Social VR Applications in Indoor 5G Networks: Modelling, Feasibility and Optimisation

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

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*—This …**

Keywords—Social XR, VR, 5G, modelling

I. Introduction (HANS; 3/4 page)

It is widely recognized that emerging 5G network technology will boost the development of new, highly innovative applications in virtually all domains. One of the most challenging application classes targeted by 5G 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. 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 enabling virtual meetings, exploitation of remote expertise or skills e.g. in smart industry context (collaborative working, inspection, maintenance), remote education and training, etc. <<<Internet Skills>>>

<<<Cloud>>> To support such so-called Social VR applications, with compelling visual, haptic, audio and possibly even olfactory experiences for remote users, the networking infrastructure has to handle extremely high-bandwidth streams, while keeping the end-to-end latency low. Furthermore, the infrastructure should provide powerful (in-network) processing capabilities, reducing the need for heavy computing devices at the client, thus increasing the clients' flexibility and mobility. Actually, natural Social XR experiences can only be realized when all components in the capture, transmission and rendering pipeline optimally work together. This requires dynamic orchestration of these components and allocation of resources in the networking and compute infrastructure, exploiting the connect-compute trade-off (end-to-end latency matters and computing latency can be traded for bit rate and vice versa) in order to optimize QoE. In particular, dynamic orchestration and resource allocation should also be able to cope with the intermittent nature of high-frequency radio channels to be used in 5G to satisfy the huge capacity and throughput requirements.

Other (beyond) 5G radio access features and capabilities to support extremely demanding wireless applications, like Social XR, are the use of beamforming (incl. D-MIMO) to overcome the associated attenuation challenges, the high numerologies, mini-slot and self-contained subframe features supporting low latencies, and the use of channel-aware delay-based packet schedulers to achieve efficient use of resources while satisfying latency requirements.

Ultimate goal: Design autonomously managed 5G network slice to support Social XR applications requiring the development of AI/ML-based methods for dynamic orchestration and allocation of processing and transmission resources.<<<op radio richten >>>

This paper focuses on radio access challenges for supporting Social VR applications. In particular, we describe the modelling of all relevant aspects that need to be considered for an adequate feasibility assessment and performance optimisation of 'Social VR' applications in indoor office scenarios. The aspects considered in our study comprises the specifics of the use case (office scenario, application), network deployment (xxxx), propagation environment (xxx) and key 5G traffic handling mechanisms (xxxx)

Met het model kunnen we concreet dit en dat doen (sensitivity analyses,

Haalbaarheid (qua QoS reuirements) van scenarios – voor lagere en hogere frequenties sub/sup 6GHz

Hoeveel resources daarvoor nodig?

Impact gebruikersgedrag

Invloed spectrum, numerology, TDD split, MU/D MIMO, scheduler.

These models provide also insights needed to develop effective (AI/ML-based) methods for dynamic resource allocation and tuning of radio access parameters. □ Design autonomously managed 5G network slice to support Social XR applications ("Social XR network slice").

II. Challenges and Objectives (HANS/REMCO; ½ page)

Elaborate on the intrinsic challenges of these scenarios, e.g. deployment, (cross-layer/end-to-end optimisation, connect-compute trade-off. What else?

Explain the objective/contribution of the paper, i.e. to model the handling of social XR applications in an indoor 5G network to enable a feasibility assessment of different scenarios in terms of a.o. number of participating users, application characteristics and available spectrum/BS resources; and to optimise traffic handling in the radio access network, incl. packet scheduling and D-MIMO management.

III. RELATED WORK (JOÃO/SJORS; ¼ PAGE)

Cutting the Cord in Virtual Reality [2] discusses the impact of blockage by hand, head and body. Although Wi-Fi in mmWave bands is considered, it's relevant for our research.

Toward Low-Latency and Ultra-Reliable Virtual Reality [3] discusses the challenges and enablers for ultra-reliable and low-latency VR. Required for VR is the C3 paradigm: (edge) computing, caching, and communication resources. With respect to communication, multi-connectivity (MC) is used as means to mitigate the blockage and disturbance caused by *impulse actions*. A case study is worked out, showing the latency and delay performance for sixteen players being served by four APs, both for the average and the 99th percentile values.

- Papers on the **challenges of mmWaves?** Rappaport has a few very cited ones:
    - Millimeter-Wave Cellular Wireless Networks: Potentials and Challenges (https://ieeexplore.ieee.org/document/6732923)
    - Millimeter Wave Mobile Communications for 5G Cellular: It Will Work! (https://ieeexplore.ieee.org/document/6515173)
    - Overview of Millimeter Wave Communications for Fifth-Generation (5G) Wireless Networks—With a Focus on Propagation Models (https://ieeexplore.ieee.org/document/7999294)
- **SRS vs GoB: http://ma-mimo.ellintech.se/2020/10/02/reciprocity-based-massive-mimo-in-action/**

Our contribution: different scenario (office instead of gaming arcade), comparison of mmWave to 3.5 GHz, sensitivity analysis of BS array, number of virtual/physical users, I/P frames characteristics.

IV. MODELLING (REMCO; 1½-2 PAGES)

In this extensive modelling section, we describe the modelling of all relevant aspects that need to be considered for an adequate feasibility assessment and performance optimisation of social VR applications in indoor office scenarios. Consecutively, we address the application scenario or use case, the network deployment, the propagation environment and key traffic handling mechanisms. REFER TO TABLE FOR CONCRETE PARAMETER SETTTINGS.

A. Application scenario

As a use case we consider the use of 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 centre of the room. Uniformly spread around the conference table are $N = N_P + N_V$ chairs, seating N_P participants that are physically present in the room, and N_V 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 – …).

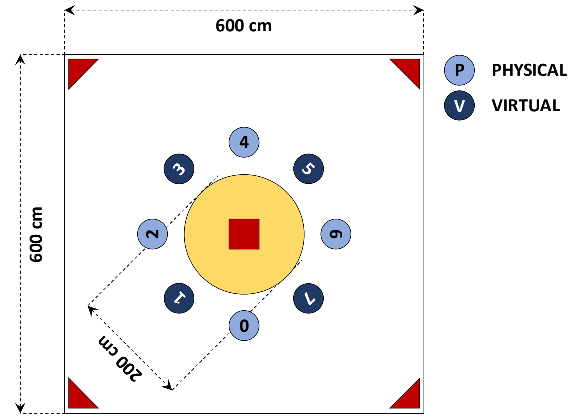

Figure 1: Visualisation of the 'Social VR' application scenario with eight physically ('P') or virtually ('V') meeting participants.

As we concentrate on the communication challenges in the considered conference room, for some modelling aspects we specifically focus on the participants that are physically present. Each such participant wears a VR headset towards which downlink images are transferred. Furthermore, two on-table video capture devices are installed and pointed to the participant to capture images for uplink transfer.

Each physically present participant is and remains seated in his chair, but the participant's head is modelled to exhibit some degree of randomised *head motion* in terms of changes in position and orientation:

- The change of a head's *position* is 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 μ_z 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 and 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 $0.1 \times \alpha$ m/s, where $\alpha \in \{0, 1, ..., 7\}$ is the head motion index. 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 'yes', 'no' and 'may be' axes perpendicularly cutting through the head, assuming uniform angular distributions on the ranges $[-\beta_Y, \beta_Y]$, $[-\beta_N, \beta_N]$ and $[-\beta_M, \beta_M]$, respectively. Whenever the speaker changes, such offsets are sampled and the corresponding change in orientation is modelled to take $1.5 – 0.2 \times \alpha$ seconds, with α 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.

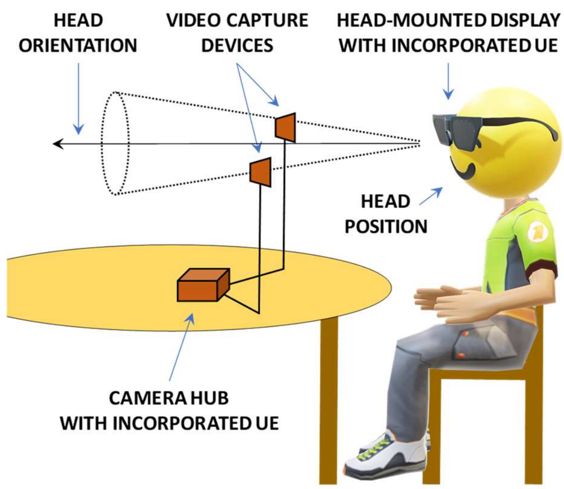

Figure 2: TBD

Aside from substantial changes in the head orientation that are due to speaker changes, the randomisations of the head position and its orientation effectively constitute some degree of 'head wobbling' in 3D, with seven distinct intensities configured by head motion index α. In general, the relevance of modelling the head motions 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.

The *meeting dynamics* are modelled by traversing through a pre-generated speaker list, letting each designated participant speak for a deterministic time of τ seconds, and then switch to the next speaker on the list. For the simulated scenario with $N = N_P + N_V = 4 + 4 = 8$ participants, as visualised in Figure 1, the speaker list is given by '0 – 5 – 2 – 7 – 4 – 1 – 6 – 3', where we choose to alternate between physically and virtually present users.

The *VR application* is modelled as a persistent real-time video streaming application, where each up- or downlink stream is a effectively a sequence of GoPs ('Groups of Pictures') and each GoP comprises a single I ('intra-coded') and M_P P ('predicted') video frames. The size of the relatively large fresh I and the differential P frames is denoted S_I and S_P, respectively. A typical frame generation rate of R_F fps (frames per second) is assumed, which is noted to apply at the source. Given this frame rate and the choices of M_P, S_I and S_P, the application-level bit rate is equal to $(8 / 1000000) \times R_F \times (S_I + M_P S_P) / (1 + M_P)$ Mb/s. Given the relative negligibility of the audio component in terms of required bit rates, we limited our modelling to include the transfer of VR images only.

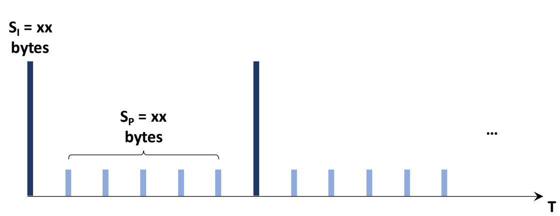

Figure 3: Visualisation of the VR frame generation process at the source for a choice of R_F 30 fps, $M_P = 5$, S_I = xx bytes and S_P = xx bytes, corresponding with an application-level bit rate of xx Mb/s.

For the considered indoor office scenario, in the *uplink* these sources are two video capture devices per person, with some degree of local pre-transmission processing and also an aggregation of both streams before the *uplink* transfer. The source of the *downlink* transfers depends on whether the viewed speaker is physically or virtually present. In the former case, the source of the downlink transfer is xx, while in the latter case the source is the remote video capture device recording the virtually present speaker at its remote location. Generated as application-level frames at the source, the video information arrives in the form of 1500-byte IP packets at the uplink (UE (User Equipment)) or downlink (BS (Base Station); gNodeB) transmission buffer, with the packets dispersed in time due to the variabilities of packet latencies on the path from source to transmit buffer. For the uplink streams, with the video capture devices directly connected to a 5G UE, no such time dispersion is assumed, and the IP packets are available in the UE's transmit buffer immediately upon frame generation. The time dispersion for the downlink streams is modelled by a burstiness parameter. To be continued …

The QoS requirement of the VR application is given by a maximum tolerable end-to-end one-way frame-level latency of Δ_{E2E} (in ms), of which we assume a budget of $\Delta_{RAN} < \Delta_{E2E}$ is assigned to the Radio Access Network (RAN). Any VR frame delivered with a latency exceeding the end-to-end requirement is considered useless.

B. Network aspects

The meeting room is adequately equipped to support social VR meetings, including, besides audio/video capturing and audio playback devices, also advanced wireless networking hardware. More specifically, we assume that $K_{BS} \in \{1, 2, 4, 5\}$ massive MIMO antenna arrays are deployed, where the different values of K_{BS} indicate distinct deployment scenarios. For $K_{BS} = 1$, we assume a single antenna mounted at the ceiling centre and pointed downwards; for $K_{BS} = 2$ we assume two downtilted antennas mounted in adjacent (scenario '2a') or in opposite (scenario '2o') ceiling corners; for $K_{BS} = 4$ we assume four downtilted antennas mounted in all four ceiling corners; lastly, for $K_{BS} = 5$ we assume a combination of the antennas at the centre and in the corners of the ceiling. The different deployment scenarios are visualised in Figure 4.

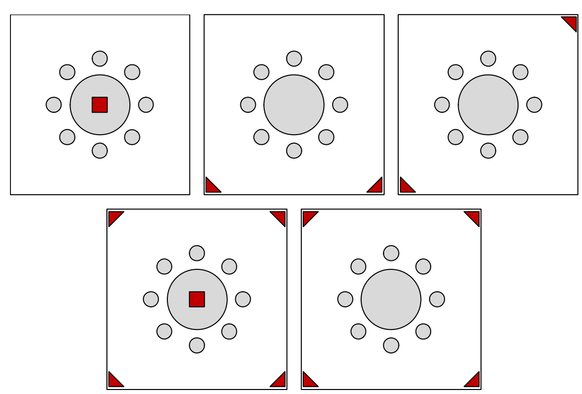

Figure 4: Network deployment scenarios: clockwise, starting from top-left, we have scenarios with K_{BS} = 1, 2 ('2a'), 2 ('2o'), 4 and 5.

Each BS antenna is assigned a single TDD carrier of B MHz in either the 3.5 GHz or the 26 GHz (mm-wave) band, which constitute distinct assessment scenarios. To enable a fair comparison the band-specific assumptions regarding the design of the antenna arrays are limited by a maximum form factor of xx × xx cm. For the 3.5 GHz scenario, this allows an antenna array comprising xx × xx pairs of cross-polarised AEs (Antenna Elements) with a half-wavelength inter-AE spacing in both dimensions. As also visualised in Figure XX, we assume xx × xx-sized subarrays fed with independent transmission/reception chains, hence effectively constituting an xTxR antenna. For the 26 GHz scenario, the form factor limitation allows an antenna array comprising xx × xx pairs of cross-polarised AEs with half-wavelength inter-AE spacings. With an assumption of xx × xx-sized subarrays we again end up with an xTxR antenna (see Figure 5). The maximum transmit power and receiver noise figure are denoted $p_{MAX,BS}$ (in dBm) and NF_{BS} (in dB), respectively.

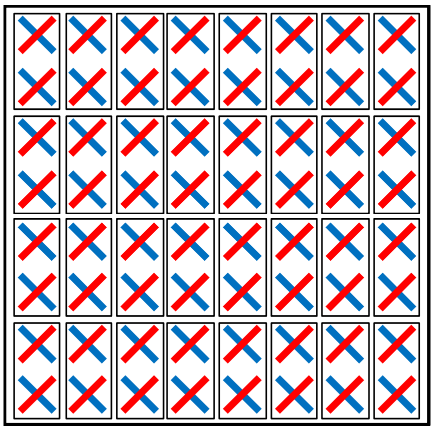

Figure 5: TBD

At the UE side, a distinction is made between *(i)* the UEs incorporated into the VR headset, whose communication role is downlink- and hence reception-oriented; and *(ii)* the UEs attached to the video capturing devices placed in front of each physically present participant on the meeting table, which are primarily taking care of uplink transmissions. Each UE is equipped with xx or xx antenna elements for the 3.5 GHz and 26 GHz scenarios, respectively, which are distributed in cross-polarised pairs as visualised for both cases in Figure xx. The maximum transmit power and receiver noise figure are denoted $p_{MAX,UE}$ (in dBm) and NF_{UE} (in dB), respectively.

Propagation environment (SJORS)

The characteristics of the propagation environment (in terms of path loss, shadowing, and multipath fading) are taken from the *indoor office* scenario specified by [3] with implementations provided by Quadriga [5]. Since Quadriga does not implement the blockage models from 3GPP, this feature is implemented on top of it. We consider three types of LOS blockage: self-blockage by head, self-blockage by hand/arm and blockage from another person. For these types of blockage, we use an average loss in received power of XX dB, XX dB, and XX dB, respectively. The blockage events are modelled as an …

Figure 1 shows the impact in channel gain that events like human blockage and orientation change can have at 26 GHz.

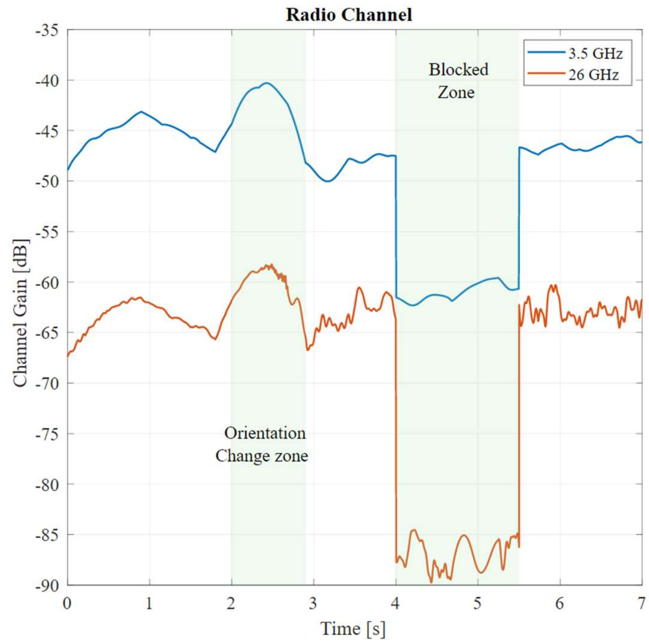

Figure 6: Impact of Blocking and Orientation in Channel Gain

C. *Traffic handling*

The up/downlink (UL/DL) social VR traffic is handled over a 5G NR radio access network utilising the 3.5/26 GHz carrier and the BS/UE antennas as described above. At each BS antenna a Grid of Beams (GoB) comprises a set of pre-defined transmission beams (precoders) in a three-dimensional angular grid. 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 grid of CSI-RS beams has an angular granularity of xx degrees in both dimensions and a half-power beamwidth of xx degrees.

Each UE selects the serving BS antenna towards which it experiences the strongest average channel gain as its serving cell, while its serving cell cluster may be extended beyond this 'best server' by including additional BS antennas towards which the UE experiences an average channel gain no more than x dB lower to towards the best server. With configuration parameter x set to a value greater than 0 dB this implies D-MIMO (Distributed MIMO) where a UE may be concurrently served by the aggregation of multiple BS antenna arrays.

At the serving cell, the UE selects the strongest CSI-RS beam, indicated via the CSI-RS Resource Indicator (CRI) feedback, with a correspondingly precoded Physical Downlink Shared CHannel (PDSCH) traffic beam used for downlink transmission. 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'. Given the use of a TDD carrier, 'beam correspondence' is assumed and the same beam pair is applied for uplink communications.

The adaptive modulation and coding scheme estimates the Signal-to-Interference-plus-Noise-Ratio (SINR) at the receiver 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 [XX]. To adequately compensate for feedback/measurement delay-induced errors in deriving the optimal transmission parameters, a dynamically tuned UE-specific outer-loop link adaptation scheme is applied.

A QoS-oriented packet scheduler governs UL and DL transmissions, given the resources configured by the applied TDD duplexing scheme, characterised by a xx-slot frame size and a xx:yy UL/DL split, as visualised in Figure 7. The scheduler applies head of line packet dropping in the transmit buffers at the BS and UE in case the packet's corresponding VR frame is determined not to meet the aforementioned latency budget Δ_{RAN}.

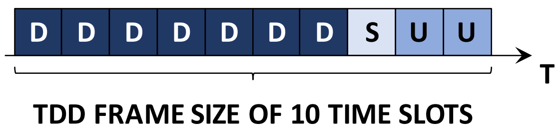

Figure 7: TBD

V. Initial Assessment (HANS; 1-2 pages)

Intro text … Mention implementation in simulator here …

A. Scenarios and KPIs

Text …

TABLE I. Simulation scenarios.

APPLICATION SCENARIO	N_P	4	NETWORK ASPECTS	$p_{MAX,BS}$	X dBm
	N_V	4		$p_{MAX,UE}$	X dBm
	μ_z	1.2 m		NF_{BS}	X dB
	σ_x	0.067 m		NF_{UE}	x dB
	σ_y	0.067 m			
	σ_z	0.033 m			
	β_Y	m			
	β_N				
	β_M				
	τ	xx s			
	M_P				
	S_I	xx byte			
	S_P	xx byte			
	R_F	30 fps			
	Δ_{E2E}	Xx ms			
	Δ_{RAN}	Xx ms			

B. Numerical results

Text …

VI. Concluding Remarks (1/4 page)

Recap …

Open challenges, incl. AI/ML-based slice optimisation and cross-layer adaptation, incl. adapting application layer to radio network capabilities and optimising the connect-compute trade-off. Predictive beam handover using AI/ML (anticipate blocker-induced beam failure: timely beam change or beam addition).

References (1/4 page)

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

[2] O. Abari, D. Bharadia, A. Duffield and D. Katabi, 'Cutting the Cord in Virtual Reality', Association for Computing Machinery, New York, USA, 2016

[3] M. S. Elbamby, C. Perfecto, M. Bennis and K. Doppler, 'Toward low-latency and ultra-reliable virtual reality', *IEEE Network*, vol. 32, no. 2, pp. 78-84, March-April 2018,

[4] 3GPP, 'Study on channel model for frequencies from 0.5 to 100 GHz', TR38.901, v16.1.0, 2019.

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

[6] 3GPP, 'Extended Reality (XR) in 5G', TR26.928, v16.0.0, 2020

[7] ...