed happens closer to zero PLR because the higher average throughputs lead to less frequent packet accumulation.

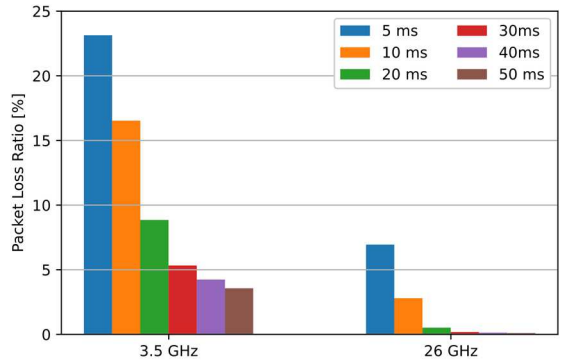

Figure 9: PLR for different RAN latency budget requirements.

D. Performance impact of the meeting setting

In the previous subsection we showed how PLR performance is impacted by the two most relevant service requirements, bit rate and latency. In this section we focus on the impact of the number of meeting participants. The objective is to assess how meetings of different sizes require different quantities of resources and how those quantities vary with the specific service requirement. To do that, the vertical axis of plots in this section is the minimum bandwidth (in MHz) required to obtain a PLR less than 5% for a given scenario. Figure 10 and show the required bandwidths as function of the number of participants, for distinct bit rates and latency budgets, respectively.

As intuitively clear, the required bandwidth decreases in the latency budget and grows in both the application bit rate and the number of users. Perhaps even more interesting is to observe (see Figure 11) the degree to which a more relaxed latency requirement allows the packet scheduler to achieve a higher resource efficiency, translating to significantly lower eventual bandwidth requirement. This demonstrates the benefits of reducing transmission or (perhaps primarily) processing delays elsewhere in the end-to-end chain in order to allow a somewhat more generous latency budget in the radio access network.

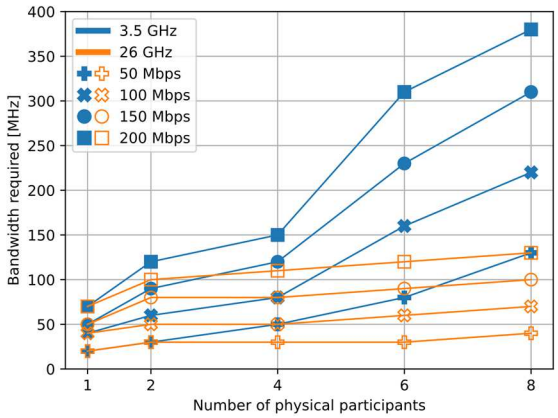

Figure 10: Bandwidth required to support different bit rates and number of participants, with Δ_{RAN} = 10 ms.

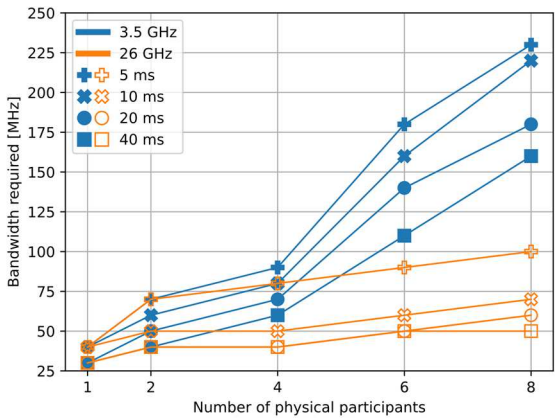

Figure 11: Bandwidth required to support different latencies and number of participants, with R_B = 100 Mbps.

Furthermore, for a given bit rate or latency requirement, 26 GHz practically always requires less bandwidth than 3.5 GHz. Independently of the scenario, the increased directivity of mmWaves leads to less interference and, hence, higher average SINRs; so higher MCSs are supported, and less bandwidth is required. The only exception to the rule is in case of a single physical participant (communicating with a single virtual participant), because there are no other users in the room, hence no interference, and no advantage of mmWaves since the maximum MCS is used in both frequencies. Note that for 26 GHz the bandwidth requirements grow slower than for 3.5 GHz, when the number of users increase. Again, since the beams are narrower, there is less interference, so the system is more resilient to the increase in user proximity. Furthermore, the maximum carrier bandwidth allowed in the sub-6 GHz band is 100 MHz [27], which implies that required bandwidths exceeding 100 MHz require the use of multiple carriers.

Other observations that can be made include *(i)* a meeting with N = 8 participants supported on a mmWave carrier can support per-participant bit rates of 200 Mbps, which is 4× that supportable on a 3.5 GHz carrier with the same bandwidth; and *(ii)* given the same bandwidth, in meetings with N = 4 participants, use of the mmWave carrier can support a RAN latency budget < 5 ms, while the 3.5 GHz carrier supports a latency budget > 20 ms.

VII. Concluding Remarks

We have established an integral modelling framework for the feasibility assessment and performance optimization of the radio access network for Social VR applications in indoor office scenarios, noting that the framework provides a good basis for extension to handle other use cases. The modelling framework covers in detail all relevant aspects regarding meeting room layout, number of participants, human behavior, application-layer specifics, network deployment and configuration, signal propagation and resource management. Using this modelling framework, we have conducted an extensive simulation study to assess the performance impact of these scenario specifics for two frequency bands, viz. 3.5 GHz and 26 GHz. In particular, it is shown how the required carrier bandwidth for supporting Social VR depends on the number of participants and the available RAN latency budget. Regarding the latter, our results demonstrate the potential significance of the increased radio network efficiency when e.g. accelerated processing elsewhere in the end-to-end chain allows a somewhat more generous latency budget in the RAN. Also, the behavior that each user exhibits impacts performance more prominently in higher frequencies. Results further reveal the benefits of the smaller beamwidths enabled in the 26 GHz case in lowering interference the average user experiences, and thus improving the efficiency of settings with high spatial re-use of the radio resources. Additionally, results support that RAN parameters, such as the period with which channel state information is acquired, can be adjusted to mitigate the impact resultant from user behavior or different carrier frequencies.

A topic for future research is the development of (AI/ML-based) methods for dynamic configuration of RAN parameters to support Social VR applications. This involves cross-layer adaptation, e.g. the adaptation of application-layer data rates (hence QoE) to actual radio network capabilities, optimization of the connect-compute trade-off in the radio access (incl. edge cloud), and pro-active beam management based on human movement predictions. Ultimately, we envision the design of a self-configuring 5G network slice, autonomously managed to support Social VR applications.

References

[1] H. Holma, A. Toskala and T. Nakamura, *'5G technology - 3GPP New Radio'*, John Wiley & Sons, Hoboken, USA, 2020.

[2] 5G PPP, *'5G innovations for new business opportunities'*, white paper, see https://5g-ppp.eu/wp-content/uploads/2017/01/5GPPP-brochure-MWC17.pdf, 2017.

[3] M. Latva-aho and K. Leppänen (eds.), *'Key drivers and research challenges for 6G ubiquitous wireless intelligence'*, white paper, University of Oulu, Finland, 2019.

[4] 3GPP, *'Extended Reality (XR) in 5G'*, TR26.928, v16.0.0, 2020 .

[5] F. Hu, Y. Deng, W. Saad, M. Bennis and A. H. Aghvami, 'Cellular-connected wireless virtual reality: requirements, challenges, and solutions', *IEEE Communications Magazine*, vol. 58, no. 5, 2020.

[6] E. Bastug, M. Bennis, M. Medard and M. Debbah, 'Toward interconnected virtual reality: opportunities, challenges and enablers', *IEEE Communications Magazine*, vol. 55, no. 6, 2017.

[7] S. Dijkstra-Soudarissanane, K. El Assal, S.N.B. Gunkel, F. Ter Haar, R. Hindriks, J.W. Kleinrouweler and O.A. Niamut 'Multi-sensor capture and network processing for virtual reality conferencing', *Proceedings of MMSys '19*, Amherst, USA, 2019.

[8] T.V. Doan, D. You, H. Salah, G.T. Nguyen and F. H.P. Fitzek, 'MEC-assisted immersive services: orchestration framework and protocol', *Proceedings of BMSB '19,* Jeju, South-Korea, 2019.

[9] M.S. Elbamby, C. Perfecto, M. Bennis and K. Doppler, 'Edge computing meets millimeter-wave enabled VR: paving the way to cutting the cord', *Proceedings of WCNC '18,* Barcelona, Spain, 2018.

[10] M. S. Elbamby, C. Perfecto, M. Bennis and K. Doppler, 'Toward low-latency and ultra-reliable virtual reality', *IEEE Network*, vol. 32, no. 2, 2018.

[11] S. Barbarossa, E. Ceci and M. Merluzzi, 'Overbooking radio and computation resources in mmW-mobile edge computing to reduce vulnerability to channel intermittency', *Proceedings of EuCNC '17*, Oulu, Finland, 2017.

[12] M. Chen, W. Saad and C. Yin, 'Virtual reality over wireless networks: quality-of-service model and learning-based resource management', *IEEE Transactions on Communications*, vol. 66, no. 11, 2018.

[13] O. Abari, D. Bharadia, A. Duffield and D. Katabi, 'Cutting the cord in virtual reality', *Proceedings of HotNets '16*, Atlanta, USA, 2016.

[14] M. Huang and X. Zhang, 'MAC scheduling for multiuser wireless virtual reality in 5G MIMO-OFDM systems', *IEEE International Conference on Communications Workshops,* Kansas City, USA, 2018.

[15] S. Fremerey, A. Singla, K. Meseberg and A. Raake, 'AVtrack360: an open dataset and software recording people's head rotations watching 360° videos on an HMD', *Proceedings of MMSys '18*, Amsterdam, The Netherlands, 2018.

[16] E.J. David, J. Gutiérrez, A. Coutrot, M. Perreira Da Silva and P. Le Callet, 'A dataset of head and eye movements for 360° videos', *Proceedings of MMSys '18*, Amsterdam, The Netherlands, 2018.

[17] S. N. B. Gunkel, H. M. Stokking, M. J. Prins, N. van der Stap, F.B.T. Haar and O.A. Niamut, 'Virtual reality conferencing: multi-user immersive VR experiences on the web', *Proceedings of MMSys '18*, Amsterdam, The Netherlands, 2018.

[18] S. Mangiante, G. Klas, A. Navon, Z. GuanHua, J. Ran and M. Dias Silva, 'VR is on the edge: how to deliver 360° videos in mobile networks', *Proceedings of VR/AR Network '17*, Los Angelos, USA, 2017.

[19] 3GPP, *'Study on channel model for frequencies from 0.5 to 100 GHz'*, TR38.901, v16.1.0, 2020.

[20] 3GPP, *'Home Node B (HNB) Radio Frequency (RF) requirements (FDD)'*, TR25.967, v16.0.0, 2020.

[21] Fraunhofer, *'Quadriga – The next generation radio channel model'*, www.quadriga-channel-model.de, 2020.

[22] F. Afroz, K. Sandrasegaran and P. Ghosal, 'Performance analysis of PF, M-LWDF and EXP/PF packet scheduling algorithms in 3GPP LTE downlink', *Proceedings of ATNAC '14*, Melbourne, Australia, 2014.

[23] J.-H. Rhee, J.M. Holtzman and D.-K. Kim, 'Scheduling of real/non-real time services: adaptive EXP/PF algorithm', *Proceedings of VTC '03 (Spring)*, Jeju, South-Korea, 2003.

[24] 3GPP, *'NR; User Equipment (UE) radio access capabilities'*, TS38.306, v16.4.0, 2021.

[25] S. Pratschner, B. Tahir, L. Marijanovic, M. Mussbah, K. Kirev, R. Nissel, S. Schwarz and M. Rupp, 'Versatile mobile communications simulation: the Vienna 5G link level simulator', *EURASIP Journal on Wireless Communications and Networking*, 2018.

[26] A. O. Adeyemi-Ejeye, M. Alreshoodi, L. Al-Jobouri, M. Fleury and J. Woods, 'Packet loss visibility across SD, HD, 3D, and UHD video streams', *Journal of Visual Communication and Image Representation*, 2017.

[27] T. Rappaport *et al.*, 'Millimeter wave mobile communications for 5G cellular: it will work!', *IEEE Access*, 2013.

[28] 3GPP, *'NR; User Equipment (UE) radio transmission and reception; Part 1: Range 1 Standalone'*, TS38.101-1, v17.1.0, 2021.