CORPORATE PARTICIPANTS

Lilian Veras
Moderator

Richard Mills
Sky VR Studios – Technical Director

Matt McCartney
Sky VR Studios – Head of Immersive Technology

Rob Koenen
Tiledmedia – Founder and CBO

PRESENTATION

Lilian Veras

Welcome, everyone, to the Intel Network Builders webinar program. Thank you for taking the time to join us today for our presentation titled “Producing Live VR Events at the Highest Quality”. Before we get started, I want to point out some of the features of the BrightTALK tool that may improve your experience. There's a Questions tab below your viewer. I encourage our live audience to please ask questions at any time. Our presenters will hold answering them until the end of the presentation. Below your viewing screen, you will also find an Attachments tab with additional documentation and reference materials, including a number of websites and documents mentioned in this presentation. Finally, at the end of the presentation, please take the time to provide feedback using the Rating tab. We value your thoughts and we'll use the information to improve our future webinars.

Intel Network Builders webinar series takes place live twice a month, so check the channel to see what's upcoming and access our growing library of recorded content. In addition to the resources you see here from our partners, you'll also find a link to our comprehensive NFV and SDN training program through Intel Network Builders University.

Today we're pleased to welcome Richard Mills and Matthew McCartney from Sky VR and Rob Koenan from Tiledmedia. Richard Mills is technical director at Sky VR in the UK. He has over 30 years’ experience and a proven track record in creating high-end imaging solutions for the film and broadcast industries. Richard has held senior operational, technical, and management posts, and has award-winning experience of Shoot, Post, and VFX supervision. At Sky UK, Richard has been responsible for setting technical standards, best practice, and devising workflows for 360 and VR content. He has worked extensively to integrate virtual and immersive content within a broadcast framework. In addition to this technical work, Richard has contributed in the fields of cinematography, direction, and Shoot and Post supervision on numerous VR projects. Matt McCartney is Head of Immersive at Sky. His role is to explore how AR, VR, and other immersive technologies can enhance or augment the audience's viewing experience. Matt is an expert in immersive technology with a core focus on VR and AR experience creation, and content delivery with a major focus on live sports content. Matt’s background is in technical delivery. He has previously-- He was previously a program director, head of delivery, with extensive OTT broadcast platform experience.

Rob Koenan is a co-founder of Tiledmedia, the leading VR streaming company. Rob is Tiledmedia’s Chief Business Officer. Rob has over 30 years of experience in research and development of novel multimedia technologies. He held many leading roles in multimedia industry groups, including MPEG, where he initiated the MPEG-I standard with I for Immersive. Rob also co-founded the VR Industry Forum, VRIF, and he was VRIF’s first president.

Welcome, everyone, and thank you for taking the time to join us today. Matt, over to you.

Matt McCartney

Hi, thank you. I'm going to spend a couple of minutes just talking through the context of the bulk of the talk that Rob and Richard are going to pick up on a little bit later, and help set the scene as to why Tiledmedia and Sky are working together, and a little bit of background around the Sky Worlds project, which is the front-facing VR product where all this innovation and all this technology is being surfaced.

So, back-- I say back in the day. A few years ago, when we set out with a mission statement to build Sky Worlds and targeting virtual reality as the platform we wanted to go out on, it was very clear from the outset that video quality was probably our highest-- highest on our agenda to hit. I think with anyone that's familiar with VR sort of four or five years ago, I think will be honest that the picture quality, certainly from a livestream, certainly wasn't as good as it could be, and one of the things that working at Sky is you're always held to a very high standard for picture quality. So, when we were looking to build Sky Worlds, and we knew that we wanted to put not just linear content in there, and flat 16:9 content, we actually wanted 180 and 360 spatial livestreaming, we knew we had to hold ourselves to the same goal that the main guys got held to, i.e., everything had to be getting as close to 1080p or UHD as possible.

And that core objective led us to meet Rob and Tiledmedia, and very quickly, very instantly, we realized these were guys that were actually going on exactly the same path we wanted to go on, and were able to bring in expertise to help us achieve that picture quality that I think is very core to a media viewing platform within virtual reality.

And then off the back of that, we've been able to build additional features, and I'm sure Rob will talk about some when we talk about synchronization with multi-streams coming in, layering co-viewing on top of it, and as we've gone on over the last year or so, out to market, we've been able to push lots of world firsts in terms of the types of content, and sport particularly, that we've been able to put into the application, which is very exciting.

So, when we talk about that, and we'll see from a clip which will illustrate in a few seconds far clearer than I'll be able to articulate, we've actually managed to bring, I think, the netball, The Hundred cricket, and the Extreme E are all world firsts in terms of those formats going out into the application, which has been amazing. Obviously, with the Premier League as well, the football is a major draw. I think, again, being able to bring multi-camera switching into all those sports in different formats has been fantastic. And then on top of that, we were actually able to launch every single one of Sky Sports' linear channels, and again at full HD quality, which is fantastic, and they look really rather nice in the application. And again, with all of this as well mixed into the picture quality, we've been able to stick to industry standard DRM and security protocols as well, which was another layer of communication overhead within the stream traffic that we had to manage, and I think the folks at Tiledmedia within the workflow have shown how that can be done really well.

And then a little bit beyond sports as well. We've been able to bring cinematic content into the app as well, and that's just been a slightly different format and a slightly different way, and we've recently launched our first non-sports channel Sky Arts, which has been quite interesting and quite good as well.

But rather than listen to me talk about it, if I can-- Here we—

Voice-Over

We set up cameras inside and feeds of sound, also pictures as well, and we put those into an overall experience to go into the app.

Vox pop

Welcome to the home of the champions.

Vox pop

Let's do this. Live football, on a Sunday, in VR. Yes, please.

Vox pop

Oh! Good goal.

Vox pop

Yeah!

Vox pop

Behind me, whoa!

Vox pop

It's like being in the stadium. This gives you a one up on everybody else that's watching it on TV.

Vox pop

Getting ready for the Vitality Netball. Here we've got camera control and sound mixing. And over here we've got visual mixing and also replay for inserts and commercials.

Vox pop

I am at the goal end of Strathclyde Sirens. Vinkhumbo's just come through for an interception.

Vox pop

Hi, Ebb's!

Vox pop

Hey, Da!

Voice-Over

Welcome to Extreme E.

Vox pop

Here we go!

Vox pop

This is going to be Molly Taylor winning the first ever Extreme E X Prix. She crosses the line!

Vox pop

Can I get some virtual reality popcorn?

It's a totally immersive movie watching experience. Basically, you get your own personal cinema and the best seat in the house. You're all together watching the movie, in this cinema.

It can be a social experience, even if you're not in the same room.

Vox pop

Oh my god, I've got all the Sky Sports channels in one place. Wow!

Vox pop

Wow, you lucky Sky VIP customers.

Matt McCartney

So, hopefully that gives you guys an indication of what the product is about, and from an end user perspective, it was very much about looking for the features that we can give people. I think we saw in the clip the jumbotron, which worked for football and netball as a single screen, but you saw from the Extreme E clip there, we were actually able to give customers the choice of what livestream they were putting into that jumbotron. So, that interactivity was very important as we were building the product.

There was a lot of synchronization within that. Co-viewing sits very much at the heart of the product, so when we think about how we're serving up multiple streams, how folks switch between those streams, and I think for a football match for example, we could be at any one of four different camera positions with up to four people all co-viewing, so that's a lot of traffic to keep synchronized and no one's missing the action, everyone's seeing the right frame at the right second, because there's nothing worse than, in live sports, someone you're watching it with sees a goal five, ten seconds before you do and it ruins the experience. So, there were definitely technical challenges to overcome as part of this content contribution pathway that we've then built, which is very exciting.

And then obviously moving up towards data and statistics, and more information that we can layer in that all sync to the live sport itself. So, things like the Opta feeds were coming into the football experience, and again, making sure that they're surfaced in line with the media streams has been very important to us.

And to talk through some of that path end-to-end, I'll hand over to Rob.

Rob Koenen

Yes, and very briefly, because I'll hand over to Richard in a second, but will be-- for this experience to come together, obviously, there's an end-to-end chain, which you see here. Production, Contribution, Transcoding, to an Origin server on a CDN, a Distribution, and to an Application.

And also, because this is an Intel Network Builders webinar, I'd just like to highlight a little bit where Intel sits in all of this, and I think the most important part here is the transcoding, which is where-- there's actually two elements. The transcoding happens in the cloud, and I'll speak a bit more about that later, but it relies on massively parallel transcoders, like hundreds of cores per camera, and it also-- and which run on Xeon processors, and it also uses Intel's open source SVT-HEVC encoding, which was specifically adapted for our use, but again, we'll talk about that in a little bit. And then there's also Intel servers in the Sky control room, master control room, where we pull our feeds to our cloud platform.

But more about that later. Let's move to Richard who will talk about the first part which is the production.

Richard Mills

Thanks, Rob. On the production side, we had a period of research and development to ascertain which type of camera and which production system would be the best to employ for this application. We had to start with a high resolution image, good quality lenses, but also have a feeling of immersivity. So, we looked at 360 cameras, the arrays of stitched cameras. We looked at 180 degrees stitched, a fisheye lens as an option, and we also discussed whether the capture should be mono or stereo.

So, for practical purposes, the camera system was the most important aspect of it, and we settled on the Blackmagic URSA Mini 4.6K, coupled with a Canon 8-15mm lens. This gave us excellent resolution, and a combination of the resolution and a wide angle lens. The resolution in the headset is the equivalent of 8k equirectangular projection, and we found this camera matched that extremely well.

Here you've got an overview of what happens on location. You can see some shots here of the experience. Also, where cameras are rigged, we saw a bit of that in a clip in the video, and how production was organized. Basically, it's a miniature remote, or outside broadcast as we call it in the UK, feeding bespoke camera feeds back to Sky's master control room from distribution.

Now here's an idea of the video pathways. We take a camera, it comes in over either hybrid-- SMPTE hybrid fiber cable, or alternatively, fiber-optic. Goes into a processing system where sound is added. What we do is we do sub-mixes for each camera. So, we've got a mixture of overall sound and the announcers, the commentators, but we also add in local effects as well. So, when you switch in the app between the various cameras, you get a variation in the local effects. We also take a feed from the main remote production vehicle, which is the new HD broadcast, and we add that into a jumbotron in one or more of the cameras. These feeds are then encoded. We use a pair of encoders, 75 Mbit/s per stream. That then goes into a switch, which is passed down a broadcast fiber network back to Sky. From Sky, that's passed on to Tiledmedia for the ClearVR process.

There's a lot of experimentation when it came to camera layouts and positioning. Certainly, in the development side of things, we looked at a number of different camera positions. It was good to have a high wide angle which gave you, particularly for soccer, an overall view of the pitch corner to corner so you wouldn't miss any of the action. Bear in mind that we then have a picture-in-picture of the main broadcast. At either end of the pitch, we had cameras behind the goal, sometimes low, sometimes high, and then a fourth camera, which could be near the players and coaches' technical area.

And we had a variety of different types. A lot depends on how big the football ground is. Also, a particular sport. When we were doing netball, the coverage was different. So, sometimes it was advantageous to put a camera on the corner. That way, you can see half of the pitch and the-- or the action in the end zones.

Another alternative for small grounds was to have the main camera again high, but also have high wide cameras on the 18 yard line. This enabled the viewer to switch between cameras as play advances up and down the pitch. In some cases, the technical area's on the opposite side, so a camera was positioned there.

Here is another option again, behind the goal high rather than low.

And now I'll hand over to Rob now for the contribution aspects of things.

Rob Koenen

Thanks very much, Richard. Yes, so we've seen the content being produced, and we see that it needs to go to our ClearVR Cloud. We do all our tiled transcoding, and I'll explain tiled transcoding in a bit. We do that in the cloud. So, the contribution, and Richard already alluded to this, is set at 70 or 75 Mbit/s for the 4K signal, and what's really important is that it has accurate timing information embedded, because this allows us to do not just-- we don't just sync the people on the playback using our client-side technology, but what also happens is we sync all the cameras frame accurate. So, when a user switches from one camera to another, the sound continues seamlessly, and the video also switches frame accurate. So, you don't jump up and back and forth in the timeline when you switch. We use HLS pull, and then as Matt already alluded to, the whole system is secure end-to-end, which means we also use VPN and/or VPN encryption and a firewall to make sure that only we can pull the data from Sky MCR.

So, next, we-- I'll spend a little more time on explaining the transcoding, or at least the basic principles of it. So, that's it's here in the chain. So-- and we are called Tiledmedia for a reason, which is because we do a tiled processing of the video. Now, if I take a typical 360 picture, and the principles are the same for 360 and 180, a typical 360 picture looks a little bit like this. If you make it flat, that's called an ERP or an equirectangular projection, and you only see a very small part of this, what's called the viewport, and typically that viewport port is only one-eighth. So, what we do in our cloud platform is we cut it all up in titles, and then we determine at the client-side, by the way, which tiles are in view, and then the client only retrieves those tiles that are in view. So, this means all these tiles are encoded individually, and they're reassembled at the client-side. And we also send a very low resolution background to make sure if people move their head quickly, they always have a-- they always have some picture, they don't see black, but the typical update time is so fast that you don't see this low resolution background consciously. You will see it, but you will not notice it. High resolution imagery appears almost instantly.

And one of the clever things in our system is that we don't need to discard everything if you move your head. We just discard the red tiles and we reload the green tiles, and all these tiles that are in the middle can just stay in the buffer.

So, that was very quick on how tiled encoding works. We do it actually on what's called a Cubemap, which is a-- it's a better representation than an ERP for doing tiling. It's a little more efficient. It's a bit of a technical detail.

So, to get the details, and there will be a link to the presentation in the attachment to this webinar so you can review it, and this is, again, where Intel comes in. So, we do parallel encoding of all these individual titles using independent encoders, and this is literally hundreds of cores per camera. So, we have several thousands of cores running during a four-game-- four-camera match. And these are all packaged in an MP4 format, and we also have it-- for those that know image coding, we have multiple GOP representations, groups of pictures, which means we can optimize between switching speed and quality. If you move your head, you go to a short GOP, which is slightly lower quality, and then pretty soon you will switch to a longer GOP, which is higher quality and lower bitrate. And then the total bitrate for all the tiles, which is not the bitrate that the user sees, but that's the bitrate that goes from our cloud platform to the CDN, is between 120 and 200 Mbit/s for each camera, and it depends a little bit on whether we do 180 or 360. Obviously, it's a little lower for 180.

So, we need to contribute this to an origin server. In this case, we use Akamai, and in this case, we distributed over a number of Akamai ingests, each taking up to a 45 Mbit/s, which means if we have four cameras, we need 24 ingests. Again, this is all quite a bit of detail. So, I will not spend too long, but do feel free to ask your question if you're interested.

And then the distribution is the last part, how do we get this to the consumer? So, this is the clever bit in that it uses pure HTTP distribution. So, it's very, very scalable. It's also very flexible because the client-- we show this to headsets, but you can also distribute this to mobile devices. The client knows exactly where the user is watching, what their viewport is, what the network condition is. So, they can-- the client has all the logic and the smarts to retrieve the right tiles at the right quality and the right bitrate.

And then, what's also interesting if some tiles are missing, it doesn't matter. It's robust to that. It just... What arrives gets decoded and gets presented to the user, which means it's very error-robust, and one of the clever tricks here is that all these tiles, everything gets rewritten into a single bitstream, and what this means is that every-- you only need one single decoder in the device, the hardware decoder in the headset in the mobile device. So, no matter how many tiles you have, you don't need multiple decoders. It can-- so basically runs on any device that has an HEVC decoder, and this whole principle was adopted in an MPEG standard called OMAF, for Omnidirectional MediA Format.

The last part here is that the use of a bitrate then doesn't-- is not 120 megabit, or the 200 megabit that I mentioned earlier, but only between-- somewhere between 16 and 50 Mbit/s, which depends a little bit on how often users switch their camera or how often they move their head. And again, for people that are interested in the technical details, the tile switching latency, which we also call motion to high resolution, basically I move my head, how long does it take for a high resolution tile to become visible, is somewhere below 100 milliseconds on what I would call normal network conditions, and this means it's basically invisible to a user. That's below what you can observe. Camera switching latency is a little higher, but the image, if you use your-- if you switch your camera, say, the high camera will continue until we have the feed acquired for the <<over the goal camera and then it switches seamlessly, so you never miss any of the action, which was a very important consideration for Sky. Users should never miss any of the action, and the end-to-end latency is about 40 seconds, which means end-to-end or glass-to-glass or camera-to-headset is about 40 seconds.

So, as already mentioned, by Matt as well, it's end-to-end secure, which means the delivery is secure. We encrypt all the tiles when they're put on the CDN and it uses a regular DRM system. It's Widevine in this case, which is fully integrated with the player, so, which is why we can do high quality, high value content in this app without any problems.

And lastly, there's a... There's an application, obviously, which was built by Cosm and Sky together. The application is Unity-based. It's for Quest 1 and the Quest 2, and it has everything that Matt already mentioned, and you see a little bit of the user interface here in the screen. It takes care of the application, takes care of all the syncing, it takes care of keeping all the people in sync, it takes care of joining people in rooms. You can make a room, you can leave a room, you can chat with your friends, and it also takes care of keeping all the feeds in sync, all the cameras in sync.

And with that, I would like to hand back to Matt, because there's more to come.

Matt McCartney

Thanks, Rob. Yes, I mean, I think if we look at what we've just done, and what Rob's just talked through, it's amazing, and it's really great, but it's also just the beginning. So, certainly, from a Sky perspective, one thing we're keen to do is just to begin unlocking more sport, more live content, and actually stretching beyond sports. We pushed the boundaries over the last 12 months when we picked up things like the Vitality Netball and Extreme E, things that have never been done before in this format, and in a second, I'll just talk specifically about the cricket because we just had a crazy four weeks with The Hundred in the UK, and in Sky Worlds, and through this pathway we've just talked through, we've been able to serve up a sort of experience for that sport that, again, had never been done before, which is fantastic. And when we look at the linear, we've got 12 sports channels, which is great, but I think Sky has considerably more channels than 12. So, really, what we want to start looking at is scale, and how we can start putting more content, more live events, more linear events through this pathway at a robust scale, which is great. And the other thing that I think sort of got alluded to in the video a little bit is movies. We're looking to up our movie schedules and bring in more long form studio content to the application, which is great.

I think specifically, the one event that I think is going to lay a little bit of a template for spatial sport consumption has been what we've done for the cricket. What you can see from the image here, and there's a clip in a few seconds that shows it quite nicely, is we had a large wide angle livestream shot as your foreground. It wasn't using an immersive camera, as Richard talked about before. It was a standard camera with a 4:3 wide angle lens on the end of it. But for cricket, because it's set back from the pitch, that gave us a wonderful panorama of the stadium, and then over the top of that we were able to overlay picture-in-picture, the jumbotron, and for the first time, as you can see from the bottom image there, there was four feeds coming through, and viewers had a chance to basically select which one of those mini feeds was in the main screen, and really starting to take a directorial control over how they watch the event, and what they're seeing, and all of that was stitched together with the broadcast audio. So, you still got the punditry and the commentary over the top of it, no matter what feed you were watching.

I think that worked really well for the cricket, and I think it started to show a template certainly for what I refer to as larger geography sports. So, when you think about golf, or Formula One, or sailing, or even marathon running and cycling events that cover large amounts of ground, you're not as-- it's not as practical to put multiple cameras in the way that you could for the netball or the way that we did for the Premier League. So, actually just making use of existing feeds that are coming back into certainly Sky's control room, and being able to repurpose them and serve them up in a way that hasn't previously been done before, is, A, quite innovative, but it actually has the added benefit of fitting within the current broadcast chain, which means we're able to serve up these immersive experiences without actually adding any production overhead to the guys on the ground that are rigging cameras and setting this up to start with. I think that's quite important because that's what's going to allow us to go at scale over the next couple of months and into next year, as we look to bring the next season of Premier League into the app.

And if I show it again, you folks should be able to see—

[Clip playing]

And that's it. That was our cricket, which is fantastic, and I think it brings us to the end of our slideshow. So, if anybody has any questions that they'd like to ask, fire away, please.

Lilian Veras

Fantastic. Well, thank you all for such a great presentation. We do have a few questions that have come in while you were presenting. So, I'll start off with the first one. First question we have here is for Matt, how do your customers like this?

Matt McCartney

Speaking for them broadly, I think quite well. The product was well received. Again, when we actually set out to build it, and I talked at the beginning about picture quality being very key to it, the other driving thing that we thought about is why would people go to use this

app. I mean, obviously, I think the best way to watch live sports is still in a stadium. You're never going to escape the thrill of being there, surrounded by people and shouting and cheering, and I think that's fantastic. The next best thing is to use our app, because we've been able to replicate that and being able to bring that experience into the home.

Rather unfortunately, the last couple of years has just not gone well for the world, and when we launched Sky Worlds in October 2020, it was at the height of the UK lockdown and everything that was going on, which was rather unfortunate, but by launching the product when we did, we were actually able to allow folks that couldn't, at that time, go anywhere near stadiums, let alone do live sport with friends and family, to actually start to do that, and I think in the early days, we got a lot of positive feedback about that being something that we were able to release at that time, and I think folks needed it.

I would say on the other side, it's an innovation product, and I think some of the application journey is still a bit funny. We have some fun and games with login from time to time because we have to authenticate, and obviously, as Rob said, it's premium content behind paywall, and I think those journeys through virtual reality aren't always great, but the in-app experiences has been pretty well received, which is great.

Rob Koenen

Are you still there, Lillian?

Lilian Veras

I'm here. Sorry, my microphone was on mute. Thank you, Matt, for the answer. There is another question here from an attendee who's saying he's a Sky customer and he has tried and loved Sky Worlds, but he doesn't ever watch a whole 90 minutes VR game. So, on average, what is the average in-headset session duration of the viewer?

Matt McCartney

The honest answer is it varies sport to sport. Forgive me if I don’t go too detailed on some of the stats, but I think for football, I think our average is around about 35 to 40 minutes per session. We do see around 20, 25% of folks breaking the hour mark, which I think is very good for a headset. I won't sit here and say that the Oculus Quest is the most comfortable thing to spend hours and hours and hours, and we all know it’s a headset and it gets quite heavy and it does get quite warm, but I think generally, we see, certainly as an average session, above 30 minutes. I think for the cricket recently, again, we've started getting close to that 45, 50 minutes as well at a time, which is great when you think about what, you know, having those things stuck on your head is kind of like. So, we think we get quite good figures in terms of average per session.

Lilian Veras

That's great. I have another question here, this time for Richard. Are you doing any remote production and has that been helpful?

Matt McCartney

You're on mute, Richard.

Richard Mills

The system was designed from the outset to encompass production. Reason being is that the equipment at the front end is a fairly small part of the operation and if you can reduce the crew footprint on location, that does certainly help with the overall logistics of getting people around from ground to ground. Remote production, however, has suddenly taken a huge leap, and most, or pretty much all, of Sky Sports are now produced as remote productions. So, cameras are encoded on location, sent back over fiber networks back to a control room at Sky where the show is produced, and this fact that all the cameras coming back has actually been very beneficial to the app. It means that we have at-- back at Sky, we have all the cameras for cricket, Premier League, from all the sports available to us to actually

put into the VR application. So, that's given us a huge benefit, but our particular operation was designed for remote production from the get-go.

Lilian Veras

Fantastic. Thank you, Richard. I have a question here for Rob now. An attendee is asking if the jumbotron and other camera feeds baked into the VR feed, or if they are streamed separately and added at client end?

Rob Koenen

Yes, that's a really good question. What you saw for the football, it was baked in, but we've known, and Sky has known, from, well, I would say well over a year ago that we wanted to change this at some point. We wanted to make this more flexible, because if it's not baked in, we can more easily put it over any camera, we don't have to bake it into four camera signals. We can let the user make it appear and disappear. We can make them move it around and make it smaller or larger. So, the user preference becomes an option.

What you've seen for the cricket, it actually is not baked in anymore. This is actually truly streamed separately. Still synchronized with the main background feed, but it's streamed as separate feeds and which allows us a lot of flexibility, and you saw the instant camera switches from a-- you click on a thumbnail and it becomes the large jumbotron. This is something we could never do with baked-in content, obviously. So, we're moving towards more flexibility from having something, because I think we realized that having a jumbo was really useful for a user, but we wanted to make it more flexible, and it's now becoming more flexible.

Lilian Veras

That was superb. Thank you so much, guys, for this presentation. That was our last question. I want to ask our live audience to, please, do not forget to give our team a rating for the live recording. Richard, Matt, and Rob, thank you so much. I'll see you next time.

Matt McCartney

Fantastic, thank you.

Rob Koenen

Thanks very much, Lilian. Thanks to the audience.

Richard Mills

Thank you very much, indeed. Bye all.

Lilian Veras

Thank you all and this concludes our webcast today.