VIRTUAL AND AUGMENTED REALITY – HOW DO THEY AFFECT THE CURRENT SERVICE DELIVERY AND HOME AND NETWORK ARCHITECTURES?

CHARLES CHEEVERS, MARK BUGAJSKI, AJAY LUTHRA, SEAN MCCARTHY, PAUL MORONEY, KEVIN WIRICK

ARRIS™

TABLE OF CONTENTS

INTRODUCTION 3

DESCRIPTION OF THE VIRTUAL REALITY SOLUTION AND AREAS OF IMPORTANCE FOR THE SERVICE PROVIDER 5

- New Camera Technologies 7
- New Storage Technologies 7
- Content Stitching, Encoding, Compression and Mapping 8
- The Delivery Network 10
- The FOV Headset Scene Generator aka PC Game Console or the Future VR STB 10
- The VR Head Mounted Device (HMD) 11

OPPORTUNITIES FOR THE CABLE INDUSTRY AND VIRTUAL REALITY SOLUTIONS 14

- VR and AR Content Delivery 15
  - Avid VR Users will be More Likely to Subscribe to Gbps Broadband Services 15
  - VR Turbo Download Mode 15
  - VR File or Pre-emptive Scene Caching 16
  - Multicast Carousels for Popular VR Title Download 16
  - Live VR Unicast Solutions 17
  - Hybrid Multicast Solution for Live VR Efficiencies 22
- In Home Connectivity to HMD and VR Server/Rendering Device 22
  - Symbiotic Wi-Fi Delivery and QoS/QoE for Virtual Reality 22
  - 60 GHz Routing Added to Service Providers Gateway or In Room Device Strategy 26
  - Pre-emptive Caching Opportunity for Next Rendered VR Scene 28
  - Merging the VR and TV Experience – HMD and/or TV and/or PC/Tablet 29

CONCLUSION 30

ABBREVIATIONS 32

RELATED READINGS 34

REFERENCES 35

INTRODUCTION

As we move forward with new forms of immersive entertainment, as well as interactivity with cloud services and everyday tasks, one of the key technologies emerging is the bending of reality as our eye and brain perceives it to create another new level of entertainment and interaction with our everyday surroundings. This technology comes in a number of forms, but we typically group the technologies and directions into two main categories:

(i) **Virtual Reality (VR)** – The digital creation of a fabricated immersive world, typically via a headset technology, that generates all the photons that the eye sees.

(ii) **Augmented Reality (AR)** – The digital creation of a fabricated set of objects that can be interspersed with real world elements, usually through a headset that overlays the objects on the lens, as the user also views their real surroundings.

A third variant has been defined by several new players to the market as:

(iii) **Mixed Reality** – This is a fusion of both virtual reality and augmented reality.

From the 2015 published Gartner Hype Cycle curve [1] – we can see that virtual reality is at a very interesting stage of its life. VR is about to enter the 'Slope of Enlightenment' from the 'Trough of Disillusionment' – after being around in one shape or other for 20+ years. However, with the improvement in technology and emergence of a new VR Head Mounted Device (HMD), VR is potentially ready to become a significant service in early adopter households in 2016 and into 2017.

Augmented reality is a little further behind virtual reality; it's potentially about to make the turn upwards to the 'Slope of Enlightenment Phase' as we leave the current era of experimentation and now significant Venture Capital investment and move into the first real products and content for VR, AR, and MR and the new experiences created may influence eventually more aspects of a person's daily activities. The implementations of AR have the additional technical challenge of using cameras and sensors to map and sense room surfaces and objects to overlay the augmentation elements, as well as tracking the six degrees of freedom from the eye – trying to create a blended experience of what is really in view and what is overlaid so that the eye finds it hard to discern the difference. This paper will focus on VR more than AR given VR's present relevance for use in gaming, adult content viewing, and the potential viewing of live sporting events.

The stickiness of the technology has to be proven, yet there is a growing belief that we are now on the path to making both virtual and augmented reality useful in our lives. In particular, there has been good progress with trialing of prototypes of Virtual Reality devices. This has stimulated much interest in driving the concepts to products and this year a number of serious players in the gaming and entertainment areas are bringing their products to market, as well as Augmented and Virtual Reality Venture Capital investment rising to levels where there is likely to be significant product and content breakthroughs over the next 2-5 years.

If the physics of the technology can make the brain believe what the eye is taking in from the artificial device and world – and allow the user to stay in this world for more than 15 minutes – then the technology may stick and evolve to be a useful part of our work and entertainment lives. While the initial set of devices launching this year will no doubt fall a little short of all the parameters required to make VR and AR work, it is expected that these devices will also serve to fuel enough acceptance to drive further generations of device and delivery. This in turn will force the acceptance of the technology through higher resolution, higher frame rates, lower latency for pitch, yaw, and roll derived scene generation, non-blurring photons in the Head Mounted Device screen, lighter HMD, and non-tethered use of the technology.

What does this new technological experience mean for the cable and connectivity industry? This paper will attempt to define the technology and the cause & effect opportunity of VR and AR on the delivery network and the home network. This paper will focus on the bandwidth required to make the technology successful and on some of the ideas that will make VR and AR more symbiotic in the service provider network. It also proposes some ways the service provider can work with game console and headset manufacturers to provide the best chance for this technology to succeed. Live VR and cloud to headset AR solutions will require the content delivery company, the service provider, and the headset manufacturers to all work together and understand the end-to-end solution metrics and quality of experience. This is something we explore more of in this paper.

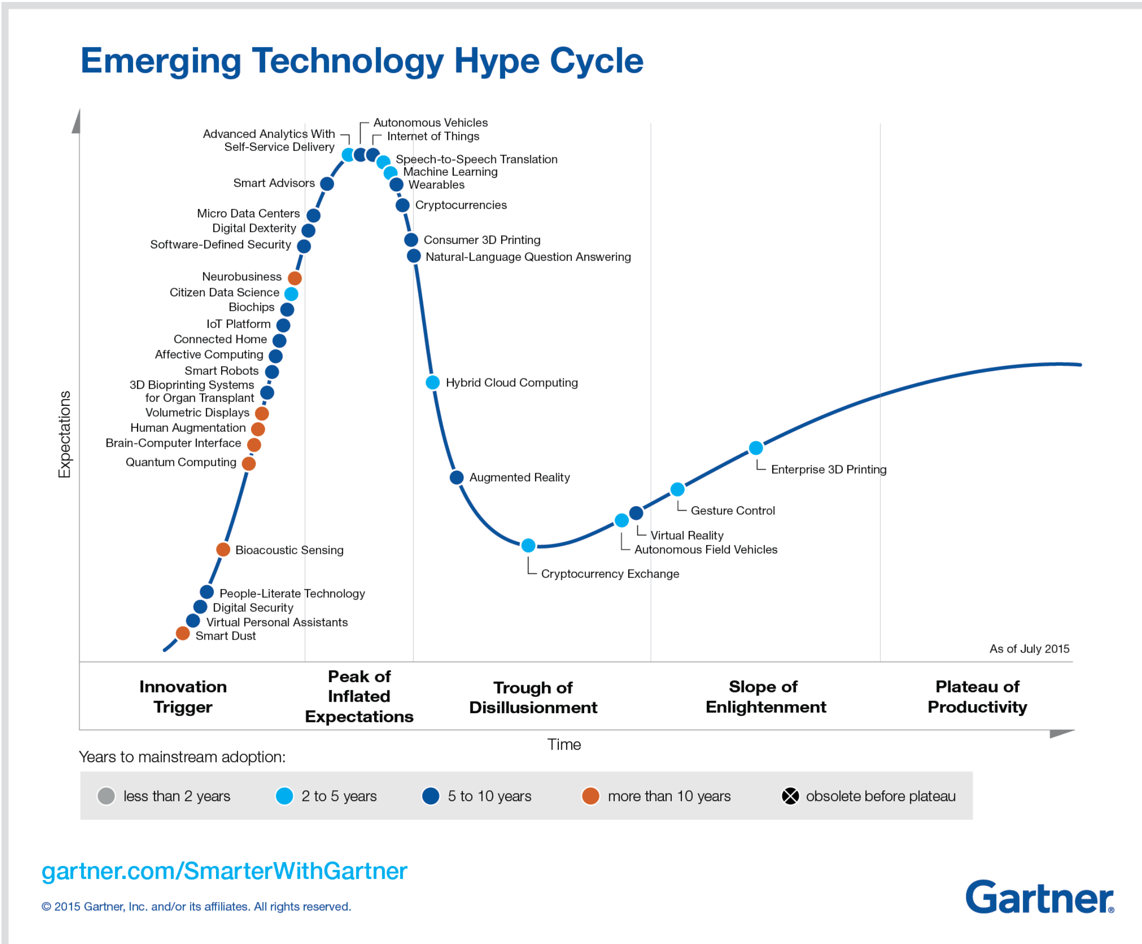


Figure 1 - Gartner 2015 Hype Curve

DESCRIPTION OF THE VIRTUAL REALITY SOLUTION AND AREAS OF IMPORTANCE FOR THE SERVICE PROVIDER

Virtual reality immersive multimedia, or computer-simulated reality, is a computer technology that replicates an environment, real or imagined, and simulates a user's physical presence and environment in a way that allows the user to interact with it. VR typically requires a headset that contains a screen and lens to immerse the user into the simulated environment. It currently needs a dedicated PC or game console to generate the 3D simulated graphics and video for the headset although there has been streaming content available for some time and companies now experimenting with live VR transmission.

Figure 2 - Conceptual View of What Most People Think of VR Experience

One key element of this paper is to explore how VR affects the service provider both from the obvious additional bandwidth of large files being sent to consoles and PCs and also if live VR can generate opportunities for new solutions like multicast VR or source to headset allowing new caching, pre-emption, and stitching. The migration of headsets from HDMI/USB tethered connections to a wireless connection is also something that may offer an opportunity for service providers to add in support for headsets as part of the wireless devices supported from their gateway and extenders. This may be done primarily in the 802.11ax ramp or potentially using 60 GHz technology – which offers a huge opportunity to service the headset with 20 Gbps of wireless connection with new IEEE802.11ay underway as the next generation to current 802.11ad.

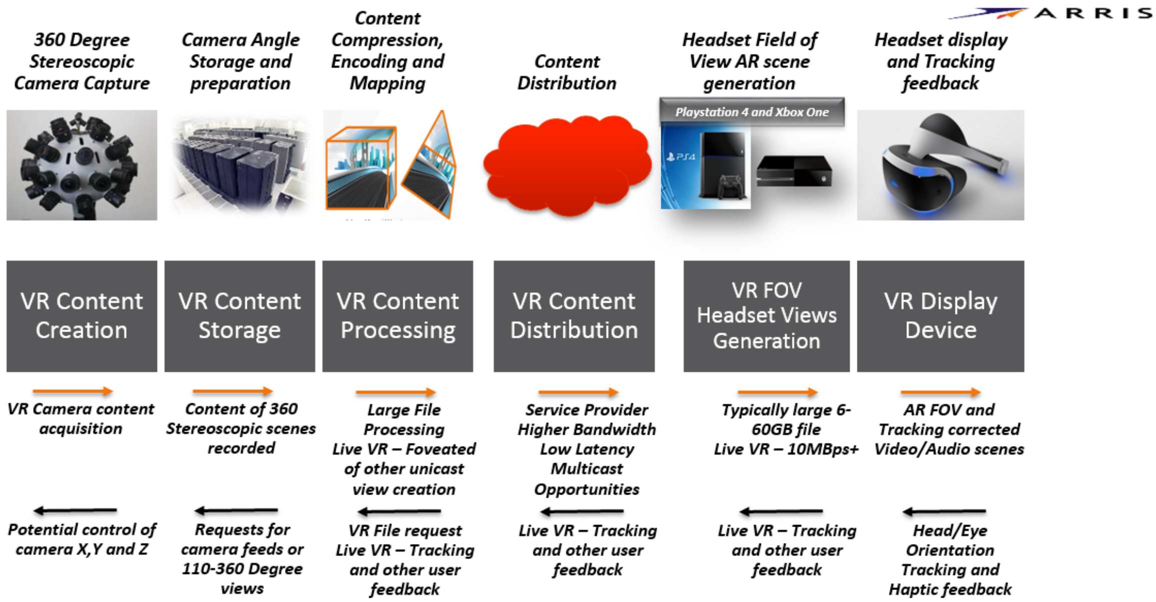


Figure 3 - Overall VR Acquisition and Delivery System

The virtual reality end-to-end solution chain is defined below in Figure 3. It comprises:

New Camera Technologies

Cameras that can capture the scene in stereoscopic 360 degree formats and at the highest resolution and frames per second possible are being developed. Most deployments today use a concave view with multi-lens cameras. VR video today can be 180 degrees or 360 degrees. For 360 degree systems the field of view ratio to the captured video is about 20% with 180 degree system its closer to 33% offering higher resolutions.

New Storage Technologies

Storage technologies focused on solid state devices to allow the high uncompressed video to be recorded and ready for compression process. For live VR it may require that everything from stadiums to outside broadcast trucks have both new storage solutions for substantially more video content, and higher speed network connections to other data centers that also support processing the raw video feeds for either unicast foveated video distribution or potentially multicast streaming of all cameras to other locations for VR scene processing to the individual users headset.

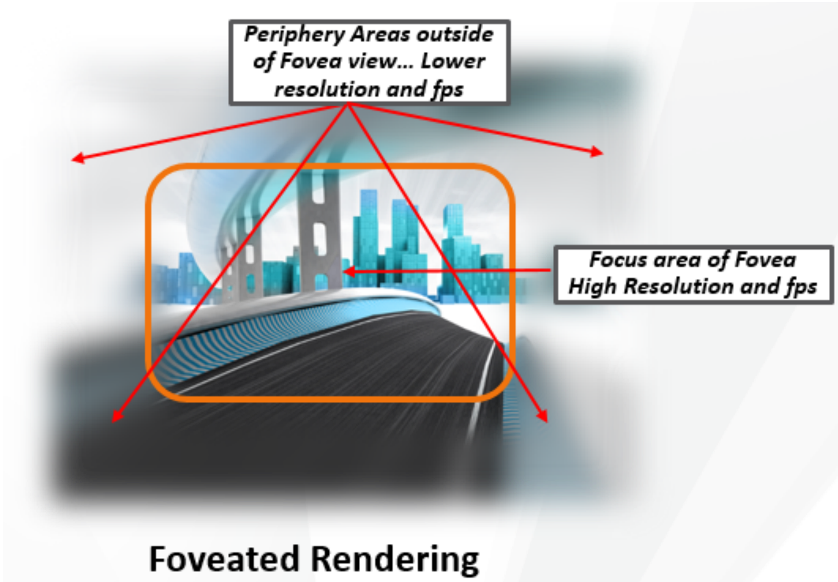


Figure 4 - Foveated Rendering of VR Scene

Content Stitching, Encoding, Compression and Mapping

For processing foveated scenes for the live VR streaming of non-live file based VR experiences, the individual camera streams are encoded and formatted to allow the headset attached PC or gaming device to create the required scenes including the associated 3D audio where supported. Live VR also requires the stitching of the camera feeds to present the view where the end user is looking – called the Region of Interest (ROI). Stitching in real-time, either in the cloud or on the game console/PC, is typically dimensioned at the moment for 4-6 1080p streams into one video stream of up to 2160p60 8-bit resolution.

ROI Mapping is another area of avid research with mapping the 180 or 360 degree views to equirectangular projection or polygons. For example, some companies are looking at pyramidal mapping claiming 80% improvements in bandwidth. This invention is making it more difficult to ensure interoperability. With ITU, MPEG, and DVB also getting involved with VR standards, it will be interesting to see where innovation and standards intersect. Encoding tries also to follow the ROI and the mapping algorithms to improve real-time speeds and the relevance of the output VR stream.

Many believe that for VR to be successful, the entire viewable area needs to have more than 16k x 12k or even 32k x 24k pixels. Consider an example case of one high quality VR system with a viewing range of 150 x120 degrees. It may need more than 16,000 x 12,000 pixels to cover that range. In the delivery architecture where the entire VR scene is sent to a client, assuming non-stereoscopic (2D) capture, 10 bits per pixel and 60 fps, the scene will have 16,000 x 12,000 x 60 x 30 (> 300 Gbps) bits per sec of raw (uncompressed) bit rate. Even if it is assumed that a compression ratio of about 300:1 is used (this is of about the same order as used in digital TV distribution), it gives about > 1 Gbps compressed bit rate (far higher than what is getting used today in low quality VR).

In reality, in that approach that rate may be even significantly higher as a quick response with low latency requirement may prohibit that high (~ 300) of a compression ratio. On other hand, in a very low end system as done today in many demonstrations, the entire VR scene is limited to 4k and the bit rates used are of the same order as in current 4kTV (10 to 20 Mbps). However, it is widely believed that the current versions of VR delivery approaches and platforms are very inadequate to provide the much needed quality of experience for it to be successful.

There are some studies which show that many people feel uncomfortable beyond 15 min of viewing VR scenes. One key concern is “simulator sickness” which includes vertigo, discomfort, headaches, and motion-sickness like symptoms. Simulator sickness is fairly common with the current generation of VR content and gear. However, it is quite possible that simulator sickness issues will go away with time as the content and display industries learn the craft of creating good VR. It is known that a key factor in avoiding simulator sickness for longer duration (> 15 min) viewing is to have very good visual quality.

This leads to requiring significantly higher spatial resolution and higher frame rates in comparison to what is done today in many demonstrations. 47,000 pixels per degree is being proposed by some as the minimum required pixels/degree. In addition, to minimize the simulator sickness, the delay in response to the head movement needs to be very small. It is not clear exactly how small that needs to be, but it is believed by several experts that it needs to less than about 30 milliseconds (ms) which is far less than the 100s of ms in many demonstrations shown today. The delay may even need to be significantly smaller than 30 ms, but it is hard to quantify this number due to the lack of comprehensive formal studies on this topic. This 30 ms time is something that can potentially be accomplished from the network with fiber and even DOCSIS latency being able to support. However, it assumes that not only the network, but also potentially stitching and even Wi-Fi and TCP latency issues have to be really understood and tuned.

Since the majority of viewers are now used to seeing high resolution video, to avoid simulator sickness high visual quality is necessary. Hence, the bit rates required to stream VR video, in the delivery architecture where entire VR scene is downloaded, will be several hundreds of megabits per second or even gigabits per second. This will likely require alternative delivery architectures where entire VR scenes are not downloaded. There are various options in that regard being discussed, but they are in very embryonic stage – typically focusing on pre-emption of head movement or tracking user motion. Discussions include concepts like streaming a selected part of the scene depending upon what is the viewing angle or having a variable resolution of the scene depending upon the viewing angle and hybrid delivery network architectures where some part of the scene is cached locally and others are streamed as needed.

The MPEG group has also already started the work item of carrying VR streams in MPEG DASH. The group is exploring the topic of what kind of VR related signaling need to be included. MPEG has also initiated MPEG Omnidirectional Media Application Format (OMAF) standardization for storage and delivery.

The Delivery Network

Whether content is live or not – networks will experience an increase in traffic as VR headsets are adopted and used. This will have an effect on the core, edge devices, the network itself, and perhaps even the gateways in the home. (That last topic will be discussed further in this paper.) VR broadcast might use transport stream delivery which will allow a regular set-top box (STB) to display VR on a TV. Today, this is possible with the DVB UHD-1 specification. Unicast of VR is expected to use DASH ISO BMFF for equirectangular and tiled transmission. There has been some discussion of having something like 30 camera streams with 5 ABR profiles – creating as many as 150 feeds or files to support – particularly for streaming over the unreliable internet. On net solutions could leverage the quality of service (QoS) to have a single profile and leverage DOCSIS QoS mechanisms. This profile could be affected when the headset device is un-tethered while using Wi-Fi or other Wireless solutions. This could lower available airtime when the video scene is transmitted.

The FOV Headset Scene Generator aka PC Game Console or the Future VR STB

The field of view (FOV) device can process all the video camera angles and produce the 110 degree low latency, pitch, and yaw corrected video for the headset. It has to stitch each camera view together for display on the headset under 30ms and closer to 15ms.

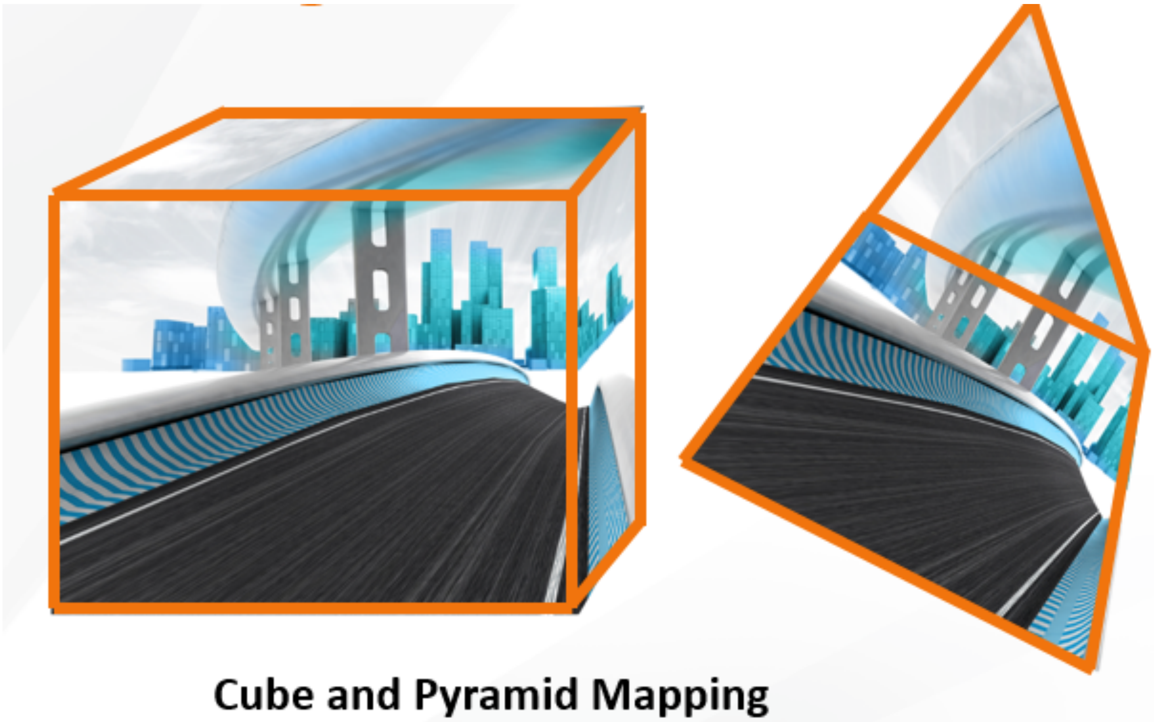


Figure 5 - Cube and Pyramidal Mapping of 360 Degree Scenes

The VR Head Mounted Device (HMD)

Two of the main complaints against head mounted devices for VR are that they are heavy and wired. Various companies are working on addressing these issues. There are a number of issues to solve including:

1. What wireless technology will support the bit rates required? Does the PC/console do the rendering or the HMD?
2. What will be the high speed video wireless transmission solution? The JPEG committee's most recent effort towards developing a new standard for this, JPEG XS, has this application (among others like mobile device or set-top box to TV wireless connection, etc.) in the crosshairs where low cost, low power, and low delay compression can help towards achieving wireless head mounted display. Some experiments are also being conducted to show the VR video on a TV screen. They demonstrations currently look more like multi-view TV or free-view TV. The much larger number of deployed TVs than the number of deployed VR headsets and lack of viewing sickness problem in TV viewing is encouraging some to push farther in this direction.

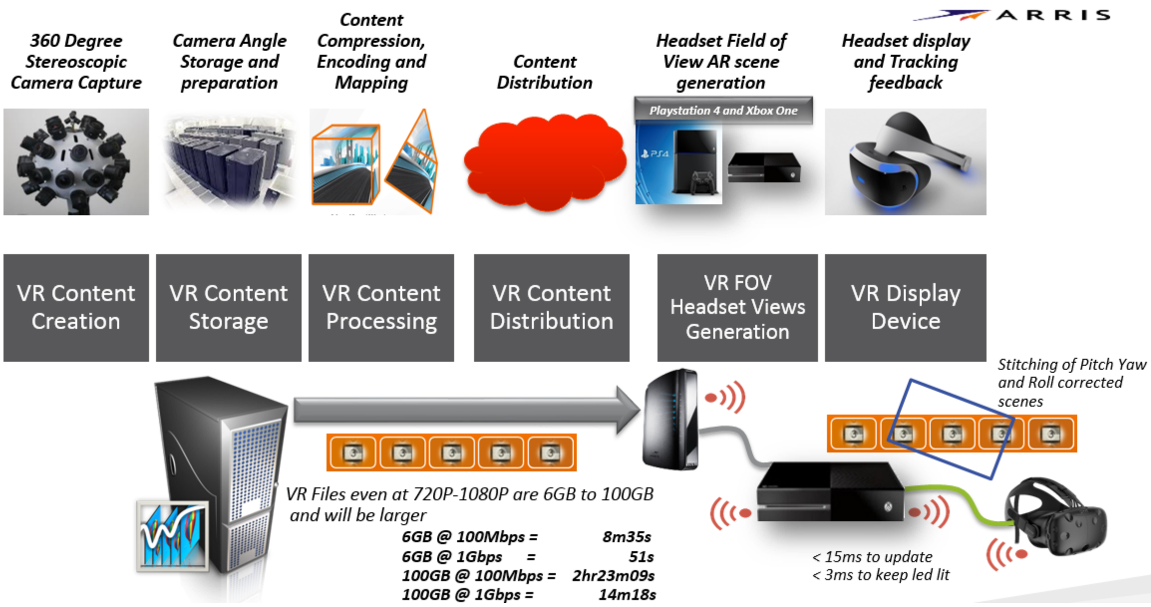


Figure 6 - Large VR File Download Experience

The current VR experience can be probably classified as two-fold:

(i) A download of large file(s) to the game console or PC

(ii) Streaming video from sites like YouTube; these are often encoded with ABR profiles to allow for use over mobile and constrained wireless networks including Wi-Fi, typically at 720P and sometimes 1080p resolutions.

As you can see from the numbers in Figure 6, the experience of a typical user today:

- At the smaller side of VR experience – 6 GB – it can take over 8 minutes to download the file – even on 100 Mbps services
- If we assume a reasonable level of 2017 VR experience at 100 GB of content with high quality video as well as local image generation – the experience at 100 Mbps would be very poor having to wait almost 2.5 hours before enjoying the VR experience generated from the local system. Even at 1 Gbps of network bandwidth – the file would still take over almost 15 minutes to download – so even then, it's too much time to wait
- Live VR and streaming the content at 100 Mbps – if using pre-emption of movement and foveated rendering – can still only cope with moderate sub 4K resolutions and we may be moving to solutions where Gbps speeds can really only give that truly 'presence' based experience and a solution that allows immersion for more than 15 minutes without nausea
    - It is also unclear if live sports will translate well to consumers. There has been great progress in camera capture technology, but it seems that the

ultimate sports experience may be the combination of immersive action from multiple camera locations (including interpolating within the action itself) and what the director shoots in 2D.

For example, watching a tennis game from the best seat at the center line – looking left and right to track the ball and then cutting to the close up of the losing player showing disappointment – would be a combination of VR and inserting 2D close ups into the HMD. Football, soccer, or any sport becomes an opportunity for promoters to sell prime seats to any number of fans. The opportunity may also translate to plays, concerts and any other forms of entertainment

Figure 7 - Best Seat in the House Delivered to Your (and 1 Million) Other Sofas

  - The other experience considered is using the flat 2D TV to allow the user to pan around the game and control the view using some hand held gyrometer or hand gestures. It remains to be seen whether this offers a user experience that will capture enough of the viewing market
- There are also attempts to develop parallel VR experiences in addition to the regular 2D TV viewing. This parallel use of VR to provide engagement snippets of 15 minutes or less, also needs to be proven with the consumer. The goal is to give viewers additional immersion in the original show by capturing elements in VR and then allowing the viewer to immerse themselves into key areas and events. In each snippet, the viewer could potentially discover additional features, plot points, story aspects, or even gain the ability to visit where the characters

have been. This could lead to an increase in viewer loyalty and present opportunities for contests, as well as, perhaps even meeting the actors "live" via VR.

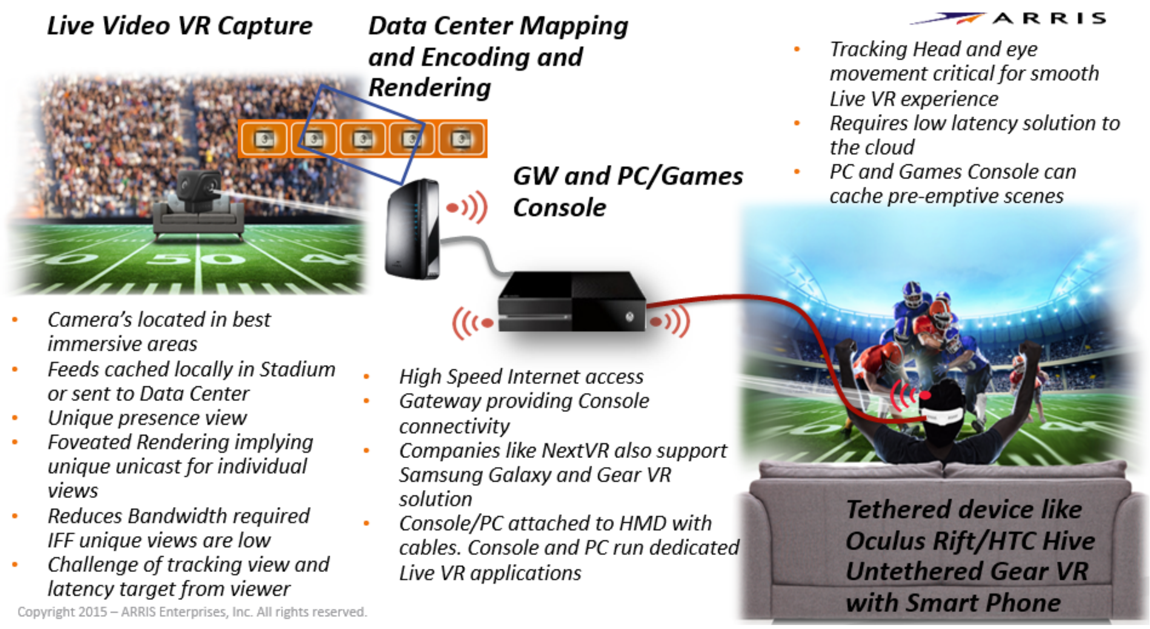


Figure 8 - End-to-End Live VR: Key Elements

Figure 8 highlights some of the key challenges of end-to-end live VR – in particular, as the resolution increases and providing 180 or 360 stereoscopic views. At the moment, the solutions are unicast to the VR application which is loaded on something like a smart phone using a VR headset, or a PC/game console connected to a tethered HMD. There are a number of companies emerging trying to specialize in the transmission of live VR and this technique and technology is set to improve as the "Live Experience" is explored for video, surround audio, and tactile/haptic experiences.

OPPORTUNITIES FOR THE CABLE INDUSTRY AND VIRTUAL REALITY SOLUTIONS

It's time to review the potential opportunities for virtual reality and augmented reality applications in the MSO provided home user experience. The assumption here is that the solutions proposed will be made to work and the issues of the experience and quality will improve – particularly, with symbiotic relationships between content creators, VR solution providers, and the service providers.

This opportunity potentially breaks out into the following areas:

(i) VR and AR Content Delivery
   a. Improvements in download speeds for large VR files
   b. Improvements in delivery and latency for live VR
   c. Unicast and Multicast VR – is there opportunity there

(ii) In Home Connectivity to HMD and VR server (PC/console)
   a. Symbiotic Wi-Fi delivery and QoS/QoE for virtual reality
   b. 60 GHz routing added to service providers gateway or in room device strategy
   c. Pre-emptive caching opportunity for next rendered VR scene
   d. Merging the VR and TV experience – HMD and/or TV and/or PC/tablet

VR and AR Content Delivery

As shown above, the file sizes for VR files are larger than has ever seen before for full HD video movies or the latest 2D or 3D games. As such, if VR or AR is successful, many multiple versions of large files will be downloaded. The experience at 100 Mbps waiting almost 2.5 hours to download will drive VR gamers to services that improve the experience.

Avid VR Users will be More Likely to Subscribe to Gbps Broadband Services

An obvious correlation to VR and HSD services is that an avid VR user will likely pay for the lowest latency and high bandwidth tiers to improve the overall time to optimize download of VR files and future live VR experience to best possible levels. This should prove to be interesting in the future to correlate VR and AR usage against the HSD tier levels offered and in the future potential wireless router/spectrum collaboration between the Service Provider and the HMD or Game Console providers. Mapping QoS from the network to the end photon will certainly have the benefit of tuning the overall VR experience for lowest possible latency and best possible enjoyment.

VR Turbo Download Mode

For those subscribers not on Gbps speeds who want their files downloaded quicker – an MSO could offer a special additional fee service to allow additional burst speeds for the download of the file. We'll call it Turbo Mode for VR instantaneous gratification.

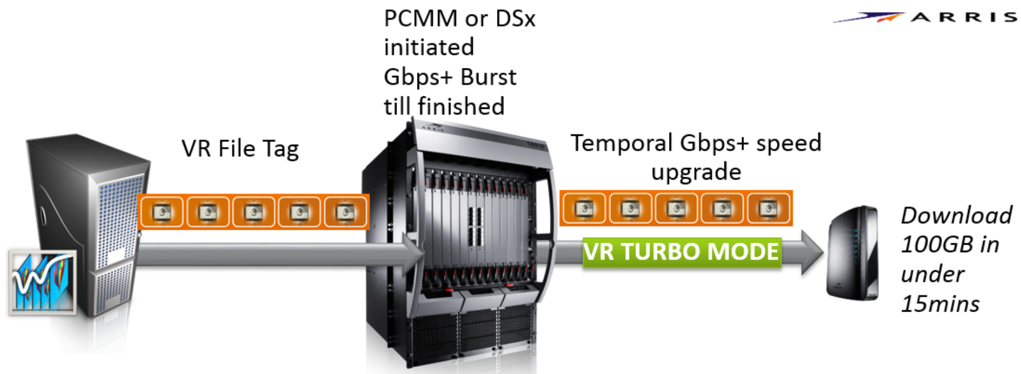


Figure 9 - VR Turbo Mode of DOCSIS QoS

VR File or Pre-emptive Scene Caching

With large multiple file downloads there has always been the potential to cache the files locally rather than pull them multiple times from the source through an expensive CDN. With VR, the solution looks even more attractive to cache VR files locally in the service provider network.

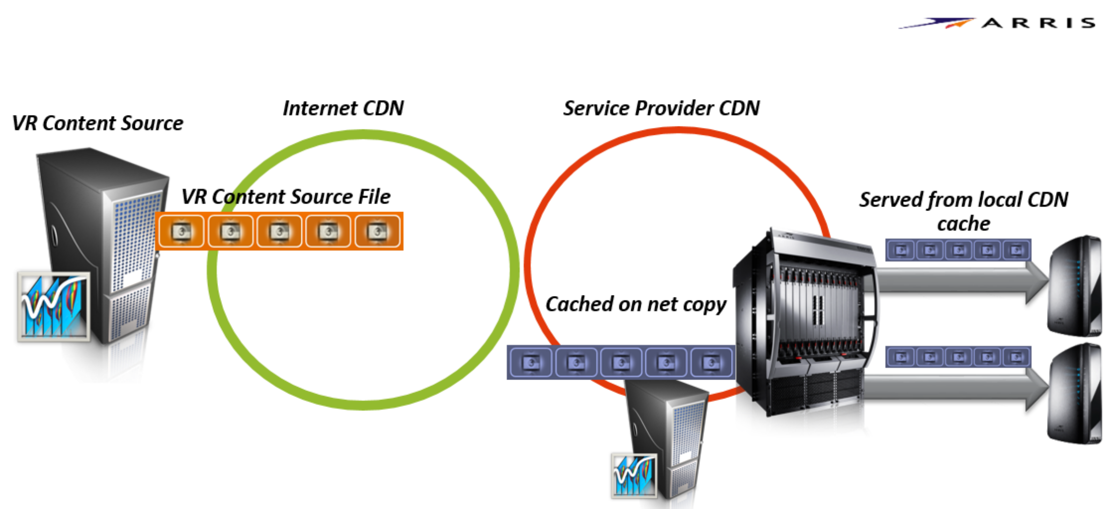


Figure 10 - Local CDN caching of VR files

Multicast Carousels for Popular VR Title Download

Consider creating a scope for the release of new VR titles – to also offer a multicast window on the network for half a day – to allow users to download even more efficiently. This seems to be worth looking at, especially if the gaming market for VR has the first day and week release ramp as the current high profile games launches.

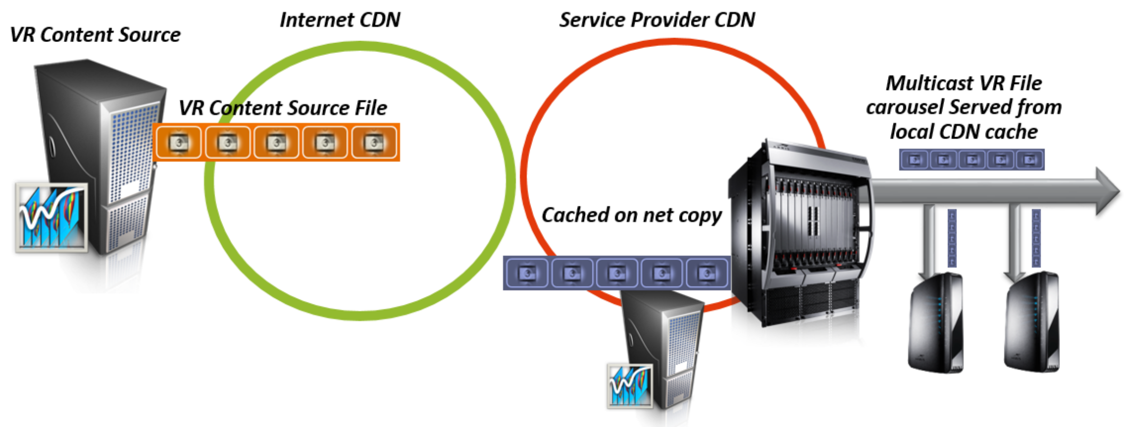


Figure 11 - Multicast VR File Carousel

Live VR Unicast Solutions

Live VR is extremely dependent on low latency solution and architecture. Can the DOCSIS network be tuned to support a low latency path from HMD tracking to a new VR scene sent to the headset < 20-15 ms through the network and over wired or wireless connection? The answer is still unclear. It's probably, "Yes," for low resolution and frame rates, but as we move to higher resolutions and frame rates then it will certainly have to track lower node subscriber counts and in the era of full duplex DOCSIS and fiber PON growth to 64 or less homes. In the meantime, for 720p - 1080p live VR transmissions - DOCSIS QoS mechanisms will have to be used to ensure this low latency. DOCSIS best effort solutions are often the only configuration that an operator sets up outside some specific treatment of voice calls.

Let us step back and take a look at the elements required for VR to work – to create an immersive and 'present' experience while reducing the likelihood of simulator sickness:

- Resolution to satisfy the fovea and periphery of the retina are being researched
    - The high-acuity center of vision can resolve up to about 60 pixels per visual degree (equivalent to a 1080p HD display spanning about 30 degrees of one's visual field, which is the typical HDTV viewing condition). The field of view for each of our eyes is ~170 degrees horizontally and ~150 degrees vertically. The central binocular overlap of our eyes' field of vision is ~120 degrees, and the total monocular field of vision is ~210 degrees. The current generation of HMDs presents a ~100x100 degree view to each eye. People typically use only their eyes to look around the central ~40 degrees of the visual field, and use combined head and eye motions to look around larger areas [2]

- Thus, it is probably safe to estimate that a ~2k x 2k display (40 x 40 degree at 60 pixels per degree) would be just sufficient for binocular high-acuity viewing and eye-only movements. (If the same pixel density were used to cover the entire ~210 x 150 degree field of vision, a 12k x 9k display would be needed, though the encoded image could have lower effective resolution outside the central 40x40 degree region.) Recent work indicates that a dynamic FOV reduces simulator sickness [3]
- Currently most of the devices use 1280 x 720p displays to present a ~100 x 100 degree view to each eye. Thus, the resolution at the level of the retina is 7 to 13 pixels per degree at best – which is well below the 60 pixels per degree need to make pixels essentially disappear

- Low pixel persistence
  - To avoid blurring or streaking with eye motion requires that the pixel persist (remains lit) for no more than 3 ms. This is more relevant for VR HMD, because of the relation of the eye movement to the distance from the VR HMD. The longer pixels persist the farther they smear across the retina and the blurrier the scene becomes
  - Unfortunately, there is a requirement to decrease the pixel persist time with higher resolutions with a 1k x 1k supporting 3 ms and higher pixel densities requiring shorter persistence
- Refresh rate at a level to make the eye comfortable
  - Correlated with pixel persistence, this is necessary to avoid flicker when persistence is low
  - To gain presence 95 MHz works, but it presents significant challenge to maintain high refresh rates for stereo rendering when resolutions increase
- Tracking of HMD and eye at accurate levels
  - Has to support accurate tracking to allow image translation to x, y, z
  - Millimeter or less precision and ¼ degree accurate rotation is required
- Optical lens and calibration
  - One or two lenses per eye to try and maximize focus on retina of the projected images. It is also harder to achieve multiple lenses with the restrictions of weight and industrial design and the viewing distance of the HMD screen
  - This still leaves a problem with focal length, distortion, and aberrations
  - A process that accurately characterizes the lenses and correcting the rendered image is required in the solution
- Low latency
  - We have been using 15ms as the benchmark throughout this document, but the results are reasonable at 20ms. This means the virtual image needs to be in the right place at the right time. This latency measurement

is a function from the time the head and HMD move to the time the last photon is emitted (resulting in painting that new image over the motion)

- Positional tracking
    - As you move around in VR scenes, there needs to be a way to report your location to the scene generator accurately. There are a number of schemes being considered – some include positioning markers around the VR gaming area such as marking or allowing position to be triangulated. These schemes may use either RF location or camera recognition of the location. In the future, better depth cameras may be able to work with less aids pre-located

Figure 12 - Critical Parameters of Immersive Presence based VR

Let us now look at the DOCSIS latency capabilities. As we discussed, VR will prompt the use of QoS policies to DOCSIS either to improve download speed and experience; more importantly, if we are to send live VR to the console and attached headset. As you can see from Figure 13, the latency sensitive loop is the connection from the PC/console/VR STB to the HMD. There is latency in encoding, mapping the foveated video stream, and the internet adds 30-50 ms with additional time added by DOCSIS (could be as low as 2 ms and as average at 15 ms). Additionally, wireless transmission time, as well as the computing a frame of video to HDMI, could increase delay by 16-32ms.

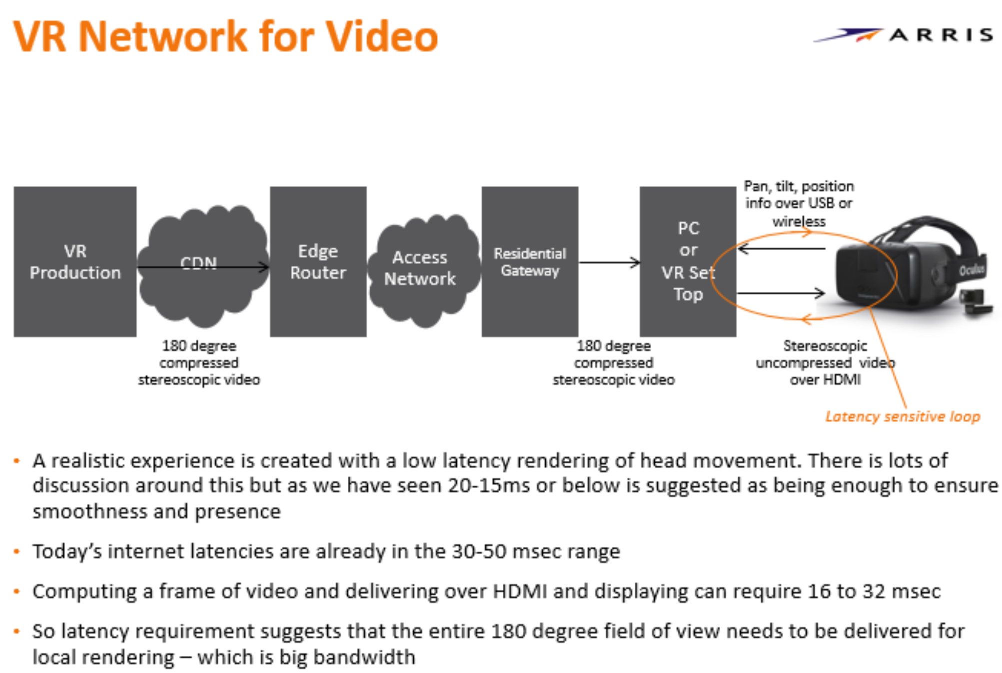


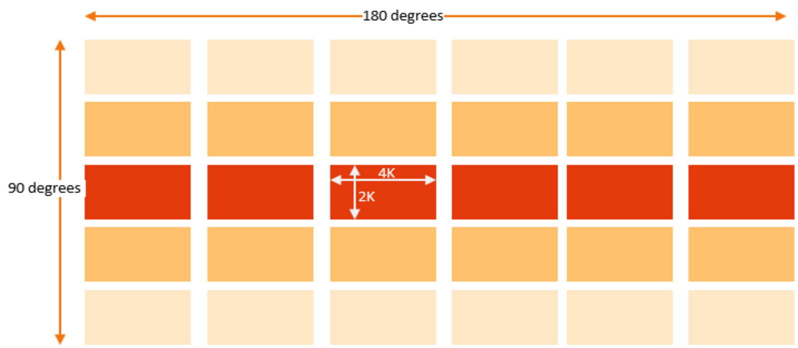
Figure 13 - Latency Factors from VR Content to HMD Device

Figure 13 suggests that the entire 180 degree field of view needs to be delivered for local rendering; which requires using lots of bandwidth, as well as leveraging the pre-emptive caching of predictive next scene (left/right of tennis game). Figure 14 and 15 illustrate these projected bandwidths.

Headset	Release date	Resolution		Compression benchmark for equivalent 2k x 4k p60 (Mbps)	Duration of Frame (msec)	180 x 90 degree compressed data rate	Total for 180 x 90 degree with 20% additional data for stereoscopic (Mbps)	Per Frame (Mbp17ms)
Sansung Gear VR and Oculus Rift DK2	12/'14 and 7/'14	Low resolution: 1280 x 1440 in 96 degress	15 pixels per horizontal degree	25	17	14	17	0.167
HD resolution future headset	2-4 year future	HD level 1K x 2K in 30 horizontal degrees	64 horizontal pixels per horizontal degree	20	17	160	192	3.2
Ultimate resolution future headset	> 5 years in the future	Limit of Human visual perception: 2K x 4K in 30 horizontal degrees	120 pixels per horizontal degree	15	17	480	576	9.6

Figure 14 - Bandwidth for current and future stereoscopic VR projections

How much data?



Headset	Release date	Resolution		Compression benchmark for equivalent 2k x 4k p60 (Mbps)	180 x 90 degree compressed data rate	Total for 180 x 90 degree with 20% additional data for stereoscopic (Mbps)
Sansung Gear VR and Oculus Rift DK2	12/'14 and 7/'14	Low resolution: 1280 x 1440 in 96 degress	15 pixels per horizontal degree	25	14	17
HD resolution future headset	2-4 year future	HD level 1K x 2K in 30 horizontal degrees	64 horizontal pixels per horizontal degree	20	160	192
Ultimate resolution future headset	> 5 years in the future	Limit of Human visual perception: 2K x 4K in 30 horizontal degrees	120 pixels per horizontal degree	15	480	576

Figure 15 - 180 x 90 Degree FOV SD to UHD VR bandwidth

Using DOCSIS and unsolicited grant service flow setup, these bandwidths are feasible. For the minimum latencies for a single frame (17 ms for 60 fps) immersion can be achieved and there is scope to increase the burst profile to yield 2x to 10x, or more, of cached scenes locally. This will improve the latency at the expense of network bandwidth and decreased statistical multiplexing to larger subscriber groups. xPON fiber solutions with their small PON service groups should also be able to achieve round trip latency.

Companies specializing in live VR have been working to reduce the bitrates for use in current access networks; particularly over LTE and other constrained networks. From captures of 6K @ 80 fps they can encode down to 4 Mbps and still claim some quality of immersive and presence experience. These technologies focus on foveated rendering and prediction of the motion of the head or eye. Foveated rendering, as explained previously, understands the area that the user is looking at as it relates to the focus area or fovea of the retina – producing a high resolution and high frame rate version of this area of the scene while reducing the periphery of the view to lower bit rates. These differential compression and transmission techniques allow different portions of the 360 to be transmitted at varying frame rates and resolutions.

Requiring a unique view per consumer reduces the bandwidth to the HMD, but will be bandwidth intensive if the unicast count of unique views increases. Of course, there will be common view areas as people focus on the action of, for example a live sports game. There should be a large statistical overlap for encoding, mapping, and processing of the video. However, the higher the resolution, frame rate, and wider field of view coupled with the < 20 ms latency requirement to present the first new frame of the scene seems to require that the entire 180 degree (potentially 110 degree) FOV needs to be delivered together for local rendering vs stitched from the PC or VR set-top. This implies bandwidths reaching almost 600 Mbps for 4kp60 stereoscopic VR. This also assumes reasonable improvements in compression and mapping. We expect the industry to drive even more innovation to constantly improve this bandwidth/latency/quality equation. Time will tell if the current lower resolution experiences will work or if they will need improvement to become a real sticky user application.

Hybrid Multicast Solution for Live VR Efficiencies

While it seems most of the focus on making live VR viable is to work through foveated rendering and pre-emptive tracking of the next focus points of the eye. There may be scope to look at multicasting for high viewership events. While each user can have their own specific visual experience – there is likely to be a common cluster of head movement in directions of play and action. This tracking analysis of the feedback from the HMDs could generate multicast flows of the compressed scenes to get them cached at the edge or at the rendering device (gateway, PC, game console, set-top, etc.). Audio cues and other video points of interest could also be candidates to multicast these scenes for stitching and rendering locally.

In Home Connectivity to HMD and VR Server/Rendering Device

Symbiotic Wi-Fi Delivery and QoS/QoE for Virtual Reality

The higher performance HMDs use wired HDMI and/or USB 3.0 connections. Note some prototypes of devices have with up to 5 cables! The decoding of the scene is typically done on a PC or game console and then sent to HMD uncompressed for direct rendering over HDMI. Lower performing solutions, like smartphone based VR, can operate the scene decoding locally on the smart phone itself. However, this is limited in capability for higher resolution/frame rate and screen latency and it's even more limited on the bandwidth from the scenes to the phone over Wi-Fi and LTE or 3G. These solutions usually support Wi-Fi only for more deterministic performance. They also rely on downloaded video, particularly for 360 and stereoscopic viewing, and also play VR videos directly from YouTube at lower resolution; typically supporting ABR profiles to deal with latency and congestion issues.

For smartphone based VR solutions and similar wireless solutions there are a number of opportunities for symbiotic and improved experiences that a service provider can explore to help untether the HMD. This is a major technological challenge and one that does not have a simple solution path.

Figure 16 - Wi-Fi and VR HMD - Video Decoding Latency Prevents Its Use

The current approach has a PC render the virtual world, and also use Wi-Fi to stream the rendered images to the HMD using Wi-Fi direct, Steamlink, casting solutions, etc. The problem with this is, again, latency. In this scenario, video streaming can get down to a 1 ms in latency for compressed MPEG video, but an uncompressed video stream of 1440P 90Hz video is almost 8 Gbps of data. A good Wi-Fi average in a household is closer to 100 Mbps today, even though we aspire to Gbps in the future! Uncompressed scenes over Wi-Fi do not work and compression is required. Because the compression and decompression process will add anywhere up to 80 ms of latency depending on solution – it falls way outside the simulator sickness window of < 20 ms desired.

As we have explored throughout this paper, more bandwidth can solve the problem. From server to renderer, future wireless technologies like 5G and millimeter wave are aiming for 10 Gbps and Terabit numbers. This is not yet practical. There are a few technologies such as Li-Fi which may find a good niche to transmit 10 Gbps at low latency especially in the downstream direction. One could envision a bank of LEDs on the server/PC/STB, and using the surface of the HMD as the optical receiver aperture, it could be an application that works.

Another approach, which is hugely complex, is to accept the latency and try to correct the old image to the new head position. This still produces some delay, but avoids the nausea created by head motion lag. There is a term for this called asynchronous time warp; where a small graphics processing unit (GPU) on the HMD warps the old image into the new perspective allowing a fluid update despite the delay. The problem with this is that it transfers simulator sickness to what is called 'positional judder' where as you move your head side-to-side or do a simple rotation of the head – you will see judder on objects in the image that are close to you. Increasing GPU correction, you can implement a more comprehensive warping which does both translation and orientation correction to the original rendered frame. This needs to include the depth and object displacement which in itself creates artifacts at object edges with gaps and spaces. Even further GPU logic could be applied to fill these gaps interpolating what to put there.

There are a few techniques being considered to send more information to try and fill these gaps. These include

- Render the near and far field objects separately and send separate streams for both
- Render a light field which can be re-projected by large head offsets without dis-occlusion. How much this can minimize artifacts is still a research project
- Use prediction and motion models to predict how to minimize dis-occlusion errors

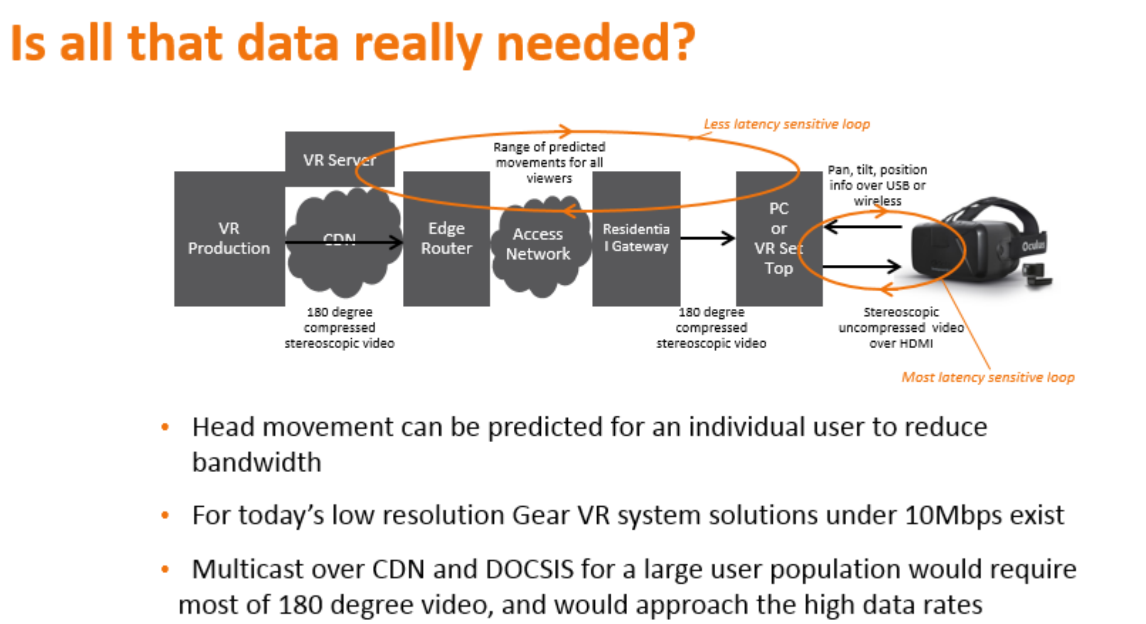


Figure 17 - Trying to Create Predictive Solutions for Lower Latency

Rendering pushed as close to the HMD as possible minimizes latency. Ultimately having the HMD render itself potentially offers lowest latency and allows wireless transmission. Smartphone based VR works this way, but it does not have the horsepower of multi-processor game consoles or PCs to deal with higher resolution VR scenes.

The technology is improving and products like some of the augmented reality devices in prototype stage at the moment also does the rendering locally. These smartphone level GPUs run at their limit for mediocre VR and generate heat. Investment is going into companies that are also trying to develop HMD specific GPUs for VR by specifically targeting the image processing and rendering required for the VR experience and trying to keep the heat down.

There may be two elements to the HMD where the GPUs are contained somewhere else on the body (on the waist or in your pocket) and the display of the HMD only needs to be on your head. This could also be wired. Many of the current VR gamers today actually use a backpack for their PC with battery and cramp the cable on their body to allow freedom of movement for games like first person shooters.

One immediate wireless technology that could have real uses for VR is 60 GHz. Sometimes called WiGig and standardized as 802.11ad. This technology was originally conceived with one use case to be wireless HDMI – to replace the HDMI HD cable. It has been commercialized and can run at 5-7 Gbps in the current 802.11ad generation. It offers this performance with line of sight and suffers sharp drop off in performance when obscured by objects including the human body and head. The future versions of the standard to define 802.11ay will drive the bandwidth even further to 25 Gbps. The 802.11ad/ay standards allow for a low latency potential at under 20ms. However, one challenge can be reformatting the video to resend to a wired HDMI input.

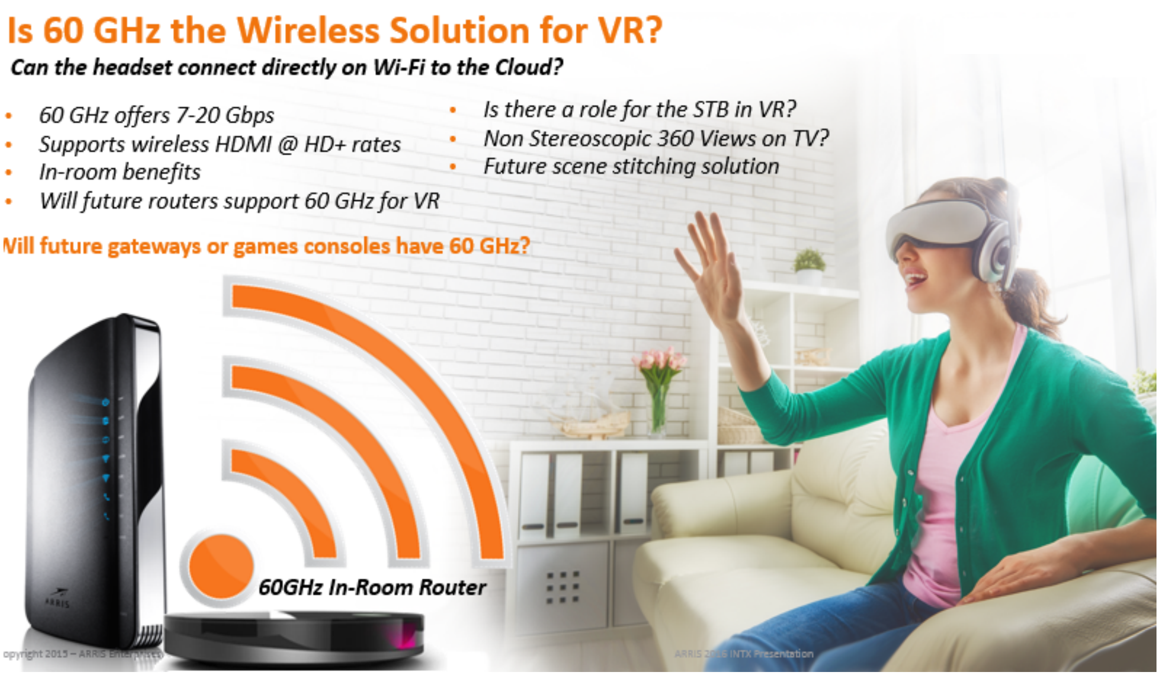


Figure 18 - 60 GHz Could be the Key to Untethered VR HMD Connection Technology

Applying QoS tags to VR video sessions can be done either directly from the network side or by recognizing the video flows and/or the sink device types. Gamers are always keenly interested in the lowest latency and also the best user experience. They might pay more to apply QoE improvements to their experiences.

We already discussed this with file download times, for Wi-Fi based services MSOs could also apply airtime polices for VR to increase the QoS and decrease the latency. This will be a key part of a service provider addition to VR delivery and a relationship opportunity with VR content experience providers both live and file based.

Using local SSD based cache devices in the gateway (GW) there is the potential to cache pre-emptive scenes for local stitching and rendering. Again, the SD read/write technology does introduce latency, but the technology today is sub 2 ms and improving. Figure 19 shows a trivial home wireless architecture highlighting these ideas; here WAN based DSCP tags for VR flows could be used to set up priority queues in Wi-Fi and 60 GHz airtime managers. 802.11ax may prove to be a strong improvement for compressed video VR transmission as it offers much better deterministic performance and allows the VR rendering device to be scheduled over Wi-Fi. It also improves the overall bandwidth capabilities with 160 MHz channels and increased client Wi-Fi antenna. As suggested before, 60 GHz offers a better scope for uncompressed video transmission to the HMD to decrease latency. Assuming the GPU performance of the HMD still lacks the ability to render uncompressed HDMI over 60 GHz in a shorter timeframe.

Figure 19 - Potential Future VR capable GW/Router for Wireless HMD Connection

60 GHz Routing Added to Service Providers Gateway or In Room Device Strategy

As we note in Figure 19, 60 GHz transmissions to the HMD seem to be something to explore for an untethered VR experience. This has a potential to affect the overall home architecture with the use of 60 GHz being a potential inclusion in the service providers devices – for both VR inclusion – and for in room 5 Gbps+ private area network capacity. Initial 60 GHz deployments for VR will probably be dedicated direct link from VR rendering device to the headset and will be capable of sending both uncompressed and compressed scenes to the HMD. However, as 60 GHz devices become more prevalent (see Figure 20) there will be a requirement to add 60 GHz routing to the GW and the extender/Hub. This will allow for non-direct server to a single client device and to share the 60 GHz ISM radio bands of spectrum between VR and other applications.

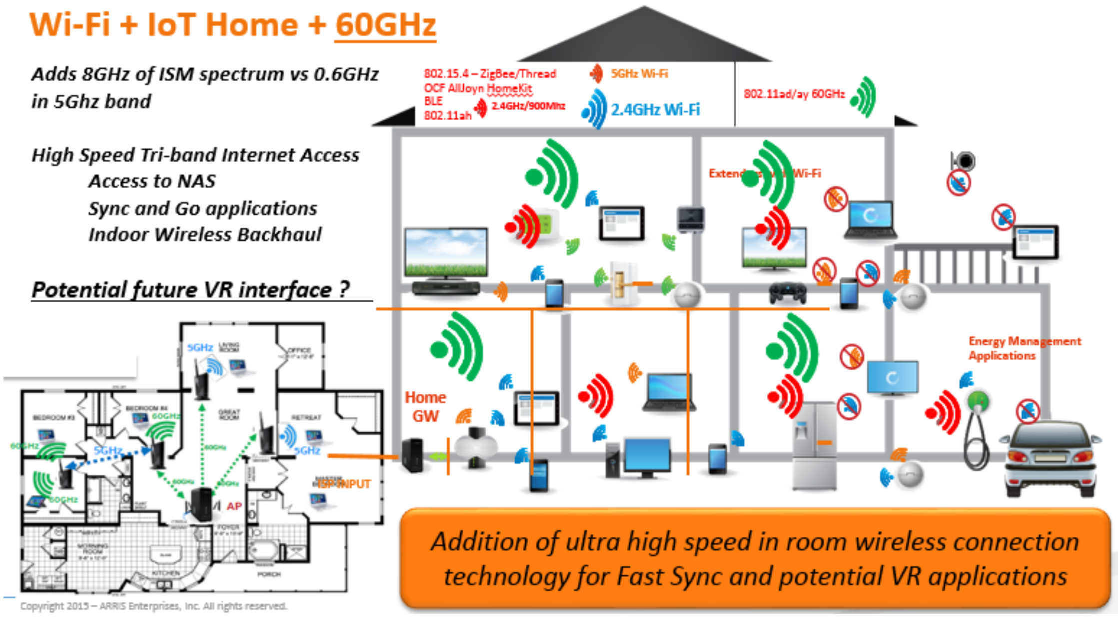


Figure 20 - 60 GHz Additions to the Home: Potential for In Room & Intra Room Networking

This evolution to higher bandwidth, wireless connections in the home for un-tethered VR may drive the home architecture to something like that illustrated in Figure 21. In this scenario, every room has a high speed wireless micronode that may be uncompressing or compressing video to end user devices. As the HMD itself tracks technology improvements in GPUs for power and energy, there may be more compressed video usage. However, the delays in performing the decompression of the VR scene need other techniques to deal with latency issues.

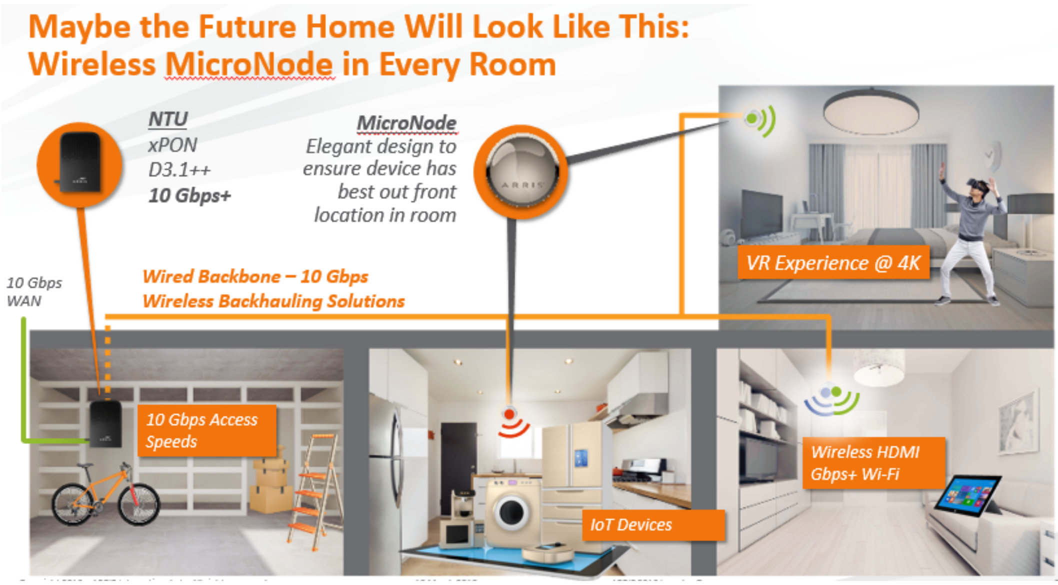


Figure 21 - Future 10 Gbps+ Wireless Home with VR HMD Capability

Pre-emptive Caching Opportunity for Next Rendered VR Scene

With the massive bandwidths required to send a full 360 degree stereoscopic 4Kp60 90Hz scene fast enough to the HMD to not incur simulator sickness, it is clear that intelligent algorithms must be used to minimize the high resolution, field of view area of the scene. Figure 22 illustrates some of the items that may be used. There will be

- Tracking head and eye movement, and keeping the images coming in pre-emptively, in the direction of the tracking on the headset
- Using visual and audio cues. In the example below, the speakers will trigger viewing focus as will the changes to new slides/projections. Sporting events will be similar with triggers such as a player’s running direction, strike of the ball, etc.

Figure 22 - Caching the most likely next focus area to improve latency

This offers the potential for the GW and/or the VR rendering device (PC or console or VR STB) to cache the next scenes to either process them or send them uncompressed to the HMD.

Merging the VR and TV Experience – HMD and/or TV and/or PC/Tablet

There is also the opportunity to merge the VR experience with the TV experience. There are a few combinations of this that have been discussed. These are

(i) Moving the TV experience to a VR HMD experience: The movie or TV show is captured in VR and the user can experience more immersion in the action or scene. Some scenes could be VR and some could be regular 2D. This places new costs on the generation of the movie content (full not partial movie sets would have to be made and administered) or drives the industry to all green screens with computer additions. This would occur later with augmented reality.

(ii) Combining the TV experience with an out-of-band VR experience: Some content producers may create a separate parallel VR immersive experience that can be investigated later after (or before) watching the regular 2D show on the TV. This additional VR session may give the viewer more insights to key scenes in the program and would allow them to discover additional elements of a plot or experience being in a scene.

(iii) Looking at non-stereoscopic 110 or 360 degree FOV scenes on the TV: Using an attached hand held device (gyrometer) to navigate around the scene (like in Figure 22) and have the image display on the static, always-in-front-of-you, TV device. This may be useful for exploring locations like those in news, discovery, or nature programs. For most fiction-based movie productions, the director usually wants the user to stay within a certain frame view as part of the storytelling. However, there may be new takes on directing VR movies that give each viewer a different perspective on what is happening and from where in their own 360 degree FOV.

Figure 23 - Using the HMD to Watch all TV or “Made with VR” Programs

CONCLUSION

We have to wait to see if VR, AR, and MR will become the next version of our entertainment experience and have other far reaching applications, like replacing all the screens we carry with us with one headset/glasses/contact lens projection. Imagine an entertainment experience where you feel as though you are in the scene – whether it’s a movie, TV show, or exploring some foreign land. Now imagine a working environment where you don’t need your PC or tablet anymore. Both are projections on your retina coupled with either a haptic feedback solution for typing in the air or using your eye tracking, voice, or thoughts to initiate actions.

In the near realistic timeframe, VR, AR, and MR look to have great promise for gaming, adult content, and sports viewing. As we have shown in this general tutorial paper – the key elements of the sheer size of the scenes needed to create immersion and the low latency for these scenes to be displayed – the requirements are not yet fully available with the current network, encoding, video processing, GPU, Wi-Fi, HDMI, screen technology etc. However, we are taking baby steps and working on lower resolution solutions, as well as initially tolerating tethered connections to the HMD.

We have shown that there will be a role for the service provider to embrace VR and to add value to the VR chain around improved QoS of file and live delivery. We have defined a symbiotic wireless home solution that does not compete with the VR device companies trying to carve out their own spectrum (to ensure quality and low latency) at the expense of shared spectrum problems in the home.

Figure 24 - Our new virtual work and experiences await us

There is a lot more to understand and develop to perfect this technology, yet we are well on the way to the Star Trek holodeck experience. If you watch the many VR videos on YouTube, as well as some of the concept videos from Magic Leap, you can see that there is a lot of work already done to determine our immersive and presence based future.

ABBREVIATIONS

2D	Two dimensional
ABR	Adaptive bitrate
AR	Augmented reality
bps	Bits per second
CDN	Content data network
DASH	Dynamic Adaptive Streaming on HTTP
ISO BMFF	ISO base media file format
DOCSIS	Data Over Cable System Interface Specification
FOV	Field of view
GB	Gigabyte
Gbps	Billions of bits per second
GHz	Gigahertz
GPU	Graphics processing unit
GW	Gateway
HD	High definition
HDMI	High-Definition Multimedia Interface
HDTV	High definition television
HFC	Hybrid fiber-coax
HMD	Head mounted display
Hz	Hertz
ISBE	International Society of Broadband Experts
ISM radio band	Industrial, scientific, and medical (ISM) radio band
JPEG	Joint Photographic Experts Group
Li-Fi	Light fidelity
LTE	Long-term evolution
Mbps	Megabits per second
MHz	Megahertz
MPEG	Moving Picture Experts Group
MPEG OMAF	MPEG Omnidirectional Media Application Format
ms	Millisecond
PON	Passive optical network
QoS	Quality of service
SCTE	Society of Cable Telecommunications Engineers
SD	Standard definition
STB	Set-top box
TCP	Transmission control protocol
TV	Television
USB	Universal serial bus

VR	Virtual reality
WiGig	Wireless Gigabit Alliance

RELATED READINGS

- **The Edge Resource Center: Leveraging NFV and SDN for High Availability/High Performance Network Functions** – This paper covers the concept of an Edge Resource Center, a small-scale virtual machine interface (VMI) that is co-located with the CCAP device in a cable head-end, and supports high availability and high capacity SDN/NFV resources.

- **Service and Management Orchestration in Distributed High Speed Data Networks** **–** This paper describes a cloud-based orchestrator that leverages SDN principals and works in conjunction with purpose built applications in high-speed data networks. This orchestration capability can unburden the existing network elements in the collection and management of network data, and help create value added services for network configuration, network optimization and network planning.

- **Advanced QoE Monitoring Techniques for a New Generation of Traffic Types Carried by DOCSIS** – this paper explores future improvements to DOCSIS QoE monitoring tools, addressing metrics based on traffic types, SLAs, active subscriber counts, CMTS scheduling, requested loads and sampling rates.

MEET ONE OF OUR EXPERTS: Charles Cheevers

Charles Cheevers, Chief Technical Officer, Customer Premises Equipment, is responsible for the two to five year technology vision of ARRIS CPE Business and the evolution of home gateways, set-top boxes, and user interaction with all digital services. He is responsible for defining architectures for CPE devices and the cloud to ground solutions and evolving customer digital life requirements.

Previously, Charles served as Chief Technology Officer of Europe at ARRIS Group, Inc. and served as Vice President of Product Line Management, Multimedia and was also responsible for ARRIS Edge QAM business.

Since joining ARRIS in 2003, Charles has developed DVB, DOCSIS headend, and CPE solutions and products. He has also focused on convergent solutions for the network and home through all IP delivery. He is a recipient of the Cable and Satellite Euro50 award for his contributions to cable technology.

REFERENCES

(1) Gartner Hype Cycle; Gartner, Inc. online at http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp

(2) "How Independent are HDR, WCG, and HFR in Human Visual Perception and the Creative Process"; Sean T. McCarthy; SMPTE Motion Imaging Journal 125(4):24-33; April 2016

(3) "VR motion sickness alleviated with field of view alterations"; Jack Loughran; The Institution of Engineering and Technology; June 2016; online at http://eandt.theiet.org/news/2016/jun/vr-motion-sickness-alleviated.cfm