

# Holographic communication in 5G networks

Once confined to the realm of science fiction, holographic communication now ranks as one of the most wanted 5G-enabled applications by both consumers and enterprise users, according to recent research.

ALI EL ESSAILI, SARA THORSON, ALVIN JUDE, JÖRG CHRISTIAN EWERT, NATALYA TYUDINA, HÉCTOR CALTENCO, LUKASZ LITWIC, BO BURMAN

**Holographic communication refers to real-time capturing, encoding, transporting and rendering of 3D representations, anchored in space, of remote persons shown as stereoscopic images or as 3D video in extended reality (XR) headsets that deliver a visual effect similar to a hologram.**

■ After several years of experience with video calls on smartphones and tablets, many users report that they are eagerly awaiting the chance to meet others digitally using immersive communication services such as 3D holographic augmented reality (AR) calls over 5G [1]. Compared with flat screen video, holographic communication can convey the subtleties of non-verbal communication and provide a sense of presence and immediacy that enhances the quality of human interaction. More than 50 percent of smartphone users [2] expect this technology to be available to them within a few years.

Consumers are not the only ones longing for more authentic forms of digital communication. A recent study showed that the main barrier to remote working was the need for social interaction [3]. With the time spent working outside the office expected to increase in the coming decade, many office workers will require more immersive forms of digital interaction.

Starting with AR glasses and holographic calls with spatial audio, office workers around the world expect to benefit from the emergence of haptic technology, which adds tactile sensing of digital objects [3]. More than half of them say they would like a digital workstation with multisensory presence at the office when working remotely. Similarly, in a recent online survey of 7,115 self-defined early adopters aged 15-69 in 14 major cities, 80 percent of the respondents said they expect to have telepresence facilities to socialize with remote colleagues by 2030.

The potential for realizing holographic communication use cases in the years ahead is high. They are already anticipated in numerous consumer and enterprise areas ranging from attending family events as a hologram or meeting a doctor from home to remote presence at the office, expert help at a factory and immersive marketing. As soon as the ecosystem is ready to deliver these new experiences at a good price point and quality of experience (QoE), our research suggests that both private consumers and enterprises will be keen to use it.

The ability to make holographic communication an everyday reality is dependent on three key factors. Firstly, there needs to be a desire that drives the behavioral change (the human factor). Secondly, the appropriate AR devices need to become available. Lastly, mobile networks must have the ability to support the holographic communication pipeline.

### The human factor

Users are the primary beneficiary of any type of human communication technology, and holographic communication is no exception. It is therefore essential to understand how they perceive it, along with the benefits from their perspective and their opinions about how to improve the user experience.

The two major user groups for holographic communication (enterprises and consumers) have different priorities. Enterprises will use holographic communication if it satisfies productivity goals better than existing tools. These can be defined as effectiveness, efficiency, and satisfaction in a specified context of use [4].

Consumers, on the other hand, tend to select the communication option that best satisfies their fun goals, characterized by hedonic, emotional and experiential perspectives [5]. Emotions are known to be an important guide to decision making [6]. It is therefore crucial that people's feelings are understood throughout the development process, which is achievable through user studies.

In an internal study on the topic, we discovered that a key emotion people feel about holographic communication is excitement. They are excited about seeing a holographic version of people they

## ●● A KEY EMOTION PEOPLE FEEL ABOUT HOLOGRAPHIC COMMUNICATION IS EXCITEMENT ●●

know in the same room as themselves, and they are excited about the future of this technology. Such a response is expected due to the novelty of the interaction, but as the novelty (or "halo") effect diminishes, people tend to prioritize factors such as usability, usefulness and familiarity [7].

While some QoE metrics already exist for XR communication [8], several other tools are also available. Design thinking [9] can be used early on to ensure the best solution is built, while qualitative methods such as interviews can be used with early prototypes to provide deeper insights into how people feel. Other human-centered topics such as ethics, consent and accessibility should be considered and regularly monitored. Involving experts in human-computer interaction, ergonomics, psychology and user experience early in the design process will ensure that the final product is in line with the intended users' expectations.

It is essential that the human factor is prioritized in the development of holographic communication. One way to facilitate this would be to start referring to the end-to-end (E2E) pipeline as human to human rather than glass to glass.

### Holographic communication

The area of **holographic communication** is divided into three paradigms: (1) avatars, (2) professional-quality digital representations of users and (3) consumer-friendly digital representations of users. "State-of-the-art avatar" technologies use artificial intelligence (AI) to generate a photorealistic representation of a user. Professional-quality digital representations are created through the use of multiple cameras and real-time studio recordings. Consumer-friendly digital representations are generated with the help of an AI-enabled, 3D-capturing setup on consumer grade phones or tablets. The third paradigm is the focus of this article.

|                         | Use case                  | Typical usage (hours) | Resolution | Battery life | Mobility |
|-------------------------|---------------------------|-----------------------|------------|--------------|----------|
| Enterprise              | Retail shopping           | 1-4                   | +++        | ++           | +++      |
|                         | Logistics (picking)       | 4-8                   | +          | +++          | +        |
|                         | Augmented education       | 4-8                   | +++        | +++          | +        |
|                         | Remote support            | 1-2                   | +++        | ++           | +++      |
| Consumer                | AR gaming                 | 1-2                   | +++        | ++           | +++      |
|                         | AR navigation             | 1-4                   | +          | ++           | +++      |
|                         | AR event (stadium)        | 1-2                   | +++        | ++           | +++      |
| Enterprise/<br>Consumer | Holographic communication | 1-4                   | +++        | +            | +++      |

Figure 1 AR use cases and their requirements on the glasses ranked by importance (+ to +++)

### Augmented reality glasses and other devices

Early AR glasses were connected by a cable, followed by the usage of Wi-Fi technology that limited mobility to home and office use in most cases. Over the last five years, however, the AR glasses market has been characterized by strong technology evolution with respect to all of the most important parameters: weight, field of view, resolution, battery life and mobility.

The first generation of lightweight AR glasses, which are connected through USB-C to 4G- and 5G-enabled smartphones, were launched two years ago. The next generation is expected to have built-in cellular connectivity, similar to today's smart watches. Figure 1 highlights some of the enterprise

and consumer use cases that will benefit most from next-generation AR glasses.

There are two different AR device types: those with SLAM (simultaneous localization and mapping) support and those without it. SLAM uses one or more front cameras to create and update a map of the surroundings that allows the device to localize itself and to anchor virtual content (including holograms) on fixed locations in the environment. A true holographic impression is generated by the ability to see the hologram from different angles, which can be achieved with SLAM.

The ongoing evolution of both mobile and stationary holographic displays as well as AR-enabled contact lenses will also contribute to

### Terms and abbreviations

AI – Artificial Intelligence | AR – Augmented Reality | E2E – End-to-End | FPS – Frames Per Second | IEC – International Electrotechnical Commission | ISO – International Organization for Standardization | LIDAR – Light Detection and Ranging | MPEG – Moving Picture Experts Group | MS – Millisecond | NR – New Radio | QoE – Quality of Experience | RAN – Radio Access Network | SLAM – Simultaneous Localization and Mapping | ToF – Time of Flight | V3C – Visual Volumetric Video-based Coding | V-PCC – Video-based Point Cloud Compression | XR – Extended Reality

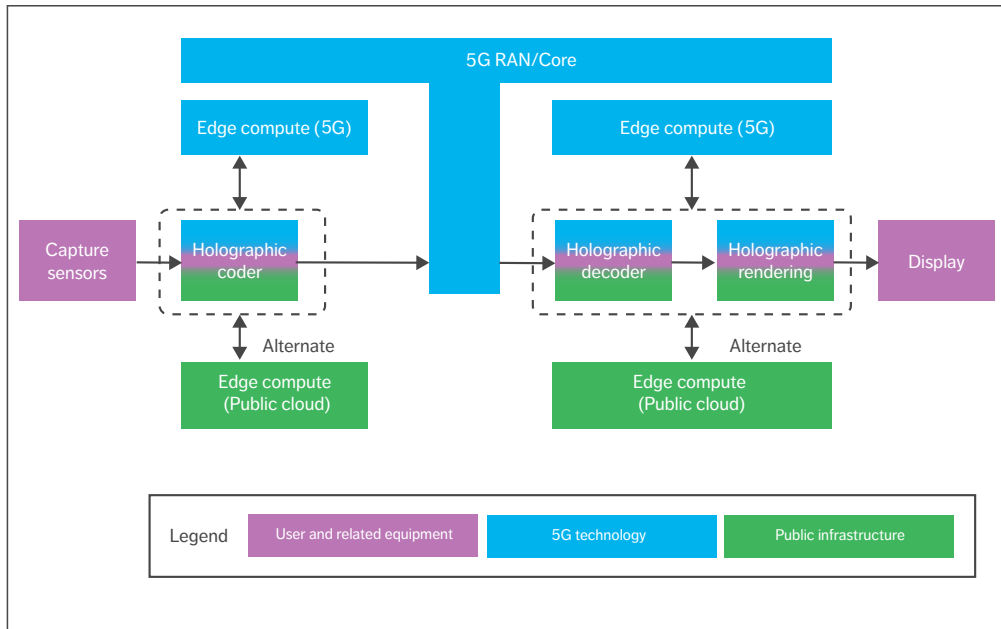


Figure 2 End-to-end pipeline for holographic communication

the uptake of holographic communication. AR-enabled tablets and phones allow a lower barrier to entry in cases where hands-free operation is not required.

### Holographic communication pipeline

Our proposed architecture to support holographic communication is depicted in [Figure 2](#). In it, capturing sensors provide a real-time representation of the human face and body. Format conversion and filtering is applied before encoding to reduce the bitrate requirements on the network. The compressed hologram is transmitted to an XR device over a low latency reliable transport network such as 5G. On the XR device, the compressed hologram is decoded and processed before rendering in the consumer environment. The rendering engine considers positioning and semantic information of the device and the rendered scene. The virtual human representation is displayed on the XR device.

The network can act as a compute platform for the encoding, decoding and rendering of holographic data to offload heavy processing from the devices.

### Capturing technologies

Holographic capturing is the process of creating a measurable 3D representation of an object, person or environment. This process is divided into four steps:

1. Acquisition
2. Depth estimation
3. Data fusion
4. Post-processing.

The acquisition step makes use of visual sensors to capture a volume of interest. Several different visual sensor technologies are used for 3D capturing. Nowadays, the most common mechanism is the usage of time-of-flight (ToF) sensors (such as

## MESHES CAN BE SIGNIFICANTLY SMALLER THAN POINT CLOUDS, INCREASING THE SPEED OF STORAGE, TRANSMISSION AND RENDERING

LiDAR), which measure distance by the time it takes a light pulse to travel to its destination and back.

In the depth estimation step, sensor streams are used to compute depth. ToF sensors can provide depth information directly, while stereo cameras and multi-camera systems estimate depth by capturing the person from different angles.

In the data fusion step, the depth information or depth map from different perspectives is fused into a single stream of 3D points by matching key points and calculating the optimal geometric transformation from the different views.

The post-processing step reduces the data size of the stream of 3D points by cleaning up redundant points, noise and outliers. The resulting 3D representation can be delivered in various visual media formats such as point clouds or meshes.

A point cloud is a collection of points that represent the captured volume. Each point contains information on its location, as well as the option of the red, green and blue color model and intensity value at a specific frame. A mesh unites those points with triangles, disregarding redundant points and filling any holes. Meshes can be cleaned and reduced further by decreasing the number of vertices. Depending on their resolution, they can be significantly smaller than point clouds, increasing the speed of storage, transmission and rendering of meshes in comparison with point clouds.

### Rendering and display

Rendering is the process of generating an image of a scene or a model from a given viewpoint using computations. A scene is a container object describing a volume and its contents. Sources are

objects located in the scene to be rendered. A camera is an instance that renders the viewpoint and consists of a location, focus, orientation and resolution.

Engines render the content based on rendering pipelines. The rendering pipelines take care of culling, rendering and post-processing. There are several types of pipelines with different capabilities and performance characteristics that are suitable for different applications and platforms. A general-purpose render pipeline is optimized to process graphics across a wide range of platforms, while a high-fidelity graphics pipeline is suitable for high-end platforms.

Additional techniques and approaches are available to improve the rendering process in various ways, such as enabling faster rendering with better QoE, smoothing squared edges or increasing the quality of objects in a scene. Furthermore, AI algorithms can recreate an object of a scene or create a photorealistic representation [10] such as an avatar. The photorealistic representation contains a model: a visual representation or mesh model and configurations. Configurations are the bones or access points of a model. These points are animated during the rendering process with preloaded animations.

Real-time data captured from a depth camera takes more computation for rendering compared with an avatar representation. The rendering of all parts of the mesh is needed for every frame for a hologram, compared with rendering a difference for an avatar, such as updating facial expressions like the blinking of an eye on an existing avatar model.

Split rendering is an approach that offloads rendering capabilities to an edge cloud. The key idea is that the six degrees of freedom position and

## REAL-TIME DATA CAPTURED FROM A DEPTH CAMERA REQUIRES MORE COMPUTATION FOR RENDERING

orientation of the XR device in a scene are tracked, and constant feedback is provided to the edge in real time. The rendering of a 3D scene happens in a cloud. Based on the position of a user, the cloud streams back a 2D video to a scene from the user's angle. For this approach, an end-user device does not need to have high-end characteristics. However, good QoE in this scenario requires low latency communication between the edge and the device.

When it is ready, the rendered stream is transmitted to a device with the ability to provide users with a holographic experience. There are four device types: handheld devices (such as smartphones and tablets), holographic displays, AR glasses and virtual reality glasses. Using these devices, it is possible to place a hologram in a room and move around it.

### Media formats and codecs

The delivery of holographic communication will require the processing and transfer of various visual media formats. Those formats aim to represent more realistic and interactive visual representations of humans and/or an environment than established 2D video formats used in traditional videoconferencing, yet the increased informational load may put high pressure on transmission bitrates across the whole communication chain.

For example, point cloud data representing a single human can typically consist of 100,000 to 1 million points and beyond per single time instance (video frame). Streaming such data with 30 frames per second (fps) – the typical videoconferencing streaming rate – would require roughly between 300Mbps to 3Gbps of available bandwidth. Such uncompressed bitrates are not feasible today, and current real-life videoconferencing systems employ media codecs to reduce the bandwidth requirements to single Mbps (1-6Mbps) offering compression ratios from 250:1 up to 1000:1.

Methods to offer similar compression ratios for immersive 3D representations such as point clouds are therefore needed to enable the deployment of holographic communication services. In terms of handling 3D visual formats and their decoding on

## IT IS POSSIBLE TO PLACE A HOLOGRAM IN A ROOM AND MOVE AROUND IT

XR devices, two scenarios can be considered.

In the first, which is typically applied in the case of split rendering, the processing and decoding of 3D formats is done at the edge, while XR devices decode pre-rendered 2D video with traditional 2D video codecs such as ITU-T (International Telecommunication Union Telecommunication Standardization Sector) and ISO/IEC (International Organization for Standardization/International Electrotechnical Commission) MPEG (Moving Picture Experts Group) High Efficiency Video Coding/H.265 or Versatile Video Coding/H.266 codecs.

In the second scenario, the processing and decoding of 3D formats is done on XR devices requiring the support of additional immersive decoders on the device itself. One approach standardized in ISO/IEC MPEG-I visual volumetric video-based coding (V3C) and video-based point cloud compression (V-PCC) employs 3D to 2D projection algorithms that create intermediate 2D video representations. These representations can be decoded with multiple instances of 2D video codecs such as H.265 or H.266.

An alternative approach would be to employ “native” 3D codecs such as the ISO/IEC MPEG-I geometry-based point cloud compression. In this codec, compression tools operate directly on a 3D point cloud representation. Unlike for the V3C/V-PCC codec, this approach would require adding support for such native point cloud codec(s) in mobile hardware chipsets.

### Test results

To determine the feasibility of using existing 5G technology to build an E2E pipeline that enables high-quality holographic communication, we tested the approach presented in Figure 2 in a live 5G New Radio (NR) network.

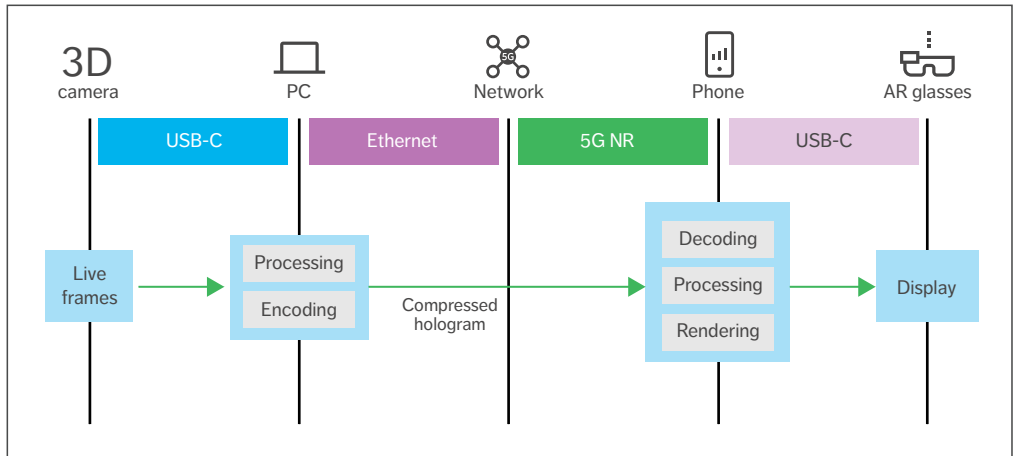


Figure 3 The test setup in the lab

The 3D frames were captured by a single 3D camera connected to a computer, which was connected by ethernet to the 5G network. The captured stream was compressed and sent over the 5G network to a 5G mobile phone with connected AR glasses. The decoding and rendering were executed on the mobile phone and displayed in the AR glasses. The SLAM functionality made it possible to walk around the 3D representation. The measurement setup is shown in [Figure 3](#).

Factors that affect QoE in holographic communications include bandwidth, latency and quality. These three are collectively referred to as the trade-off triangle because only two of them can be optimized at the same time, and it will always be at the cost of the third. In our measurements, we chose to change the quality and bandwidth.

**FACTORS THAT AFFECT QOE IN HOLOGRAPHIC COMMUNICATIONS INCLUDE BANDWIDTH, LATENCY AND QUALITY**

The quality was changed by adjusting the number of points in the point cloud (10,000 points/30fps to 1 million points/30fps). The compression rate was kept constant resulting in bandwidth of between 5Mbps and 100Mbps.

The total latency in the phone is determined by two factors: the decoding and the rendering. As expected, the latency increases when the quality and bandwidth increase. It is clearly visible that the compute in the mobile phone is reaching its limit for the high quality, causing a latency of approximately 170ms (milliseconds). The first experiments show that a move of the decoding compute from the phone to the edge cloud can reduce the latency to 70ms.

[Figure 4](#) shows the bitrate, round-trip time and phone latency results for four quality levels (Q1-Q4) all with 1024x780 resolution and 30fps. The average number of points for each was Q1 (900,000), Q2 (225,000), Q3 (100,000) and Q4 (36,000).

### Conclusion

Our research indicates that both consumers and enterprises understand the potential of immersive, augmented reality (AR) applications, and holographic communication in particular, to improve their lives not only in terms of workplace productivity,

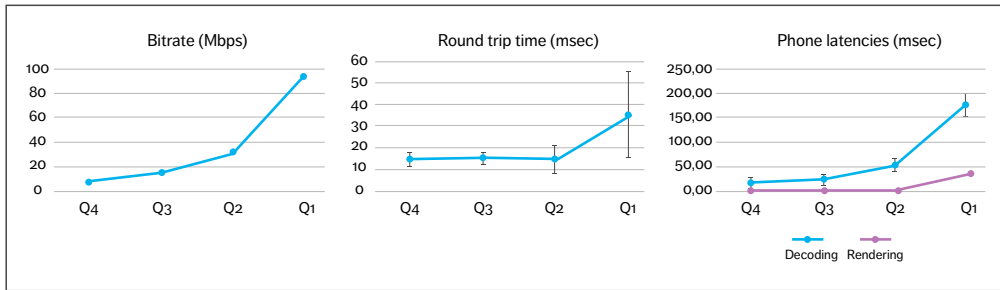


Figure 4 5G NR measurement results

but also when it comes to social interaction and entertainment. The emergence of lightweight AR glasses and powerful 3D compression algorithms have made it possible to start deploying AR use cases using existing 5G technology.

An architecture with edge compute functionality across the full chain is necessary to deliver the 3D captured streams and the rendering of such applications anywhere, while simultaneously meeting device requirements on size, weight and energy consumption. This approach makes it possible to move high performance computing to the network and thereby reduce both the energy consumption of mobile devices and the end-to-end latency.

The introduction of new media formats such as point clouds into extended reality (XR) communication scenarios will greatly enhance the attractiveness, usefulness and efficiency of the information conveyed between communicating parties. Successful communication requires a joint understanding of the communication format, which is typically solved through standardization. In Release 18, the 3GPP has taken on the challenging task of enhancing 5G to offer more efficient support for XR services. Since various aspects of XR are already being addressed in other standardization forums, the intention is to reuse as much as possible.



## Further reading

- » **Ericsson Technology Review, XR and 5G: Extended reality at scale with time-critical communication**, available at: <https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/xr-and-5g-extended-reality-at-scale-with-time-critical-communication>
- » **Ericsson Technology Review, Versatile video coding explained**, available at: <https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/versatile-video-coding-explained>
- » **Ericsson blog, Welcome to XR immersive experiences**, available at: <https://www.ericsson.com/en/blog/2021/11/welcome-to-xr-immersive-experiences>
- » **Ericsson, AI in networks**, available at: <https://www.ericsson.com/en/ai>

## References

1. **Ericsson Consumer and Market Insight report, Five ways to a better 5G**, available at: <https://www.ericsson.com/en/reports-and-papers/consumerlab/reports/five-ways-to-a-better-5g>
2. **Ericsson Consumer and Market Insight report, Busting the myths around the value of 5G for consumers**, available at: <https://www.ericsson.com/en/reports-and-papers/consumerlab/reports/5g-consumer-potential>
3. **Ericsson Consumer and Market Insight report, The dematerialized office**, available at: <https://www.ericsson.com/en/reports-and-papers/industrylab/reports/the-dematerialized-office>
4. **International Organization for Standardization, ISO 9241-11: Ergonomics of Human System Interaction, 1998**, available at: <https://www.iso.org/obp/ui/#iso:std:iso:9241-11:ed-1:v1:en>
5. **IGI Global, Hedonic, emotional, and experiential perspectives on product quality**, in *Encyclopedia of human computer interaction, 2006*, Hassenzahl, M., available at: <https://www.igi-global.com/chapter/hedonic-emotional-experiential-perspectives-product/13133>
6. **Annual review of psychology, Emotion and decision making, 2015**, Lerner, J. S; Li, Y; Valdesolo, P; Kassam, K. S., available at: <https://doi.org/10.1146/annurev-psych-010213-115043>
7. **British Journal of Psychology, Most advanced, yet acceptable: Typicality and novelty as joint predictors of aesthetic preference in industrial design, 2003**, Hekkert, P; Snelders, D; Van Wieringen, P. C., available at: <https://bpspsychub.onlinelibrary.wiley.com/doi/abs/10.1348/000712603762842147>
8. **IEEE MultiMedia, Towards a QoE model to evaluate holographic augmented reality devices, 2018**, Zhang, L; Dong, H; El Saddik, A., available at: <https://doi.org/10.1109/MMUL.2018.2873843>
9. **Interaction Design Foundation, What is Design Thinking?**, available at: <https://www.interaction-design.org/literature/topics/design-thinking>
10. **Facebook 2020 Research: Photorealistic Avatars & Full Body Tracking**, available at: [https://www.youtube.com/watch?v=Q-gse\\_hFkJM](https://www.youtube.com/watch?v=Q-gse_hFkJM)

## THE AUTHORS



### Ali El Essaili

◆ joined Ericsson in 2014 and currently serves as a master researcher at Ericsson Research. His current focus is on XR and enabling immersive media applications on 5G. He holds a Ph.D. in electrical engineering from Munich University of Technology in Germany.



### Sara Thorson

◆ joined Ericsson in 2019 and currently serves as head of concept development at Ericsson Consumer and Industry Lab. Her work focuses on emerging use cases in XR and 6G, with a special interest in sustainability. Thorson holds an M.Sc. in international marketing from the School of Business, Economics and Law at the University of Gothenburg, Sweden.

### Alvin Jude

◆ joined Ericsson in 2014 where he currently serves as a senior researcher at Ericsson Research. Jude holds an M.S. in computer science from Baylor University in Waco, Texas, USA. He specializes in human-computer interaction.



### Jörg Christian Ewert

◆ joined Ericsson in 1999 as a system manager. In 2005, he joined the core network product management team. At present, he is responsible for 5G communication services in Business Area Digital Services. Ewert studied business administration at the FernUniversität in Hagen, Germany, and holds a Ph.D. in physics from the University of Göttingen, Germany.



### Natalya Tyudina

◆ joined Ericsson in 2019 and currently serves as a technology developer at Business Area Digital Services. Her work focuses primarily on XR and Infrastructure-as-a-Service. Tyudina holds an M.Sc. in information systems and technologies from Don State Technical University in Rostov-on-Don, Russia.



### Lukasz Litwic

◆ joined Ericsson in 2007. In his current role, he is the research leader of the Visual Technology team at Ericsson Research. Litwic received an M.Eng. from Gdańsk University of Technology, Poland, in 2005 and a Ph.D. in electronics from the University of Surrey in Guildford, the UK, in 2015.



### Bo Burman

◆ joined Ericsson in 1996. After 14 years at Ericsson Research Media Technologies, he now serves as a senior specialist at Ericsson Digital Services. Burman holds an M.Sc. in computer technology and engineering from the Institute of Technology at Linköping University, Sweden.



The authors would like to thank Chris Phillips, Du Ho Kang, Esra Akan, Fredrik Alriksson, Jonas Kronander, Johan Lundsjö, José Araújo, Jose Luis Pradas, Kenneth Wallstedt, Maria Edvardsson, Maria Serra, Martín Ek, Mauricio Aracena, Michael Björn, Michael Meyer, Nils-Erik Gustafsson and Peter Marshall for their contributions to this article.