Comparison between 4:2:2P and 4:2:0
– for 525- and 625-line pictures

L. Cheveau
Head of Transmission Technologies,EBU

A. Caruso
CBC

This article reports on the results of tests carried out by the EBU and CBC on 4:2:2P and 4:2:0 codecs, using the newly-available Tektronix Picture Quality Analyzer. CBC were responsible for the 525-line tests and the EBU for the 625-line tests.

1. Introduction

Prior to the digitalization of the EBU's *Eurovision* network in August 1998 [1], extensive tests had been carried out on different bit-rates and coding schemes. For operational reasons, two bit-rates – 24 Mbit/s and 8 Mbit/s – were selected for use on the network. Furthermore, the technical people recommended that 4:2:2 codecs should be used *only* at the 24 Mbit/s bit-rate, and that (4:2:0) MP@ML codecs should be used at 8 Mbit/s.

However, when making real use of the MPEG equipment after August 1998, it soon became apparent to the *Eurovision* engineers that, at 8 Mbit/s, the quality of the picture was better with 4:2:2 coding than with MP@ML coding. The EBU Technical Department tried to find some clear evidence that 4:2:0 might be superior to 4:2:2, but was unable to find any picture sequence that could prove this. On the contrary, it became evident that 4:2:2 was always better, and the decision was taken to use 4:2:2 systematically for all the *Eurovision* exchanges.

Around the same time, CBC was also looking to digitalize its network, using two different bit-rates, i.e. 8-10 Mbit/s for distribution and 18-20 Mbit/s for collection-contribution. The question of 4:2:0 or 4:2:2 was raised and CBC and the EBU joined their efforts to organize some tests. The first subjective tests were performed by CBC, with 525-line material. The results were presented at IBC 98 in Amsterdam and also in **EBU Technical Review** [2].

The conclusion was that 4:2:2 is always better than 4:2:0, even at very low bit-rates (i.e. 2 or 3 Mbit/s), where the limit is the encoding equipment rather than the quality of the pictures.

Since then, additional measurements have been made by the authors, using the picture-quality analyzer developed by Tektronix. The CBC tests were made on three different encoders

Original language: English
Manuscript received: 9/4/99.

and the results confirmed that there was no cross-over point where, at a given bit-rate, 4:2:0 becomes better than 4:2:2.

Similar tests were made by the EBU Technical Department using an additional test sequence at 625-lines and a different GoP structure.

The EBU results were slightly different and suggest that, for 625-line systems, there may be a cross-over point at 4 or 5 Mbit/s, but the discussion here shows that additional evidence may be needed, to prove the existence of such a cross-over point.

2. Subjective evaluation of 525-line pictures by CBC

The material used for the CBC subjective evaluation was prepared in the NDS laboratories, Southampton (UK), in early November 1997. NDS provided the necessary hardware and engineering assistance to set up two parallel MPEG-2 systems. The tests were performed with the participation of experienced engineering personnel from CBC, the EBU and NDS.

A complete description of the set-up and the results of this evaluation were reported extensively at IBC 98 [3] and has also been published in **EBU Technical Review** [2].

A total of eight videotapes (Digital Betacam format) were produced and included all 48 combinations of the following parameters:

- 4:2:2 P@ML and 4:2:0 MP@ML (Main Profile at 4:2:0);
- one and two cycles of encoding-decoding;
- composite (525-NTSC) and component (SDI, SMPTE 259 M) interfaces;
- variable bit-rates down from 15 Mbit/s: (15, 12, 10, 8, 6 and 4).

When preparing the tests at this time, it was a common belief that the cross-over point was around 10 Mbit/s and the recordings were limited at 4 Mbit/s.

The source material used for all tests amounted to a total duration of 10 minutes. It consisted of 25 excerpts of a large variety of 525-line scenes, including sport, film, computer-generated scenes and EBU reference test material.

The subjective assessment of the tapes was performed at a later date (early 1998) in Canada, utilizing the double-stimulus comparison method described in ITU-R Recommendation BT.500.

The evaluation was limited to a comparison between the 4:2:2 and the 4:2:0 codec pictures. The results for each configuration and bit-rate tested were averaged over the number of participating persons in the viewing session, and were summarized in 24 tables, one for each configuration and bit-rate tested.

The most important conclusions of this investigation were:

- After (a) one cycle of compression/ decompression with either SDI or composite NTSC used for interfacing the equipment, and (b) two cycles of compression/decompression

with the SDI interface only — the video that was encoded using 4:2:2 P@ML remained on the whole "*The Same*" as the video encoded using (4:2:0) MP@ML. A few sequences encoded using 4:2:2 P@ML were deemed on average to be "*Slightly Better*" than the reference MP@ML.

This was true for all the bit-rates tested.

⇨ However, after two cycles of compression / decompression with the composite NTSC interface, the video encoded using the 4:2:2 P@ML remained on average "*Slightly Better*" than the reference MP@ML, for all the bit-rates tested.

⇨ At bit-rates of 6 Mbit / s and below, video artefacts were observed on certain programme sequences. However, these artefacts were equally observed on both profiles.

As a final conclusion, it was stated that the sequences encoded and decoded using the 4:2:2 P@ML were never rated "*Slightly Worse*" than those coded using the MP@ML, at all the bit-rates investigated. This clearly indicated that a "cross-over point" where 4:2:0 is better than 4:2:2 did not exist, at least above 4 Mbit / s and for 525-line pictures.

3. Tests using Tektronix PQA equipment

3.1. 525-line signals

After the intensive subjective evaluation of the CBC tapes in Montreal, the CBC had the opportunity to use the brand new Tektronix Picture Quality Analyzer, PQA 200.

This equipment makes a comparison of two short sequences, pixel-by-pixel, and provides a number, called the *PQR value*, which is supposed to measure the degradation between the original and the tested sequence. For this, Tektronix have used a proprietary algorithm which is still being discussed in the standardization bodies.

The results given by the Tektronix equipment are expressed in *PQR units*.

According to the manufacturers, the PQR [1] value can be loosely interpreted as follows:

⇨ *A PQR rating of 1 indicates impairments that have a small perceptual impact*;

⇨ *A PQR rating of 3 indicates impairments that are almost always observable but not strong*;

⇨ A PQR rating of 10 indicates impairments that are clearly observable.

For **PQRy** (y = luminance), measurements are made only on the luminance component of the video signal, to provide a faster analyzis of the video test sequence. Measurement of the luminance differences provides a good basis for the comparison of picture differences in the majority of cases.

For **PQRyc** (luminance and chrominance), measurements – similar to the luminance measurements described above – are made on both the luminance and the chroma components of the video signal. Chrominance differences have a small impact on the PQR of a scene, and the

1. Extract from the user manual of the PQA 200 Picture Quality Analyzis System.

measurement takes more time than making the luminance-only measurement. However, it does provide a means of detecting errors in the colour channel.

For this series of tests, we have of course chosen to use the PQRyc indicator.

A series of tests was organized using the sequence *Mobile and Calendar* (in its 525-line version) which was contained on the CD-ROM (Mobile +) that accompanied the Tektronix equipment.

Tests were repeated with three codecs supplied by three different manufacturers. The results are reproduced in *Fig. 1* for each of the tested codecs.

It can be noticed that although the three codecs have a different behaviour, the most striking result is that the 4:2:2 results are always better than 4:2:0 above 2 or 3 Mbit / s.

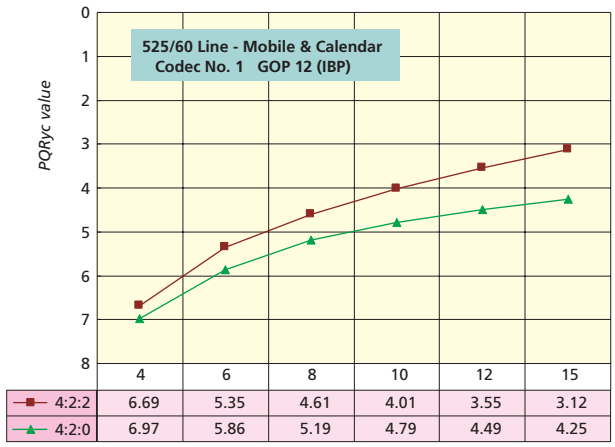
3.2. 625-line signals

At the beginning of 1998, the EBU was able to use a comparable Tektronix PQA system to make similar measurements on 625-line pictures. In order to be able to compare these measurements with the CBC results, a similar signal was first analysed: Mobile +, with a GoP of 12 and an IBP structure. This is shown in *Fig. 2(a)*.

Only one coder-decoder pair was available during these tests, so the EBU decided to analyze other configurations (IBBP and IBP structures) *Fig. 2(b)* gives the "Mobile +" analyzis for an IBBP structure.

The EBU also decided to use a second reference sequence, "*Flower Garden*", for a series of measure-

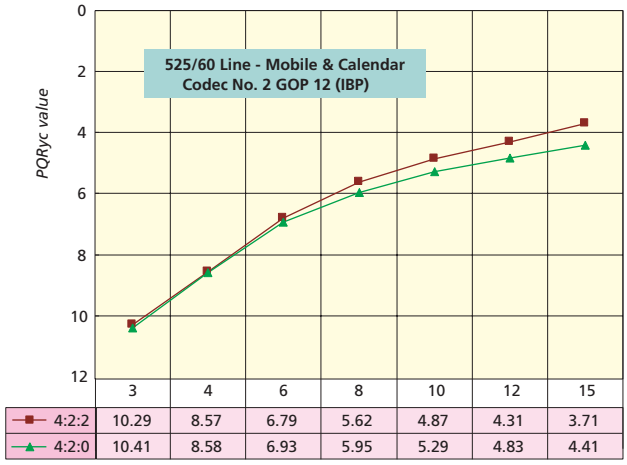
(a)



	4	6	8	10	12	15
4:2:2	6.69	5.35	4.61	4.01	3.55	3.12
4:2:0	6.97	5.86	5.19	4.79	4.49	4.25

Mbit/s

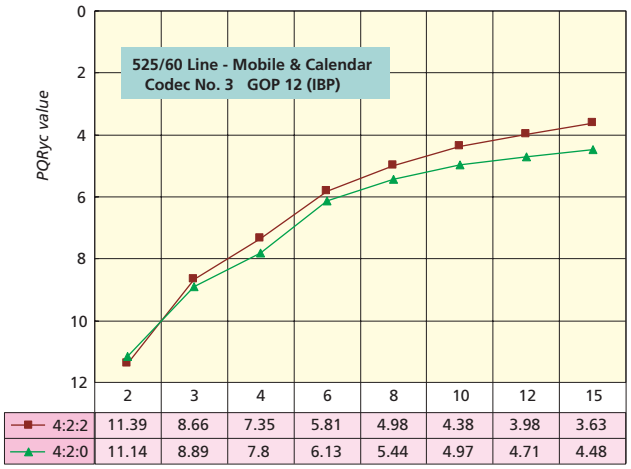
(b)



	3	4	6	8	10	12	15
4:2:2	10.29	8.57	6.79	5.62	4.87	4.31	3.71
4:2:0	10.41	8.58	6.93	5.95	5.29	4.83	4.41

Mbit/s

(c)

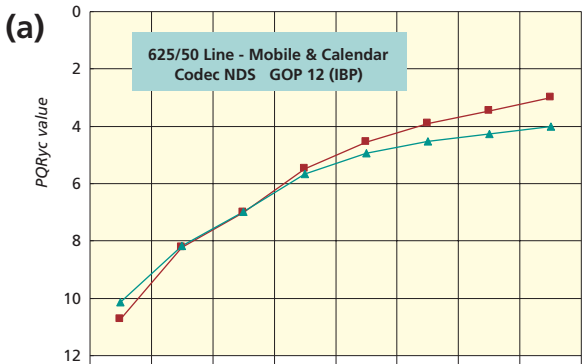


	2	3	4	6	8	10	12	15
4:2:2	11.39	8.66	7.35	5.81	4.98	4.38	3.98	3.63
4:2:0	11.14	8.89	7.8	6.13	5.44	4.97	4.71	4.48

Mbit/s

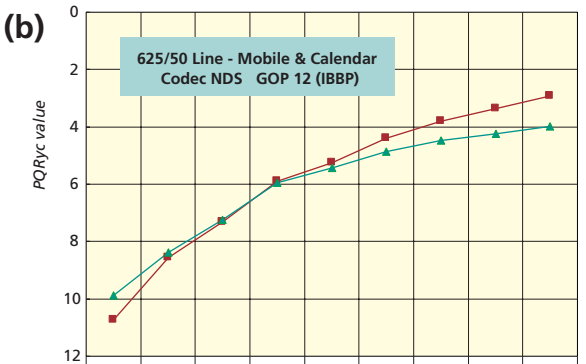
Figure 1
Test results on three different codecs.

(a)



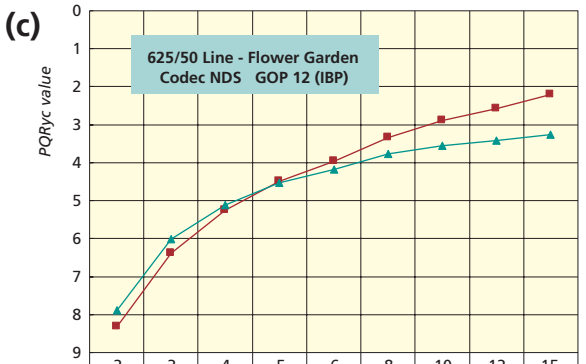
	2	3	4	6	8	10	12	15
4:2:2	10.72	8.22	7.02	5.49	4.54	3.91	3.46	3.01
4:2:0	10.15	8.16	6.99	5.66	4.95	4.53	4.27	4.00

(b)



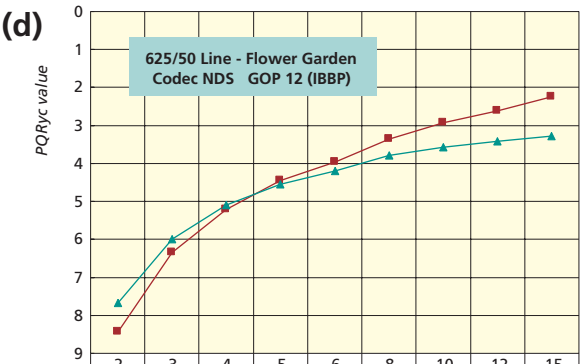
	2	3	4	5	6	8	10	12	15
4:2:2	10.73	8.55	7.32	5.89	5.24	4.40	3.79	3.37	2.93
4:2:0	9.88	8.39	7.23	5.94	5.44	4.87	4.47	4.25	3.99

(c)



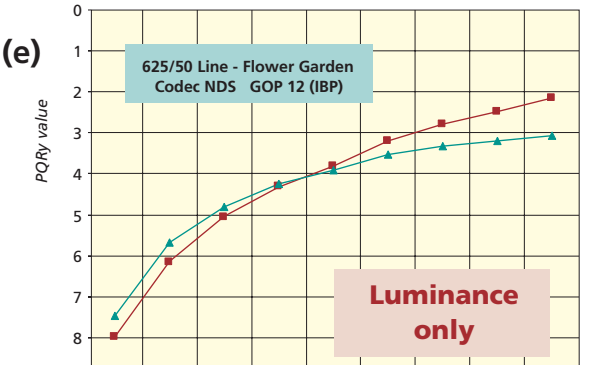
	2	3	4	5	6	8	10	12	15
4:2:2	8.31	6.39	5.25	4.49	3.96	3.33	2.89	2.57	2.21
4:2:0	7.88	6.01	5.11	4.52	4.17	3.77	3.55	3.41	3.27

(d)



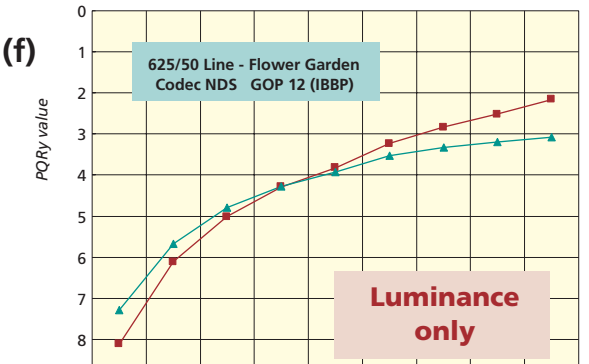
	2	3	4	5	6	8	10	12	15
4:2:2	8.43	6.35	5.21	4.46	3.97	3.35	2.93	2.61	2.24
4:2:0	7.67	5.99	5.09	4.54	4.19	3.78	3.57	3.42	3.28

(e)



	2	3	4	5	6	8	10	12	15
4:2:2	7.99	6.14	5.04	4.31	3.80	3.20	2.79	2.48	2.14
4:2:0	7.46	5.67	4.81	4.24	3.91	3.53	3.32	3.19	3.07

(f)



	2	3	4	5	6	8	10	12	15
4:2:2	8.11	6.11	5.01	4.29	3.82	3.22	2.82	2.52	2.16
4:2:0	7.28	5.67	4.79	4.27	3.93	3.53	3.33	3.20	3.07

Figure 2
Codec test results for 625-line pictures, using the Tektronix PQA analyzer.

ments. *Figs. 2(c)* and *2(d)* give the *Flower Garden* analyzis for the same two structures of the GoP.

Finally, we took the opportunity to compare the PQRy (luminance only) values with the PQRyc values. This was done only for *Flower Garden* and is shown in *Figs. 2(e)* and *2(f)*.

3.3. Preliminary conclusions

3.3.1. Cross-over value

It appears that there is no cross-over value for 525-line systems or, if it exists, it is lower than 2 Mbit/s.

For 625-line systems, the cross-over value is about 4 Mbit/s.

This difference of behaviour between the 525- and 625-line systems will be discussed in *Section 4*.

3.3.2. Interpretation of the PQR value

The use of the Tektronix PQA system has been of tremendous value in making all these tests. Standard subjective assessments according to ITU-R rules would have been much too long to perform with this kind of tests. However, the drawback when using the Tektronix equipment is how do we interpret the PQR value. This is also discussed in *Section 4*. We will simply note here that the PQRy values are systematically 0.2 units worse than the PQRyc values, as shown in *Fig. 2*.

3.3.3. Influences of the GoP

Although it was not the scope of this study, it seems that the GoP structure (IBBP or IBP) has no measurable impact on the results of the tests.

Abbreviations

4:2:2 P@ML	(MPEG-2) 4:2:2 Profile at Main Level
CBC	Canadian Broadcasting Corporation
IBC	International Broadcasting Convention
GoP	Group of pictures
ITU-R	International Telecommunication Union, Radiocommunication Sector
ITU-T	International Telecommunication Union, Telecommunication Standardization Sector
MP@ML	(MPEG-2) Main Profile at Main Level
MPEG	(ISO/IEC) Moving Picture Experts Group
NTSC	National Television System Committee (USA)
PAL	Phase alternation line
SDI	Serial digital interface
SMPTE	Society of Motion Picture and Television Engineers (USA)

Louis Cheveau *qualified as a Physics Engineer from the University of Liège, Belgium, in 1967 and obtained a Ph.D. in Physics from the University of Montreal, Canada, in 1974. That year, he joined the EBU Technical Centre in Brussels as head of the computing department and, initially, worked in the field of terrestrial television broadcasting. In 1977, the emphasis of his work changed to satellite broadcasting.*

In 1984, Dr Cheveau was detached for two years to CBC in Canada. There, he worked in International Relations with a special emphasis on satellite broadcasting and HDTV matters. In 1986, he returned to the EBU Technical Centre, this time to work on Eurovision transmissions. Since 1989, he has been Head of Transmission Technologies within the EBU Technical Department in Geneva.

Anthony Caruso *obtained a degree in Telecommunications Engineering at Buenos Aires University in 1972 and joined the development group of PYE Telecommunications Argentina, as an engineering assistant. He was transferred to PYE Telecommunications Canada in 1975 where he continued his post-graduate studies. He then joined the Engineering Department of the Canadian Broadcasting Corporation in 1981.*

Mr Caruso – a professional engineer licensed in Canada – is presently a Senior Specialist Engineer in a group within the Delivery Technology Department of CBC. He is concerned with the investigation of new digital technologies that are applicable to the contribution, distribution and transmission networks of CBC. He represents CBC in such matters on several international technical committees.

Anthony Caruso's most recent achievement was the planning and deployment of the first MPEG-2 4:2:2 Profile system to be used for the satellite / terrestrial transmission of the Winter Olympic Games (Nagano 98, Japan). He is currently responsible for the digitalization of CBC's terrestrial and satellite networks.

4. Technical considerations

In this section, we will try to analyze the reasons which could explain the most immediate results of these tests:

⇨ differences in behaviour between the 525- and 625-line systems.

⇨ the very low bit-rate value at the quality cross-over point (of 4:2:2 and 4:2:0 coding).

4.1. Analyzis of results

All the comparisons here are based on the sequence *Mobile and Calendar*, with GoP = 12, IBP structure and with the NDS codec at 15, 10 and 6 Mbit/s.

The first observation is that the PQRyc difference ("VAR") between 4:2:2 & 4:2:0 for the 625/50 system is smaller than the difference for the 525/60 system, as follows:

a) 525/60 system, Mobile & Calendar, GoP = IBP, N = 12

15 Mbits/s VAR: 1.13

10 Mbits / s VAR: 0.78

6 Mbits / s VAR: 0.51

b) 625/50 system, Mobile & Calendar, GOP = IBP, N = 12

15 Mbits / s VAR: 0.99

10 Mbits / s VAR: 0.62

6 Mbits / s VAR: 0.17

The reasons that could explain these results must be related to the fact that the PQRyc rating for 4:2:0 on the 625-line system renders a higher quality than 4:2:0 on the 525-line system (like PAL compared to the NTSC system).

On the other hand, the PQRyc variation between the 525- & 625-line systems for the 4:2:2 profile is smaller than the variation between the 525- & 625-line systems for the 4:2:0 profile, as follows:

⇨ 525 / 60 4:2:2@15 Mbits / s
PQRyc = 3.12
625 / 50 4:2:2@15 Mbits / s
PQRyc = 3.01, **VAR = 0.11**

⇨ 525 / 60 4:2:2@10 Mbits / s
PQRyc = 4.01
625 / 50 4:2:2@10 Mbits / s
PQRyc = 3.91, **VAR = 0.10**

⇨ 525 / 60 4:2:2@6 Mbits / s
PQRyc = 5.35
625 / 50 4:2:2@6 Mbits / s
PQRyc = 5.49, **VAR = – 0.14**

The PQRyc measurements at 4:2:0 on the same codec with the same parameters have a larger difference between 625 / 50 & 525 / 60, as follows:

⇨ 525 / 60 4:2:0@15 Mbits / s
PQRyc = 4.25
625 / 50 4:2:0@15 Mbits / s
PQRyc = 4.00, **VAR = 0.25**

⇨ 525 / 60 4:2:0@10 Mbits / s
PQRyc = 4.79
625 / 50 4:2:0@10 Mbits / s
PQRyc = 4.53, **VAR = 0.26**

⇨ 525 / 60 4:2:0@6 Mbits / s
PQRyc = 5.86
625 / 50 4:2:0@6 Mbits / s
PQRyc = 5.66, **VAR = 0.20**

This could explain the cross-over between 4:2:2 & 4:2:0 for the 625 / 50 system. The 4:2:0 profile renders lower PQRyc figures (higher quality) for the 625-line than for the 525-line system. This

is equivalent to pushing the 4:2:0 curve upwards and relatively closer to the 4:2:2 profile curve, which remains almost unchanged.

Another factor could be that the PQA 200 relates the subjective viewing into an "objective result". If the empirical process that relates the subjective picture quality assessment into an objective figure was based on the 525/60 system, then it cannot yield exactly the same for the 625/50 system. In other words, Tektronix should perhaps use two different algorithms (one for the 525-line system and another for the 625-line system) to relate the subjective assessments into objective results.

From the practical point of view, the difference in PQRyc numbers between the 4:2:2 & 4:2:0 profiles for the 625-line system below 5 Mbits/s are very small (less than 0.3 points) and almost negligible, subjectively speaking.

Therefore, as long as the correlation between PQR numbers and ITU-R picture-quality grading is not established, it will be difficult to define a minimum significant PQR grading differential. In our opinion, this difference cannot be less than something between 0.5 and 1.

4.2. The cross-over point

The value of the cross-over point (2 Mbit/s for 525-line systems and around 4 Mbit/s for 625-line systems) is surprisingly lower than expected. It was a common belief that this point should occur somewhere between 10 and 15 Mbit/s. The slight difference of behaviour between 525- and 625-line systems is not really significant and we need a more general explanation of this effect. To understand the basic difference between 4:2:0 and 4:2:2, we have to refer to the MPEG-2 specification, where we can find the description of the positioning of the chrominance and luminance samples (see *Fig. 3*, which is taken from ITU-T Recommendation H.262).

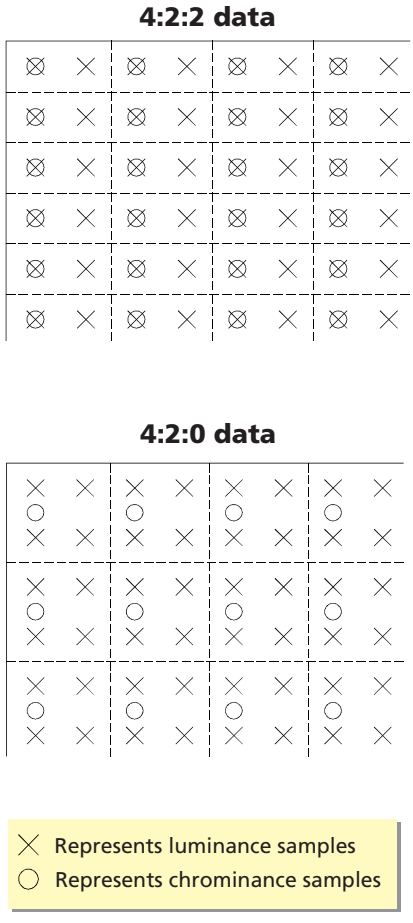


Figure 3
Positioning of chrominance and luminance samples in 4:2:2 and 4:2:0.

Looking carefully at both representations, it appears that in the 4:2:0 format, the samples of luminance and chrominance are not co-sited, contrary to the 4:2:2 format.

The main reason for not co-siting the sample is due to the request of MPEG to treat in a similar manner the chrominance samples in the frame, whether the frame is represented as a single frame-picture or two field-pictures (interlaced).

This means, in a sense, that in a 4:2:2 environment – which was the case in our tests since we used the SDI interface (SMPTE 259 M) – a double interpolation has to occur: one at the encoder to create the non co-sited sample and a second one at the decoder to recreate the co-sited sample.

The degradation caused by this interpolation has apparently been captured by the Tektronix PQA 200, which analyzes the picture, pixel by pixel.

Conclusions

Although some work remains to be done, we can already draw some definite conclusions:

- If we can trust the results given by the Tektronix PQA system, in order to achieve the same picture quality in the 4:2:0 format as in the 4:2:2 format at say 8 Mbit/s – you need to operate 4:2:0 codecs at bit-rates of 9, 10 or even 11 Mbit/s in 525-line systems, and at 10 Mbit/s in 625-line systems.

  Thus, 4:2:0 is 25% more costly to operate than 4:2:2.

- The *cross-over point* (i.e. the bit-rate at which the 4:2:0 system becomes better than 4:2:2) does not exist for 525-line systems (or is lower than 2-3 Mbit/s).

- For 625-line systems, the cross-over point seems to be around 4-5 Mbit/s.

  This result, however, does need clarification to ensure that it was not due to the measuring equipment used for the tests.

- Additional tests, using standard ITU-R subjective assessment methods, will be necessary.

Bibliography

[1] L. Cheveau: **Switching to digital**
EBU Technical Review No. 276, Summer 1998.

[2] A. Caruso, L. Cheveau and B. Flowers: **MPEG-2 4:2:2 Profile – its use for contribution/ collection and primary distribution**
EBU Technical Review No. 276, Summer 1998.

[3] A. Caruso, L. Cheveau and B. Flowers: **The use of MPEG-2 4:2:2 profile for contribution and primary distribution**
IBC 98, Amsterdam.