AVC/H.264 encoding

Operators can realize 50 percent efficiency gains over MPEG-2.

BY PIERRE LARBIER

Using 4:2:2 10-bit encoding has become the de facto standard for professional video because it is captured and transmitted over SDI, allowing the entire production chain to use at least a 10-bit signal. In broadcast contribution, however, encoders and decoders are still limited to 8-bit signals. As a result, picture information can get lost, and quality can suffer when transmitting video.

This article demonstrates the advantages of processing video in its native SDI format using AVC/H.264 4:2:2 10-bit encoding. AVC/H.264 has historically been used at low bit rates for distribution applications. The 50 percent efficiency gains over MPEG-2 allow increased channel density, wider distance reach and reduced transmission costs. However, there is a growing need for production and contribution applications with higher standards of video quality.

Introduction

HD contribution applications are varied but share some common characteristics, including bit rates from 20Mb/s to 60Mb/s and above; low to moderate end-to-end latency, typically less than one second down to 250ms; and the need to decode and re-encode video several times before reaching the viewer.

For more than 10 years, production and contribution applications have used the MPEG-2 4:2:2 profile, but early design stages showed the potential for AVC/H.264 to be an improvement over MPEG-2. All MPEG-2 features were included, with the notable exception of an easy transrating process. Still, the majority of today's AVC/H.264 encoders and decoders are limited to relatively low bit rates and lack specific tools mandated by production and contribution applications.

As illustrated in Figure 1, most current AVC/H.264 broadcast contribution systems are based on existing distribution encoders and decoders.

The efficiency gains over MPEG-2 allow increased channel density, wider distance reach and reduced transmission costs.

Handling only High Profile requires downscaling to 4:2:0 8-bit before encoding, and upscaling back to 4:2:2 10-bit after decoding. This approach is also limited to less than 30Mb/s, which impedes the highest video quality applications in HD.

As technology matures, many products are now implementing the High 4:2:2 Profile, which is a superset of High Profile with two new tools that avoid the downscale/upscale stages shown in Figure 1, and feature 4:2:2 processing and up to 10-bit pixel bit depth handling.

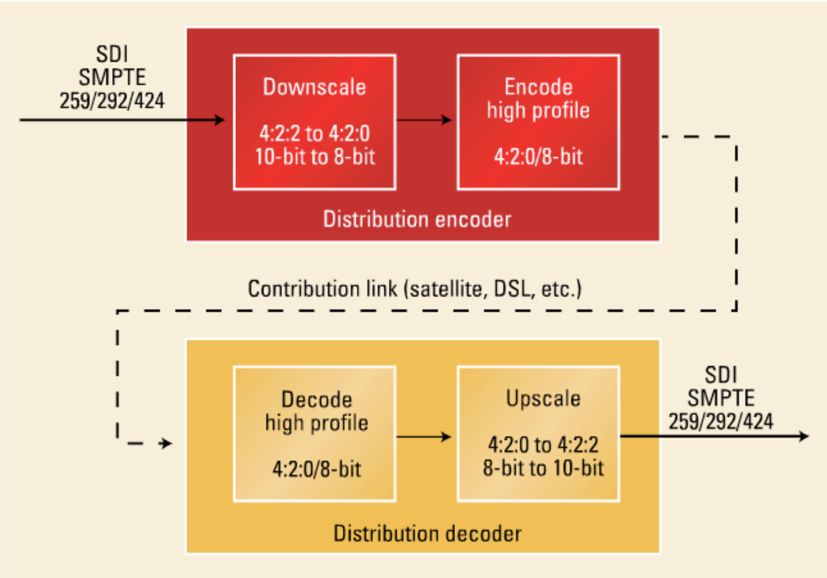

Figure 1. Architecture with distribution encoder/decoder

To show actual data versus theoretical upper bounds that could never be obtained in real time, three products were used to gather results: a contribution encoder for AVC/H.264 4:2:2 8-bit measurements, a real-time HD encoder for AVC/H.264 4:2:2 10-bit measurements and a software file encoder for MPEG-2 measurements.

Using PSNR metrics to evaluate quality

Peak signal-to-noise ratio (PSNR) measures the difference between the source and the decoded pictures of a video sequence.

A known problem with PSNR is a lack of correlation with the human

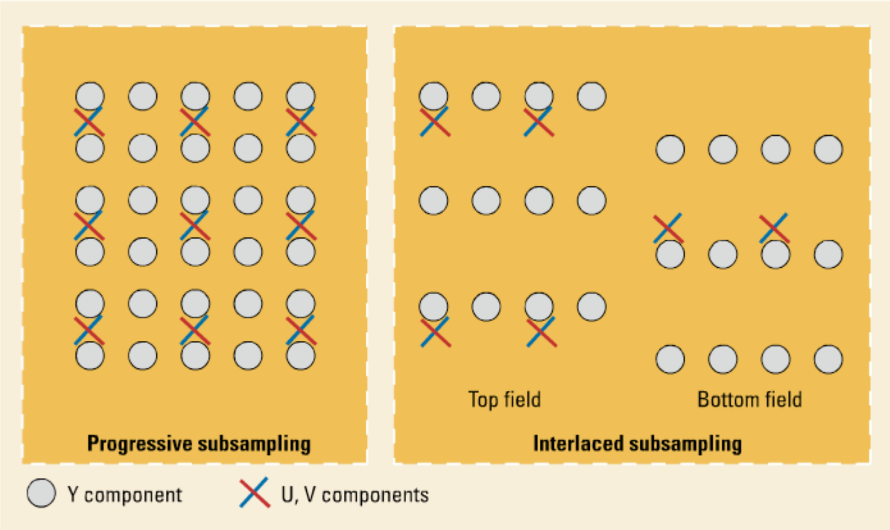

Figure 2. Chroma 4:2:0 subsampling locations

visual impression. For example, the same PSNR of 30dB for one sequence could look very good, while another could look very poor. Thus, two PSNR measurements on two different sequences are almost meaningless when measuring video quality.

However, two PSNR measurements on the same sequence performed with PSNR-optimized configurations can reveal a lot. In this case, the encoder capable of providing the highest PSNR will also be able to provide the best video quality.

When evaluating coding efficiency, it is customary to use the PSNR of the luma component only. If chroma has to be taken into account, a combined PSNR metric is often used:

$$Combined_{PSNR} = 0.8*Y_{PSNR} + 0.1*U_{PSNR} + 0.1*V_{PSNR}$$

This metric gives a good idea of the overall coding efficiency while keeping the chroma visual importance.

Encoder configurations

AVC/H.264 encoders are configured either in High Profile or High 4:2:2 Profile. AVC/H.264 profiles below the High 4:2:2 Profile process the video as 4:2:0. Because SDI transports 4:2:2 signals, chroma components need to be subsampled vertically prior to encoding and upsampled after decoding. The intent was to simplify the design as well as lower the bit rate needed to transmit compressed video.

A side effect of this process is reduced chroma detail. This is usually not a problem, however, because the human eye is not very sensitive to color information.

Even though the AVC/H.264 standard allows for six possible locations for the chroma samples relative to the luma samples, only the standard MPEG location is widely used. As shown in Figure 2, two schemes are available to handle progressive and interlaced sources.

Artifacts introduced by 4:2:2 to 4:2:0 conversions

The AVC/H.264 standard does not precisely define how the chroma subsampling or upsampling should be performed, leaving this decision to manufacturers. Thus, there can be a mismatch between the downsampling filter in the encoder and the upsampling filter in the decoder. Misinterpretation of the progressive or interlaced nature of the video can lead to faulty decoding of the whole chroma plane.

Video quality has to be kept at the highest possible level to handle several encoding-decoding steps. A mismatch in the chroma sampling can introduce color degradations that worsen with each generation, including color bleeding, loss of color contrast and details, chroma displacement relative to luma, and creation of interlaced (or progressive) color artifacts on progressive (interlaced, respectively) pictures.

It should be noted that an interlaced (alternatively, progressive) chroma artifact might confuse encoders in the cascading process, significantly reducing their coding efficiency and introducing luma degradation.

Figures 3 and 4 give an example of such problems after only five generations. The only defect introduced was

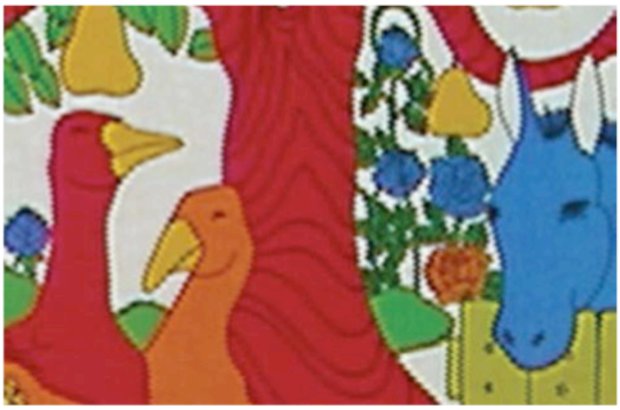

Figure 3. Source picture (mobile and calendar)

Figure 4. Source picture after five 4:2:2 to 4:2:0 conversions

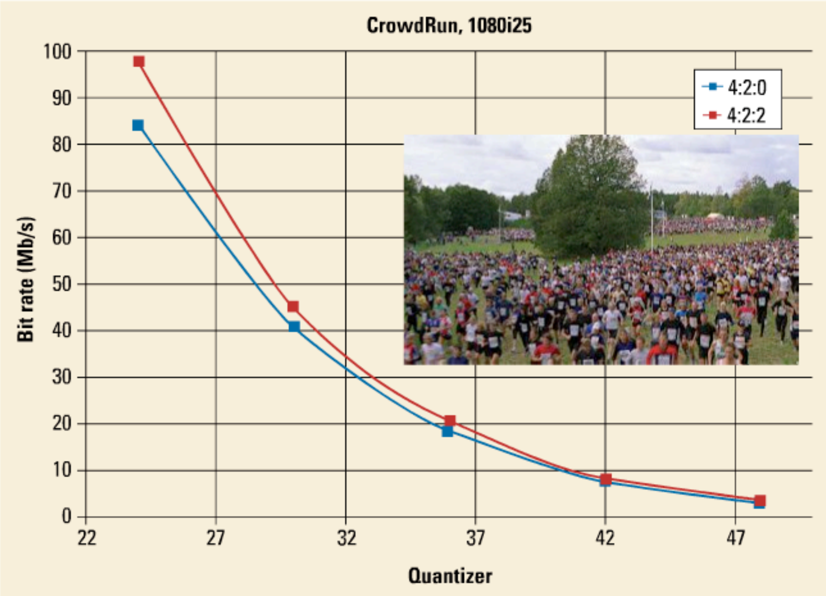

Figure 5. 4:2:2 vs. 4:2:0 bit rate comparison

a mismatch in the chroma resampling filters — polyphase bicubic downsampler before encoding and simple tent upsampler after decoding.

4:2:2 processing

The only way to avoid those artifacts is to process the video in its original color format (4:2:2). This is possible using the AVC/H.264 High 4:2:2 Profile.

The drawbacks of encoding in 4:2:2 include a moderate bit rate increase (for a given quantizer) relative to 4:2:0 encoding. (See Figure 5.) This bit rate increase does not lead to a loss of video quality with the first generation; in fact, the perceived quality is roughly the same.

As shown in Figure 6, objective measurements such as PSNR reflect this subjective impression.

10-bit versus 8-bit video compression

All AVC/H.264 profiles above High Profile encode pixels with a bit depth greater than 8 bits:

- High 10 Profile: 8 bits up to 10 bits
- High 4:2:2 Profile: 8 bits up to 10 bits
- High 4:4:4 Predictive Profile: 8 bits up to 14 bits
- High 10 Intra Profile: 8 bits up to 10 bits
- High 4:2:2 Intra Profile: 8 bits up to 10 bits
- High 4:4:4 Intra Profile: 8 bits up to 14 bits
- CAVLC 4:4:4 Intra Profile: 8 bits up to 14 bits

The bit depth increase provides improved accuracy to the AVC/H.264 compression scheme, including motion compensation, intra prediction and in-loop filtering. Figure 7 on page 22 illustrates the gains that can be achieved using greater than 8-bit processing (measured in 4:2:0 with an 8-bit source upscaled to 10, 12 or 14 bits).

Extensive experimentation demonstrates that the coding efficiency gains are highest on videos that contain shallow textures and low noise.

Figure 8 on page 22 illustrates PSNR improvement gained from increasing the bit depth to 10 or 12 bits on relatively noisy, textured standard sequences.

These curves illustrate that the gain is smaller as the bit rate is reduced, but cannot be considered negligible, making this feature attractive for low bit rate applications.

Encoding in 10 bits can achieve a PSNR increase of more than 1dB on some natural sequences and measures an average of 0.25dB at 60Mb/s on a varied test set of broadcast HD sequences. This translates to an average savings of about 5 percent and up to 20 percent, while retaining the same video quality.

However, further testing shows that increasing the bit depth to 12 bits (or even 14 bits) provides a much smaller coding efficiency gain (up to about 1 percent in bit rate savings), but again, no loss greater than 8 or 10 bits.

Lastly, there is no relation between 10-bit encoding and the frame format. The advantages are the same

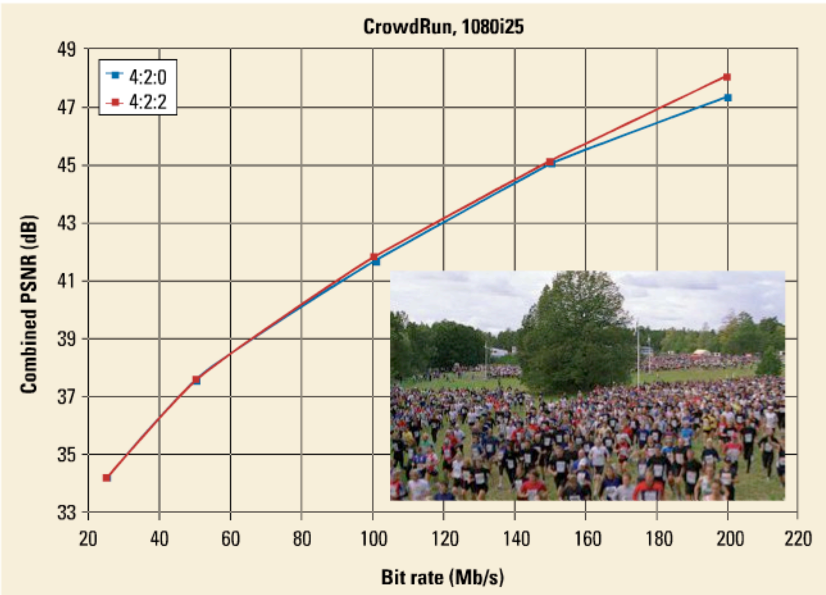

Figure 6. 4:2:2 vs. 4:2:0 quality comparison

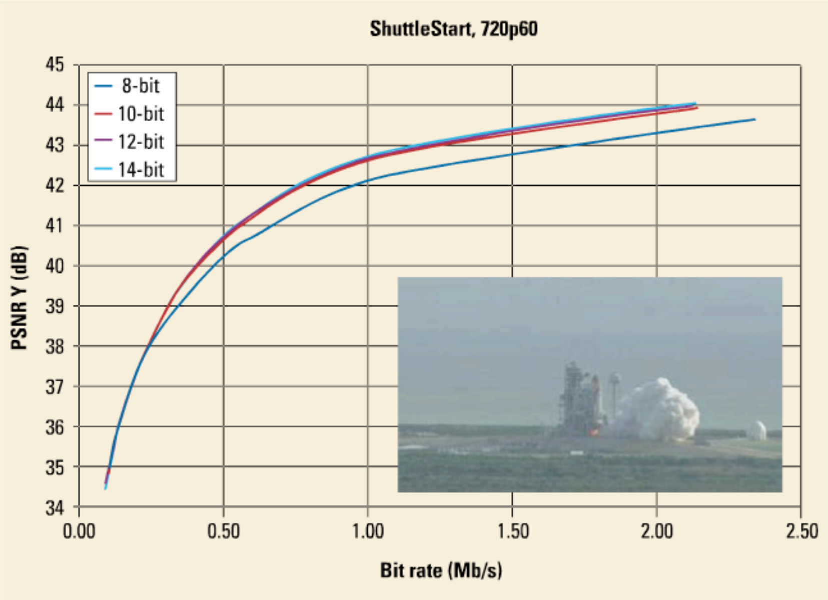

Figure 7. Coding efficiency gain using greater than 8-bit

whether the source video is HD, SD, progressive or interlaced.

Beyond coding efficiency improvements

A noteworthy benefit of 10-bit processing is perceivable gains in the reduction of three kinds of artifacts: contouring, banding and mosquito noise.

This gives a better aspect to plain surfaces and shallow textured areas (smoke,clouds,sky,sunset,etc.) because it slightly improves object edges. These impairments are otherwise difficult to reduce using traditional tools.

If the source is not too noisy and the plain area is not too large relative to the picture surface, lowering the local quantizer produces an effect close to the one achieved with 10-bit processing. This has several negative impacts, the most important one being a strong reduction of the coding efficiency and a degradation of the rate-control stability.

Another approach is to hide the defects by adding noise during the encoding process, but the amount of added noise needed to achieve the same visual improvement is significant. Even at high bit rates, this can lead to an unacceptable reduction in coding efficiency.

The gains are the result of increased accuracy in internal computations; 8-bit video sources also show improvements. Interestingly, the reduction of artifacts provided by 10-bit processing is perceivable even on standard (8-bit or dithered 6-bit) LCD panels.

10-bit compression in contribution applications

Given that high bit rates benefit most from using 10-bit compression, production and contribution applications are the best candidates for using this tool. Furthermore, it provides the opportunity to keep the original pixel bit depth all along the processing chain.

AVC/H.264 for contribution applications

The AVC/H.264 High 4:2:2 Profile enables high maximum bit rates for the video coding layer:

- 40Mb/s for 525i and 576i (Level 3)
- 200Mb/s for 720p and 1080i25/30 (Level 4.1)
- 200Mb/s for 1080p50/60 (Level 4.2)

HD encoding at about 50Mb/s provides quasi-transparency for the vast majority of broadcast content. However, measurements show that up to 150Mb/s (35Mb/s in SD) might be needed to achieve 43dB, which is a common definition of “true” transparency. The High 4:2:2 Profile bit rate capabilities can cover the full range of production and contribution applications, including those that require advanced archiving and mezzanine format support.

MPEG-2 versus AVC/H.264

Today, HD contribution is mostly performed with MPEG-2 using 422P@HL. This profile offers 4:2:2

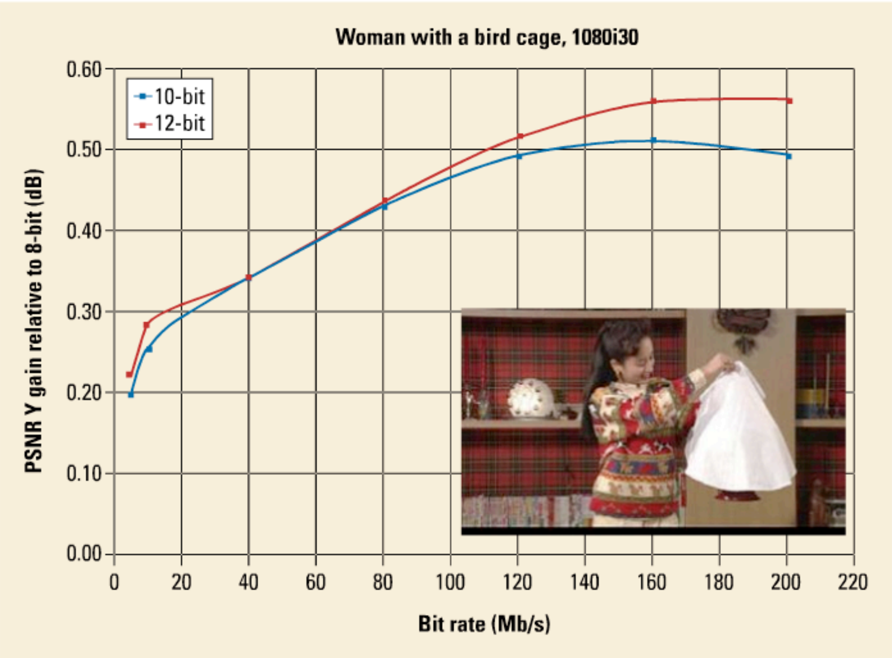

Figure 8. Coding efficiency gain using greater than 8-bit coding

processing but is limited to 8-bit pixel component bit depth. As illustrated by Figure 9, AVC/H.264 High 4:2:2 Profile offers important savings when compared to MPEG-2, even at the highest bit rates.

These HD examples allow us to draw some conclusions verified by subjective measurements:

- AVC/H.264 offers a bit rate gain of roughly 50 percent, below 15Mb/s.
- Above 30Mb/s, AVC/H.264 produces results comparable in quality to MPEG-2 with a 20Mb/s increase. For instance, MPEG-2 quality at 60Mb/s is achieved by AVC/H.264 at only 40Mb/s or less. At very high bit rates, this rate saving can sometimes be even greater.
- Above the 50Mb/s mark, the quality provided by AVC/H.264 increases linearly with the rate, showing that most of the encoder "effort" is spent coding nonredundant information like noise. Because the human eye is not very sensitive to noise fidelity, most sequences look quasi-transparent above this rate.

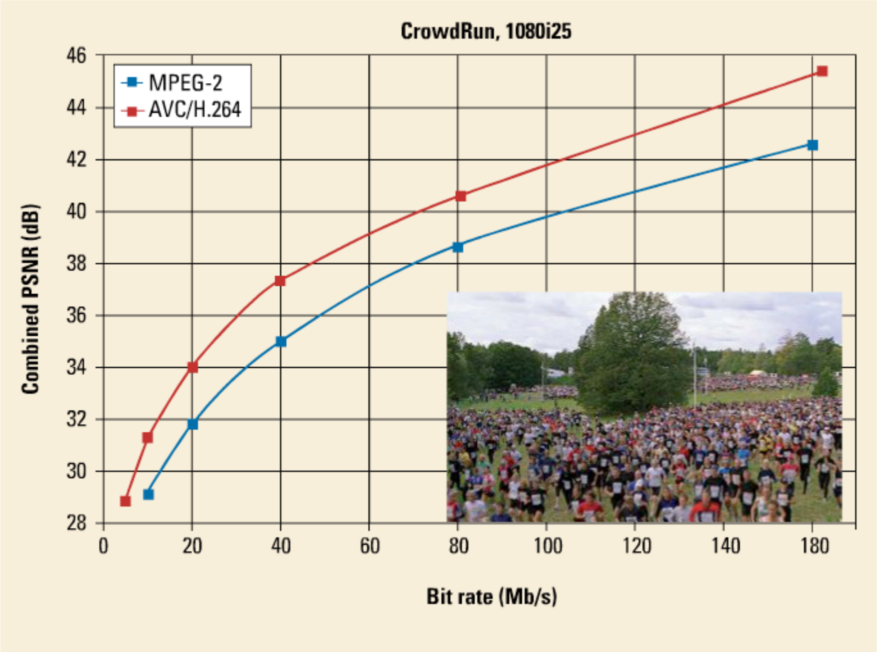

Figure 9. AVC/H.264 H422P compared to MPEG-2 H422P

Summary

There are significant advantages to using the AVC/H.264 High 4:2:2 Profile. Using 4:2:2 10-bit coding provides the most compelling solution for production and contribution applications.

Encoding with 4:2:2, 10-bit or a combination of the two will always present a gain over High Profile because all subjective and objective measurements exhibit a quality increase for the same bit rate.

In addition, the AVC/H.264 High 4:2:2 Profile offers important rate savings over MPEG-2 even at the highest bit rates, allowing the user to significantly lower transmission costs, keeping the same visual quality, or to greatly improve the video quality using existing transmission links.

This year will be a turning point for contribution applications as encoders and decoders exploit the full potential of High 4:2:2 Profile. Furthermore, relying on a highly standardized bit stream syntax guarantees products from different manufacturers to be easily interoperable. BE

Pierre Larbier is CTO at ATEME.

Video interfaces

Learn how to maximize interoperability and minimize errors.

BY ALDO CUGNINI

Professional digital interfaces continue to evolve in parallel with higher-resolution professional and consumer video formats. The once simple arena of analog connections now encompasses both compressed and uncompressed serial digital interfaces (SDI), with ever-higher transfer speeds. (See Figure 1.)

The SDI is a serial link originally standardized by SMPTE 259M. It is used to transmit uncompressed digital video over 75Ω coaxial cable and has a maximum data rate of 360Mb/s. HD-SDI is the second-generation version of SDI and allows transmission of HD (1080i and 720p) signals over the same 75Ω cables as SD-SDI. It handles rates up to 1.485Gb/s, and is defined by SMPTE 292M and ITU-R BT.656. Dual-link HD-SDI, defined by SMPTE 372M, uses two coaxial cables to provide up to 2.97Gb/s throughput while supporting up to 1080p resolution.

3Gig-SDI is the latest version of SDI and allows transmission of 1080p60 HD signals (with a 4:2:2 sampling structure at 10 bits per sample) over a single 75Ω coax cable. Defined by SMPTE 424M, the single-link studio interface can reach a maximum bit rate of 2.97Gb/s and can be thought of as a multiplexed version of HD-SDI. A companion standard, SMPTE 425M, specifies the mapping of different video signals to the 3G physical interface, as well as the supported video formats, which are listed in Table 1 on page 26. A single, high bit rate video signal can be carried over the 3Gig-SDI link and is referred to by some users as Level A, for direct image format mapping. Alternatively, two lower bit rate signals could be carried, such as two SMPTE 292 HD-SDI, and referred to as Level B. However, these specific levels are not strictly a part of the standard.

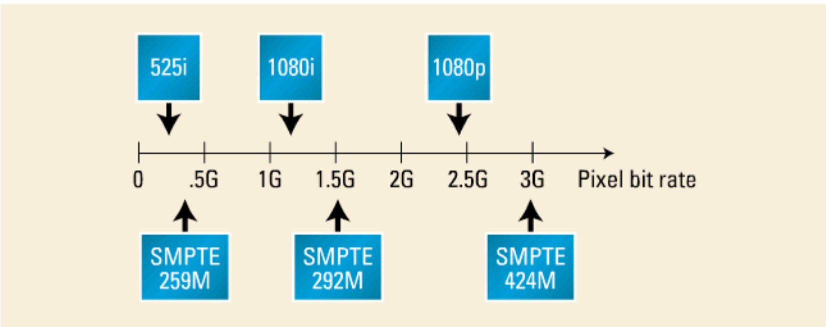

Figure 1. Serial digital standards are pushing ever-increasing bit rates.

This single-coaxial-cable solution, however, is currently limited to a cable length of about 90m, so SMPTE 435-3-2007 defines a 10.692Gb/s optical fiber interface for longer runs. Nonetheless, a demonstration in February at the Hollywood Post Alliance showed a 140m length of coax using 3Gig-SDI. This suggests the potential for new circuit designs to advance the state of the art.

Several interfaces that originated in consumer equipment are beginning to find their way into broadcast equipment.

An alternative high-speed solution, proposed by the BBC, is to apply a mild, mezzanine compression to the HD-SDI signal, compressing a 1080p60 signal by a factor of about 2.5:1, so that it can be mapped into a legacy 1.5Gb/s HD-SDI infrastructure. This approach has led to the development of a related SMPTE standard, currently in the ballot process (at press time). The compression is essentially artifact-free, but it introduces a small delay of a few video lines and requires a simple encoder and a decoder at each interface point.

An interesting feature of the proposed mezzanine codec is that the two most-significant bits of the 10-bit output can be used to carry a compatible interlaced version of the input picture. In this way, the compressed signal can be viewed as if it were standard 1080i video, making the content recognizable, although corrupted by noise. This makes it possible to monitor the signal (without a decoder) for the purposes of identification, and to provide a confidence check that the encoder is working. With full decoding, of course, the noise is not present.

Crossover applications

While the aforementioned interfaces are used almost exclusively in professional equipment, several interfaces that originated in consumer equipment are beginning to find their way into broadcast and production equipment. While these are not likely to replace SDI and similar interfaces, their growing popularity in display products are making them appear on an increasing number of professional displays.

IEEE 1394 (sometimes trademarked as i.Link or Firewire) was originally designed to support bit rates of up to 400Mb/s, but newer versions of the standard support speeds as high as 3.2Gb/s. Designed as a networked interface, IEEE 1394 is not limited to carrying video, and had been the interface of choice for DV-based cameras and prosumer HD equipment. IEEE 1394 now appears to be losing traction, however, to HDMI (High Definition Multimedia Interface) and USB.

HDMI, Digital Visual Interface (DVI) and DisplayPort are becoming de-facto display standards in consumer electronics and appear to be replacing IEEE 1394 for short-distance interconnects. HDMI1.3a supports a bandwidth of up to 10.2Gb/s, and optionally supports Deep Color with

Picture format	Signal format	Progressive frame rate	Interlaced field rate
1920 x 1080	10-bit 4:2:2 (Y'C'BC'R)	60, 59.94, 50	
	10-bit* 4:4:4 (R'G'B') 4:4:4:4 (R'G'B' + A) 4:4:4 (Y'C'BC'R) 4:4:4:4 (Y'C'BC'R + A)	30, 29.97, 25, 24, 23.98	60, 59.94, 50
	12-bit 4:4:4 (R'G'B') 4:4:4 (Y'C'BC'R) 4:2:2 (Y'C'BC'R)*		
1280 x 720	10-bit 4:4:4 (Y'C'BC'R) 4:4:4:4 (Y'C'BC'R + A) 4:4:4 (R'G'B') 4:4:4:4 (R'G'B' + A)	60, 59.94, 50, 30, 29.97, 25, 24, 23.98	
2048 x 1080 (D-Cinema)	12-bit 4:4:4 (X'Y'Z')*	24	

** Includes PsF (Progressive segmented Frame) format*

Table 1. 3Gig-SDI supported video formats, per SMPTE-425M

30-bit, 36-bit — and 48-bit $Y'C_BC_R$, *xvYCC* and *sRGB* — signals. HDMI has gotten a strong foothold in the consumer market, and this is a big plus, considering the long lead times of CE companies investing in new technologies. And while HDMI runs are usually limited to about 15m, twisted-pair (Cat 5) can extend this to about 50m, and fiber can push this to over 330m.

DVI was developed to support high-resolution PC monitors. A single DVI link consists of four twisted pairs of wires, each carrying one red, green, blue and clock signal, supporting up to 24 bits per pixel. In addition, DVI is probably the only popular standard that optionally supports both analog and digital signals over the same connector.

With a single DVI link, the largest resolution possible at 60Hz is 2.75 megapixels. Higher resolutions are possible by using a dual-DVI link (over a single cable), with alternate pixels transmitted on each link. The maximum bit rate in single link mode is 3.96Gb/s and in dual link mode is 7.92Gb/s. A single link can therefore carry a 1920 x 1080p60 signal, and a dual link can support a 2560 x 1600p60 (WQXGA) display. Most high-end PC graphics boards supply one or more single or dual-link DVI connections.

DisplayPort is a relatively recent license-free digital audio/video interconnect, proposed by VESA to succeed DVI; it's primarily intended to be used between a computer and its display monitor. Because it's based on a micro-packet protocol, DisplayPort may have an advantage over other interfaces by easing future expansion of the standard. The interface consists of a unidirectional main link for connecting A/V streams from source devices to sink devices, and a half-duplex bi-directional auxiliary channel used for realizing plug-and-play features such as equipment control. The main link may have one, two or four differential signal pairs (or lanes), offering a total raw capacity of up to 10.8Gb/s. Computer giant Dell is a strong DisplayPort supporter, and rumors have it that the company will include the interface on all products within two years.

Traditional USB 2.0 (high-speed USB) has already become the interface of choice for many computer peripherals, including storage media. It supports a maximum transfer rate of 480Mb/s. The new USB 3.0 specification (Super Speed) has already been completed, and USB 3.0 products are expected in late 2009. This version will increase transfer rates up to 5Gb/s.

Despite the proliferation of displays with HDMI or DVI interfaces, studios will still need to test video quality on analog monitors.

Where's it all going?

Most production and broadcast video systems now include SD/HD-SDI, which supports up to 720p/1080i (uncompressed), and marketplace evidence suggests that video equipment is transitioning from HD-SDI to 3Gig-SDI. With 1080p video becoming the dominant video format in the future (and possibly the current de-facto production standard), transmission speed will be the bottleneck in any video interface. Research over the last two years has shown that 1080p can be done at an equivalent bit rate to current 1080i broadcasts, provided MPEG-4 (H.264/AVC) is used.

Despite the proliferation of displays with HDMI or DVI interfaces, studios will still need to test video quality on legacy analog TVs and monitors. So, don't count out analog component video interfaces — yet. **BE**

Aldo Cugnini is a consultant in the digital television industry.