Toward Interconnected Virtual Reality: Opportunities, Challenges, and Enablers

Ejder Baştuğ, Mehdi Bennis, Muriel Médard, and Mérouane Debbah

The authors discuss the importance of VR technology as a disruptive use case of 5G (and beyond) harnessing the latest development of storage/memory, fog/edge computing, computer vision, artificial intelligence, and others. In particular, the main requirements of wireless interconnected VR are described, followed by a selection of key enablers; then research avenues and their underlying grand challenges are presented.

Abstract

Just recently, the concept of augmented and virtual reality (AR/VR) over wireless has taken the entire 5G ecosystem by storm, spurring an unprecedented interest from academia, industry, and others. However, the success of an immersive VR experience hinges on solving a plethora of grand challenges cutting across multiple disciplines. This article underscores the importance of VR technology as a disruptive use case of 5G (and beyond) harnessing the latest development of storage/memory, fog/edge computing, computer vision, artificial intelligence, and others. In particular, the main requirements of wireless interconnected VR are described followed by a selection of key enablers; then research avenues and their underlying grand challenges are presented. Furthermore, we examine three VR case studies and provide numerical results under various storage, computing, and network configurations. Finally, this article exposes the limitations of current networks and makes the case for more theory, and innovations to spearhead VR for the masses.

Introduction

Leveraging recent advances in storage/memory, communication/connectivity, computing, big data analytics, artificial intelligence (AI), machine vision, and other adjunct areas will enable the fruition of immersive technologies such as augmented and virtual reality (AR/VR). These technologies will enable the transportation of ultra-high resolution light and sound in real time to another world through the relay of its various sights, sounds, and emotions. The use of VR will go beyond early adopters such as gaming to enhancing cyber-physical and social experiences such as conversing with family and acquaintances, business meetings, and disabled persons. Imagine if one could put on a VR headset and walk around a street where everyone is talking Finnish and interact with people in Finnish in a fully immersive experience. Add to this the growing number of drones, robots, and other self-driving vehicles taking cameras to places humans could never imagine reaching; we shall see a rapid increase of new content from fascinating points of view around the globe. Ultimately, VR will provide the most personal experience with the closest screen, providing the most connected, most immersive experience witnessed thus far.

AR and VR represent two ends of the spectrum. On one hand, AR is based on reality as the main focus, and the virtual information is presented over the reality, whereas VR is based on virtual data as the main focus, immersing the user into the middle of the synthetic reality virtual environment. One can also imagine a mixed reality where AR meets VR, by merging the physical and virtual information seamlessly. Current online social networking sites (Facebook, Twitter, and the like) are just precursors of what we will come to truly witness when social networking encompasses immersive VR technology. At its most basic, social VR allows two geographically separated people (in the form of avatars) to communicate as if they were face to face. They can make eye contact and can manipulate virtual objects that they both can see. Current VR technology is in its inception since headsets are not yet able to track exactly where eyes are pointed, by instead looking at the person to whom one is talking. Moreover, current state-of-the-art VR technology is unable to read detailed facial expressions and senses. Finally, and perhaps the biggest caveat, is that most powerful VR prototypes are wired with cables because the amount of transmitted high-resolution video at high frame rates simply cannot be done using today's wireless technology (4G/LTE), let alone the fact that a perfect user interface (the VR equivalent of the mouse) is still in the making.

These shortcomings have spurred efforts to make social VR happen in the near future. One of a number of startup companies, Linden Lab (a screen-based simulation), is getting ready to roll out a new platform called SANSAR [1], which is a host for user-created virtual experiences and tools for VR headsets, standard computer monitors, and mobile devices. Similarly, the SANSAR world will function much like Second Life, with people leasing space for their virtual creations, rendered in 3D and at a high frame rate. Likewise, BELOOLA [2] is building a virtual world designed for social networking. These recent trends are a clear indication that the era of responsive media is upon us, where media prosumers will adapt content dynamically to match consumers' attention, engagement, and situation. While some of the VR technologies are already emerging (VR goggles, emotion-sensing algorithms, and multi-camera systems), current fourth generation (4G) (or even pre-5G) wireless systems *cannot cope* with the massive amount of bandwidth and latency requirements of VR.

Digital Object Identifier: 10.1109/MCOM.2017.1601089

Ejder Baştuğ and Muriel Médard are with Massachusetts Institute of Technology; Ejder Bastug is also with Université Paris-Saclay; Mehdi Bennis is with the University of Oulu; Mérouane Debbah is with Huawei France R&D and also with Université Paris-Saclay.

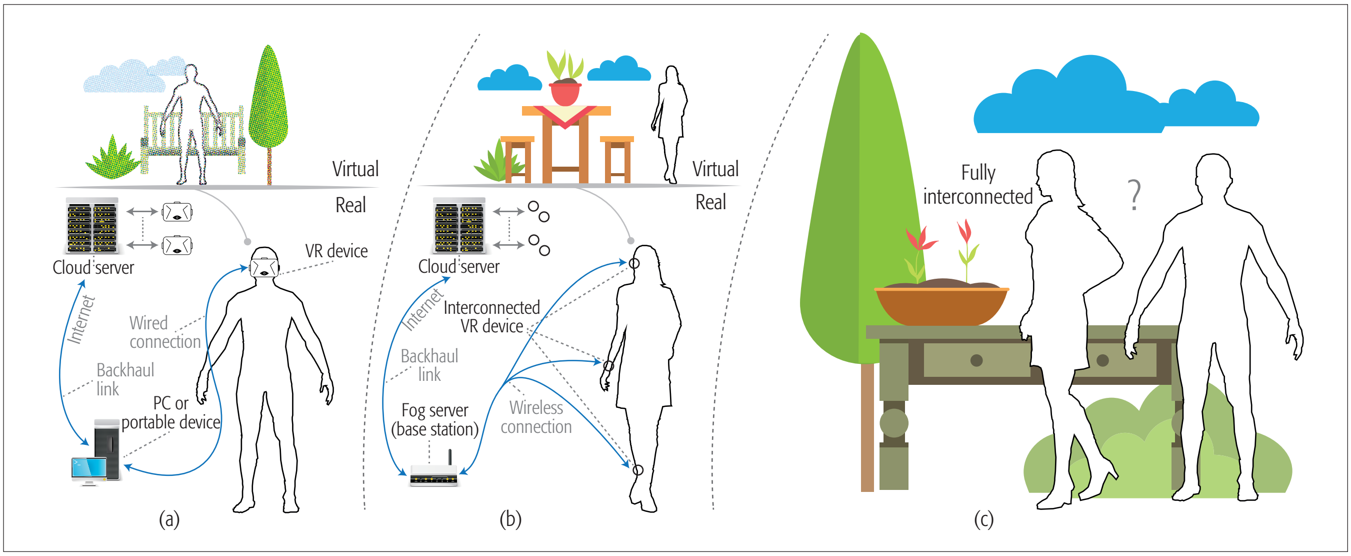

Figure 1. An illustration of virtual reality scenarios: a) current virtual reality systems; b) interconnected; c) ideal (fully interconnected) systems.

The goal of this article is to discuss current and future trends of VR systems, aiming to reach a fully interconnected VR world. It is envisaged that VR systems will undergo three different evolution stages as depicted in Fig. 1, starting with current VR systems, evolving toward interconnected VR (IVR), and finally ending up with the ideal VR system. The rest of this article is dedicated to a discussion of this evolution, laying down some of the key enablers and requirements for the ultimate VR technology. In this regard, we discuss current VR systems and limits of human perception prior to shifting toward interconnected VR and related technological requirements. Key research avenues and scientific challenges are then detailed. Several case studies (with numerical results) are then given. Finally, we debate whether an ideal fully interconnected VR system can be achieved and what might be needed in this regard.

Toward Interconnected VR

The overarching goal of VR is to generate a digital real-time experience that mimics the full resolution of human perception. This entails recreating every photon our eyes see, every small vibration our ears hear, and other cognitive aspects (touch, smell, etc.). Quite stunningly, humans process nearly 5.2 Gb/s of sound and light. The fovea of our eyes can detect fine-grained dots, allowing them to differentiate approximately 200 distinct dots per degree (within our foveal field of view) [3, 4]. Converting that to pixels on a screen depends on the size of the pixel and the distance between our eyes and the screen, while using 200 pixels per degree as a reasonable estimate (see Fig. 2 for an estimate). Without moving the head, our eyes can mechanically shift across a field of view of at least 150° horizontally (i.e., 30,000 pixels) and 120° vertically (i.e., 24,000 pixels). This means the ultimate VR display would need a region of 720 million pixels for full coverage. Factoring in head and body rotation for 360° horizontal and 180° vertical amounts to a total of more than 2.5 billion (giga) pixels. Those are just for a static image.

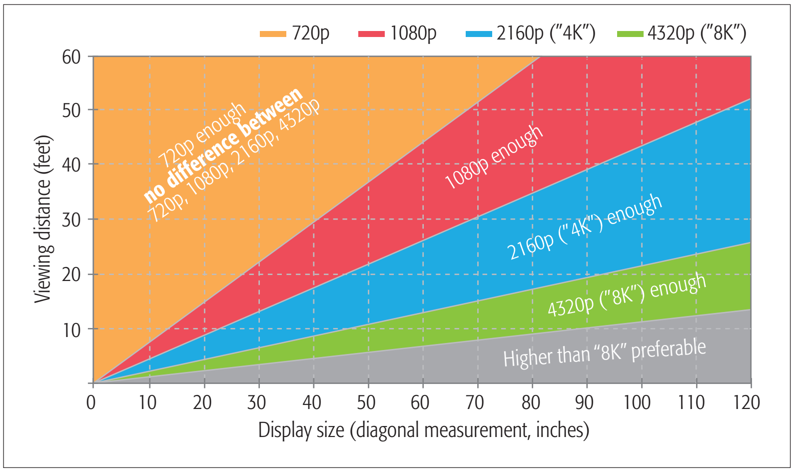

Figure 2. Display size vs. viewing distance (see [5] for an interactive example).

For motion video, multiple static images are flashed in sequence, typically at a rate of 30 images per second (for film and television). But the human eye does not operate like a camera. Our eyes actually receive light constantly, not discretely, and while 30 frames/s is adequate for moderate-speed motion in movies and TV shows, the human eye can perceive much faster motion (150 frames/s). For sports, games, science, and other high-speed immersive experiences, video rates of 60 or even 120 frames/s are needed to avoid motion blur and disorientation. Assuming no head or body rotation, the eye can receive 720 million pixels for each eye, at 36b/pixel for full color and at 60 frames/s, amounting to a total of 3.1 trillion (tera) bits! Today's compression standards can reduce that by a factor of 300, and even if future compression could reach a factor of 600 (the goal of future video standards), that still means 5.2 Gb/s of network throughput (if not more) is needed. While 8K cameras are being commercialized, no cameras or displays to date today can deliver 30K resolution.

As a result, media prosumers are no longer using just a single camera to create experiences.

This year's Super Bowl, for example, was covered by 70 cameras, 36 of which were devoted to a new kind of capture system which allows freezing an action while the audience pans around the center of the action. Previously, these kinds of effects were only possible in video games because they require heavy computation to stitch the multiple views together.

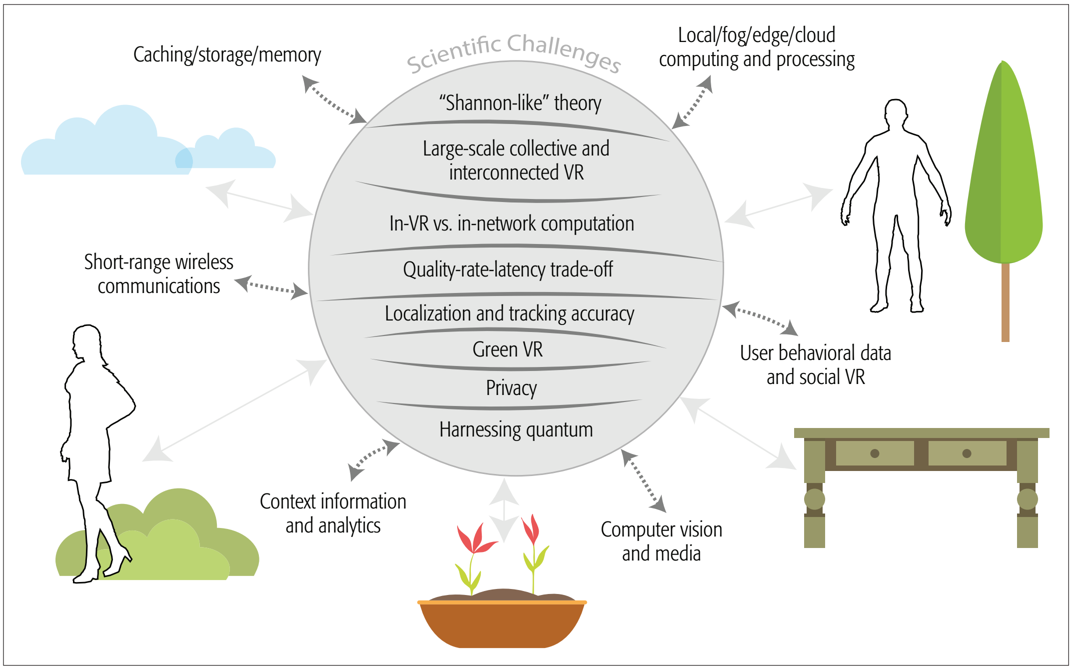

Figure 3. Research avenues and scientific challenges for interconnected VR.

At today's 4K resolution, 30 frames/s and 24 b/pixel, and using a 300:1 compression ratio, yields 300 Mb/s of imagery. That is more than 10× the typical requirement for a high-quality 4K movie experience. While panorama camera rigs face outward, there is another kind of system where the cameras face inward to capture live events. This year's Super Bowl, for example, was covered by 70 cameras, 36 of which were devoted to a new kind of capture system that allows freezing an action while the audience pans around the center of the action. Previously, these kinds of effects were only possible in video games because they require heavy computation to stitch the multiple views together. Heavy duty post-processing means such effects are unavailable during live action.

As a result, 5G network architectures are being designed to move the post-processing to the network edge so that processors at the edge and the client display devices (VR goggles, smart TVs, tablets, and phones) carry out advanced image processing to stitch camera feeds into dramatic effects.

To elaborate the context of current networks, even with a dozen or more cameras capturing a scene, audiences today only see one view at a time. Hence, the bandwidth requirements would not suffice to provide an aggregate of all camera feeds. To remedy this, dynamic caching and multicasting may help alleviate the load by delivering content to thousands from a single feed. In a similar vein with the path toward user equipment (UE) centricity, VR will instead let audiences dynamically select their individual points of view. This means that the feed from all of the cameras needs to be available instantly and at the same time, meaning that conventional multicast will not be possible when each audience member selects an individualized viewpoint (unicast). This will cause outage and users' dissatisfaction.

Technological Requirements

In order to tackle these grand challenges, the 5G network architecture (radio access network [RAN], edge, and core) will need to be much smarter than ever before by adaptively and dynamically making use of concepts such as software defined network (SDN), network functions virtualization (NFV), and network slicing, to mention a few, facilitating more flexibly allocating resources (resource blocks [RBs], access points, storage, memory, computing, etc.) to meet these demands. In parallel to that, video/audio compression technologies are being developed to achieve much higher compression ratios for new multi-camera systems. Whereas conventional video compression exploits the *similarity of the images* between one frame and the next (*temporal redundancy*), VR compression adds to that and leverages similarity among images from different cameras (including the sky, trees, large buildings and others, called *spatial redundancy*) and use intelligent slicing and tiling techniques, using less bandwidth to deliver full 360° video experiences. All of these advances may still not be enough to reach the theoretical limits of a fully immersive experience. Ultimately, a fundamentally new network architecture is desperately needed that can dynamically multicast and cache multiple video feeds close to consumers and perform advanced video processing within the network to construct individualized views.

Immersive technology will require massive improvements in terms of *bandwidth*, *latency*, and *reliability*. A current remote reality prototype (MirrorSys [6]) requires 100–200 Mb/s for a one-way immersive experience. While MirrorSys uses a single 8K, estimates on photo-realistic VR will require two 16K × 16K screens (one for each eye). Latency is the other big issue in addition to reliability. With an AR headset, for example, real-life visual and auditory information has to be taken in through the camera and sent to the fog/cloud for processing, with digital information sent back to be precisely overlaid onto the real-world environment, and all this has to happen in less time than it takes for humans to start noticing lag (no more than 13 ms [7]). Factoring in the much needed high reliability criteria on top of these bandwidth and delay requirements clearly indicates the need for interactions between several research disciplines. These research avenues are discussed in the following.

Key Research Avenues and Scientific Challenges

The success of interconnected VR hinges on solving a number of research and scientific challenges across network and devices with heterogeneous capability of storage, computing, vision, communication, and context awareness. These key research directions and scientific challenges are summarized in Fig. 3 and discussed as follows.

Caching/Storage/Memory

The concept of content caching has recently been investigated in great details [8], where the idea is to cache strategic contents at the network edge (at a base station [BS], devices, or other intermediate locations). One distinguishes between reactive and proactive caching. While the former serves end users when they request contents, the latter is proactive and anticipates users' requests. Proactive caching depends on the availability of fine-grained spatio-temporal traffic predictions. Other side information such as the user's location, mobility patterns, and social ties can be further exploited especially when context information is sparse. Storage will play a crucial role in VR where, for instance, upon the arrival of a task query, the network/server needs to swiftly decide whether to store the object if the same request will come in the near future or instead recompute the query from scratch if the arrival rate of the queries will be sparse in the future. Content/media placement and delivery will also be important in terms of storing different qualities of the same content at various network locations [9, 10].

Local/Fog/Edge/Cloud Computing and Processing

Migrating computationally intensive tasks from VR devices to more resourceful cloud/fog servers is necessary to increase the computational capacity of low-cost devices while saving battery energy. For this purpose, mobile edge computing (MEC) will enable devices to access cloud/fog resources (infrastructures, platforms, and software) in an on-demand fashion. While current state-of-the-art solutions allocate radio and computing resources in a centralized manner (at the cloud), for VR both radio access and computational resources must be brought closer to VR users by harnessing the availability of dense small cell base stations with proximity access to computing/storage/memory resources. Furthermore, the network infrastructure must enable a fully distributed cloud immersive experience where a lot of the computation happens on very powerful servers that are in the cloud/edge while sharing the sensor data that are being delivered by end-user devices at the client side. In the most extreme cases, one can consider the computation at a very local level, say with fully/partially embedded devices in the human body, having computing capabilities. This phenomenon is commonly referred to as "skin computing."

Short-Range Wireless Communications

Leveraging short- range communication such as device-to-device (D2D) and edge proximity services among collocated VR users can help alleviate network congestion. The idea is to extract, stitch, and share relevant contextual information among VR users in terms of views and camera feeds. In the context of self-driving vehicles equipped with ultra high definition (UHD) cameras capturing their local neighborhood, the task for the vehicle/robot is to not only recognize objects/faces in real time, but also decide which objects should be included in the map and share it with nearby vehicles for richer and more context-aware maps.

Computer Vision and Media

The advent of UHD cameras (8K, new cameras with 360° panoramic video) has enriched new video and media experiences. At the same time, today's media content sits at two extreme ends of a spectrum. On one hand, one distinguishes "lean-back experiences" such as movies and television where consumers are passive and are led through a story by content authors/producers. On the other hand are "lean-forward" experiences in the form of games in which the user is highly engaged and drives the action through an environment created by content authors/producers. The next generation of "interactive media" where the narrative can be driven by authors/producers will be tailored dynamically to the situation and preferences of audience and end users.

Context Information and Analytics

Use of context information has already been advocated as a means of optimizing complex networks. Typically, context information refers to in-device and in-network side information (user location, velocity, battery level, and other medium access control [MAC]/high layer aspects). In the context of VR, the recent acquisition of Apple of Emotient, a company using advanced computer vision to recognize the emotions of people, serves as a clear indication that context information will play an ever more instrumental role in spearheading the success of VR. In order to maximize the user's connected and immersive experience, the emotional, user switchiness, and other behavioral aspects must be factored in. This entails predicting users' disengagement and preventing it by dynamically shifting the content to better match individual preferences, emotional state, and situation. Since a large amount of users' data in the network can be considered for the big data processing, tools from machine learning can be exploited to infer from the context information of users and act accordingly. Of particular importance is the fact that deep learning models have recently been on the rise in machine learning applications, due to their human-like behavior in training and good performance in feature extraction.

Leveraging short- range communication such as device-to-device (D2D) and edge proximity services among collocated VR-users can help alleviate network congestion. The idea is to extract, stitch and share relevant contextual information among VR users in terms of views and camera feeds.

With users contributing to the world with different content and having multiple views from billions of objects and users, the issue of privacy naturally takes center stage. Intelligent mechanisms that automatically preserve privacy without overburdening users to define their privacy rules have yet to be developed.

User Behavioral Data and Social VR

A by-product of the proliferation of multiple screens is that the notion of switchiness is more prevalent, in which users' attention goes from one screen to another. Novel solutions based on user behavioral data and social interactions must be thought of to tackle a user's switchiness. For this purpose, the switchiness and screen chaos problems have basically the same answer. An immersive experience is an integrated experience which needs a data-driven framework that takes all of the useful information a person sees and brings it to a single place. Today that integration does not happen because there is no common platform. VR mandates that all of these experiences take place in one place. If one is watching a movie or playing a game and gets a phone call, the game (or movie) is automatically paused, and the person need not have to think about pausing the movie and answering the phone. Considering that a common data-driven platform is taking in one place, big data and machine learning tools will play a crucial role in bringing the immersive experience to users.

Scientific Challenges

The goal of this subsection is to lay down the foundations of VR, by highlighting the key different research agenda and potential solution concepts for its success.

Need for a "Shannon-Like" Theory: For a given VR device of S bits of storage, E joules of energy, and C Hertz of processing power, how do we maximize the user's immersive experience or alternatively minimize VR users' switchiness? The answer depends on many parameters such as the VR device-server air link, whether the VR device is a human or a robot, network congestion, in-VR processing, VR cost (how much intelligence can be put at the VR headset), distinction between massive amount of VR devices transmitting few bits vs. a few of them sending ultra-high definition to achieve a specific task. In this regard, haptic code design for VR systems, code construction to minimize delay in feedback scenarios [11], source compression under imperfect knowledge of input distribution, and granularity of learning the input distribution in source compression become relevant. Moreover, Nyquist sampling with no prior knowledge, compressed sensing with partial structural knowledge, and source coding with complete knowledge are some key scientific venues that can address many challenges in VR networks.

Large-Scale Collective and Interconnected VR: The analysis of very large VR networks and systems, most of them moving, is also of high interest. With so many different views and information, lots of redundancy and collective intelligence is open to exploitation for the interconnected VR.

In-VR vs. In-Network Computation: This refers to where and to which level should the decoupling between the in-VR headset and in-network computing happen. This depends on the bandwidth-latency-cost-reliability trade-offs, where computing for low-end and cheap headsets needs to happen at the network side, whereas for more sophisticated VR headsets computing can be carried locally.

Quality-Rate-Latency Trade-off: Given an underlying network topology, and storage and communication constraints, what is the quality level per content that should be delivered to maximize the quality of an immersive VR experience? This builds on the works of Bethanabhotla *et al.* [9] by taking into account the video size and quality as a function of the viewing distance. Moreover, for a given latency and rate constraints, what is the optimal payload size for a given content to maximize information dissemination rate (in the case of self-driving vehicles). Moreover, machine learning is key for object recognition and stitching different video feeds. For self-driving vehicles, given an arbitrary number of vehicles, this involves network congestion and wireless links among vehicles, a central processing unit (CPU), storage constraint, and vehicles aiming to exchange their local maps. Fundamentally speaking, for a fixed packet size of L bits, what objects need to be recognized/quantized and included in the map? For example, the map should store popular objects that have been requested a lot in the past.

Localization and Tracking Accuracy: For a fully immersive VR experience, very accurate localization techniques are needed, including the positions of objects, tracking of human eyes (i.e., gaze tracking), and so on.

Green VR: For a given target VR user's immersive experience, the goal is to minimize the power consumption in terms of storage, computing, and communication. With the green interconnected VR, the notion of "charging" the equipment should disappear or at least be minimized since this operation does not exist in the virtual world. Therefore, smart mechanisms for seamless charging of VR devices (i.e., wireless power transfer/charging and energy harvesting) are promising.

Privacy: With users contributing to the world with different content and having multiple views from billions of objects and users, the issue of privacy naturally takes center stage. Intelligent mechanisms that automatically preserve privacy without overburdening users to define their privacy rules have yet to be developed. New emerging concepts such as "collective privacy" are interesting [12].

Harnessing Quantum: Exploiting recent advances in quantum computing could enable this giant leap where certain calculations can be done much faster than any classical computer could ever hope to do.For VR, quantumness could be leveraged for:

- Bridging virtual and physical worlds, where the classical notion of locality no longer matters
- In terms of computation power, where instead of serial or even parallel computation/processing, quantum allows to calculate/compute high-dimensional objects in lower dimensions, exploiting entanglement and superposition

This can be instrumental for self-driving vehicles where latency is crucial; therein, quantum computing empowers vehicles to recognize and categorize a large number of objects in a real-time manner by solving highly complex pattern recognition problems on a much faster timescale.

Numerical Results

In this section, in light of the aforementioned challenges, we examine a number of case studies focusing on some of the fundamentals of AR/VR. Let us suppose that an arbitrary number of AR/VR devices are connected to M fog servers (or base stations) via wireless links with total link capacity of L_{wi} Mb/s. These fog servers are connected to a cloud computing service (and the Internet) via backhaul links with total link capacity of L_{ba} MB/s. Each AR/VR device and fog base station has computing capabilities of C_{vr} and C_{fg} GHz, respectively, and the cloud has computational power of C_{cl} GHz. In the numerical setups of the following case studies, the arrival process of AR/VR devices shall follow a Shot Noise Model [10] with a total time period of T_{max} hours. This model conveniently aims to capture spatio-temporal correlations, where each shot is considered as a VR device that stays in the network for a duration of T ms, and each device has μ mean number of task requests drawn from a power-law distribution [13] with exponent α. Requested tasks are computed at different locations of the network, which could be locally at the VR device or (edge) fog server or globally at the cloud. Depending on where the requested task is computed, computational and delivery/communication costs are incurred, following power-law distributions parameterized by means μ_{co} giga cycles, μ_{de} MB, and power-law exponents (or steepness factors) α_{co} and α_{de}, respectively. Moreover, computation and delivery of a task incur delays. As a main performance metric, the *immersive experience* is defined as the percentage of tasks that are computed and delivered under a specific deadline, where each deadline is drawn from a power-law distribution with mean μ_{dl} = 10 ms and a steepness factor α_{dl}. Such a definition of immersive experience is analogous to coverage/outage probability used in the literature, where the aforementioned target task deadline with mean of 10 ms is imposed for users/humans to avoid noticing lag (no more than 13 ms in reality [7]). A set of default parameters[1] is considered throughout the case studies unless otherwise stated. These parameters are to set these values such that a realistic network with limited storage, computation, and communication capacities is mimicked.

Case Study I: Joint Resource Allocation and Computing

The goal is to maximize a user's immersive experience by minimizing a suitable cost function. This optimization problem hinges on many parameters such as the wireless link between the VR device and the server (or cluster of servers), whether the VR device is a human or a robot, network congestion, in-VR processing power/storage/memory, and cost (how much intelligence can be embedded in the VR device).

The evolution of the immersive experience with respect to the arrival density of VR devices is depicted in Fig. 4. The tasks are computed at three different places (i.e., locally at VR devices, at fog base stations, or globally at the cloud) with different percentages, in order to show the possible gains. As the arrival density of tasks increases, one can easily see that the immersive experience

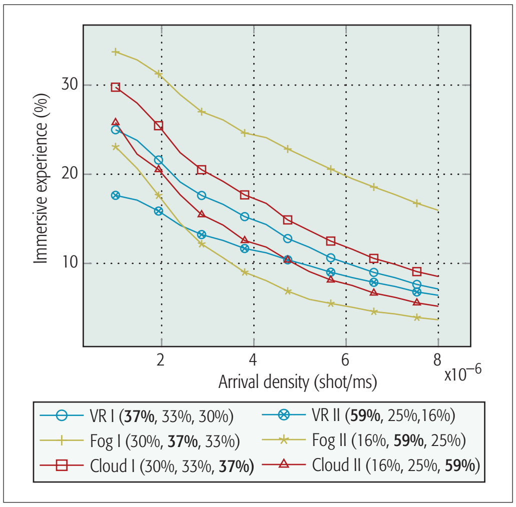

Figure 4. Evolution of the immersive experience with respect to the load, with different configurations of VR, fog, and cloud-centric computations: VR I (C_{vr} = 2 × 3.2 GHz), VR II (C_{vr} = 1 × 3.2 GHz), Fog I (C_{fg} = 256 × 4 × 3.4 GHz, L_{wi} = 1024 Mb/s), Fog II (C_{fg} = 16 × 4 × 3.4 GHz, L_{wi} = 256 Mb/s), Cloud I (C_{cl} = 1024 × 4 × 3.4 GHz, L_{ba} = 512 Mb/s), Cloud II (C_{cl} = 128 × 4 × 3.4 GHz, L_{ba} = 16 Mb/s). The triple (., ., .) given in the legend represents the percentage of tasks computed at the VR devices, fog base stations, and cloud, respectively.

There exist regimes where VR-centric computations outperform others (i.e., VR II vs Fog II), especially for higher task arrival densities. The results indicate the need for a principled framework that jointly allocate resources (radio, computing) in various network locations subject to latency and reliability constraints.

decreases due to the limited computing and communication resources in the network causing higher delays. In this configuration with 10 ms average delay deadline/requirement, computing at the fog base stations outperforms other approaches, as seen in the figure. For instance, with an arrival density of 0.42 VR devices/ms, Fog I provides 16 percent more immersive experience gains compared to other configurations. However, there are regimes where VR-centric computations outperform others (i.e., VR II vs. Fog II), especially for higher task arrival densities. The results indicate the need for a principled framework that jointly allocates resources (radio, computing) in various network locations subject to latency and reliability constraints.

Case Study II: Proactive vs. Reactive Computing

Related to the local vs. edge computing challenge above, the goal of cloud service providers is to enable tenants to elastically scale resources to meet their demands. While running cloud applications, a tenant aiming to minimize her/his cost function is often challenged with crucial tradeoffs. For instance, upon each arrival of a task, an application can either choose to pay for CPU to compute the response, or pay for cache storage to store the response to reduce future compute costs. Indeed, a reactive computing approach would wait until the task request reaches the server for computation, whereas the proactive computing approach proactively leverages the fact that several requests/queries will be made for the same computation, and thus it stores the result of the computation in its cache to avoid recomputing the query at each time instant. This fundamental observation is analyzed next.

[1] M = 4, L_{ba} = 512 Mb/s, L_{wi} = 1024 Mb/s, C_{vr} = 4 × 3.4 GHz, C_{fg} = 128 × 4 × 3.4 GHz, C_{cl} = 1024 × 4 × 3.4 GHz T_{max} = 1 hour, T = 4 ms, μ = 4 tasks, α = 0.8, mco = 100 Giga cycles, α_{co} = 0.48, μ_{de} = 100 MBit, α_{de} = 0.48, μ_{dl} = 10 ms, α_{dl} = 0.48.

Self-driving or autonomous vehicles represent one of the most important use case for 5G where latency, bandwidth and reliability are prime concerns. Self-driving vehicles need to exchange information derived from multi-resolution maps created using their local sensing modalities, extending their visibility beyond the area directly sensed by its own sensors.

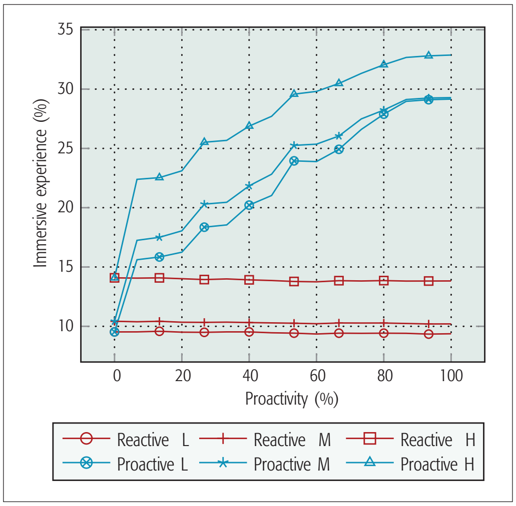

Figure 5. Evolution of the immersive experience with respect to the proactivity. Low (L), medium (M), and high (H) homogeneity settings for reactive and proactive computation of tasks at the fog servers are considered: Reactive L (α = 0.1, L_{ba} = 64 Mb/s) Reactive M (α = 0.6, L_{ba} = 64 Mb/s) Reactive H (α = 0.8, L_{ba} = 64 Mb/s) Proactive L (α = 0.1, L_{ba} = 64 Mb/s) Proactive M (α = 0.6, L_{ba} = 64 Mb/s) Proactive H (α = 0.8, L_{ba} = 64 Mb/s). The place of computations for all settings is fixed to (16 percent, 25 percent, 59 percent).

The evolution of the immersive experience with respect to the level of proactivity at the fog base stations is shown in Fig. 5. We assume proactive settings with storage size of S = 0 percent in case of zero proactivity and S = 100 percent in case of full proactivity, whereas the computation results of popular tasks are cached in the fog base stations for a given storage size. As seen in the figure, proactivity substantially increases the immersive experience, and further gains are obtained when the computed tasks are highly homogeneous (i.e., Proactive H). The gains in the reactive approaches remain constant as there is no proactivity, whereas a slight improvement in highly homogeneous case (i.e., Reactive H) is observed due to the homogeneous tasks that are prone to less fluctuations in deadlines. As an example, 80 percent of proactivity in Proactive H yields higher gains up to 22 percent as compared to Reactive L. This underscores the compelling need for proactivity in VR systems.

Case Study III: AR-Enabled Self-Driving Vehicles

Self-driving or autonomous vehicles represent one of the most important use cases for 5G where latency, bandwidth and reliability are prime concerns. Self-driving vehicles need to exchange information derived from multi-resolution maps created using their local sensing modalities (radar, lidar, or cameras), extending their visibility beyond the area directly sensed by its own sensors. The problems facing the vehicles are many-fold:

- How to control the size of the message (payload) exchanged with other vehicles based on traffic load, interference, and other contextual information
- How to control the content of the message (at what granularity should a given object be included in the message, the most popular object? the least requested object? at what timeliness? etc.)
- How to recognize objects and patterns reliably and in real time

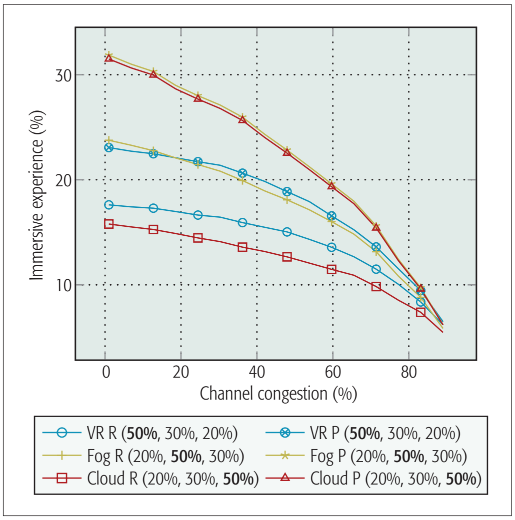

Figure 6. Evolution of the immersive experience with respect to the channel congestion, where fully reactive (R) and proactive (P) configurations of VR, fog, and cloud-centric computation are considered. Fully reactive configuration has S = 0 percent, L_{ba} = 64 Mb/s; and the proactive configuration has S = 80 percent, L_{ba} = 64 Mb/s.

The evolution of the immersive experience with respect to the wireless channel link congestion between base stations and AR-enabled self-driving vehicles is depicted in Fig. 6. The fact that higher channel congestion degrades the immersive experience in all settings is evident; however, proactivity can still provide additional improvements as compared to the reactive settings. Proactive cloud and fog oriented computation yield gains up to 11 percent when the congestion is 42 percent. This shows the need for proactivity in self-driving vehicles as well as dynamic placement of computation depending on the AR/VR network conditions.

Delving into these case studies, which show the potential of interconnected VR, we finally come to the following question.

Are We Going to Live in the "Matrix"?

One speculative question that can be raised is whether an interconnected VR can reach a maturity level so that no distinction between real and virtual worlds are made in human perception, making people end up with the following question: Are we living in a computer simulation?

Despite historical debates, several science fiction movies have raised similar points (e.g., *The Matrix*); many philosophical discussions have been carried out [14], concepts like "simulated reality" have been highlighted [15], and despite all of these, many technical and scientific challenges remain unclear/unsolved. In the context of VR, we call this unreachable phenomenon *ideal* (*fully interconnected*) VR. In fact, in the realm of ideal VR, one might think of living in a huge computer simulation with zero distinction/switching between real and virtual worlds. In this ideal VR environment, the concepts of skin/edge/fog/cloud computing might be merged with concepts like quantum computing.

Indeed, in ideal VR with no distinction between real and virtual worlds, we are not aiming to introduce a paradoxical concept and provide recursive arguments with a mixture/twist of ideas. Instead, we argue whether we can reach such a user experience with VR, therefore achieving an ideal (fully-interconnected) case. Despite the fact that we do not know the exact answer, we keep the ideal VR as a reference to all interconnected VR systems. Undoubtedly, the future lies in interconnected VR, despite its research and scientific challenges, which will continue to grow in importance over the next couple of years.

Despite the fact that we do not know the exact answer, we keep the ideal VR as a reference to all interconnected VR systems. Undoubtedly, the future lies in interconnected VR, despite its research and scientific challenges which will continue to grow in importance over the next couple of years.

Acknowledgment

This research has been supported by the ERC Starting Grant 305123 MORE (Advanced Mathematical Tools for Complex Network Engineering), the U.S. National Science Foundation under Grant CCF-1409228, and the Academy of Finland CARMA project. Nokia Bell-Labs is also acknowledged for its donation to the "FOGGY" project.

References

[1] Project SANSAR; http://www.projectsansar.com, accessed Feb. 15, 2017.
[2] Beloola; http://www.beloola.com, accessed Feb. 15, 2017.
[3] A. König, "Die abhängigkeit der sehschärfe von der beleuchtungsintensität, sitzgsber. preuss," Akad. Wiss., Physik.-math. Kl, 1897.
[4] M. H. Pirenne, *Vision and the Eye*, Chapman & Hall, vol. 47, 1967.
[5] RTRings, "TV Size to Distance Relationship"; http://www.rtings.com/tv/learn/size-to-distancerelationship, (Accessed on 15-02-2017).
[6] "Mirrorsys;" available: http://www.huawei.com/minisite/mwc2015/en/mirrorsys.html, accessed Feb. 15, 2017.
[7] M. C. Potter *et al.*, "Detecting Meaning in RSVP at 13 ms per Picture," *Attention, Perception, & Psychophysics*, vol. 76, no. 2, 2014, pp. 270–79.
[8] E. Bastug, M. Bennis, and M. Debbah, "Living on the Edge: The Role of Proactive Caching in 5G Wireless Networks," *IEEE Commun. Mag.*, vol. 52, no. 8, Aug. 2014, pp. 82–89.
[9] D. Bethanabhotla, G. Caire, and M. J. Neely, "Adaptive Video Streaming for Wireless Networks with Multiple Users and Helpers," *IEEE Trans. Commun.*, vol. 63, no. 1, Jan. 2015, pp. 268–85.
[10] G. Paschos *et al.*, "Wireless Caching: Technical Misconceptions and Business Barriers," *IEEE Commun. Mag.*, vol. 54, no. 8, Aug. 2016, pp. 16–22.
[11] D. E. Lucani, M. Medard, and M. Stojanovic, "On Coding for Delay-Network Coding for Time-Division Duplexing," *IEEE Trans. Info. Theory*, vol. 58, no. 4, Apr. 2012, pp. 2330–48.
[12] A. C. Squicciarini, M. Shehab, and F. Paci, "Collective Privacy Management in Social Networks," *Proc. 18th ACM Int'l. Conf. World Wide Web*, 2009, pp. 521–30.
[13] M. Leconte *et al.*, "Placing Dynamic Content In Caches with Small Population," *Proc. IEEE INFOCOM*, 2016.
[14] N. Bostrom, "Are We Living in a Computer Simulation?" *Philosophical Quarterly*, vol. 53, no. 211, 2003, pp. 243–55.
[15] J. D. Barrow, "Living in a Simulated Universe"; ahttp://goo.gl/jTKNng, 2007, accessed Feb. 15, 2017).

Biographies

Ejder Baştuğ is currently a postdoctoral researcher at MIT and CentraleSupélec, working with Muriel Médard and Mérouane Debbah, respectively. He was the recipient in 2015 of the Supélec Foundation's Best Ph.D. Thesis/Publications Prize, the 2016 IEEE ComSoc Best Tutorial Paper Award, and the 2017 EURASIP Best Paper Award. His research interests are in the field of information theory, machine learning, and wireless communications, with particular focus on functional compression and distributed caching.

Mehdi Bennis is an adjunct professor at the University of Oulu, Finland, and an Academy of Finland research fellow. In 2004, he joined the CWC at the University of Oulu as a research scientist. His main research interests are in radio resource management, heterogeneous and small cell networks, game theory, and machine learning. He has published more than 90 research papers in international conferences, IEEE journals, book chapters, books, and patents.

Mérouane Debbah [F] is a full professor at CentraleSupelec, France. Since 2014, he is also vice-president of the Huawei France R&D Center and director of the Mathematical and Algorithmic Sciences Lab. His research interests lie in fundamental mathematics, algorithms, statistics, and information and communication sciences research. He is a WWRF Fellow and a member of the academic senate of Paris-Saclay. During his career, he has received 17 Best Paper Awards for his research contributions.

Muriel Médard [F] is the Cecil H. Green Professor in the EECS Department at MIT and leads the Network Coding and Reliable Communications Group at the RLE at MIT. She has served as an Editor for many publications of the IEEE, and she is currently Editor-in-Chief of *IEEE JSAC*. She has served as TPC Co-Chair of many of the major conferences in information theory, communications, and networking.