Research White Paper

WHP 159

November 2007

Open Technology Video Compression for Production and Post Production

Tim Borer

BRITISH BROADCASTING CORPORATION

Open Technology Video Compression for Production and Post Production

Tim Borer

Abstract

This paper describes an Open Technology video codec aimed at production and post production applications. The codec is being standardised by the SMPTE as VC-2 and is also known as Dirac Pro. VC-2 is part of a family of interoperable codecs, which also includes the Dirac video codec. This paper provides an overview of the history and background of VC-2, how it works, typical production applications and some a guide to its features and performance.

Additional key words: open source

White Papers are distributed freely on request.

Authorisation of the Head of Research is required for publication.

OPEN TECHNOLOGY VIDEO COMPRESSION FOR PRODUCTION AND POST PRODUCTION

Tim Borer

BBC Research, UK

INTRODUCTION

With HDTV and D Cinema production enormous amounts of data must be transported around studios and production facilities. Video compression, a necessary evil in achieving this in a practical and economic way, is more important now than ever before. Previously video compression technologies have tended to be proprietary or have involved onerous licensing conditions and the payment of royalties. This makes implementation of production systems complex and expensive, and degrades picture quality by cascading multiple types of codecs.

Recently the BBC has developed a royalty free family of video codecs, which is being standardised through the SMPTE as VC-2. VC-2, also known as Dirac Pro, is a royalty free technology that anyone can use. VC-2 provides efficient compression but is simple and cost effective to implement in hardware and software for a wide range of applications.

This paper describes how the VC-2 technology works. It aims to indicate how compression parameters can be chosen to optimise VC-2 for different applications in terms of factors such as latency, compression performance, and complexity (e.g. ease of implementation and cost). Key features of the technology are discussed including low latency, multi-generation coding, lossless coding, coding of RGB and wide dynamic range video. First generation hardware and software implementations are discussed including some performance indicators.

THE EVOLUTION OF VC-2 AND DIRAC

The BBC has worked in the field of video compression for more than 50 years (Gourier (1))! For several years we have worked on the Dirac video codec, intended for low bit rate distribution applications. During the same period the BBC's desire to move towards HDTV production led us also to work on a high quality codec for production applications. At the beginning of 2006 the project teams merged. Rather than pursue two separate codecs we chose to create a family of interoperable codecs that spanned the whole range of compression applications from low bit rate mobile to lossless D-Cinema. This family of codecs is known as VC-2 and Dirac. VC-2 is also sometimes known as Dirac Pro.

Both VC-2 and Dirac are Open Technologies. There is a published specification (2) and there are no licence fees or royalty payments to be made for implementing or using them. This has important commercial implications and fits well with the business model of public service broadcasters such as the BBC. The rationale and implications of making VC-2/Dirac an open technology are discussed in more detail in Borer et al (3).

VC-2 is the professional end of the codec family, aimed at production and post production. It provides high quality pictures, is low complexity and low latency. Dirac is at the distribution end of the family but is also well suited to contribution links between production centres. VC-2 and Dirac share the same signal processing and stream syntax. The essential difference between them is that VC-2 is, currently, a purely intra frame codec whereas Dirac may also use motion compensated prediction to increase compression. So VC-2 is a subset of the Dirac codec.

VC-2 is aimed at applications that require only modest compression but *do* require high quality, low latency and low complexity. High quality implies from visually lossless to mathematically lossless. The compression factor may be between 2:1 and 16:1. Latency in VC-2, through coder and decoder together, varies from 6 HDTV lines (about 0.25ms) to a few ms. Having low complexity means that equipment is low cost, low power, physically small and can be easily implemented in real time in hardware and software. Typical applications of VC-2 include desktop production over IP networks, reducing disk storage and bandwidth in D-Cinema production and moving HD video over legacy infra-structure (such as standard definition SDI cables and equipment).

Dirac is aimed at lower bit rates (higher compression factors), comparable to those achieved using MPEG-4 AVC/H264 (hereafter “AVC”). Its applications range from contribution codecs (e.g. 50Mbit/s for HDTV) to distribution applications such as web streaming and IP TV. It can be used for the same applications as AVC and for other applications as well. Over the past two years the compression performance of Dirac has approximately doubled. It can now code 1080 line HDTV pictures at broadcast quality using 12Mbit/s or less.

VC-2 is being standardised through the SMPTE, and various drafts of the specification having been submitted over the past year. In addition to the core compression standard we are also working towards standardising the transport of VC-2 over SDI links and producing a recommended practice for the appropriate coding parameters. The specification of Dirac (2), which has been available for more than a year, closely tracks the VC-2 specification, and adds the features required for motion compensated prediction.

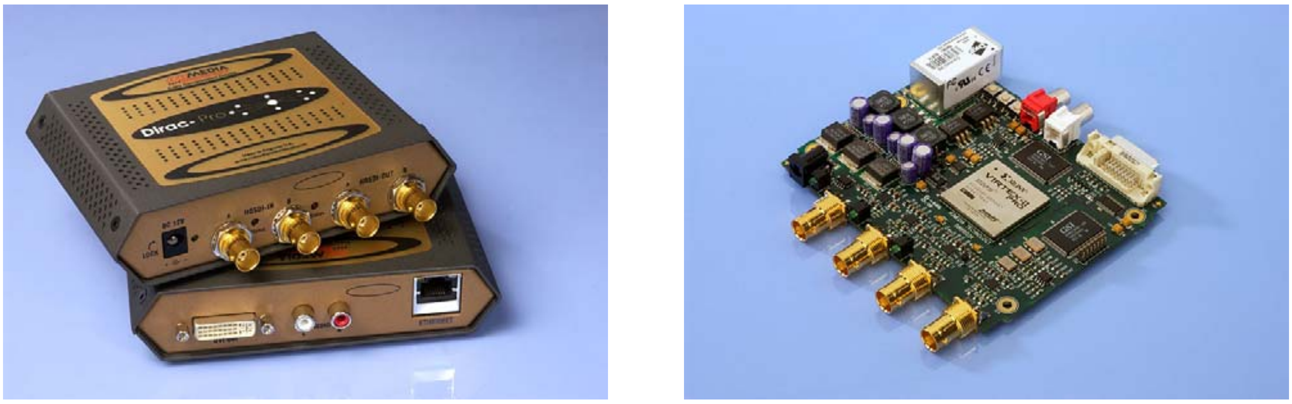

Figure 1 - VC-2 Hardware

Various implementations of VC-2 and Dirac currently exist and others are in development. We have initially focused on two applications for VC-2. The first, called Dirac Pro 1.5, is 5:2 compression of full HDTV (1080P50/60) for transport over (1.5Gbit/s) HD-SDI links. The second application, called Dirac Pro 270, is the compression of 1080I50/60 for transport over standard definition (270Mbit/s) SDI links. Reference code for these VC-2 applications is available to download (4). Commercial VC-2 hardware for Dirac Pro 1.5 was launched in April 2007 and hardware for Dirac Pro 270 will be launched soon (5). The original Dirac software (6) is approaching version 1.0, which means it will comply with the specification and be essentially bug free. An alternative, optimised, software implementation, called Schrodinger (7), should also be available by the time this paper is published. Dirac hardware is in development.

THE COMPRESSION TECHNOLOGY (HOW DOES IT WORK?)

VC-2 was designed to be a practical, royalty free, codec. Simplicity and low complexity had higher priority than ultimate compression performance. Nevertheless its compression performance is similar to other comparable codecs. A critical feature of VC-2 is its low latency; Dirac Pro 1.5 has a latency of only 6 HDTV lines (about 0.25ms). For simplicity VC-2 operates on pictures independently (I frame coding), which limits compression but minimises complexity.

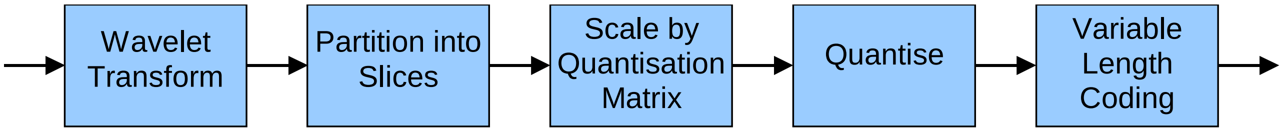

Figure 2 - VC-2 Signal Processing

Figure 2 shows a signal processing block diagram of a VC-2 encoder. The principle components are the wavelet transform, quantisation and entropy coding. The quantisation factor is transmitted, along with the quantised coefficients, in the output stream. Unlike many video codecs VC-2 uses wavelet transforms rather than block transforms. Wavelets have proved more effective than block transforms for still image compression and are used in the JPEG2000 still image compression standard (8). The artefacts induced at restricted bit rates tend to be more benign than those due to block transforms, without any "blocking" artefacts.

The wavelet transform is constructed by repeated filtering of signals into low and high frequency parts (and subsampling). For two dimensional signals, this filtering occurs both horizontally and vertically. At each stage, the low horizontal / low vertical frequency sub-band is split further, resulting in logarithmic frequency decomposition into sub-bands.

Figure 3 - Wavelet Transforms

Figure 3 shows an original image on the left, a first level wavelet transform in the middle and a second level transform on the right. In the transforms mid-grey represents zero. The transforms contain the same number of samples as the original image. But, in the first level transform, most of the energy is compacted into the top left quarter of the transform. In the second level transform the energy is compacted further into the top left sixteenth of the transform (although some energy is now visible elsewhere). Compacting the energy allows compression using variable length codes described below.

The wavelet transform is implemented using lifting. Each level of the transform has one lifting step horizontally and one vertically. So a two level transform has a total of four lifting steps. Figure 4 shows a forward lifting step, used in the encoder, on the left, and an inverse lifting step, used in the decoder, on the right. The input is first split into even and odd samples, that is, splitting produces two sub-sampled signals. P and U indicate prediction and update stages, which are simple transversal (FIR) filters. The inverse transform reconstructs the original signal.

The lifting approach to wavelet transforms is simple to implement. In the absence of quantisation is guarantees perfect reconstruction, even with finite precision arithmetic and rounding or truncation in the filters. This feature is not true for conventional DCTs or for some modes in JPEG2000. This perfect reconstruction property means that VC-2 may be used for mathematically lossless coding, for example for archive purposes.

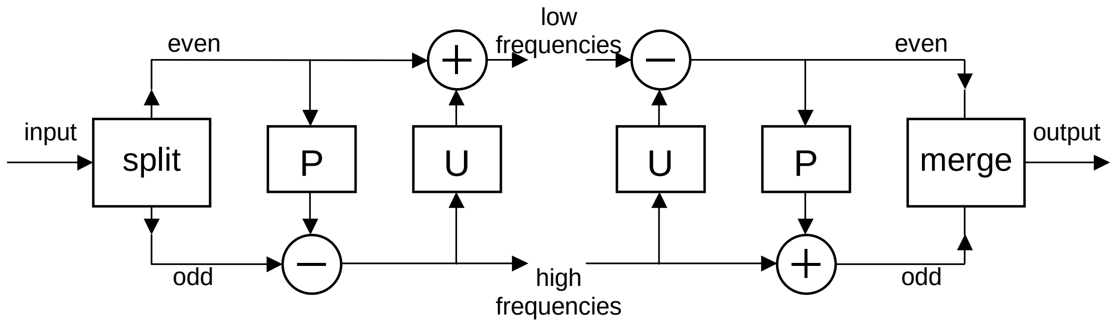

Figure 4 - Lifting Implementation of the Wavelet Transform

VC-2 provides a range of options for the wavelet filter and allows a variable number of transform levels. This enables the codec to be tailored for specific applications, by trading off complexity, compression efficiency and latency. Dirac Pro 1.5 uses a 2 level Haar wavelet transform. Haar is the simplest possible wavelet, giving low complexity and, significantly, very low latency. Dirac Pro 270, which requires a higher compression ratio, uses a three level LeGall 5,3 wavelet transform (also used in JPEG2000). This transform gives better compression performance at the expense of slightly greater complexity and higher (though still low) latency. For details of the wavelet filters see reference (2).

To improve compression further the lowest frequency ("DC") coefficients (corresponding to the top left in the transforms of figure 3) are predicted from their neighbours.

After performing a wavelet transform (and prediction) the coefficients are divided into slices. This is different to block codecs such as MPEG, in which the picture is partitioned before it is (block) transformed. The difference is a key feature of VC-2. It avoids the blockiness caused by rigid slice boundaries but still supports low latency.

Each slice contains one or more of the lowest frequency (i.e. "DC") wavelet coefficients (that is a sample from the top left of the wavelet transform in figure 3). It also contains associated higher frequency wavelet transform coefficients. Each slice corresponds approximately to a small region of the picture. In Dirac Pro 1.5 each slice contains 4 (luminance) DC coefficients in a horizontal row (plus 60 higher frequency coefficients). So each slice corresponds to a region of the picture 16 pixels wide by 4 lines high. A slice in Dirac Pro 270 contains only two (luminance) DC coefficients (plus 126 higher frequency coefficients). So, in Dirac Pro 270, each slice, roughly, corresponds to a region of the picture 16 pixels wide by 8 lines high. The coefficients within a slice are ordered from low to high frequency within the slice, which is somewhat similar to zig-zag scanning in block transform codecs.

The wavelet coefficients in each slice are scaled by a quantisation matrix depending on their frequency. The primary function of the quantisation matrix is different than in block codecs. In block codecs a quantisation matrix is used to apply frequency dependent psycho-visual weighting to the transform coefficients. In VC-2 it can be used this way too. But its primary purpose is to compensate for filter gain in the lifting stages. Even without psycho-visual weighting a quantisation matrix is still needed in VC-2.

In VC-2 each slice is quantised to fit within a fixed number of bytes. The number of bytes per slice determines the compression ratio. For Dirac Pro 1.5 each slice occupies 64 bytes and for Dirac Pro 270 slightly fewer bytes (because a higher compression ratio is required). The quantisation factor used in each slice is chosen to ensure that quantised coefficients fit within the allocated number of bytes. This method of quantisation ensures that plain regions of the picture are lightly (or losslessly) coded. Most quantisation takes place in the detailed regions of the picture where quantisation noise is masked by the picture content. This is in contrast to many codecs in which quantisation is more uniform, which can result in compression artefacts in plain areas where they are very visible.

The final stage in coding is lossless entropy coding using a variable length code. Rather than use conventional Huffman coding VC-2 (and Dirac) use a modified form of expGolomb coding. This is a simple deterministic code, which is easy to implement.

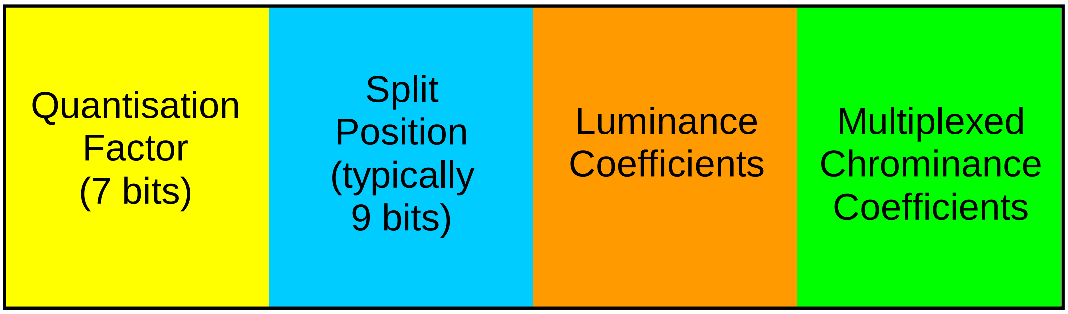

Figure 5 - Data Packing in a VC-2 Slice

The entropy coded coefficients are packaged in a slice as shown in figure 5. The chrominance coefficients are packed together with alternate U and V coefficients. The number of bits allocated to luminance and chrominance can vary depending on picture content.

The entropy coding in VC-2 does not include run length coding, which is common in other codecs. However, VC-2 does have an equivalent. If all the coefficients in a slice, above a certain frequency, are zero they don't need to be coded at all! Their presence can be inferred because the number of luminance (and chrominance) coefficients is known. Since runs of zero value, high frequency, coefficients can be coded in zero bits, the compression in VC-2 is greater than might be naively assumed considering variable length coding alone.

VC-2 Core Syntax

The description above describes the "low latency" mode of VC-2. VC-2 also has a "core syntax" mode, which is basically I frame Dirac coding. The signal processing in the two modes is identical, but there are two essential differences between the modes. In the core syntax, the wavelet coefficients are packaged as whole pictures rather than as slices. Secondly, arithmetic coding is used to increase the compression ratio in the core syntax mode.

In the low latency mode a compression ratio up to about 8:1 can be achieved. With arithmetic coding the achievable compression ratio doubles to about 16:1. But this increase in compression efficiency comes at the price of increasing both complexity and latency. Nevertheless for some applications this is worthwhile, e.g. in the compression of production quality full HDTV (progressively scanned, 1080P50/60) for transport over a standard definition (270Mbit/s) SDI link.

The arithmetic coding in VC-2 has been designed to be as simple as possible to implement and to allow a large degree of parallelism in both hardware and software. This contributes to its ease of implementation compared to JPEG2000 and I frame H264/AVC. So VC-2 can be used as a low complexity alternative to JPEG2000 and I frame H264/AVC without sacrificing picture quality, and without licensing costs.

Dirac

For even higher compression ratios motion compensation prediction is needed. Dirac extends the VC-2 codec by adding motion compensated prediction. Dirac is a hybrid motion compensated codec, similar in concept to MPEG-2 and AVC but using wavelets rather than DCTs or block transforms. The details of Dirac are described elsewhere (3) and are also on the web. Compression ratios of up to about 100:1 can be achieved with the Dirac codec for distribution applications. For production applications the compression ratio should be considerably lower. A principal application of Dirac in a production environment would contribution links and links between production centres.

PRODUCTION FEATURES OF VC-2

VC-2 was designed with production and post production applications in mind. It has a number of features that make it particularly suitable for these applications. VC-2 supports a wide range of picture formats. It can support 4K D-Cinema images and beyond. It is not limited to 8 or 10 bit signals but can support the large word widths needed in post production. All frame rates are supported. RGB and 4:4:4, 4:2:2 and 4:2:0 colour formats are all supported[1].

The VC-2 stream carries essential metadata that describes how the coded signal should be interpreted. This includes signal ranges and offsets, colour space information (primaries, colour matrices and transfer functions), pixel aspect ratios and definable "clean areas" (for pictures within larger containers). To assist editing, each frame includes a 32 bit frame number (more than 2 years of unique frame numbers at 60 frames/s!), which avoids the ambiguities of time code.

PERFORMANCE

The performance of any video codec is best judged by looking at the pictures in the context in which the compression will be used.

An objective measure of compression performance is the compression ratio achieved for lossless coding. Our experiments on lossless coding comparing VC-2 low latency, VC-2 core syntax (with arithmetic coding) and JPEG200 suggest the following conclusions. Bear in mind that lossless compression ratios can vary considerably with picture content so these conclusions are based on average compression ratios for a range of pictures. The compression ratio achieved by VC-2 core syntax was virtually identical to that achieved by JPEG2000. Perhaps this is not surprising as they used the same wavelet filter (LeGall 5,3) and both codecs use arithmetic coding. The typical compression ratio achieved, using a 3 level wavelet transform, was between 2 and 2.5 for 8 bit video. The JPEG2000 codec used did not support 10 bit video but the figures for VC-2 core syntax decreased to a

---

[1] RGB is supported in the same way it is in JPEG2000, that is by the use of a lossless colour matrix.

compression ratio of about 2 for 10 bit video. The number of levels of wavelet transform, from 1 to 4 levels, made only a few percent difference in the compressed bit count, with a 3 level transform marginally best.

Perhaps surprisingly, arithmetic coding only improved the compression ratio by about 25% for lossless coding. It seems that arithmetic coding is not very effective at high bit rates for lossless and near lossless coding. Bearing in mind difficulty of implementing arithmetic coding at bit rates of 100Mbit/s or more we decided that benefits did not justify the complexity of arithmetic coding and the higher latency for low compression ratios. Hence VC-2 does not include arithmetic coding in the low latency mode. By contrast, at high compression ratios, other experiments have shown arithmetic coding to yield a coding gain of about 3, making it essential for distribution codecs.

Lossless coding can only indicate performance at low compression ratios. Other objective measures, such as Peak Signal to Noise Ratio (PSNR), whilst extremely limited, give a broad indication of performance at lower compression ratios. The PSNR figures for Dirac Pro 1.5, with a compression ratio of 5:2, range from 62dB (with up to 75% of slices being coded losslessly) to 46dB worst case. Typical PSNR values are about 50dB suggesting visually lossless coding, which is borne out by looking at the pictures. Complete figures for Dirac Pro 270, compression ratio 5:1, are not yet available, but initial indications are that they will be about 8dB lower than Dirac Pro 1.5, yielding typical PSNR values in the low 40dBs.

Cascaded or multigeneration coding is an important aspect of codec performance, since in production the signal may be coded many times. VC-2 has been designed to minimise the effects of multiple coding. In principle the coder can be designed so that there is no loss beyond the first generation! To achieve zero second-generation loss no side channel is required for metadata; the coding works on the signal alone. The coder would be about 25% more complex to implement this feature. The first generation Dirac Pro 1.5 hardware, and the reference code do not, yet, include the feature. For difficult to code pictures they exhibit (a total of) about 0.5dB degradation in PSNR over 9 generations.

CONCLUSIONS

This paper has given an overview of VC-2 codec, which is aimed at production applications. It has described the history, technology and current status of the codec. It has indicated some of the applications. The performance figures show that VC-2 is of similar or better performance to comparable codecs. However it is also easier to implement and does not require payment of royalties or license fees.

REFERENCES

1. Gourier, G.G., 1957. Bandwidth Compression of a Television Signal. Proceedings I.E.E., Paper No. 2357 R, 104 B, pp 265-271

2. http://dirac.sourceforge.net/specification.html

3. Borer, T. & Davies, T. 2005. Dirac - Video Compression using Open Technology. EBU Technical Review No 303, July 2005

4. http://dirac.cvs.sourceforge.net/dirac/diracpro/

5. http://www.numediatechnology.com/

6. http://sourceforge.net/project/showfiles.php?group_id=102564

7. http://schrodinger.sourceforge.net/

8. http://www.jpeg.org/jpeg2000/

ACKNOWLEDGEMENTS

The author would like to thank his colleagues at the BBC and Numedia Technology for their contributions to the development of the VC-2 codec. He would also like to thank the BBC for permission to publish this paper.