5G EDGEXR

5G Edge-XR Final Report

This is the final publishable report of the 5G Edge-XR project.

It describes the project and steps through the major achievements in terms of the use cases enabled and of the technical findings with respect to 5G mobile networks, to edge based GPU rendering and to the economics of delivering such experiences at scale.

Author: Doug Williams

Reviewer: Doug Williams

Version: First draft

Date 06/04/2022 15:40:00

Document name: MS6.1 D1.4.1 5G Edge-XR Final Report V2 To DCMS.Docx

Table of Contents

List of Figures

1 Executive Summary

5G Edge-XR set out to explore the potential role of edge based GPU in the delivery of a range of XR Experiences.

The project built prototypes of seven XR experiences based on Sport, Dance-education, Health, and Construction.

Technically the experiences worked well. The benefits of high data throughput that 5G networks bring, become invaluable if these prototypes were scaled up to become live popular services. In our tests current network latencies of 25-30ms were good enough to deliver satisfying experiences with no reported instances of motion sickness. Further reductions in latency that may be possible using slicing may be beneficial but as far as we can see are not required for the experiences we trialled.

The edge based GPU delivers a visual experience (based on the ability to decode multiple video streams that may appear in the scene, to render particle effects, and perform ray tracing) that are not possible using mobile phones or tablets. User feedback suggest experiences with such visual elements can be valued by consumers.

Volumetric capture technology has been used in the production of 3D video assets that, according to those involved with the dance use case will transform dance education practice in the coming years and also be part of future sports production practice.

Whilst the edge GPU technology based on Cloud XR is good and capable of creating extraordinary and valued experiences, there is still some work to do on delivering experience at scale using edge cloud GPU infrastructure at a necessarily low cost.

Further work is required to further enhance the experiences, to increase visual fidelity of volumetrically captured images, to enable communication facilities as well as just playing out an experiences and to cost reduce the delivery infrastructure.

2 Introduction

5G Edge-XR project is exploring the role that GPU edge compute capability can take in delivering experiences that demand the time sensitive rendering of visual scenes. We describe the project using the following slide.

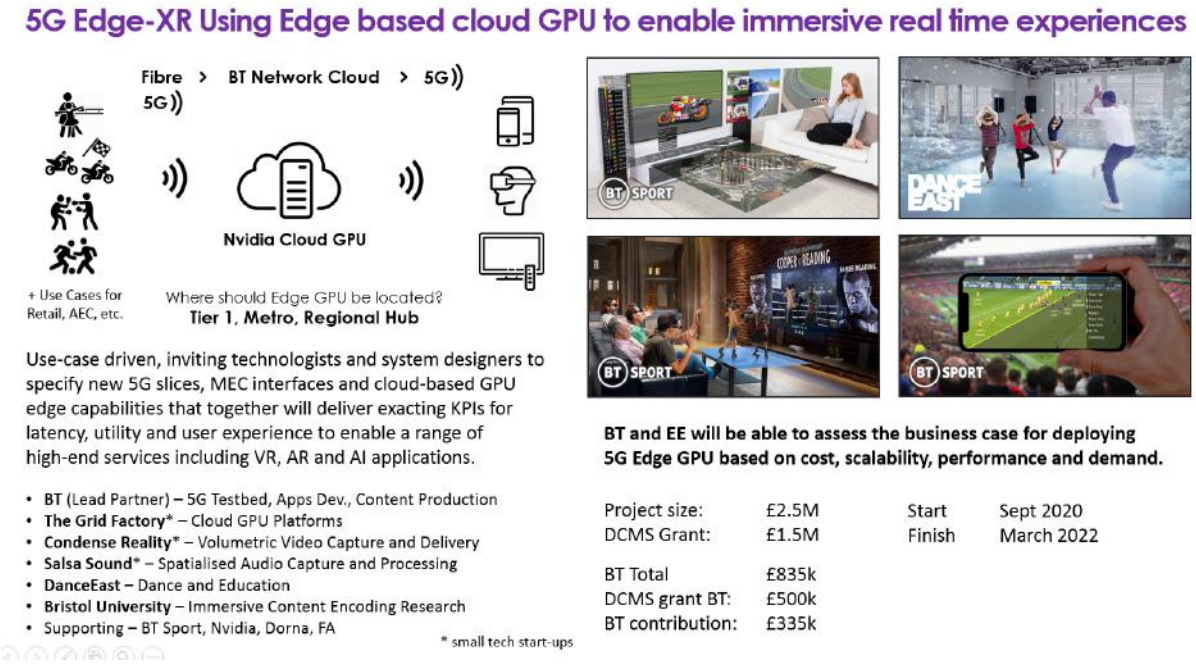

Figure 1 Single slide we use to describe the 5G Edge-XR project

Table 1 The partners in the project and the roles of each partner are described in the following table.

BT will develop use cases, offering design software and production skills to allow the use cases to be built and assessed. BT will also provide 5G network expertise to specify the network requirements. BT will enable access to 5G test networks.
The Grid Factory will, working to BT as if BT were a customer, specify, design, test and run the GPU-based compute facility that will support all the use cases. The Grid Factory have a particular role optimising the GPU cluster capability within the e2e system, responding to the requirements of the use case owners and the service provider (BT). The Grid Factory will also lead on Dissemination and Exploitation for the project with a particular focus on showcasing the use cases and exploring the on boarding of new use cases generated outside the Project.
Condense Reality will focus on the generation of volumetric capture technology to support two uses cases (Dance and Boxing) within the project. Their particular focus will be on generating broadcast quality volumetric images; a task that will be enabled through engineering (to enable newer higher fidelity broadcast cameras) and through upscaling, to enable higher fidelity at lower computational cost, and AI to enhance the image quality when occlusions and errors mean the 3D model and texture would not be drawn correctly.
Salsa Sound will be active in the technology enablers work package generating object based sound sources required for creating spatial sound required for immersive experiences. In the project they will develop AI based techniques to help automate the calibration of the microphones with a view to making the generation of sound objects simpler. Salsa Sound will also be active in the Boxing use case capturing and providing the sound objects.
Bristol University will be developing advanced coding and AI technology to support Condense Reality. Their work will focus on the research and development of new efficient

algorithms for coding and upscaling that can be used in the generation and delivery of broadcast quality volumetric video capture.
DanceEast will lead on the generation of a Dance use case, designed to examine the possibilities for teaching dance using augmented / virtual / extended reality techniques. Dance East will develop and evaluate a test lesson structure using volumetrically captured dance delivered to school children in Suffolk**.**

3 Description of what the project did

The project was organised into 6 Workpackages.

- WP1 Project Specification - This helped provide coordination between partners and to liaise with DCMS
- WP2 GPU-EC and 5G infrastructure – There were two fundamental enablers across the whole project these were the GPU Edge compute capability and the 5G network provision. Work to provide, fettle and manage these resources was controlled in WP2
- WP3 technology enablers – There were a number for technologies that needed work to support our use cased. These included Spatial sound capture processing and rendering and Volumetric capture
- WP4 Experience Production – The technologies were showcased through a range of use cases for:
    - Dance Education
    - Sport (MotoGP, Stadium, 8k 360 VR and Boxing)
    - Architecture Engineering and Construction
    - Medical Imaging and a Retail use case (these were always billed as 'nice to have' explorations with no associated commitments).
- WP5 Assessment – the use cases were assessed in different ways that were deemed appropriate for each use case. Some were assessed from a particularly technical perspective and others were assessed more from the user experience perspective.
- WP6 Exploitation and Dissemination – results from the project were presented through PR releases, through conferences, entries into competitions, and through meetings with key stakeholders.

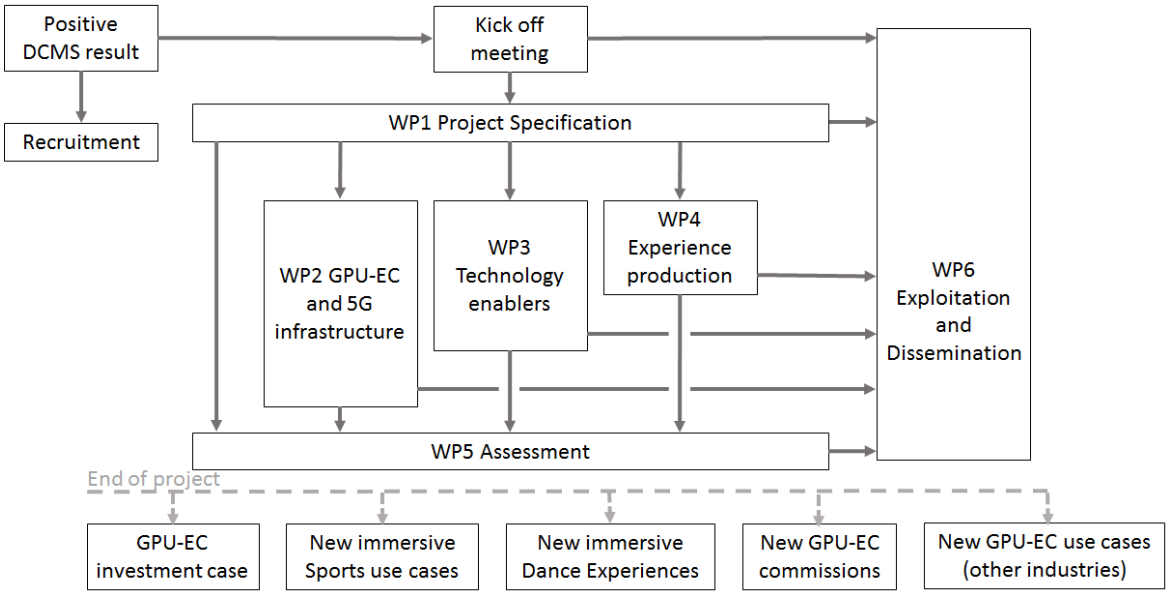

Figure 2 Illustration of the project linking key events (project start and end for example) with the key work packages within the project.

Whilst work was reported within these work packages we report the overall findings

4 Summative findings - Experiences rendered on Cloud XR cluster

The core of the project was about understanding the role of GPU clusters in the delivery of XR experiences over 5G.

The 5G Edge-XR project has shown that the use of GPU clusters can deliver extraordinary experiences in XR over 5G, of a fidelity and complexity that cannot be matched by trying to render XR experiences on the end device without support from network based GPU.

Specifically we have shown that GPU capability can support:

- AR experiences that can be viewed on multiple types of client devices including:
  - Tablets (Samsung Galaxy S7 5G)
  - Phones (Samsung S21, Pixel 3 Google (prop most modern Android phones) + iPhones)
  - Nreal AR glasses (Via Samsung S21)
- VR experiences (the 360 football experience) using an Oculus Quest 2 headset
- AR experiences with round trip times sufficiently short that users did not report lags and reported no feeling of motion sickness. This was true whether the cluster was directly connected to the same 5G RAN as the devices or whether it was accessed remotely at the equivalent of a Tier 1 node. Latency was less than 30ms.
- AR experiences including complex rendering effects such as ray tracing and particle effects have become possible at modest bandwidths (the Cloud XR rendered these experiences at between 15Mb/s and 50Mb/s). little or no discernible differences in image quality were reported, with respect to these different potential bandwidth per stream, when these images were displayed on tablets, smart phones or Nreal augmented reality glasses
- Experiences involving the simultaneous selection of up to 14 simultaneous videos decoded for selection. (this was the situation with the MotoGP experience where we happened to have 14 videos that could be selected.)

4.1 Perceived value of hard-to-render experiences

Whilst increasing complexity of the image rendering creates a problem that can be solved by a cloud GPU cluster, it doesn't necessarily create a saleable benefit. It's difficult to be definitive, but our subjective tests with MotoGP fans were encouraging.

60% of our users told us they preferred to have more than two on board bike cameras selected. (Two is this maximum number available if the app depends solely on the Samsung S21 for the rendering.)

53% of our users told us that they thought the larger room sized map (enabled by the cloud GPU) was the better feature to include in a live service (compared to the smaller table sized map that the could be rendered directly on the Samsung S21

The overall summative assessment by our users suggested that the augmented experiences were preferred to the standard BT Sport covering which scored 7.2 out of 10, and that the large room scale version was preferred to the smaller table top version which scored 9.1 and 8.3 out of 10 respectively). The sample sizes are too small to claim significance, but we are encouraged.

For 5G Edge-XR, our particular focus was on understanding whether features that have been enabled through the architecture of cloud GPU with 5G connectivity are assessed as particularly important for our evaluators. Those features that benefit from this architecture are denoted with an asterisk (*).

Perceived value of different features (1-10 scale) 0 =Not valuable, 10 = Extremely valuable	**Average**
*On board cameras	**9**
*Room sized map	**8.9**
Replays	**7.2**
*Variable viewing position	**7.7**
Tabletop Map	**7.7**
Expanding leader board	**7.6**
Spatial audio	**7.0**
Audio mix selection	**6.7**
Interactive timeline	**6.2**
*Selectable bikes (Parc Ferme)	**6.1**

The bike model when rendered at the resolution we showed, and including reflection rendered from ray tracing and particle effects (for the exhaust) is also a feature that requires the 5G GPU architecture, such detailed models cannot be rendered by end devices. However, this feature was the least valued of the ones shown – that is not to say it was not valued just that, relative to the others it was less commented upon. Whilst the selectable bike model did 'wow' and enchant many of our evaluators, some thought this Wow factor might not be sustained and that the feature might be relatively underused in regular viewing. These findings were as expected as the designers believed the Parc Ferme area would only be used for a short period at the end of the race, but felt its inclusion was necessary to showcase the GPU raytracing and particle effects capabilities in context.

4.2 Techno economic conclusion

The third aspect, after "Is it possible?" and "Is it valued?" is: "Is it economic?". As far as we can tell, based on current technology, current implementations and current risk profile, the answer is no. Technology will develop implementations will change and risks can be mitigated so this won't necessarily always be the case but the findings from the project would not support the deployment we chose for an at scale service. The inclination is to view this negatively; but the fact that the project has provided knowledge to rule an option out as uneconomic and that it has delivered a framework to consider how to assess other options is valuable.

5 WP2 Findings re Edge-GPU cluster

The GPU cluster - which (as a reminder) was based on Four Dell servers, each hosting three RTX8000 graphics cards with 48GBytes of frame buffer – was virtualised using CloudXR. CloudXR is an Nvidia software solution enabling fractionalising of GPUs, something that isn't easy to achieve. The infrastructure is owned by BT but managed remotely by The Grid Factory using their specialist systems integrator knowledge for GPU clusters.

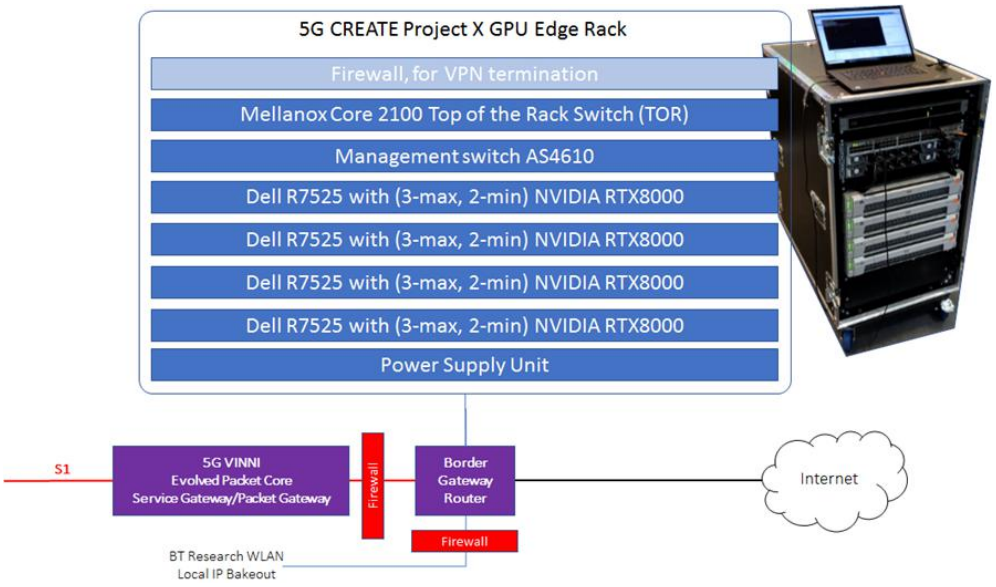

Figure 3 Schematic of the GPU cluster deployed for the 5G edge-XR project

5.1 Cluster management

Key findings:

- With the right skills, the cluster can be managed well, providing reasonably flexible deployment of virtual machines that can be dimensioned to suit the requirements of different experiences.
- The onboarding of new applications has been quite straightforward. This has enabled us to support other projects (eg Green Planet) and to trial 7 different use cases.

5.2 Cluster flexibility

As a project we feel there are significant shortcomings of the current CloudXR approach

We found some limitations which required some workarounds.

We used 3D scenes and, through interpreting SLAM data from the end user devices, allowed users to perceive the scene from a particular viewpoint. This viewpoint was generated through a virtual camera; the scene was then rendered for that viewpoint, encoded and streamed to the end user. This sounds fine, and for many applications, it is fine. But in our scenarios the pipeline exposes a limitation. In commercial deployments we would expect to have many people looking at the same model – this isn't the case for many of the use-cases for which CloudXR is used. So, to have a pipeline that requires one model to be built per user is wasteful. An architecture that would allow many (hundreds or thousands) of virtual cameras to all be viewing the same scene would be better.

Our work around has been to cache the model on the cluster and to import that into each virtual machine. This saves on input bandwidth requirements but still feels wasteful in terms of internal data transfer across the cluster.

We were targeting a multi tenanted cluster capable of dynamically supporting diverse use cases. We imagined (and hoped) that resource can be dynamically or even automatically allocated to match the changing demands from each use case. This was not the case. Instead a virtual machine was specified in terms of frame buffer/memory. The VM Ware hypervisor layer handled the sharing of resource by several VMs using different sharing scheduling protocols. The limitation we found was that only one Frame buffer split could be used for all VMs. Hence we were limited to using the VM profile that was capable of supporting the most demanding use case for all use cases. This limited density achieved on a multi tenanted platform.

The cluster was sometimes limited by the encoder capability, this meant that we were sometimes limited to only 2 instances per GPU card whilst knowing that the specifications of the card was, in other respects (such as frame buffer), more than was required to support two simultaneous experiences.

As a service provider, imagining use cases in which hundreds or thousands of instances may for short times be sharing the platform we were somewhat confounded by the current licensing model which has per user licensing. This seat licenses model is sensible in enterprise use cases (think an architect or designer accessing a Virtual machine or a shared GPU cluster in an office environment) but fails to meet the needs of the consumer use case. To be fair we know we are mis-using the GPU cluster in some respects. These issues have been reported to Nvidia; they recognise there are options, including making the system work independent of Windows and VMWare that should make the license costs less prohibitive.

There were some aspects of the way CloudXR behaved that surprised us. CloudXR works but it is a kind of ‘black box’. It is clearly making decisions but the reasoning behind those is not clear and this feels uncomfortable. For example we had three users joining the same experience on a significant 5G link (several hundred Mb/s) the video encoding would sometimes create a 50Mb/s stream and sometimes create a15Mb/s stream, and sometimes create a 5Mb/s stream. There was no obvious logic to this. The team would have liked to have been able to have more control over the encoding parameters, perhaps specifying a ceiling per instance. Nvidia have been informed of this and are trying to find a balance that would help them protect the IPR embedded in CloudXR whilst also offering more control to the users.

6 WP2 Findings - Private 5G network

6.1.1 Test network characterisation

The project used numerous network configurations based on the following RANs

- The Private 5G VINNI RAN at Adastral park
- The public EE RAN at Adastral Park
- The public EE RAN at Dance East
- The private Neutral Wireless network at StoneX

These were accessed, predominantly using Samsung S21s, and 5G CPE (WiFi routers) of various sorts. Devices were then connected via WiFi to the routers.

The following table provides an approximate characterisation of the networks used. As far as possible the device set up is fixed, on public networks, date, precise location and other network traffic on the RAN are variable. It will be these factors that account for the ranges of values recorded.

	VINNI Setup 1	**VINNI Setup 2**	**EE @ Adastral Park**	**EE @ Dance East**	**Neutral Wireless at StoneX**
Signal strength	STRONG Antenna points to indoor location.	Medium (indoor at 5G Lab)	STRONG (indoor in a steel frame metal and brick building)	WEAK (Indoor in a concrete bunker)	STRONG (devices and antenna in same room)
Centre frequency	N78 3650MHz	N78 3650 MHz	N78 3560MHz	N78 (Presumed) 3560MHz	N77 3935MHz
Spectrum	100MHz	100 MHz	40MHz	(Presumed) 40MHz	100MHz
MIMO	32T32R 8*4 Samsung Galaxy S20	Up to 4x4 on an Samsung S20	Up to 4x4 on a Samsung S21	Up to 4x4 on an Samsung S21	Up to 4x4 on a Samsung S21
UL throughput	67 – 95 Mb/s	30-50 Mbps	58-67Mb/s		
DL throughput	250 – 650 Mbps	400-550 Mbps	390-444Mb/s		340Mb/s
Latency	12 – 17 ms (round trip)	25 ms	14-22ms	14-22ms	25-30ms
Jitter		2-4 ms	2-19ms		

6.1.2 The pragmatic stuff about doing trials with private 5G networks

In the course of this work the application teams worked with the networks teams to make sure we could deliver the experiences. As we did this, we learned some of the pragmatic stuff related to delivering experiences over private 5G networks.

We were surprised by the hurdles that need to be navigated to get “off the shelf" devices that are common (i.e. mobile phones likely to bought with contract for the main UK providers), to work. Even though these technical limitations are known to MNO Operators, they are not available through the device technical specifications documents, hence we were discovering the device limitations by test and try. These hurdles were surmountable but they are obstacles. This is what we found:

- Private PLMN
    - Your 5G devices must be able to recognise the PLMNs of the private network. The PLMN is the sort of 5G equivalent of an SSID in WiFi networks. Many standard phones, bought to work on public networks (those provided by EE, Vodafone, O2 and Three), only recognise the PLMNs of those registered providers (so called “white listed PLMNs”). So will not “out of the box” connect to the private 5G network. We were able, through the relationship we (EE/BT) has as a provider, get Samsung devices re-flashed to work with private PLMNs. Others would be forced to use

phones that intrinsically support private PLMNs. Apple and Samsung devices we came across did not do this.

- Stand Alone (SA) Mode
  - Similarly to Private PLMN, the support to SA Mode is also currently restricted by the MNO "approved" OS in devices, which currently only support Non Stand Alone (NSA) Mode (because NSA Mode is what current MNO public networks support). Special OS flash was required in common devices (e.g. smartphones from Samsung and Apple).
- Shared Spectrum in N77 top end
  - Some devices will not connect in all of the frequencies of the N77 band top end, 3.8 MHz to 4.2 MHz, intended for Shared Spectrum licensed use. We discovered this limitation in two scenarios:
  - Some devices do not connect in the frequency band known as the upper end of the N77 band from 3/8-4.2GHz. They only operate in the more normally used 3.3-3.8GHz range in N78 band (which is auctioned by MNO for nation wide coverage)
  - Some devices are limited within the upper end of the N77 band used by OFCOM for test licenses. For example we were granted, by OFCOM, frequencies above 4GHz but our partner vendor 5G NR base station could not operate above 4GHz. We had to repeat our request to OFCOM and since they do not ask you to specify a preferred frequency (you just get what you are given on the assumption any frequency will be usable) this required persistence.
- MIMO in devices
  - Not all devices support MIMO operation at the same level as the 5G NR base station. To get high throughputs it is normally necessary to use MIMO, typically to 4x4. Not all devices support MIMO so you may find limitations.

It can be noted that the 5G CPE (Router) off the shelf devices offered by MNO Operators typically have the same software induced restrictions (lack of support to Private PLMN, SA Mode and N77 Frequency bands). However, unlike the current situation with the Smartphone devices, there is a growing market for the 5G CPE (Router) and 5G Dongles (USB) devices availability that do support these 3 key requirements (Private PLMN, SA Mode and N77 Frequency band 3.8 to 4.2 GHz) off the shelf. Examples are Quanta and CradlePoint among many others.

7 WP3 - Volumetric capture findings

The goal of the project has been to develop broadcast quality images captured using a volumetric capture rig in a volume 7mx7mx4m and available for use within the typical time window for use in live broadcasts (of the order 1 minute). This goal has not been achieved, but the volumetric captures have developed significantly and we have learned a lot about how to use volumetric rigs in a production like environment. Partners continue to work together towards the original goal.

The following table shows some of the developments that have taken place within the project

Volume capture
2mx2mx2m was our starting point based on a system with 8 USB 3.0 connected Kinect cameras, This set up has some flexibility and we often used a 9 camera 3mx3mx2m volume capture Limitations of the KINECT camera (Frame rate, accuracy at distance of the depth sensing tech, the inability to adequately recognise certain hair types (Afro), and instability led to a

system redsiign and upgrade based on stereo camera pairs (analysis of the image pairs led to depth characterisation). This system had much higher aggregate data rates (due to it having more cameras 32 cf 8 and higher frame rates, 120.60fps cf 30/15fps).

The new system was designed to capture larger volumes up to 7mx7mx4m and the Ethernet connectivity of the cameras (rather than USB 3 connectivity) was believed to be intrinsically more stable.

The new rig was briefly tested in the lab and then road tested at a broadcast bout at Wembley Arena in March 2022. The volume captured was 6mx6mx4m.

Image quality

Image quality is given an objective score called structural similarity. This compares the output from a virtual camera sited at the location of a real camera with the output of a real camera at that location. We, somewhat arbitrarily, suggested that a structural similarity of 80% would probably be needed before the outputs were deemed of good enough quality to be used as part of a TV broadcast production.

We found that this objective model was useful up to appoint but that certain improvement, in rendering of feet, hands and faces had, subjectively a bigger impact of quality than they did on the objective measurand.

The new larger rig based on paired camera rather than Kinect cameras was deployed for one text run at live event in March 2022 at a boxing event. This was a great learning experience, a phrase which often presages a disappointing result, as it does in this case. Captures from this first attempt suffered from a number of issues, most of which appear addressable. Some were anticipated but the mitigations put in place were not as successful as was hoped. Had we had no COVID, we may have learned these lessons earlier and been able to 'go again' within the project lifetime. But, we had Covid and the delays that caused means we have to 'go again' outside the project duration.

The issues we encountered were related to:

- Camera wobble
- Calibration
- Focus
- Exposure settings
- Inexplicable failure to capture by some cameras
- Software license support enabling certain key capabilities

What this meant was that the quality of these captures was poor. This is OK, the process is iterative, we'll get better and this was the first attempt.

AI pipeline

Artificial Intelligence with deep learning approaches was used to help improve the overall image quality. Training regimes and approaches were trialled by Bristol University to identify the best methods for improving image quality. Issues addressed in this way included:

Background removal A background removal process meant the system no longer tried to decode and render the background and thus became more effective

Occlusion management Occlusion was inevitable, parts of the volume being filmed would be obscured. AI techniques can, over several frames, fill in parts of the image by predicting what would have been there based on models learned from previous frames.

Region of interest coding

Subjectively we focus on faces hands and feet. Errors in the ways these parts of the body jar more than say the way clothing or the back is rendered. In consequence AI test showed that it was effective to identify the key body parts and to use smaller segments in these sections to create a better overall subjective impression of an encode for a given number of polygons.

Green screen The system does not need green screen to be able to extract the subject from the background, but some of our captures, for example in Dance East, took place in a green screen studio. It was not imagined that this would be problematic but it exposed a foible of

the system that meant it did not extract the subject from the green screen as well as might be anticipated. So an AI fix was required to help. It was caused by reflections and shading of the green screen.

Salt stain removal One rendering errors led to fluctuations of shades where those shades were not really fluctuating. This error looks like a "salt stain" on clothing. A system was generated to correct for salt stains recognising that they were encoding artefacts and thus leading to a more subjectively pleasing render.

Use in Production

Through using the volumetric rig "in anger", and thus necessitating rigging and de rigging a number of things were learned about the rig and about how to get the best from it.

Cable runs - The technical rig, based on the Kinect cameras is fickle. In particular there appeared to be data glitches that made the cameras appear unstable. The inclusion of USB 3.0 connected NUCs close to each camera which performed elements of the data capture (Generating jpegs) addressed this issue. It is believed that some combination of long USB cables, the data transport and processing on the fusion PC led to these glitches. Whilst it's tempting to blame the long cables, we found the system to be more stable even when the NUCs were connected using the same long cables that originally were connected directly to the fusion PC.

Ambient light levels - In conditions when ambient light levels change it can be tricky calibrating the rig and in selecting the appropriate exposure on the cameras. This was only noticed once the rig was set up in room with lots of windows. The issue can be managed by obscuring the daylight or more frequent calibration, but left un checked it can lead top poorly exposed images.

Camera movement - Camera vibration / movement when mounted on flexible surfaces. Dance and boxing both (it turns out) use floor set-ups that are not rigid. The flex in a sprung floor can lead to significant oscillating camera displacement (sway) which can affect image fidelity and calibration. To address this the team, when possible, looked for mounting position that were less susceptible to floor movement. This involved:

- taking more care with the tripod set ups. We found it useful to:
- use ballast to slow down the frequency of any sway
- dropping the central pole of the tripod to be trouncing the floor so the tripod had 4 points of contact
- using lighting truss frameworks rather than individual tripods
- using suspended lighting trusses to mount cameras isolating them from the ground movement.

And, apart from attempts to physically minimise vibration, software stabilisation of the image could, in principle at least, be employed.

Which of these is most likely to reap benefits is not clear.

8 WP3 - Spatial sound capture findings

The original plans for the spatial sound were thrown into disarray by the pandemic. Spatial sound reproduction in stadia needs matches to be played, it needs crowds in the stadia and needs testing time in stadia to prove and refine the test captures.

Whilst matches were played, it was behind closed doors. When crowds returned, production crews were, quite rightly, kept to a minimum and research captures were not possible.

Perversely, this enabled a huge amount of fundamental enabling-work to be carried out on the software. The software was re written in a more flexible format and a number of features and

capabilities were then developed against this robust software framework. When we were finally allowed to complete test captures the results were very encouraging.

The AI worked well and by the project end it was able to identify 6 audio objects each with almost with almost 90% accuracy (prior to the project it was only capable of recognising one audio object). In parallel the speed with which objects could be identified was reduced top below 40ms less than a single video frame at 25fps.

In the final days of the project an audio capture was completed at a Boxing matching. This provided apart from the audio captures which led to the generation of new identifiable sound objects, invaluable experience of the pragmatics of capturing audio at a live event.

The software developed in the project proved to be robust and has exceeding our own expectations, become a commercial product within the life of the project.

Spatial audio brings spatial video to life. Professional AI powered tools for spatial capture, audio identification and spatial rendering are now available to do this.

9 WP4/WP5 - Use case findings

There are documents covering the evaluations of each use case. The descriptions below are necessarily brief. To understand the use cases you can see these videos or refer to the more detailed document describing each use case and their evaluation.

StoneX summary video from PR day March 9th 2002:
https://www.youtube.com/watch?v=0LHwXxVhnLE

MotoGP (Tabletop)
https://youtu.be/JSad5jDY35g

Dance 'Content'
https://youtu.be/vNySW8biFIM

9.1.1 Boxing

Figure 4 Guide image for the boxing use case, showing a group of friends in a domestic setting watching a 3D rendered image of a boxing bout apparently projected onto the coffee table in front of them, with the broadcast and some additional screen apparently projected onto the wall in front of them.

This was a TRL 7 first attempt – the use of volumetric capture in a live production environment for the first time. The team learned a lot around the production details – rigging and placement of cameras etc. the consequences of having cameras and audio capture devices in the rigging way beyond the reach of humans. The instability of the floors leading to mitigations being required for camera wobble and important but mundane details like cable routing. These were useful, absolutely necessary lessons that can only be worked through by doing such trials.

- Test runs based on the smaller camera system were never of broadcast quality.
- The capture quality from the one live broadcast set up attempted was rendered offline (not in the broadcast window of about a minute). Unfortunately the captures were not usable. Some of the issues mentioned above, physical stability, and inability to control cameras remotely once set up, together with some non-optimal calibration procedures mean these first captures are often blurry, incorrectly exposed and that some captures failed altogether.

Usually a new capture technology requires several live events to perfect the process. While disappointing that our "one shot" failed, it's not entirely surprising. We will go again, beyond the project.

9.1.2 AEC

Figure 5 Guide image for the AEC use case, shows two people wearing XR headsets together looking at a model of a building.

The AEC use case feedback was very encouraging. The use case is targeted at the AEC industry (AEC stands for Architecture, Engineering and Construction) and was carefully designed to address a key problem for that industry called fee-burn. Fee-burn is the accrual of additional building costs caused by errors and late changes in requirements. The collaborative experience in which main contractors, subcontractors, clients and architects could all inhabit and explore the built space potentially overlaying BIM models onto the real world progress of a half completed building would, the stakeholders agreed help clients to understand the building they were receiving and allow them identify change requirements sooner, in anticipation of the build rather than after the event. The ability to collaborate through AR should reduce travel, save time and save on billable costs from stakeholders like Architects.

The role for 5G is quite apparent; in half built sites there is unlikely to be significant data infrastructure except for an existing public mobile networks. It is possible to install temporary WiFi networks but if 5G was available this wouldn't be required and this cost would be saved. The applications software is probably already sufficiently sophisticated though workflows and embedding this way of working will yield smoother workflows in time. The issue is 5G coverage. When it's there, it's great but if it isn't, then high bandwidth video-rich services like AR are not likely to be reliable.

9.1.3 360 football

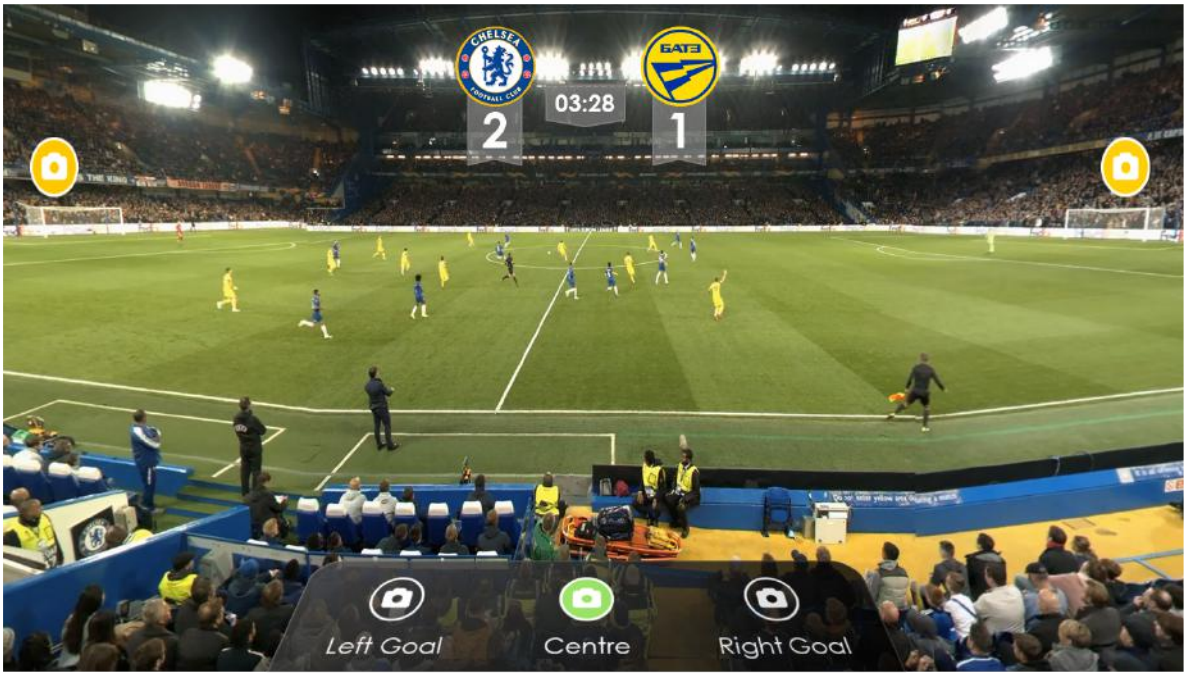

Figure 6 Guide image for the 8K 360 football use case. The illustration shows the viewpoint, that may be shown in a VR headset or a flat screen device (tablet), from an 8k 360 camera positioned at the half-way line of a football match, showing interaction options at the bottom of the screen that give the user option to 'jump' to an alternative camera position.

This use case showed how cloud GPU could change the workflow for BT Sport's 8K 360-degree immersive sports offering. The service delivered using our architecture was equivalent to the current service, but the access to the GPU compute capability significantly reduced the end to end delay.

The current BT Sport offering utilises a cloud tiling process which adds approximately 15-20 seconds delay to the live streams, meaning 'glass to glass' (camera glass to TV glass) delay is typically around 30-40 seconds from live.

The 5G Edge-XR architecture creates individual user specific field of view cropped streams as part of SteamVR and CloudXR UDP delivery process which adds negligible delay, suggesting a likely glass to glass delay of around 5 seconds from live.

The benefit here is from the Cloud GPU capability more than the 5G network. The 5G story here is the reassurance that, even when experiences create much greater demands on network bandwidth, a 5G network will be capable of supporting such extraordinary experiences even, if you use the cloud GPU delivery architecture we used, on modestly powered devices. If three people were watching together this might demand close to 45Mb/s. Thus it suggests that 5G if available is well placed to serve a role in delivering home broadband.

9.1.4 Dance education

Figure 7 Guide image for the dance education use case. The image shows children wearing AR glasses in a dance studio. The wintry scene and semi-transparent image of a dancer superimposed on the dance studio indicate the things that the children can see in their AR glasses.

The focus the project placed on the method and practice of teaching led to really encouraging outputs. A lot of thought was put into how the technology can be harnessed to achieve educational goals. This is good design practice. All projects should have good design practice.

The subjective image quality of the volumetric captures was not the best we've seen. Better image quality would be nice but a minimum threshold had been exceeded and the images were helpful in eliciting good outcomes. Practically the set up was manageable – though, as a prototype, some elements of the app could have been more stable. What was encouraging was the higher level teaching goals. The feedback is subjective but key feedback we received included:

Commentary from assessors suggest pupils learned more and concentrated more. In particular the ability to have input from a variety of different teachers and performers gave pupils the freedom to engage physically as they wished. Pupils enjoyed the variety of teachers it was noted that their movement quality improved after seeing the input of professional dancers.

Self-reported from pupils:

- 88% agreed: "Do you think you learned more than usual?"
- 81% agreed: "Did you concentrate more than usual"

In this use case when used in anger we think 5G could play an important role. If bandwidth was wholly underutilised in a school then WiFi connections may work, but in the case that the school network is often significantly utilised then 5G could be invaluable, adding significant ad-hoc bandwidth to deliver video-rich real time interactive experiences to a classful of children – but only if

the 5G coverage and the required capacity, was available. (each experience requires about 30Mb/s so, for a class of 30, 30*30 = 900Mb/s is required).

9.1.5 Stadium experience

Figure 8 Guide image for the stadium use case. The image shows a spectator in crowded rugby stadium holding their mobile phone up and displaying an image of the game captured on the mobile phone viewfinder with superimposed additional data such as player names and a menu of choices that would provide further information pertinent to the match they are watching.

It's early days for the stadium experience. Achieving the results we did was a phenomenon. This is hard. Most of the challenges came around precise alignment required to deliver the overlays graphics live onto the viewfinder image on your phone. The SLAM technology and calibration approaches we adopted worked but tended to drift so recalibration was needed.

As an experience it's early days. Impressions offered by the press following the StoneX day were positive, but it's hard to know how these AR experience will be long term.

If they prove popular, in-stadia experiences make a strong case for 5G networks, possibly private ones. If several thousand users are trying to access services, then capacity will be key and 5G networks with their efficient use of spectrum, offer that.

Technically the ball tracking capability works. The issue will be accurate SLAM (Simultaneous Location and Mapping) or device orientation accuracy. More work will be needed to deliver accuracy required to make this a reliable service.

9.1.6 MotoGP

Figure 9 Guide image for the MotoGP use case. User seated at home on a sofa and holding a remote control sees (through AR glasses, not shown) a number of screens showing video feeds and a leader boards from a MotoGP race. Appranetly floating in front of her is a map of the racing circuit showing the relative positions of the riders around the circuit.

The MotoGP use case was deliberately titled at hard to render experiences as described above. It appealed to our user survey which garnered assessments from keen MotoGp fans who already watched the BT coverage of MotoGP. As described above the evaluations, whilst not statistically significant due to the small sample size, suggest that current fans of MotoGP would prefer presentations of the form we delivered, using augmented reality with multiple selectable virtual streams including a race maps showing the positions of the riders on the track.

The role 5G plays here, in what is primarily imagined as a home based experience, is the reassurance that, even when experiences create much greater demands on network bandwidth a 5G network will be capable of supporting such extraordinary experiences even, if you use the cloud GPU delivery architecture we used, on modestly powered devices. If three people were watching together this might demand close to 100Mb/s. . Thus it suggests that 5G, if available, is well placed to serve a role in delivering home broadband.

9.1.7 Medical imaging

Figure 10 Guide image of the Medical use case. A white coated medical practitioner is shown holding a tablet device with a semi-transparent three-dimensional image of a brain floats above the tablet.

This use case has a high bar to clear if it were to become commercially available. It would need to prove improved clinical efficacy. The ability to view in VR medical images that usually presented as slices, is intriguing. Our experiments show that standard medical 3D Dicom encoded images, can be rendered as 3d models in AR/VR that can be manipulated and correlated with existing sliced images. Bandwidth requirements, given that viewers are watching a video stream, are 15-30Mb/s. The presentation is clear and fun clinical benefit is unproven.

In clinical environments 5G could be used as wireless coverage perhaps through private 5G networks but it is likely to face stiff competitions from existing site wide fixed / WiFi infrastructure provision.

10 Knowledge gained and impact of results

The 5G Edge-XR has given the partners the opportunity to refine and improve the component technologies: volumetric capture, spatial sound (capture and rendering), cloud-based GPU-rendering as well as to explore the customer experiences possible when we bring these technologies together.

The results in this project are largely existence theorem proofs, i.e. the project was not improving existing services but prototyping new services, just to see if they appeared viable. This applied to most of the use case as well as the cloud GPU deployments we used.

The knowledge gained allows stakeholders to look at these opportunities and to understand what the “levers” are that need to be moved to allow these prototypes to develop to become viable services. Partners can then decide whether to work on moving those levers.

At the end of the project the impact is that partners in the project have a better idea what to do next.

For example:

- BT will work with CR to make volumetric capture tech more robust and capable of operating in a live production environment. This will include working on mitigations against vibration, both physical and software based (image stabilisation) and conducting further tests at live venues over the course of 2022.
- BT will switch the focus of the techno economic evaluation to
    - understanding how the platform could be multi tenanted and understanding the utilisation improvements this could bring
    - looking at different options based on using hyperscalers' infrastructure
    - understanding how system design and tech evolution will impact the user density
- The Grid Factory will explore commercial models for physical cloud GPU deployment with a view to building a sustainable business serving users demanding the highest possible performance. An innovative price per GPU per month is anticipated as an alternative to the more usage based pricing of hyperscalers.
- The Grid Factory will continue to work with AEC stakeholders with a view to completing further tests hopefully in relevant industrial setting (hence moving to a higher TRL)
- Condense Reality will focus (with BT) on addressing the barriers to enabling volumetric capture in a live environment. This will include software and control systems as well as mundane work flow and process based improvement to enable all the appropriate checks and calibrations to be completed during the short deployment window prior to a live production event.
- Salsa Sound will build on their ground-breaking ability to extract sounds in real time and explore the opportunities this creates to deliver new sound based services. The work completed has provided a palette of proven possibilities; using these to generate compelling propositions for clients will be the focus of Salsa Sound's developing business.
- Dance East remains open to opportunities around AR based learning and teaching. As yet these are not deemed commercially viable; focus instead will be on exploring new models for dance education and engagement, using the experience of 5G Edge-XR as an indicator that the organisation has the capability to complete such investigations and as a pathfinder that helps steer future research both methodologically and technologically.

10.1 GPU Edge cloud

The progress regarding the experiences has been described earlier. The underlying architecture - a cloud based GPU cluster being used to perform complex rendering - was also a key part of this investigation. BT wanted to:

- Understand whether this cloud GPU approach could deliver compelling experiences The fear was that end-to-end latency may harm the experience for the end user?
- Understand some of the data inputs needed in building a business case for a cloud GPU based infrastructure play

Technically we have found the Cloud XR GPU solution works well. It delivers compelling experiences with higher visual fidelity than is achievable on mobile devices and we have been told that some features that depend upon this complex rendering are valued by our trialists.

The second part, the understanding of the business case, suggests that there is lot of work to do still. Analysis from this project advises us that, for the tpes of AR experiences we were developing, the current architecture (one based on RTX Nvidia 8000 graphics cards) can support 4 or 5 simultaneous

users per card, or 12-15 simultaneous users per server. Scaling this to a service that we hope might support say 10,000 simultaneous users for a live event, is expensive. High costs can be OK if they are met with high revenues. 5G Edge-XR has helped BT to understand the factors in the cost benefit equation: These are our overview findings

The cost basis for the licensing of the current GPU set up (Windows based, uses VM ware and Cloud XR) is too high. Lower licensing costs may be achievable by:

- Working with licensors to develop, or access, different models
- Working with open source solutions (eg Linux and a Linux based VM software technology) to deliver similar capability
- Working with other providers (hyperscalers) on a pay per use model

The number of simultaneous users on a GPU needs to be increased. This can be achieved by:

- Better software design – some applications were not designed to mirror the capabilities of the GPU cluster and used CPU more than GPU. This optimisation is complex, the defaults is to design for the dominant form of delivery which is to design for mobile phones.
- Having a more flexible way of defining the resource allocated to each instance. The fractionalisation of GPU frame buffer was defined at VM level and all VMS on a particular GPU had to be identical. It seems likely that, to avoid wasted compute resources, and achieve higher user densities a more flexible schema would be useful.
- Using newer, more powerful GPU cards, will of course support more simultaneous users, so Moore's law type effects will help

The revenue streams need to be diversified to increase utilisation and to amortise costs and reduce prices.

The 5G Edge-XR project was deliberately trying to mirror a multi tenanted world with use cases across education, AEC, Retail, medicine and Sports production. More work needs to be completed to estimate the likely take-up of AR service in these sectors and to understand whether notional benefits of using Cloud GPU can be translated into compelling value propositions. This is complex. It is also painfully recursive, as scale would help amortise costs. To invest in edge GPU clusters is not a "slam dunk" decision. It would need to be backed by significant strategic intent, and significant investment and requires more certainty about the market. The next steps are to:

- Better understand the market potential, market size, appetite, to identify key early adopters
- To look (or continue to look) at other option other architecture. Perhaps using something other than CloudXR, perhaps partnering with hyperscalers, perhaps focusing more on GPU sold with private networks as much more on premises solution.
- Consider the role for bundling Edge Cloud GPU technology with Private 5G solutions in deployments that are more constrained by geography or sector and less constrained by willingness to pay.

11 Observations and suggestions

OFCOM should update its process for applying for test spectrum license. Because not all devices serve all frequencies in the top end of n77 band there should be an option to state the frequency range within which the applicant needs to operate. For example our devices would not operate

above 4GHz but the frequency allotted first was above 4GHz. It took a lot of emails and phone calls to get this addressed.

Some of the key upsides of 5G are dependent upon providing connectivity where connectivity is poor. Currently those placed ill served with other forms of connectivity (fibre fast 4G etc.) are likely, for the very same economic and business reasons that currently make it unattractive to serve such locations, to be amongst the last places to get better 5G coverage. Any help assistance, encouragement or subsidy to accelerate 5G roll out in the areas that are otherwise likely to be last to get coverage would help deliver the potential benefits of fast 5G connectivity sooner.