Strategies for Deploying High Resolution and High Framerate Cable Content Leveraging Visual Systems Optimizations

Yasser F. Syed PhD, Dist. Eng/Applied Research
& Dan Holden, Fellow, Comcast Labs

Abstract

This paper examines how and why to deliver higher resolution and framerate content in an HFC system, especially focusing on 4K Video delivery with an advanced audio experience. It examines how to deploy this content in a bandwidth constrained environment and concentrates on improvements to the viewer's quality of experience through video compression technologies and leveraging potential video compression gains through sensitivities in the human visual system.

INTRODUCTION

The launch of higher resolution video with greater frame rates will allow MSOs to develop new business opportunities, and provide a competitive advantage against new entrants in the video marketplace. In this paper we will examine the road to better delivered video quality, especially how to leverage the existing HFC infrastructure to deliver 4k video with an advanced audio experience. The paper will concentrate on video compression technologies and the potential for leveraging the human visual system model to provide 4K video in a bandwidth constrained environment. For deployment, we will look at required upgrades to the HFC infrastructure, and what engineering requirements are needed for 4K delivery. New technologies and approaches to reduce costs will also be examined, as well as how the complexity of high-resolution video changes delivery methodology.

4k television technology was introduced at the Consumer Electronics Show in 2012. It is based on a display that has approximately 4000 pixels in the horizontal resolution. 4k differs from previous television standards (480i, 480p, 720p, and 1080P/I) in which the vertical pixel count was annotated. In a 4k display the horizontal resolution is maintained around 4000 pixels, and the vertical resolution is allowed to vary as a function of source content. This technique was adopted to allow support for various aspect ratios and letterboxing. Figure 1 shows the scale of 4K content compared to the resolutions that are supported today.

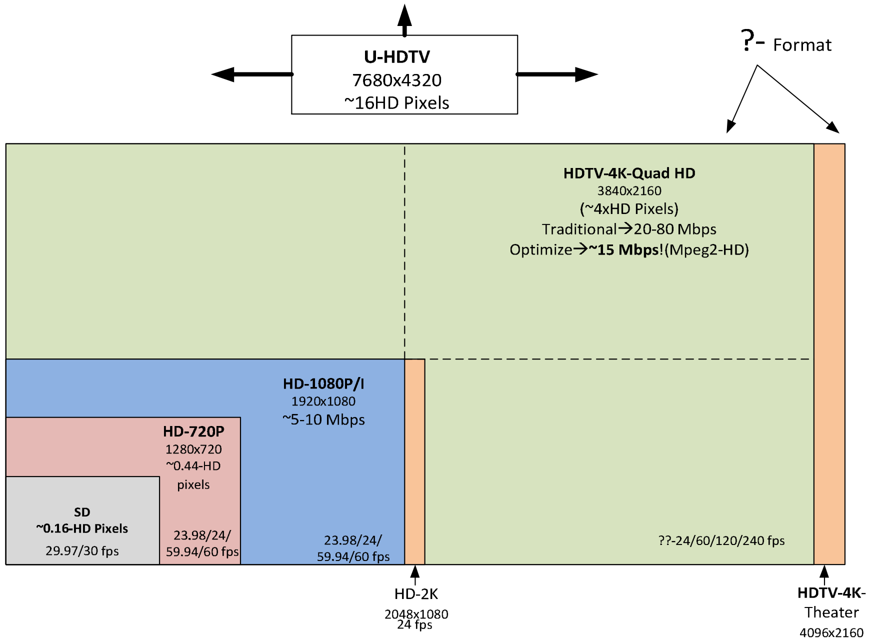

Figure 1 Comparison of 4K to Different Video Resolutions

BUSINESS OPPORTUNITY

One of the most compelling cases for higher quality video is to gain a competitive advantage in the video marketplace. 4k will require “big pipes” at a time when there is clear movement on the part of industry competitors to adopt a mobile strategy utilizing technology that will be limited by available spectrum. Newer video compression techniques will certainly reduce the size requirement for the pipes, while the demands of newer display technologies, (8k and 256 fps) will tax any future video distribution system.

Cable has a reputation for being the leader in delivering an exceptional video experience. First generation 4k delivery platforms will need a vast amount of bandwidth, which is most likely to require a full QAM in order to deliver a quality experience. If we look at historical data, early generation H.264/AVC video was around 9 Mbps for High Definition (HD) 1080i video. A few short years later, we have been able to reduce this bandwidth to 4.3 Mbps.

Until display technology retail prices drop to a reasonable level, it is expected early adopters for 4K televisions will be bars, restaurants, and high-end home theaters. Here are the key assumptions:

- Mass deployment of 4k televisions will not take place until the cost per unit is less than $3,000 per unit
- The introduction of 4k will follow the same general path as the

introduction of High Definition video, which has currently penetrated more than 70% of US households
- Adoption of 4k video will be slower than HD video
- Volume of 4k encoded VOD assets will grow exponentially over the next three years
- Studio post-production already supports a 4K workflow which can be extended to downstream VOD content delivery
- Additional revenue will be generated when customers select to watch assets in a 4k format
- 8k video will not be introduced until at least 2016
- MSOs will not simulcast 1080p60, but may select to offer this format in VOD
- Bars, restaurants, and elite home theaters offer a significant up-sale opportunity

MSOs should take the lead on the introduction of high resolution video delivery. Rather than focus solely on video, it is suggested by the paper authors that the entire sensory experience be enhanced, which includes the addition of 3D audio channels. Background noise in a bar can be very distracting, and providing a high quality audio experience will set our video offering apart from the competition. The adoption of 4k video with 3D audio will most likely not progress at the same pace as HD. HD had the added benefit of changing the format to 16:9 from 4:3, and the elimination of large cathode ray tubes, which drastically reduced the size of the television footprint in the living room. The adoption of HD televisions has been relatively quick historically, whereas, the migration to the distribution of higher resolution video has yet to be established.

ADOPTION WILL BE DIFFERENT FROM 3DTV

There have been many attempts to categorize 4k video to the 3D television experience. This type of comparison is probably not the correct model, as 4k will not suffer from the infamous 3D glasses gaffe. Additionally, massive libraries currently exist at studios that can be easily scanned or transcoded into higher resolution video for distribution. We believe comparing 4k adoption to 3D would be a mistake, since 4K will most likely follow the adoption and general operational patterns developed for HD.

The first linear 4k channel will most likely be an occasional feed that is brought up when a live 4k event is aired. Under this model, 4 HD channels would need to be taken down in order to broadcast a single 4k event. With a few enhancements to the backoffice systems, it should be possible to sell access to a 4k stream on a pay-per-view fashion. The broadcast of huge events, like the Olympics or Super Bowl, could lead to enormous up-sale opportunities.

4k VOD will most likely be the first place where we see significant inroads of high resolution video. Encoders have already been developed that can process 4k video, and it is believed VOD pumps will not have issues with the larger file sizes or MPEG-2 transport stream wrappers. Adaptive streaming technologies should also be suitable for 4k VOD distribution. The video encoding process for QAM and adaptive streaming can be identical. Fragmentors should not require modifications, unless they are "just in time," which may suffer from data transfer rates and latency. The largest gap in the distribution system will be the ability to handle 3D audio, and finding a suitable video player.

ALTERNATIVES

There are many alternatives to 4k video, including Quad HD and 8k. While Quad HD has slightly less resolution than 4k, 8k has twice the resolution and twice the bandwidth requirements. Should Quad HD TVs be introduced into the marketplace, it would be preferred if they have the capability to ingest true 4k content, as MSOs would not want to simulcast both Quad HD and 4k streams. For VOD delivery, it would be possible to support both formats, but as the VOD library grows the added storage expense would prove challenging.

Rather than rolling out 4k, another possibility is to move forward with 1080p60. Encoders and STBs were released in 2012 to support this format. Formal analysis of 1080p60 video quality is beyond the scope of this paper.

Current compression technology will most likely prevent the delivery of 8k content over a QAM, but 8k delivery could conceivably be done utilizing the CMTS and IP delivery methodologies. Both products deliver the same benefits as 4k, higher video quality.

BENEFITS OF 4K

The first implication of moving to 4k video is the size of streams and files will be massive. A single mezzanine, linear stream from a live event may reach up to 500 MBps and a stream sent to a set top box could be on the order of 38 MBps. This implies four high definition channels would need to be taken down in order to place one 4k signal on the plant (Figure 2).

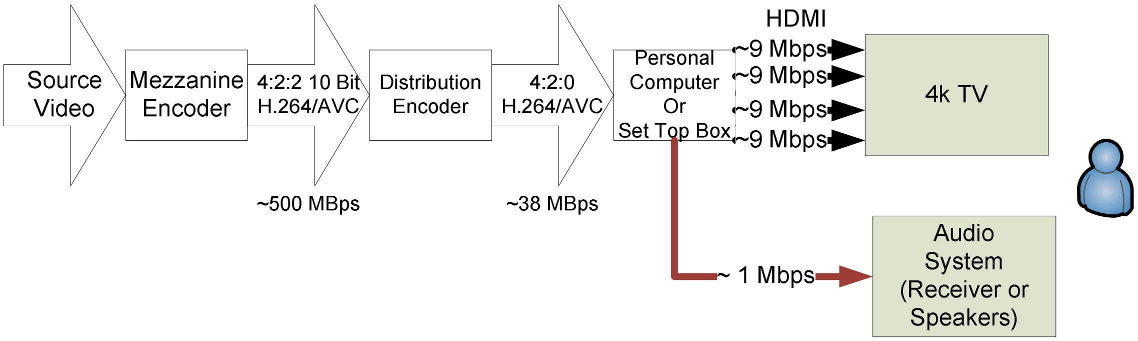

Figure 2 Delivery of 4K content from Ingestion to Consumer

Higher resolution video will allow MSOs to compete with both BluRay and local movie theaters. Many movie theaters currently delivery digital projects in 2k resolutions with a maximum audio experience of 11.1. It is theatrically possible to delivery 4k video with a 22.2 audio experience across an existing QAM to a personal computer (PC) which will replace the current functionality provided by a set top box (Table 1).

Thus, a completely optimized and compressed 4k/HEVC asset should be around 19 Mbps. When we compress this asset utilizing HEVC, we expect to gain around a fifty percent reduction in bandwidth, putting our 4k asset at approximately 10 Mbps. Next, consider that 50% of that potential gain is taken back 50% due to inefficiencies in first generation encoding technologies, frame-rate allocations, and make allowances for content types, then our 4k/HEVC asset can be distributed in the same band width as an HD asset compressed with MPEG-2 (~15Mbps).

Phase	Video Type	CODEC	Bandwidth (MBPS)	Notes
Initial	1080i	MPEG-2	19.3	19.3 was part of a specification. The first generation HD at some MSOs was set to 18 MBPS.
Today	1080i	MPEG-2	9.7	With 4:1 statistical multiplexing, it is possible to send 4 HD streams down a single QAM
Today	1080i	H.264	4.3	Average bit rate for H.264/AVC streams
Initial	4k	H.264	38	Target bit rate for lab trials
Production 4k[1]	4k/60 fps	HEVC	15	Target bit rate for 4k/60 with 22.2 audio

Table 1 Projected and Historic Bandwidth Consumption

[1] Note the projected bandwidth for a production 4k asset. The basis for the projection is calculated as follows:4k video is slightly larger than four HD signals: 4 * 4.3 ~ 18 Mbps in H.264/AVC Add additional audio bandwidth of approximately 1 Mbps for a total of 19 Mbps in HEVC

AUDIO

In addition to an enhanced video experience, the opportunity exists to upgrade the viewer’s audio experience. Cable MSOs understand that audio can enhance or detract from the video quality of experience.

Many new audio technologies are under development that will put additional audio channels into the home. Old content can be remixed to support new formats, and additional microphones can be utilized to capture a true “3D” audio experience.

In the short term, consumers will need to add additional speakers to gain the improved audio benefit; and in the near future we will see sound bars that will reduce the complexity and cost of delivering this technology into the home theater and entertainment based businesses.

A typical 22.2 audio experience would require almost 1.5 Mbps when utilizing 24 channels at 48 kbps with constant bit rate (CBR) encoding. By switching this to capped Variable bit rate (cVBR) encoding, a substantial reduction in audio bandwidth utilization will be realized. Additionally, new sound bar technologies will reduce the cost, complexity, and number of speakers required to bring a true 3D audio experience to the customer.

As part of the distribution process, care must be taken to monitor every channel and to ensure multichannel audio is down-converted to basic stereo for playback though the television speakers. While it is assumed 4k content will be viewed with enhanced audio, consumers may select to view the content while utilizing the stereo audio capabilities of the display.

COSTS

The costs to enhance the end-to-end solution for 4k can be broken into their representative components. Here is a partial list of items that may require upgrades.

Encoders – Existing VOD encoders have the ability to deliver 4k video with few modifications, while linear encoders will need to be developed that can handle massive amounts of data in very short periods. Additional modifications will be needed to handle advanced audio technologies such as Dolby Adaptive Audio, SRS Multi-Dimensional Audio, and 22.2 specifications. There will need to be a clear roadmap to get from initial 4k video with H.264/AVC encoding to HEVC. For the initial launch, a single linear 4k encoder should suffice. It will allow a MSO the ability to replace four HD streams with a single 4k stream. For VOD, it is possible to scale the number of encoders to match the size and refresh rate of the library to be converted.
SRM – A next generation SRM will need to be deployed in order to allow the VOD pump to select a 4k asset.
Metadata – New fields will need to be included to indicate the asset is 4k.
Content Encoding Profile – New profiles for 4k encoding will need to be defined.
Storage – 4 times the storage per asset, as compared to HD.
Video Player – Support for new Video and Audio formats.

Adaptive Dynamic Streaming – Support for additional audio CODECs or video CODECs.
Backoffice – Enhancements for billing.
Set Top Box – Faster single or multi-core CPUs and bigger pipes.
Direct Fiber – Larger pipes for mezzanine sources.
Mixing new audio – New mixing technologies for audio.
Trucks, Cameras, Post – Enhancements to editing systems, graphics, and source acquisition equipment for live capture content.

SERVICE AND INFRASTRUCTURE VIEWS

It has already been demonstrated in the laboratory that 4K video encoding for VOD can be accomplished on existing encoders. A single 4k transport stream is generated and sent across the plant for decoding on a Personal Computer (PC). This stream is then split into four separate streams for delivery to the display across four separate HDMI cables. Once the HDMI interface is upgraded, it is expected a single stream and HDMI cable will be attached to the television.

Linear encoding could be done by handling the encode as 4 separate HD processes that need to be synchronized (hence QuadHD) and distributed as a single stream on the wire. It is important to note this implies that within the video encoding process, the input stream could be split for processing and then combined into a single stream for transport.

While this approach may be viable, newer, multi-core CPUs will most likely be able to handle the entire encode as a single transport stream. A single stream approach across the entire plant will increase operational efficiencies and simplify the operational model. In the case of adaptive streaming, fragmenting a single transport stream would require the identification of a single boundary point in the source video.

The same intuitive logic applies to network DVR. As previously stated, utilizing the same encoding techniques for linear and VOD is optimal due to simplicity and overall operational models for distribution of 4k video (figure 3).

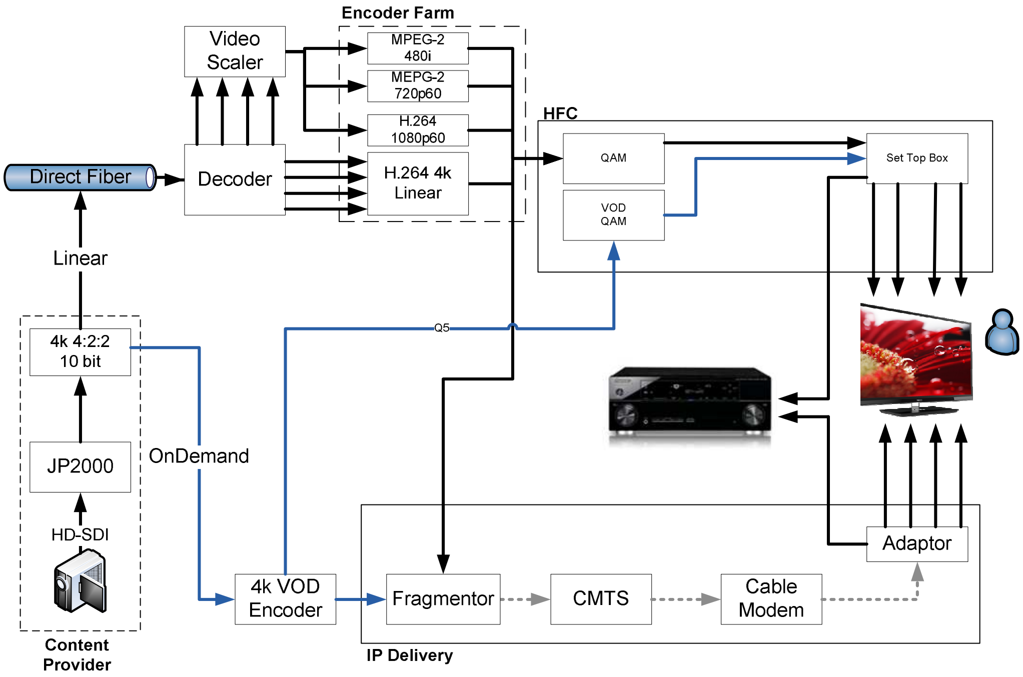

Figure 3 Operational Model for 4K Distribution Video

HD encoding of 1080I30/1080P24 using newer encoding techniques could range from 5-10Mbps when compressed with AVC/H.264. Offline VOD compression will most likely be superior due to multi-pass encoding. If 4K is supported at the same frame-rate, this could imply an encode bit rate from 20-40 Mbps in a cumulative data sense. This does not assume further compression efficiencies due to increased pixel density.

Can the infrastructure support a 40 Mbps 4K stream? A single 40 Mbps 4K channel would:

- Require the same bandwidth as 4-8 HD channels,
- Not fit into a 38.8 Mbps QAM
- And would likely not be carried by an ISP over the public internet

The bandwidth infrastructure modifications for this approach would be cost prohibitive. One bound stream could possibly be fit into a single QAM with bandwidth of 38.8 Mbps, which would replace about 2^{+} MPEG-2 HD channels (or 4 HD streams on a 4:1 Mux). To meet a 4K service for HD, each QAM would need two 4K channels. This would mean each 4K channel would need to be bounded under 19 Mbps which would be about 1 HD channels and 1 SD channel.

Is it possible to move from a 1:4 upper bound bit processing ratio to a 1:1.3 ratio? With new coding tools from MPEG such as HEVC, a 50% improvement in compression can be expected. Additionally, having greater pixel density should create some further compression efficiencies to decrease the 1:4 ratio. Even more efficiencies can be gained by the way of improvements to perceptual modeling of our visual system and applying this to coding.

There is room to create more compression efficiencies, especially since encoder design is evolving and new compression tools are becoming granular. And even if a greater frame-rate is needed, pixel processing burden would be less than expected due to increased efficiency in motion vector accuracy and longer GOP length for the same amount of time.

As we examine all of the factors of better compression, filters and modulations, it does become possible to create a 4K stream that should ultimately approach 15 Mbps in the near future.

The next part of this paper will look at potential places to leverage the human visual system model to increase compression efficiencies through perceptual coding.

PERCEPTUAL CODING AND THE HVS MODEL

HVS (Human Visual Systems) attempts to describe how we actually see [from the photoreceptors in our eyes into the visual cortex and other parts of the brain]. Perceptual video coding is used in “lossy” compression at a target bitrate to mask, transform/quantize, or conceal information that is not seen by our visual systems (psycho-visual redundancies) or is optimized to improve what we can see. This is not coding efficiencies due to manipulation of the bit-stream to improve bit/symbol rate of the stream. It attempts to narrow the total information rate to what is just needed for our visual systems.

Our eyes are made up of 127 million photoreceptors in the retina (120 million rods and 7 million cones) that feed a million neurons in the optic nerve that is connected to the brain [Figure 4]. That already represents about 127:1 convergence of information. The rest of the eye is there to focus, shape, and control the amount of light going into the retina. The rods are used for vision at very low light levels (scoptic) and do not contribute very much to color perception. However, the cones deal with vision at higher light levels (photopic) and with resolving fine spatial details and color. These cones are divided into three types (S-short, M-medium, and L-long) that are sensitive to different wavelengths of light and they are the basis for our ability to match any color through a combination of three primary colors (trichromacy) [Figure 5]. The cones are concentrated in a central part of the retina called the fovea which provides the majority of information traveling along the optic nerve. The fovea matches to what we perceive as “the center of focus” for our vision.

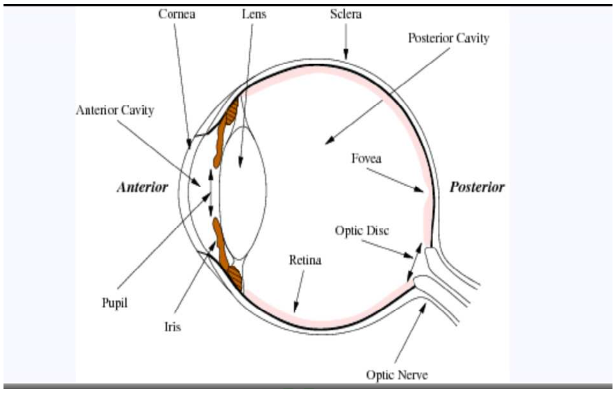

Figure 4 Eye

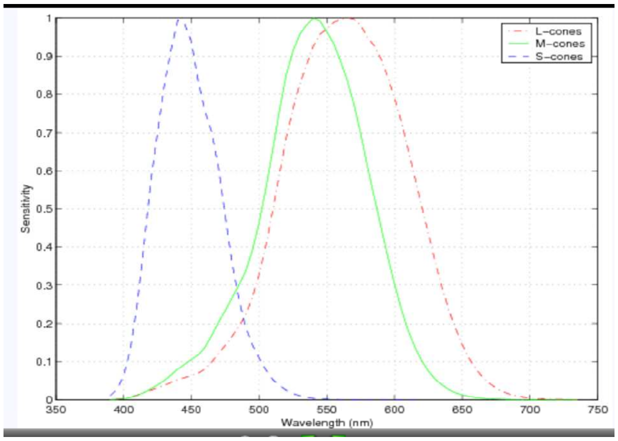

Figure 5 Different Cone Type Wavelength Sensitivity

This information electrically stimulates the optic nerve which feeds into the visual cortex of the brain for semantic and feature processing based upon differences to a windowed-steady state visual model. Eye movement, both right tied with left (saccades), is based on spatiotemporal sensitivities to capture these differences to the brain. From what we see in the human visual system, the visual cortex in the brain does not try to process all information but just what is needed to provide a semantic visual model. Perceptual coding attempts to move past the photoreceptor stage to keeping just the information that will make it into the visual cortex.

So, in trying to model HVS, it can be split up into three areas: 1) a visual attention model, 2) spatiotemporal visual sensitivity model, and 3) a visual masking model. This is basically what is interesting to see, what we can make out of it, and what we could never see at all. Our visual system is sensitive in a number of ways:

- **Contrast**- we aren't sensitive to a level of brightness, we are sensitive to differences in brightness between areas in our vision. This equates to sensitivity to edges in an image and can be affected by the brightness in the background.

- **Spatial Frequency**- as spatial frequency increases, we become less sensitive to variances in spatial details (when does edges become texture?). This can equate to tolerance in coding artifacts in high texture areas as opposed to more constant areas. In color we are even less sensitive to variances in spatial frequencies. Hence one of the reason we can sample color difference less frequently than luminosity (4:2:2 or 4:2:0).

- **Visual Acuity**- This is the ability for the eye to resolve details. One can have reduced visual acuity in fast moving objects (though eye tracking can reduce perceived motion of the object --- reduced retinal velocity). One can also reduce visual acuity by moving further from the object or screen. For ideal viewing, Viewer should be far enough away to not be able to discern pixels on the screen. Increased resolutions can allow for the observer to sit closer to the screen without being able to discern pixels [Figure 6].

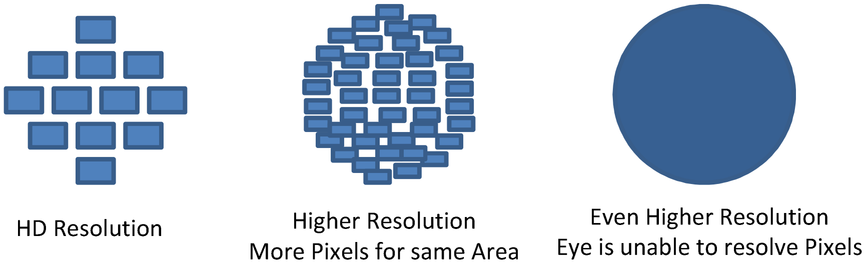

Figure 6 Visual Acuity and Denser Pixels

- **Noise**- These are unnatural changes in contrast due to the image capturing process. This could be due to the scatter on photo sensors in the CCD/ CMOS, heat on electronics carrying the pixel values, or celluloid processing leaving film grain artifacts [Figure 7]. The eye is sensitive to noise at different spatial frequencies which is why low-pass/

band-pass filtering is used as a preprocessing technique to remove these unnatural artifacts.

- **Temporal Frequency**- we are more sensitive to temporal cues rather than lack of spatial details. This is one of the reasons why interlacing can happen because it is a tradeoff of spatial frequency for temporal frequency to address bandwidth issues. It is believed below 50-60 Hz (fps), flicker can be perceived in a series of played out still frames. For this reason, 24fps material sometimes is flashed twice in frame playout on display devices and now material traditionally being shot at 24fps is being shot at 60 fps or even 120 fps for this reason. Additionally, movement that follows natural movement speed and direction is less surprising than erratic movement and speed.

- **Perceptual Uniformity**- This basically means keeping a consistent quality across a video sequence. We are sensitive to quality changes in spatial details of a moving object when viewed in the fovea area of the eye.

To mimic HVS, the attention model needs to identify areas of the image that are tracked by eye movement (saccade) to keep interesting areas in the fovea. Things outside of the fovea do not have to retain as much detail due to change blindness. Object size, and movement (predictable and unpredictable) can be used as cues to identify areas in the video sequence that need more spatial detail. Artifacts can cause a miscue in the eye to areas in the video sequence that are not natural areas of interest and need to be minimized where possible. The spatial temporal model can affect how to maintain a natural sequence with consistent quality over a content scene. Visual masking is a preprocessing function that can hide information in areas that don’t need as much spatial detail such that it is coded in a fewer number of bits.

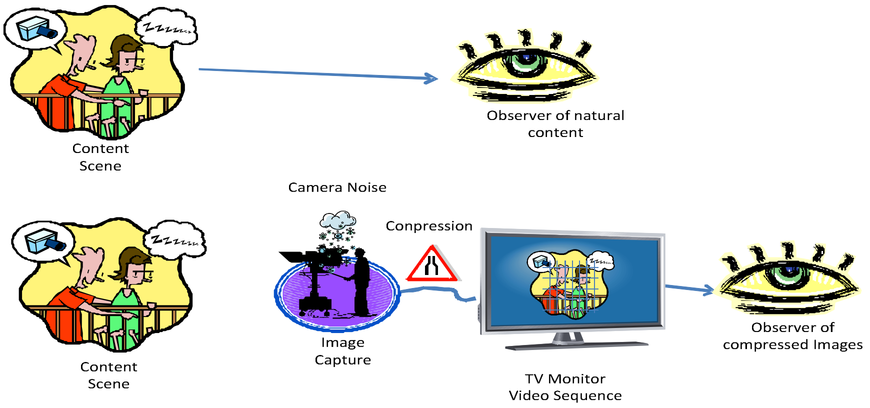

Figure 7 Capturing Natural Content on Screen

EARLY PERCEPTUAL CODING TECHNIQUES IN COMPRESSION

When we directly see a natural scene, our eyes have a filter (mentioned in the sections above) that reduces the amount of information that reaches the visual cortex. We use our eye muscles, focus and movements to change what the cones in the fovea are seeing such that attention is there for important information in the scene.

To capture the image such that we can recreate what we see (Film/ TV without compression), we represent the scene through a series of still pictures being played at a specific temporal frequency (24 fps (2x)/ 30 fps (60 fields)/60 fps). Consistent quality is maintained between each frame, and interlacing techniques are used for further reducing bandwidth using a tradeoff of spatial resolution and temporal frequency.

However, in the capturing of the image, noise is introduced into the content scene through CCD/CMOS camera devices. To avoid seeing the pixels instead of the content scene, we sit back far enough (2H-4H) such that our visual acuity cannot discern a pixel and blends them together.

With the evolution of an analog medium (6MHz analog program) to a digital medium (10 Channels in 6MHz), we now have the ability to manipulate each pixel value and only send difference information between each frame (i.e. compression). In terms of pre-processing, the noise is being removed through low-pass, band-pass, and temporal filters like MCTF. The encoder then uses block-based transforms to change the coefficient values to be measures of spatial frequency energy.

At this point, the coefficients of higher energy frequencies can be quantized with less precision and use less bits because we have less sensitivity at high spatial frequencies. Additionally, this helps with reducing data redundancies in the bit streams since many of these coefficients are quantized to zero.

In terms of motion, movement of natural objects can only move at certain speeds and are predictable which factors in to some of the coding algorithms that reduce computational complexity. This allows for a reduction of motion search space, and a reduction of number of motion vectors based on size of the object. The “errored” differences between frames can also be quantized in the same manner since errors are mostly in high spatial frequency details. In post processing, the blocking artifacts along transform boundaries can then be removed from the image.

To avoid seeing artifacts from the medium (pixels) rather than the content scene, it is important to be able to view the display screen at the proper viewing distance. If one moves in closer, visual acuity increases to the point where pixels can be discerned (visual acuity is inversely proportional to distance). In terms of monitors, we are getting larger monitors going from 40” to 50” to now 60-70” sets, and the viewing distance from these monitors is remaining mostly constant.

Additionally, we are also getting display devices like tablets and PCs that are being viewed at much closer distances than the 2H-4H recommendations.

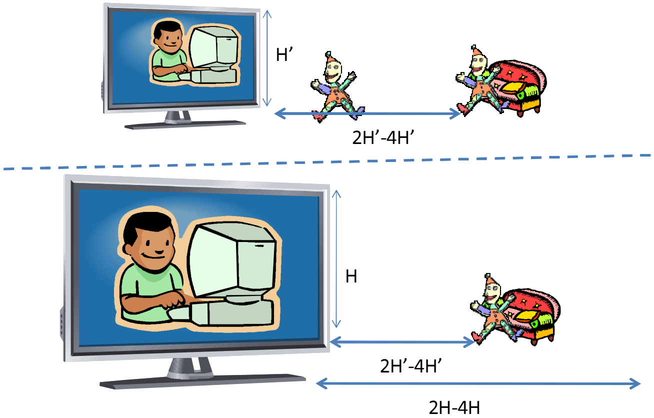

Figure 8 Screen Sizes, Distance, and Visual Acuity in Monitors and PC/Tablets

AFFECTS OF 4K AND HIGHER FRAME-RATES

Going to 4K can create more natural content scenes. Increasing pixel density does not have to create a larger picture; it creates a more densely sampled picture. Each pixel now represents a smaller area which allows for:

- Sharper Edges
  - ✓ Fonts on letters are sharper. The viewer can read documents. [It’s “Resolution-ary”].
  - ✓ Less aliasing artifacts and “jaggies” around edges
  - ✓ Textures are more detailed
- Increased pixel density
  - ✓ Approaches visual acuity limits. See less pixel definition and more of the picture at closer viewing distances and angles.
  - ✓ The Viewing distance becomes more flexible. We can get closer to pictures in both large and small displays (This aligns better with the attention model)
- Better contrast
  - ✓ Pictures look brighter/ more natural due to contrast differences and more gradient increases and decreases (This was always an issue for compression)
  - ✓ Neighboring pixels are more correlated since they represent a smaller area

Going to higher frame-rates can create more natural content scenes cues, by sampling motion in content scene to make it more linearly predictive. This is becoming more helpful as CGI (computer-generated imagery) effects in film and video content introduce faster moving objects in sequences. It is also very helpful in sports content where motion is quick and erratic. If the frame rate is too slow for the motion in the content scene, we can get "juddering" artifacts especially if the picture is flashed multiple times to simulate higher frame-rates:

- Smoother Motion
  - ✓ Movement between frames is shorter and can be predicted better
  - ✓ "juddering" can be reduced due to more sampling of motion and less repeated flashing of the picture
- Less Noise from Image Capturing devices
  - ✓ Noise is not temporally correlated and can be filtered through comparisons of sequential frames.

LOOKING AT CODING WITH RESPECT TO HIGHER RESOLUTION AND COMPRESSION

With increases in resolutions, there are going to be more pixels to process. The encoder picks a target bitrate and then tries to make decisions in coding based upon that. Generally, the encoder attempts to conduct:

1) ***Pre-filtering*****:** remove noise and apply a low-pass filter to remove information and details that would never be resolved at that bit rate anyways. Basically, to remove the information that makes the encoder work harder than it needs to be working.
2) ***Transform/Quantization*****:** change the information order of the data stream to make it more compact and quantize high spatial frequency information. Apply entropy coding to the output of this stream
3) ***Predict Subsequent Frames*****:** Use a reference frame(s) to produce a set of motion vectors and "errored" difference frames (P& B Frames). Calculation of motion vectors need to go through a motion vector search which can be a complicated encoding process.
4) ***Post processing*****:** Conceal artifacts created by the encoding process such a blocking and boundary artifacts through post filtering approaches

Places where we can improve this process, due to having higher resolutions and frame-rates, include:

1) ***Pre-filtering:*** Removing noise may be easier because it is approaching the granularity of our visual acuity while natural content scenes would not have this level of granularity. Using the stronger correlation between neighboring pixels, there can be improved techniques for filtering and dithering to handle noise. Additionally with the improvements in CMOS, we may be able to do this earlier at the point of image capture.
2) **Transform/Quantization**: The transform represents a smaller area and more correlation between the pixels which can help in energy

compaction. Some savings can be achieved as well because quantization levels can be changed for a smaller area. However, there are more transform blocks to deal with at higher resolutions.

3) **Predict Subsequent Frames**: With higher resolutions, movement can go beyond the motion search space, which would mean more bits to encode. With higher frame rates, movement is shortened between frames and is much more predictable, which could reduce the amount of bits that are expended. Objects are also bigger (have more pixel density), which would require less motion vectors to support this process. With ½ pel (pixel) motion accuracy across a smaller portion of the picture, the effect of this approach could be fewer errors in the "errored" difference frame. With more accuracy this can save on bits as well. Lastly another effect is longer GOPs over the same time period (just more frames in the same period) which can reduce the expected increase in data through temporal compression.
4) **Post-processing**: There would still be blocking artifacts that would need post processing it would just be smaller in the picture and may only need simpler post-processing techniques.

ENCODERS AND NEW CODING TOOL ABILITIES

With new demands for multiple bitrate encoders and addressing multiple devices, encoders have been evolving to output streams at multiple target bit rates. In many encoders, there is already a calculated quality metric used to make encoding decisions used for the purpose of meeting multiple target bitrates. Additionally encoders are also deploying "look ahead" to analyze the source content to optimize encoding decisions. Both these mechanisms help out in maintaining perceptual uniformity and enabling better visual masking throughout the video sequence through the use of dynamic adaptable filters.

The newer coding standards (i.e. AVC/HEVC) have also been evolving that are developing advancements in coding tools to handle each sub-area of the image and sequence in a different manner. The objective is to use as few bits to convey parts of the image or sequence that don't need as much detail such that more bits can be spent elsewhere. For instance, the background may not need as many bits as a moving object in the foreground. Also, a moving object may not need as much motion vectors since the object travels at the same relative speed against the background. Some tools being developed or refined are:

- Spatial Intra-frame compression Techniques
- Better motion pixel motion search down to ¼ or 1/8
- More granularity in quantization across coding units or transforms
- Changing the transform block size- 8x8, 4x4, 8x4, 4x8

- Changing the size of the macro-block (16x16 to 64x64)
- Changing the number of motion vectors needed for a macro-block
- Reducing the number of motion vectors needed for coding

These different tools contribute to being able to identify and handle separate areas of the image, treat specific bands of spatial frequencies with alternate options, and to code objects as separate temporal frequencies. Combine this with the ability to analyze content and a calculated quality metric in the encoder, and you have the basic tools for creating an attention model along with further refinements in the spatiotemporal sensitivity model and visual masking. From this, the HVS model used in encoding can rapidly improve encoding and reduce the amount of bits needed that can be processed by our HVS system beyond the 50 % reduction already claimed by the latest codecs.

CONCLUSION

The first phase of 4k video delivery should focus on a quality experience for the customer. It is expected that 4k will start with a single, linear occasional channel and a small library of 4k VOD assets encoded with H.264 compression techniques. Should 4k prove to be a success, it will be easy to expand the VOD library by transcoding studio content into higher resolution video.

In order to support new audio formats, assets would need to be remixed. MSOs could have a very basic 4k solution in place in the very near future; and HEVC encoding will allow a production 4k solution using substantially less bandwidth. Based on our calculations, and leveraging coding algorithms sensitive to the human visual system (HVS), 4k assets may in the near future consume the same bandwidth on the local loop as an existing HD asset encoded with MPEG-2.

References

[1] Tang, Chih-Wei, *"Spatiotemporal Visual Considerations for Video Coding"*, IEEE Trans. On Multimedia, Vol. 9, No. 2, Feb. 2007, pp. 231- 238.

[2] Naccari, Matteo and Pereria, Fernando, *"Advanced H.264/AVC-Based Perceptual Video Coding: Architecture, Tools, and Assessment"*, IEEE Trans. On CSVT, Vol. 21, No. 6, June 2011, pp.766-782

[3] Wu, H.R. and Rao, K.R. eds., Digital Video Image Quality and Perceptual Coding, CRC Taylor and Francais Group, New York, 2006.

[4] JCTVC- G1113 WD5: Working Draft 5 of High-Efficiency Video Coding, Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, 7th Meeting: Geneva, CH, 21–30 November, 2011

[5] ITU-T Rec. H.264 | ISO/IEC 14496-10, (2005), *"Information Technology – Coding of audio visual objects –Part 10: Advanced Video Coding"*

[6] "*Understanding CCD Read Noise*", www.qsiimaging.com/ccd

[7] Additional Conversations and some eye diagrams Dr. Damian Tan and Dr Henry Wu, School of Electrical and Computer Engineering, Royal Melbourne Institute of Technology, Melbourne, Victoria Australia.