

Challenges in Networking to Support Augmented Reality and Virtual Reality

Cedric Westphal
Huawei

IETF98- ICNRG Meeting - Thursday 3/30/17



UNIVERSITY OF CALIFORNIA
SANTA CRUZ



Trends in AR/VR

- **Augmented Reality and Virtual Reality are going to create tremendous growth in traffic going forward**
- **Multiple phases:**
 - **Low hanging fruits, phone-based applications**
 - **Head-mounted gear connected to local compute resource**
 - **Say, to Playstation, Xbox or computer through HDMI interface**
 - **360 video streaming, immersive video streaming**
 - **Requires more bandwidth...**
 - **Then... Mobile, networked uses of AR/VR**



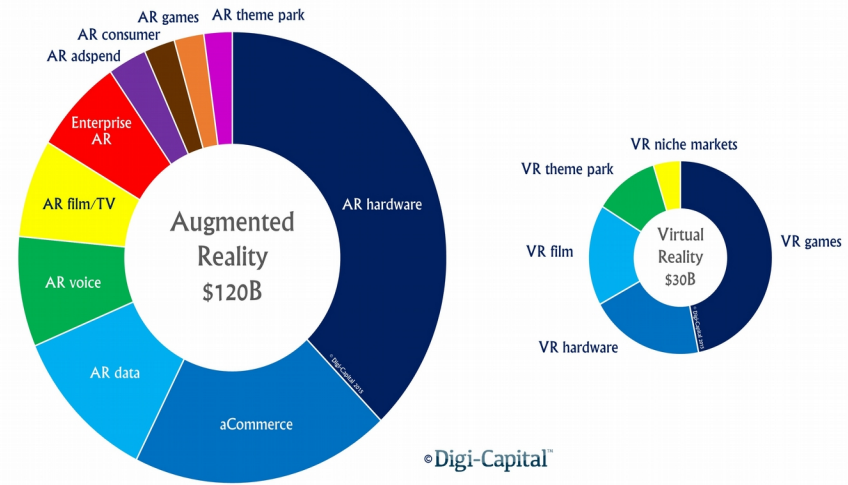
Trends in AR/VR: projections

Worldwide Augmented and Virtual Reality Hardware Shipments and 5-Year CAGR (shipments in millions)

Product Category	2016	2020	2016-2020
	Shipments	Shipments	CAGR
Virtual Reality Hardware	9.6	64.8	183.8%
Augmented Reality Hardware	0.4	45.6	189.8%

Source: IDC, April 2016

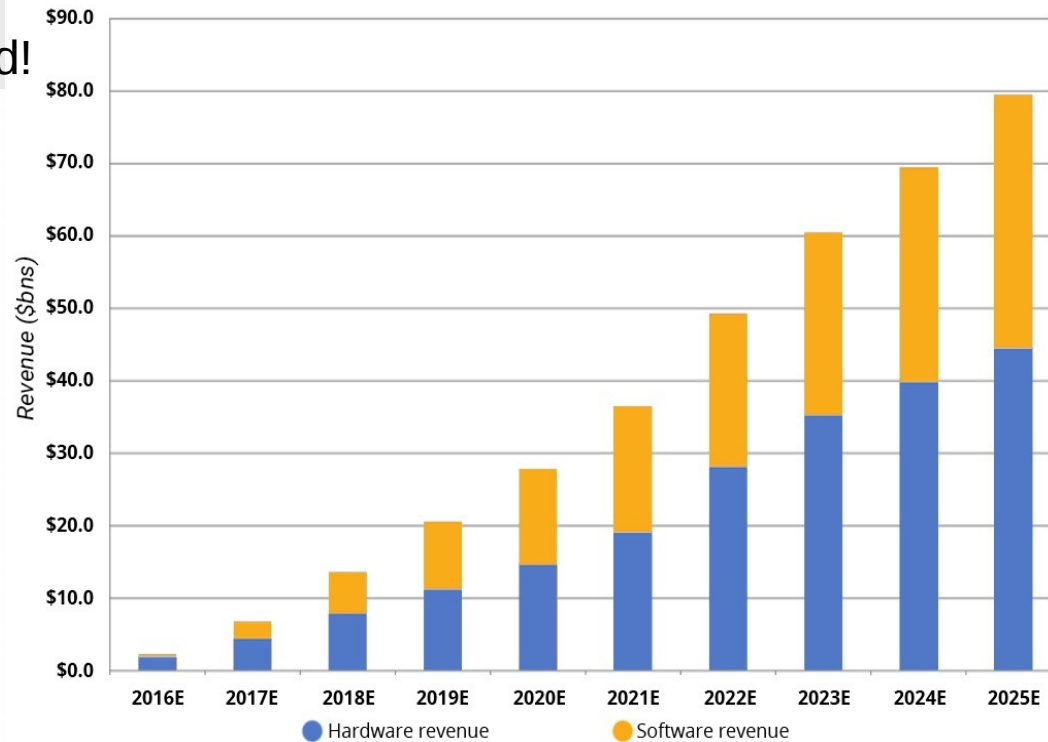
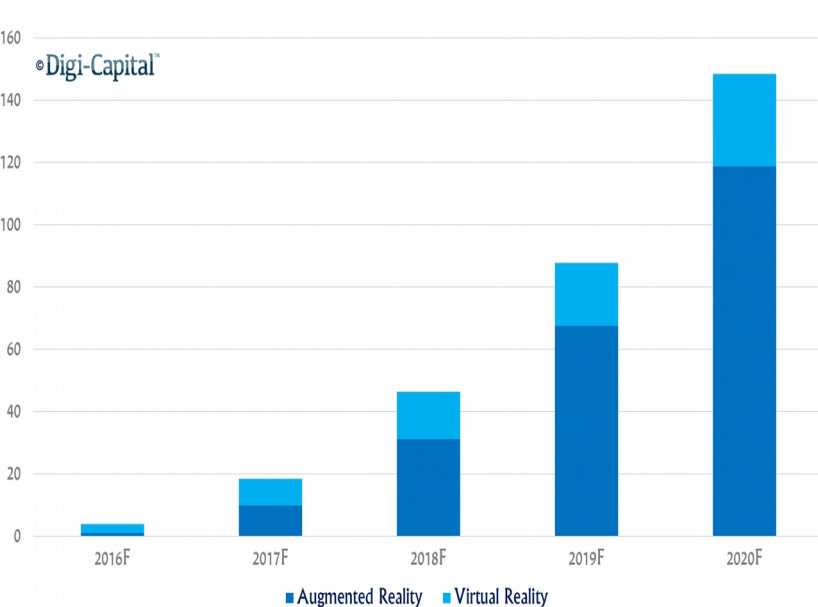
Augmented/Virtual Reality Revenue Share 2020F



VR AND AR MARKET FORECAST ACCORDING TO GOLDMAN SACHS

As sales go, so goes bandwidth demand!

Augmented/Virtual Reality Revenue Forecast (\$B)



AR/VR: definitions

Augmented Reality: an AR system inserts a virtual layer over the user's perception of the real objects, which combines both real and virtual objects in such a way that they function in relation to each other, with synchronicity and the proper depth of perception in three dimensions.

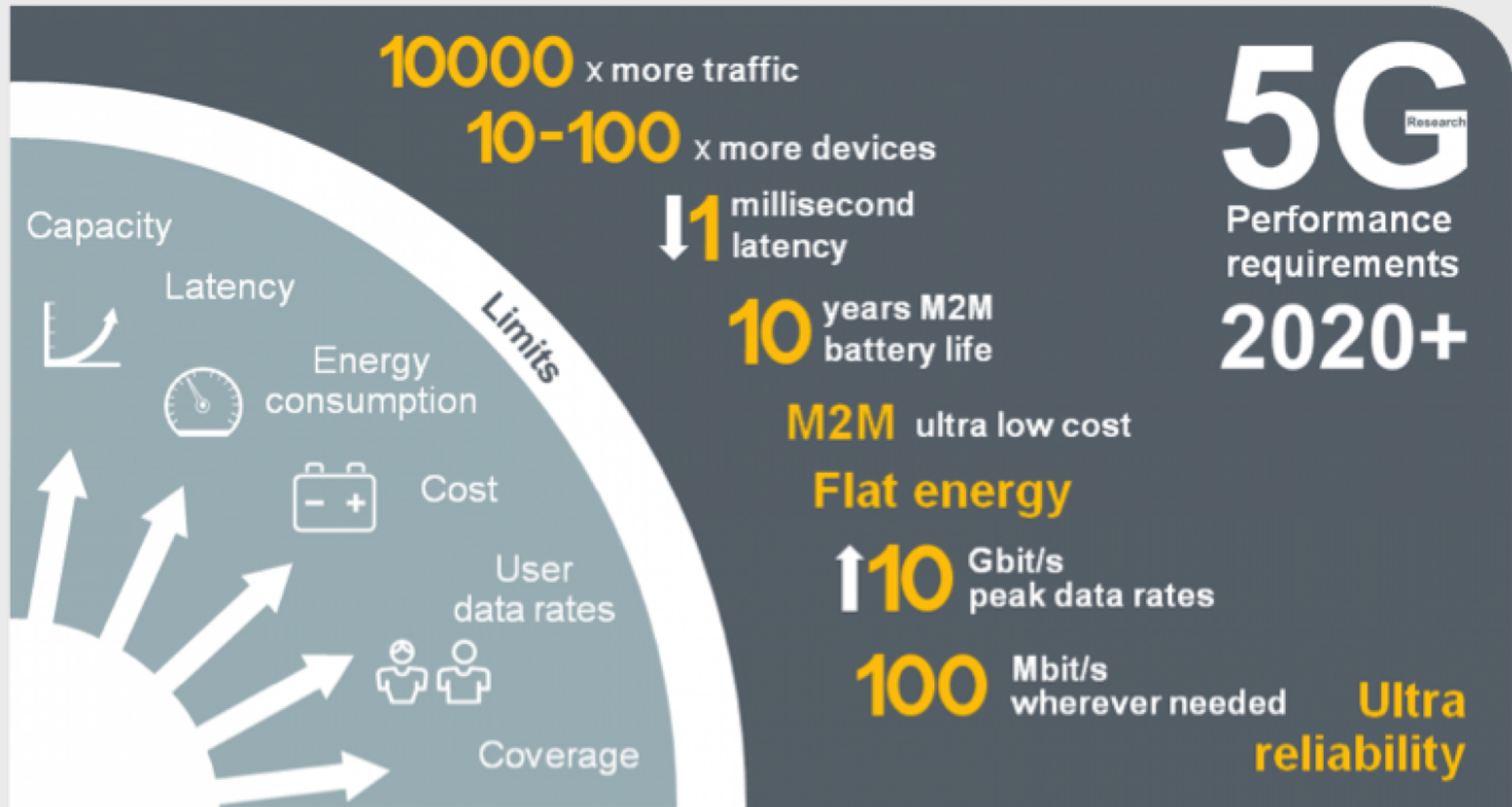
Virtual Reality: a VR system places the user in a synthetic, virtual environment with a coherent set of rules and interactions within this environment and with the other participants in this environment.

Spectrum that includes mixed reality

Use cases from a networking PoV

- **Office productivity, personal movie theater:** canonical use case, head mounted device is only a display.
 - little networking requirements, as all is collocated and could even be wired.
- **Retail, Museum, Real Estate, Education:** recreates the experience of being in a specific area, such as a home for sale, a classroom or a specific room in a museum.
 - files may be stored locally, can be processed ahead of time.
 - how to move the virtual environment onto the display. Prefetching? Local cache?
- **Sports:**
 - (pseudo-)real time, as live event; scale, as many users...
 - how to distribute live content in a timely manner that still corresponds to the potentially unique field of view of each of the users; how to scale this distribution to a large number of concurrent experiences.
- **Gaming:**
 - virtual environment, with interactions in between the different participants
 - Synchronization in between users, responsiveness, consistency
- **Maintenance, Medical, Therapeutic:** overlay instructions on top of equipment so as to assist the agent in performing maintenance.
 - Stringent synchronization and round-trip time requirements, both on the display and on the sensors capturing the motion and position.
- **Augmented maps and directions, facial recognition, teleportation:**
 - general scenario of AR: absorbs the environment of the user and annotates it
 - Patterns/people recognition; tight delay constraints; back-end processing; significant up- and down streams

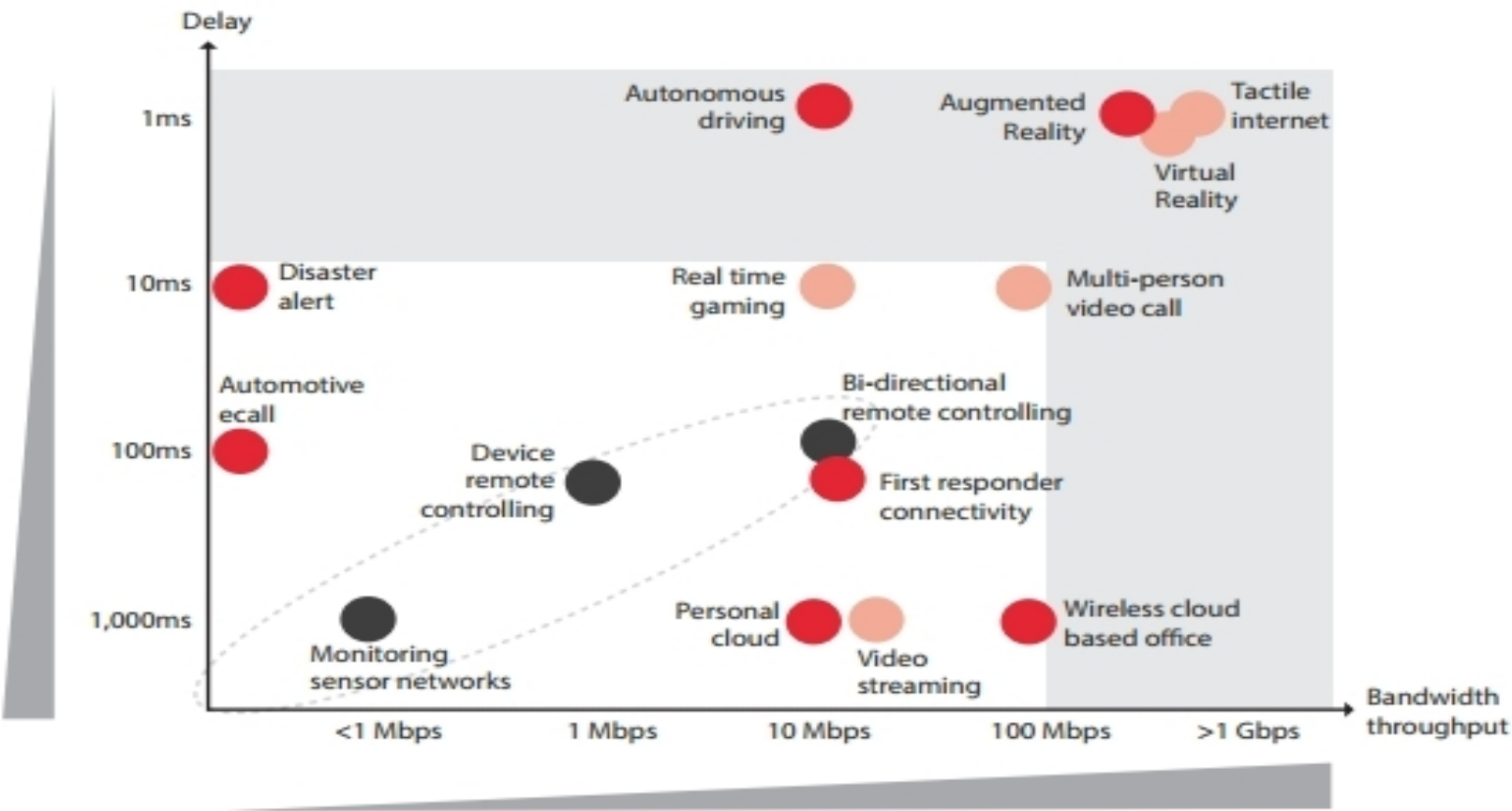
5G Networks



Source: nokia

But is it enough?

5G Networks and AR/VR



- Services that can be delivered by legacy networks
- Services that could be enabled by 5G
- Fixed
- Nomadic
- On the go
- M2M connectivity

Network Impact: Bandwidth

▫ The Math of Bo Begole:

- The fovea of our eyes can detect dots as fine-grained as 0.3 arc-minutes of a degree, that 200 distinct dots per degree within our foveal field of view... 200 pixels per degree as a reasonable estimate....across a field of view of at least 150 degrees horizontally and 120 degrees vertically within an instant (less than 100 milliseconds). That's 30,000 horizontal by 24,000 vertical pixels...a region of 720 million pixels for full coverage....add head and body rotation for 360 horizontal and 180 vertical degrees for a total of more than 2.5 billion (giga) pixels.
- Those are just for a static image... For motion video,the human eye can perceive much faster motion - some estimates are as high as 150 frames per second. For sports, games, science and other high-speed immersive experiences, video rates of 60 or even 120 fps are needed to avoid “motion blur” and disorientation.
-the eye can receive 720 million pixels for each of 2 eyes, at 36 bits per pixel for full color and at 60 frames per second: that's 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 (which is the goal of future video standards), that still means we need **5.2 gigabits per second of network throughput; maybe more.**
- If 60fps is enough and if future compression reaches a factor 600, then 5.2Gbps
- No margin of error! 5G will have a hard time supporting this in static mobility, let alone vehicular

Network Impact: Delay

- Delay requirement of < 1 ms vs 5G network typical requirement is 10ms
- Note: there are multiple types of delays, from human interactions and from human perception
 - End-to-end delay may not matter in all applications
 - For instance, immersive video conference/teleportation could have a lag of 100ms for human interactions
- However, other applications require 10ms or delay to avoid motion sickness
 - Overlaying objects on top of real world; transmitting only the proper FoV of the user and not the whole 360 environment
 - The 10ms limit includes the transmission and the rendering on the screen which by itself takes 10ms to 20ms for current LCD technology
 - (DLP displays are much faster)
- It takes 13ms for light to go from East to West Coast!

Network Impact: Bandwidth-Delay Trade-Off

- Some type of latency is not perceptible but other is
- One could try to transmit only the tiles of the FoV of the users, thereby reducing bandwidth by a factor up to 6
- However, user is very sensitive to latency in changing the FoV
 - Excellent prediction mechanisms
 - Buffer around the FoV to anticipate changes
- Latency of the streaming of a video is that of a buffer playback (a couple seconds usually) but sensitivity to FoV is $<10\text{ms}$, and no change possible once in playback buffer
- No buffer = no jitter, very short delays, but potentially less bandwidth

Network Architecture Issues

- 5G will face an uphill battle delivering AR/VR as application is resource intensive
- 5G current time line: specifications by 2020
- Integration of AR/VR into the network vs overlay play?
 - Application Driven Networks vs dumb pipe
- Guaranteed QoS, low delay, low jitter
 - 5G slices dedicated to AR/VR
- Placement of function at the edge for low latency
 - MEC, fog

Information-Centric Networking

- A network architecture which exposes content semantics to the network layer
- Network layer is aware of content, and may take action based upon the specific content requirements
- Native support for multicast - embedded in ICN - is essential for multi-party AR/VR applications
 - many such use cases, including sports events, gaming, videoconference, etc...
- Sharing of tiles in 360 videos through AR/VR specific naming schemas
- Enhanced content distribution for (potential large) 3D models for VR
- Edge caching and edge functions are prominent in ICN

- Will ICN provide better support to AR/VR than current IP?

Research Challenges I

- identify the network architecture that can delivery AR/VR capability in 5G networks;
 - in particular, which functions are necessary, and how to integrate this functions within the network architecture;
- identify the interfaces required from AR/VR to the network;
 - in particular, can AR/VR function purely as overlay, or should AR/VR require some help from the infrastructure for caching, multicasting, traffic engineering, QoS, etc.
- identify the proper naming semantics to expose at the network layer enough information to allow the sharing of data in between several VE sessions;
- identify the coding of the VE so that a VE can be packetized into multiple view "cells" units that can be recomposed into a VE sessions, and that can be shared in between different sessions;
- characterize the motion prediction and the corresponding network protocol; to assess whether this information needs to be shared with the infrastructure;

Research Challenges II

- identify rate adaptation mechanism for AR/VR, similar to, say, DASH
- identify caching policies for AR/VR content;
- specify QoS on the fly, using SDN or similar control tools; to specify what an SDN controller needs to know about an AR/VR application;
- characterize the reliability requirements of AR/VR sessions at the network layer;
- transport protocols that are designed for AR/VR; transport protocols could be adapted to the different channels from the user (say, for updating its position with very low latency; or for receiving a remote image, potentially with less stringent delay);
- identify security mechanisms;

- And many many many others!
-
- Is there interest to look into these issues?
- Draft will be submitted on the topic

Questions?