International Telecommunication Union

ITU-R

Radiocommunication Sector of ITU

Recommendation ITU-R BT.601-7

(03/2011)

Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios

BT Series

Broadcasting service (television)

International Telecommunication Union

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Annex 1 of Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from http://www.itu.int/ITU-R/go/patents/en where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Recommendations

(Also available online at http://www.itu.int/publ/R-REC/en)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	**Broadcasting service (television)**
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management
SNG	Satellite news gathering
TF	Time signals and frequency standards emissions
V	Vocabulary and related subjects

Note: *This ITU-R Recommendation was approved in English under the procedure detailed in Resolution ITU-R 1.*

Electronic Publication
Geneva, 2011

RECOMMENDATION ITU-R BT.601-7

Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios

(Question ITU-R 1/6)

(1982-1986-1990-1992-1994-1995-2007-2011)

Scope

This Recommendation also covers the pixel characteristics that represent a 525- or 625-line interlace digital television image.

This Recommendation specifies methods for digitally coding video signals. It includes a 13.5 MHz sampling rate for both 4:3 and 16:9 aspect ratio images with performance adequate for present transmission systems.

The ITU Radiocommunication Assembly,

considering

a) that there are clear advantages for television broadcasters and programme producers in digital studio standards which have the greatest number of significant parameter values common to 525-line and 625-line systems;

b) that a worldwide compatible digital approach will permit the development of equipment with many common features, permit operating economies and facilitate the international exchange of programmes;

c) that an extensible family of compatible digital coding standards is desirable. Members of such a family could correspond to different quality levels, different aspect ratios, facilitate additional processing required by present production techniques, and cater for future needs;

d) that a system based on the coding of components is able to meet these desirable objectives;

e) that the co-siting of samples representing luminance and colour-difference signals (or, if used, the red, green and blue signals) facilitates the processing of digital component signals, required by present production techniques,

recommends

that the following be used as a basis for digital coding standards for television studios in countries using the 525-line system as well as in those using the 625-line system.

1 Extensible family of compatible digital coding standards

1.1 The digital coding should allow the establishment and evolution of an extensible family of compatible digital coding standards. It should be possible to interface simply between any members of the family.

1.2 The digital coding should be based on the use of one luminance and two colour-difference signals (or, if used, the red, green and blue signals).

1.3 The spectral characteristics of the signals must be controlled to avoid aliasing whilst preserving the passband response. Filter specifications are shown in Appendix 2.

2 Specifications applicable to any member of the family

2.1 Sampling structures should be spatially static. This is the case, for example, for the orthogonal sampling structures specified in this Recommendation.

2.2 If the samples represent luminance and two simultaneous colour-difference signals, each pair of colour-difference samples should be spatially co-sited. If samples representing red, green and blue signals are used they should be co-sited.

2.3 The digital standard adopted for each member of the family should permit worldwide acceptance and application in operation; one condition to achieve this goal is that, for each member of the family, the number of samples per line specified for 525-line and 625-line systems shall be compatible (preferably the same number of samples per line).

2.4 In applications of these specifications, the contents of digital words are expressed in both decimal and hexadecimal forms, denoted by the suffixes "d" and "h" respectively.

To avoid confusion between 8-bit and 10-bit representations, the eight most-significant bits are considered to be an integer part while the two additional bits, if present, are considered to be fractional parts.

For example, the bit pattern 10010001 would be expressed as 145_d or 91_h, whereas the pattern 1001000101 would be expressed as 145.25_d or 91.4_h.

Where no fractional part is shown, it should be assumed to have the binary value 00.

2.5 Definition of the digital signals Y, C_R, C_B, from the primary (analogue) signals E'_R, E'_G and E'_B

This paragraph describes, with a view to defining the signals Y, C_R, C_B, the rules for construction of these signals from the gamma pre-corrected primary analogue signals E'_R, E'_G and E'_B. The signals are constructed by following the three stages described in § 2.5.1,2.5.2 and2.5.3. The method is given as an example, and in practice other methods of construction from these primary signals or other analogue or digital signals may produce identical results. An example is given in § 2.5.4.

2.5.1 Construction of luminance (E'_Y) and colour-difference ($E'_R - E'_Y$) and ($E'_B - E'_Y$) signals

The construction of luminance and colour-difference signals is as follows:

$$E'_Y = 0.299\,E'_R + 0.587\,E'_G + 0.114\,E'_B$$

then:

$$(E'_R - E'_Y) = E'_R - 0.299\,E'_R - 0.587\,E'_G - 0.114\,E'_B = 0.701\,E'_R - 0.587\,E'_G - 0.114\,E'_B$$

and

$$(E'_B - E'_Y) = E'_B - 0.299\,E'_R - 0.587\,E'_G - 0.114\,E'_B = -0.299\,E'_R - 0.587\,E'_G + 0.886\,E'_B$$

Taking the signal values as normalized to unity (e.g. 1.0 V maximum levels), the values obtained for white, black and the saturated primary and complementary colours are shown in Table 1.

TABLE 1

Normalized signal values

Condition	E'_R	E'_G	E'_B	E'_Y	$E'_R - E'_Y$	$E'_B - E'_Y$
White Black	1.0 0	1.0 0	1.0 0	1.0 0	0 0	0 0
Red Green Blue	1.0 0 0	0 1.0 0	0 0 1.0	0.299 0.587 0.114	0.701 −0.587 −0.114	−0.299 −0.587 0.886
Yellow Cyan Magenta	1.0 0 1.0	1.0 1.0 0	0 1.0 1.0	0.886 0.701 0.413	0.114 −0.701 0.587	−0.886 0.299 0.587

2.5.2 Construction of re-normalized colour-difference signals (E'_{C_R} and E'_{C_B})

Whilst the values for E'_Y have a range of 1.0 to 0, those for $(E'_R - E'_Y)$ have a range of +0.701 to −0.701 and for $(E'_B - E'_Y)$ a range of +0.886 to −0.886. To restore the signal excursion of the colour-difference signals to unity (i.e. +0.5 to −0.5), re-normalized red and blue colour-difference signals E'_{C_R} and E'_{C_B} respectively can be calculated as follows:

$$E'_{C_R} = \frac{E'_R - E'_Y}{1.402} = \frac{0.701E'_R - 0.587E'_G - 0.114E'_B}{1.402}$$

and

$$E'_{C_B} = \frac{E'_B - E'_Y}{1.772} = \frac{-0.299E'_R - 0.587E'_G + 0.886E'_B}{1.772}$$

The symbols E'_{C_R} and E'_{C_B} will be used only to designate re-normalized colour-difference signals, i.e. having the same nominal peak-to-peak amplitude as the luminance signal E'_Y thus selected as the reference amplitude.

2.5.3 Quantization

In the case of a uniformly-quantized 8-bit or 10-bit binary encoding, 2^8 or 2^{10}, i.e. 256 or 1 024, equally spaced quantization levels are specified, so that the range of the binary numbers available is from 0000 0000 to 1111 1111 (00 to FF in hexadecimal notation) or 0000 0000 00 to 1111 1111 11 (00.0_h to $FF.C_h$ in hexadecimal notation), the equivalent decimal numbers being 0.00_d to 255.75_d, inclusive.

In this Recommendation, levels 0.00_d and 255.75_d are reserved for synchronization data, while levels 1.00_d to 254.75_d are available for video.

Given that the luminance signal is to occupy only 220 (8-bit) or 877 (10-bit) levels, to provide working margins, and that black is to be at level 16.00_d, the decimal value of the quantized luminance signal, Y, is:

$$Y = \text{int}\{(219E'_Y + 16) \times D\}/D$$

where D takes either the value 1 or 4, corresponding to 8-bit and 10-bit quantization respectively. The operator int( ) returns the value of 0 for fractional parts in the range of 0 to 0.4999 … and +1 for fractional parts in the range 0.5 to 0.999 ..., i.e. it rounds up fractions above 0.5.

Similarly, given that the colour-difference signals are to occupy 225 (8-bit) or 897 (10-bit) levels and that the zero level is to be level 128.00_d, the decimal values of the quantized colour-difference signals, C_R and C_B, are:

$$C_R = \text{int}\{(224E'_{C_R} + 128) \times D\}/D$$

and

$$C_B = \text{int}\{(224E'_{C_B} + 128) \times D\}/D$$

The digital equivalents are termed Y, C_R and C_B.

2.5.4 Construction of Y, C_R, C_B via quantization of E'_R, E'_G, E'_B

In the case where the components are derived directly from the gamma pre-corrected component signals E'_R, E'_G, E'_B, or directly generated in digital form, then the quantization and encoding shall be equivalent to:

$$E'_{R_D} \text{(in digital form)} = \text{int}\{(219E'_R + 16) \times D\}/D$$

$$E'_{G_D} \text{ (in digital form)} = \text{int}\{(219E'_G + 16) \times D\}/D$$

$$E'_{B_D} \text{ (in digital form)} = \text{int}\{(219E'_B + 16) \times D\}/D$$

Then:

$$Y = \text{int}\{(0.299E'_{R_D} + 0.587E'_{G_D} + 0.114E'_{B_D}) \times D\}/D$$

$$\approx \text{int}\left\{\left(\frac{k'_{Y1}}{2^m}E'_{R_D} + \frac{k'_{Y2}}{2^m}E'_{G_D} + \frac{k'_{Y3}}{2^m}E'_{B_D}\right) \times D\right\}/D$$

$$C_R = \text{int}\left[\left\{\left(\frac{0.701E'_{R_D} - 0.587E'_{G_D} - 0.114E'_{B_D}}{1.402}\right)\frac{224}{219} + 128\right\} \times D\right]/D$$

$$\approx \text{int}\left[\left\{\left(\frac{k'_{CR1}}{2^m}E'_{R_D} + \frac{k'_{CR2}}{2^m}E'_{G_D} + \frac{k'_{CR3}}{2^m}E'_{B_D}\right) + 128\right\} \times D\right]/D$$

$$C_B = \text{int}\left[\left\{\left(\frac{-0.299E'_{R_D} - 0.587E'_{G_D} + 0.886E'_{B_D}}{1.772}\right)\frac{224}{219} + 128\right\} \times D\right] / D$$

$$\approx \text{int}\left[\left\{\left(\frac{k'_{CB1}}{2^m}E'_{R_D} + \frac{k'_{CB2}}{2^m}E'_{G_D} + \frac{k'_{CB3}}{2^m}E'_{B_D}\right) + 128\right\} \times D\right] / D$$

where k' and m denote the integer coefficients and the bit-lengths of the integer coefficients, respectively. The integer coefficients of luminance and colour-difference equations should be derived as per Annex 2.

TABLE 2

Integer coefficients of luminance and colour-difference equations

Coefficient bits	**Denominator**	**Luminance *Y***			**Colour-difference C_R**			**Colour-difference C_B**		
m	2^m	k'_{Y1}	k'_{Y2}	k'_{Y3}	k'_{CR1}	k'_{CR2}	k'_{CR3}	k'_{CB1}	k'_{CB2}	k'_{CB3}
8	256	77	150	29	131	−110	−21	−44	−87	131
9	512	153	301	58	262	−219	−43	−88	**−174**	262
10	1 024	306	601	117	524	−439	−85	−177	−347	524
11	2 048	612	1 202	**234**	1 047	−877	−170	−353	−694	1 047
12	4 096	1 225	2 404	467	2 095	−1 754	−341	−707	−1 388	2 095
13	8 192	2 449	4 809	934	**4 189**	−3 508	−681	−1 414	−2 776	4 190
14	16 384	4 899	9 617	1 868	8 379	−7 016	−1 363	−2 828	−5 551	8 379
15	32 768	9 798	19 235	**3 735**	16 758	−14 033	−2 725	−5 655	−11 103	16 758
16	65 536	19 595	38 470	7 471	33 516	−28 066	**−5 450**	−11 311	−22 205	33 516

NOTE 1 – The bold values indicate that the values are modified from the nearest integer values by the optimization.

To obtain the 4:2:2 components Y, C_R, C_B, low-pass filtering and sub-sampling must be performed on the 4:4:4 C_R, C_B signals described above. Note should be taken that slight differences could exist between C_R, C_B components derived in this way and those derived by analogue filtering prior to sampling.

2.5.5 Limiting of Y, C_R, C_B signals

Digital coding in the form of Y, C_R, C_B signals can represent a substantially greater gamut of signal values than can be supported by the corresponding ranges of R, G, B signals. Because of this it is possible, as a result of electronic picture generation or signal processing, to produce Y, C_R, C_B signals which, although valid individually, would result in out-of-range values when converted to R, G, B. It is both more convenient and more effective to prevent this by applying limiting to the Y, C_R, C_B signals than to wait until the signals are in R, G, B form. Also, limiting can be applied in a way that maintains the luminance and hue values, minimizing the subjective impairment by sacrificing only saturation.

2.6 Colour and opto-electronic transfer characteristic[1]

Item	Characteristics				
	Parameter	625		525	
2.6.1	Chromaticity coordinates, CIE 1931[(1)]	x	y	x	y
	Primaries Red	0.640	0.330	0.630	0.340
	Green	0.290	0.600	0.310	0.595
	Blue	0.150	0.060	0.155	0.070
2.6.2	Assumed chromaticity for equal primary signals – Reference white	D_{65}			
		x		y	
	$E_R = E_G = E_B$	0.3127		0.3290	
2.6.3	Opto-electronic transfer characteristics before non-linear precorrection	Assumed linear			
2.6.4	Overall opto-electronic transfer characteristic at source	$E = (1.099\ L^{0.45} - 0.099)$ for $1.00 \geq L \geq 0.018$ $E = 4.500\ L$ for $0.018 > L \geq 0$ where: L: luminance of the image $0 \leq L \leq 1$ for conventional colorimetry E: corresponding electrical signal.			

(1) Chromaticity coordinates specified are those currently used by 625-line and 525-line conventional systems.

3 Family members

The following family members are defined:

– 4:2:2 for 4:3 aspect ratio, and for wide-screen 16:9 aspect ratio systems when it is necessary to keep the same analogue signal bandwidth and digital rates for both aspect ratios.

– 4:4:4[2] for 4:3 and 16:9 aspect ratio systems with higher colour resolution.

Annex 1

Encoding parameters for members of the family

1 Encoding parameter values for the 4:2:2 member of the family

The specification (see Table 3) applies to the 4:2:2 member of the family, to be used for the standard digital interface between main digital studio equipment and for international programme exchange of 4:3 aspect ratio digital television or wide-screen 16:9 aspect ratio digital television when it is necessary to keep the same analogue signal bandwidth and digital rates.

---

1 It is recognized that a practice is now sometimes used by which, when programs produced in HDTV are release in SDTV, their HDTV pixel map is re-mapped onto the SDTV pixel map without changing the colorimetry of the original program.

2 In the 4:4:4 members of the family the sampled signals may be luminance and colour difference signals (or, if used, red, green and blue signals).

TABLE 3

Parameters	525-line, 60 field/s systems	625-line, 50 field/s systems
1. Coded signals: Y, C_R, C_B	These signals are obtained from gamma pre-corrected signals, namely: E'_Y, $E'_R - E'_Y$, $E'_B - E'_Y$ (see § 2.5)	
2. Number of samples per total line: – luminance signal (Y) – each colour-difference signal (C_R, C_B)	 858 429	 864 432
3. Sampling structure	Orthogonal, line, field and frame repetitive. C_R and C_B samples co-sited with odd (1st, 3rd, 5th, etc.) Y samples in each line	
4. Sampling frequency: – luminance signal – each colour-difference signal	 13.5 MHz 6.75 MHz The tolerance for the sampling frequencies should coincide with the tolerance for the line frequency of the relevant colour television standard	
5. Form of coding	Uniformly quantized PCM, 8 or 10 bits per sample, for the luminance signal and each colour-difference signal	
6. Number of samples per digital active line: – luminance signal – each colour-difference signal	 720 360	
7. Analogue-to-digital horizontal timing relationship: – from end of digital active line to O_H	16 luminance clock periods	12 luminance clock periods
8. Correspondence between video signal levels and quantization levels: – scale – luminance signal – each colour-difference signal	(See § 2.4) (Values are decimal) 0.00_d to 255.75_d 220 (8-bit) or 877 (10-bit) quantization levels with the black level corresponding to level 16.00_d and the peak white level corresponding to level 235.00_d. The signal level may occasionally excurse beyond level 235.00_d or below level 16.00_d. 225 (8-bit) or 897 (10-bit) quantization levels in the centre part of the quantization scale with zero signal corresponding to level 128.00_d. The signal level may occasionally excurse beyond level 240.00_d or below level 16.00_d.	
9. Code-word usage	Code words corresponding to quantization levels 0.00_d and 255.75_d are used exclusively for synchronization. Levels 1.00_d to 254.75_d are available for video. When 8-bit words are treated in 10-bit system, two LSBs of zeros are to be appended to the 8-bit words.	

2 Encoding parameter values for the 4:4:4 member of the family

The specifications given in Table 4 apply to the 4:4:4 member of the family suitable for television source equipment and high-quality video signal processing applications.

TABLE 4

Parameters	525-line, 60 field/s systems	625-line, 50 field/s systems
1. Coded signals: Y, C_R, C_B or R, G, B	These signals are obtained from gamma pre-corrected signals, namely: E'_Y, $E'_R - E'_Y$, $E'_B - E'_Y$ or E'_R, E'_G, E'_B	
2. Number of samples per total line for each signal	858	864
3. Sampling structure	Orthogonal, line, field and frame repetitive. The three sampling structures to be coincident and coincident also with the luminance sampling structure of the 4:2:2 member	
4. Sampling frequency for each signal	13.5 MHz	
5. Form of coding	Uniformly quantized PCM, 8 or 10 bits per sample	
6. Duration of the digital active line expressed in number of samples	720	
7. Analogue-to-digital horizontal timing relationship: – from end of digital active line to O_H	16 clock periods	12 clock periods
8. Correspondence between video signal levels and quantization level for each sample: – scale – R, G, B or luminance signal(1) – each colour-difference signal(1)	(See § 2.4) (Values are decimal) 0.00_d to 255.75_d 220 (8-bit) or 877 (10-bit) quantization levels with the black level corresponding to level 16.00_d and the peak white level corresponding to level 235.00_d. The signal level may occasionally excurse beyond level 235.00_d or below level 16.00_d. 225 (8-bit) or 897 (10-bit) quantization levels in the centre part of the quantization scale with zero signal corresponding to level 128.00_d. The signal level may occasionally excurse beyond level 240.00_d or below level 16.00_d.	
9. Code-word usage	Code words corresponding to quantization levels 0.00_d and 255.75_d are used exclusively for synchronization. Levels 1.00_d to 254.75_d are available for video. When 8-bit words are treated in 10-bit system, two LSBs of zeros are to be appended to the 8-bit words.	

(1) If used.

Appendix 1 to Annex 1

Definition of signals used in the digital coding standards

1 Relationship of digital active line to analogue sync reference

The relationship between the digital active line luminance samples and the analogue synchronizing reference is shown in:

– Figure 1 for 625-line.

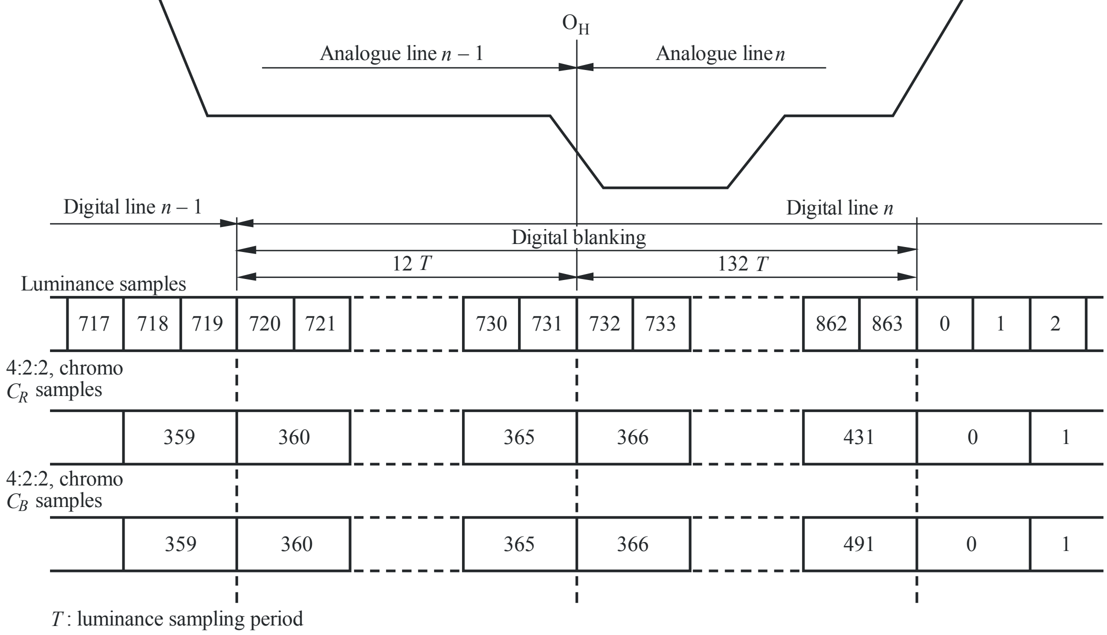
– Figure 2 for 525-line.

In the Figures, the sampling point occurs at the commencement of each block.

The respective numbers of colour-difference samples in the 4:2:2 family can be obtained by dividing the number of luminance samples by two. The (12,132), and (16,122) were chosen symmetrically to dispose the digital active line about the permitted variations. They do not form part of the digital line specification and relate only to the analogue interface.

FIGURE 1

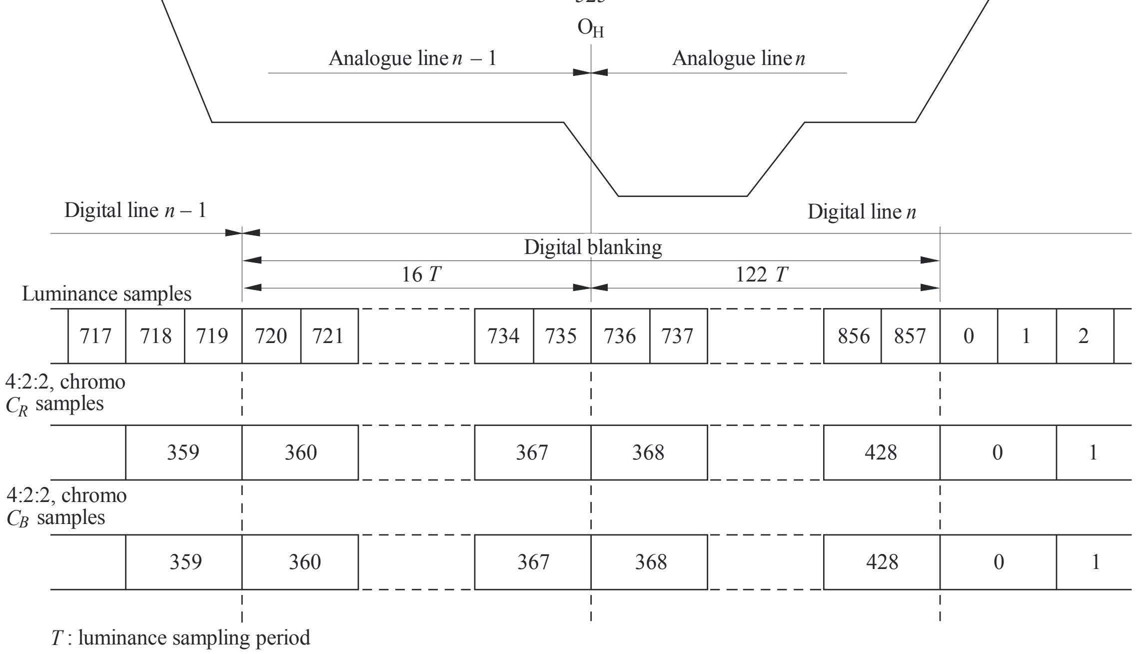
16:9 or 4:3 at 13.5 MHz
625



BT.601-01

FIGURE 2

16:9 or 4:3 at 13.5 MHz
525



BT.601-02

Appendix 2
to Annex 1

Filtering characteristics

1 Some guidance on the practical implementation of the filters

In the proposals for the filters used in the encoding and decoding processes, it has been assumed that, in the post-filters which follow digital-to-analogue conversion, correction for the ($\sin x/x$) characteristic is provided. The passband tolerances of the filter plus ($\sin x/x$) corrector plus the theoretical ($\sin x/x$) characteristic should be the same as given for the filters alone. This is most easily achieved if, in the design process, the filter, ($\sin x/x$) corrector and delay equalizer are treated as a single unit.

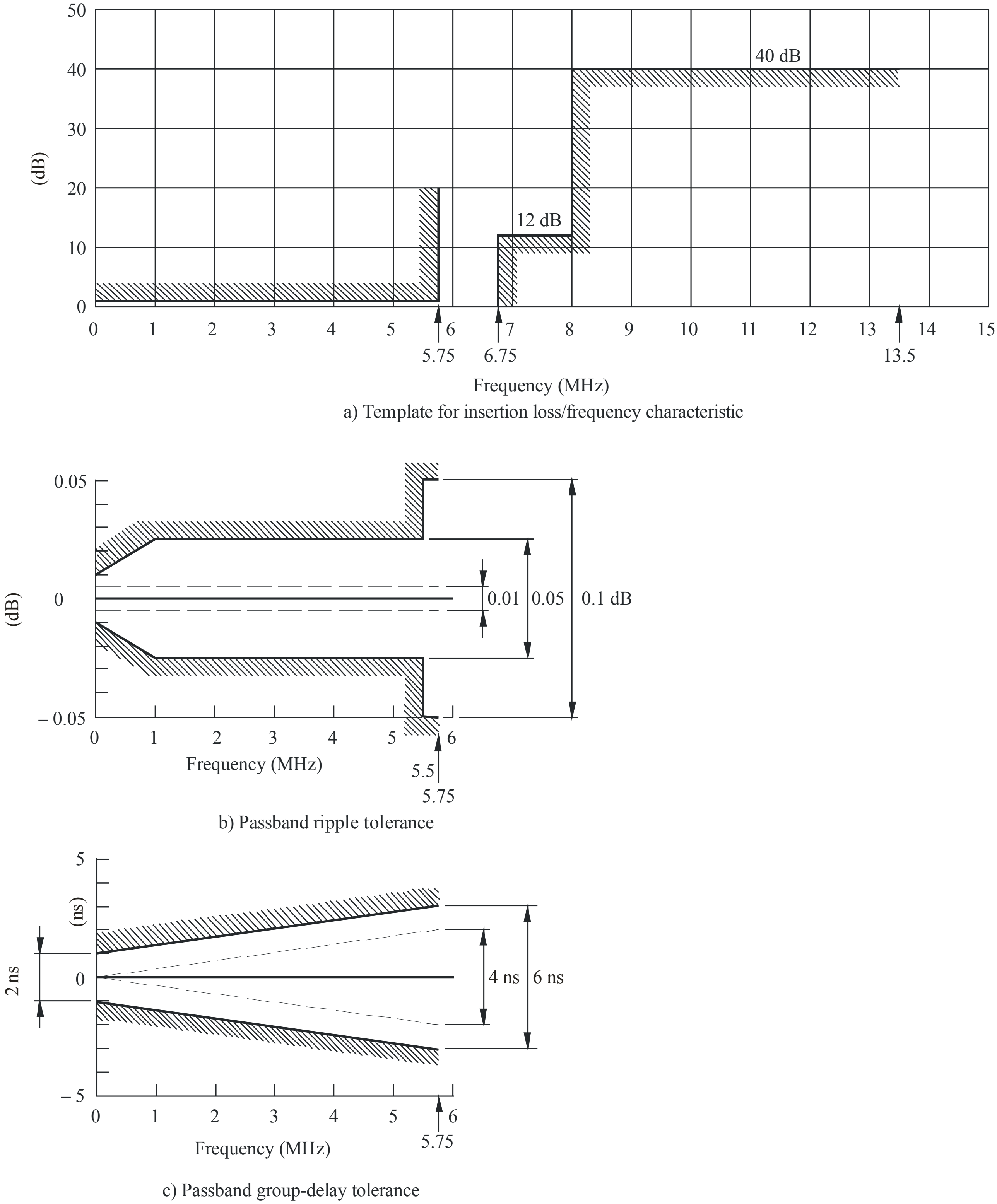
The total delays due to filtering and encoding the luminance and colour-difference components should be the same. The delay in the colour-difference filter (Figs 4a) and 4b)) is typically double that of the luminance filter (Figs 3a) and 3b)). As it is difficult to equalize these delays using analogue delay networks without exceeding the passband tolerances, it is recommended that the bulk of the delay differences (in integral multiples of the sampling period) should be equalized in the digital domain. In correcting for any remainder, it should be noted that the sample-and-hold circuit in the decoder introduces a flat delay of one half a sampling period.

The passband tolerances for amplitude ripple and group delay are recognized to be very tight. Present studies indicate that it is necessary so that a significant number of coding and decoding operations in cascade may be carried out without sacrifice of the potentially high quality of the 4:2:2 coding standard. Due to limitations in the performance of currently available measuring equipment, manufacturers may have difficulty in economically verifying compliance with the tolerances of individual filters on a production basis. Nevertheless, it is possible to design filters so that the specified characteristics are met in practice, and manufacturers are required to make every effort in the production environment to align each filter to meet the given templates.

The specifications given in Appendix 2 were devised to preserve as far as possible the spectral content of the Y, C_R, C_B signals throughout the component signal chain. It is recognized, however, that the colour-difference spectral characteristic must be shaped by a slow roll-off filter inserted at picture monitors, or at the end of the component signal chain.

FIGURE 3

Filter template for a luminance, RGB or 4:4:4 colour-difference signal

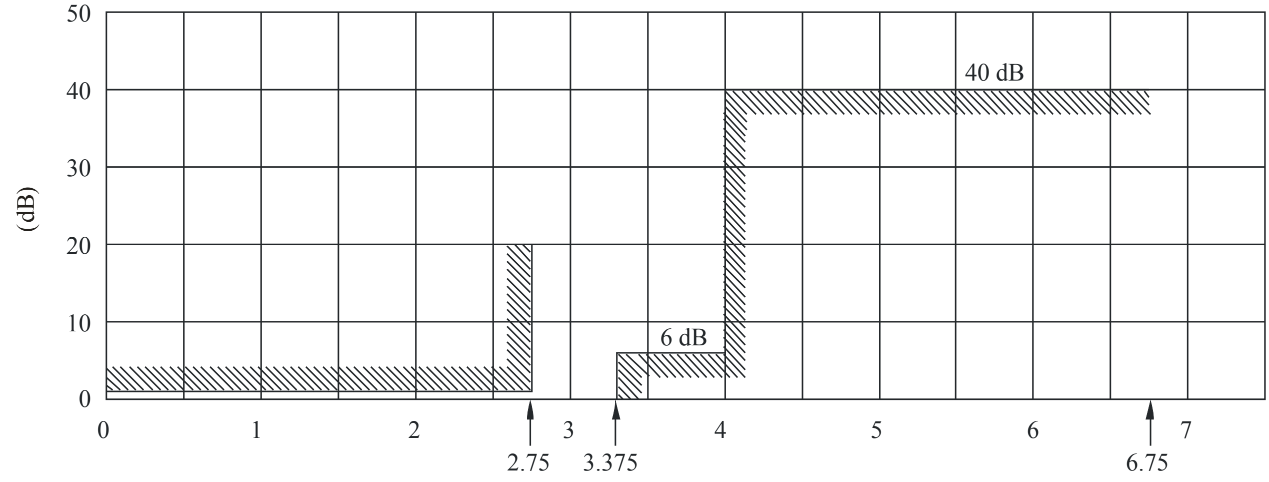


Note 1 – The lowest indicated values in b) and c) are for 1 kHz (instead of 0 MHz).

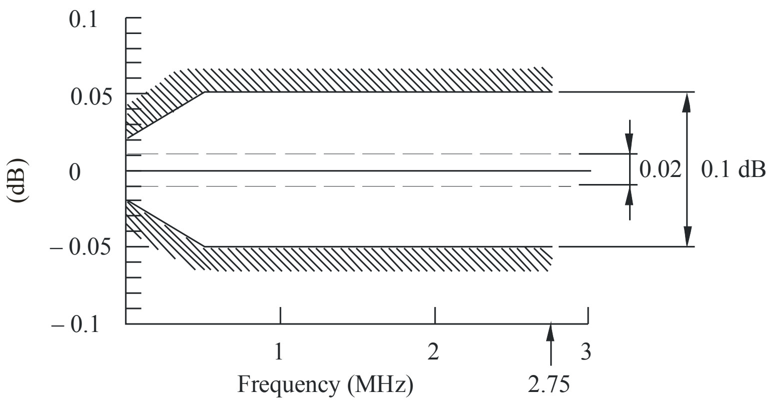
BT.601-03

FIGURE 4

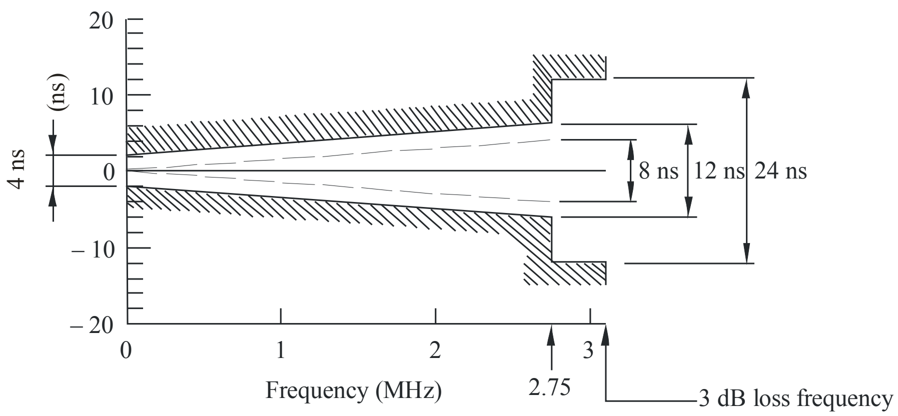
Filter template for a 4:2:2 colour-difference signal



a) Template for insertion loss/frequency characteristic



b) Passband ripple tolerance



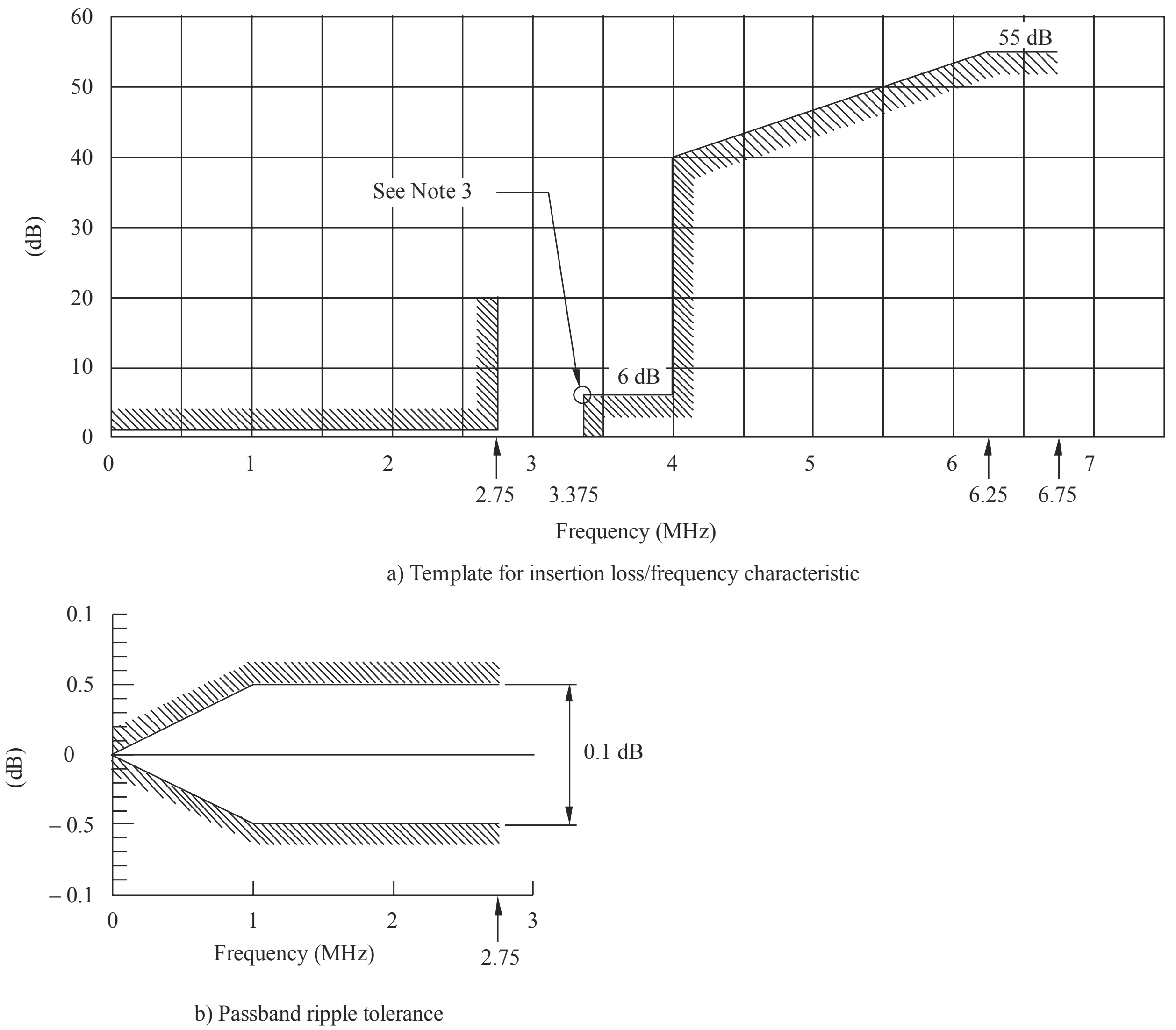
c) Passband group-delay tolerance

Note 1 – The lowest indicated values in b) and c) are for 1 kHz (instead of 0 MHz).

BT.601-04

FIGURE 5

Digital filter template for sampling-rate conversion from 4:4:4 to 4:2:2 colour difference signals



Notes to Figs 3, 4 and 5:

Note 1 – Ripple and group delay are specified relative to their values at 1 kHz. The full lines are practical limits and the dashed lines give suggested limits for the theoretical design.

Note 2 – In the digital filter, the practical and design limits are the same. The delay distortion is zero, by design.

Note 3 – In the digital filter (Fig. 5), the amplitude/frequency characteristic (on linear scales) should be skew-symmetrical about the half-amplitude point, which is indicated on the figure.

Note 4 – In the proposals for the filters used in the encoding and decoding processes, it has been assumed that, in the post-filters which follow digital-to-analogue conversion, correction for the (sin *x*/*x*) characteristic of the sample-and-hold circuits is provided.

BT.601-05

Annex 2

Derivation of integer coefficients of luminance and colour-difference equations for the conventional colour gamut system

Digital systems may introduce computation errors in the luminance and colour-difference signals due to the finite bit-length of the equation coefficients. Also, digital luminance and colour-difference signals may take slightly different values depending on the signal processing sequence, i.e. the discrepancy between signals quantized after analogue matrixing and signals digitally matrixed after quantization of RGB signals. To minimize such errors and discrepancies, the integer coefficients for the digital equations should be optimized. The optimization procedure and the resultant integer coefficients for several bit-lengths are given in the following.

1 Digital equations

In the following, m and n denote the bit-lengths of the integer coefficients and digital signals, respectively.

The digital luminance equation for the conventional colour gamut system is described as follows:

$$D'_Y = \mathrm{INT}\left[\,0.299\,D'_R + 0.587\,D'_G + 0.114\,D'_B\right] \tag{1}$$

$$= \mathrm{INT}\left[\frac{r'_{Y1}}{2^m}D'_R + \frac{r'_{Y2}}{2^m}D'_G + \frac{r'_{Y3}}{2^m}D'_B\right] \tag{2}$$

$$\approx \mathrm{INT}\left[\frac{k'_{Y1}}{2^m}D'_R + \frac{k'_{Y2}}{2^m}D'_G + \frac{k'_{Y3}}{2^m}D'_B\right] \tag{3}$$

where r' and k' denote the real values of the coefficient and the integer coefficients, respectively, given below:

$$r'_{Y1} = 0.299 \times 2^m \qquad k'_{Y1} = \mathrm{INT}\left[r'_{Y1}\right]$$

$$r'_{Y2} = 0.587 \times 2^m \qquad k'_{Y2} = \mathrm{INT}\left[r'_{Y2}\right]$$

$$r'_{Y3} = 0.114 \times 2^m \qquad k'_{Y3} = \mathrm{INT}\left[r'_{Y3}\right]$$

The digital colour-difference equations for the conventional colour gamut system are described as follows:

$$D'_{CB} = \mathrm{INT}\left[\frac{-0.299\,D'_R - 0.587\,D'_G + 0.886\,D'_B}{1.772} \times \frac{224}{219} + 2^{n-1}\right] \tag{4}$$

$$= \mathrm{INT}\left[\frac{r'_{CB1}}{2^m}D'_R + \frac{r'_{CB2}}{2^m}D'_G + \frac{r'_{CB3}}{2^m}D'_B + 2^{n-1}\right] \tag{5}$$

$$\approx \mathrm{INT}\left[\frac{k'_{CB1}}{2^m}D'_R + \frac{k'_{CB2}}{2^m}D'_G + \frac{k'_{CB3}}{2^m}D'_B + 2^{n-1}\right] \tag{6}$$

$$D'_{CR} = \text{INT}\left[\frac{0.701\,D'_R - 0.587\,D'_G - 0.114\,D'_B}{1.402} \times \frac{224}{219} + 2^{n-1}\right] \tag{7}$$

$$= \text{INT}\left[\frac{r'_{CR1}}{2^m} D'_R + \frac{r'_{CR2}}{2^m} D'_G + \frac{r'_{CR3}}{2^m} D'_B + 2^{n-1}\right] \tag{8}$$

$$\approx \text{INT}\left[\frac{k'_{CR1}}{2^m} D'_R + \frac{k'_{CR2}}{2^m} D'_G + \frac{k'_{CR3}}{2^m} D'_B + 2^{n-1}\right] \tag{9}$$

where:

$$r'_{CB1} = -\frac{0.299}{1.772} \times \frac{224}{219} \times 2^m \qquad k'_{CB1} = \text{INT}\left[r'_{CB1}\right]$$

$$r'_{CB2} = -\frac{0.587}{1.772} \times \frac{224}{219} \times 2^m \qquad k'_{CB2} = \text{INT}\left[r'_{CB2}\right]$$

$$r'_{CB3} = \frac{0.886}{1.772} \times \frac{224}{219} \times 2^m \qquad k'_{CB3} = \text{INT}\left[r'_{CB3}\right]$$

$$r'_{CR1} = \frac{0.701}{1.402} \times \frac{224}{219} \times 2^m \qquad k'_{CR1} = \text{INT}\left[r'_{CR1}\right]$$

$$r'_{CR2} = -\frac{0.587}{1.402} \times \frac{224}{219} \times 2^m \qquad k'_{CR2} = \text{INT}\left[r'_{CR2}\right]$$

$$r'_{CR3} = -\frac{0.114}{1.402} \times \frac{224}{219} \times 2^m \qquad k'_{CR3} = \text{INT}\left[r'_{CR3}\right]$$

2 Optimization procedure

Equation (3) shows the digitally matrixed luminance signal which includes computation errors due to the finite bit-length of the integer coefficients. When the coefficient bit-length is increased, the argument (the value in [ ]) of equation (3) gets close to that of equation (2), resulting in the reduced errors or discrepancies between the equations. Therefore, the difference between the arguments of equations (2) and (3) can be regarded as a measure of the integer coefficient optimization. As the difference of arguments depends on input RGB signals, "Least Square Error" optimization is defined, in which the integer coefficients are adjusted in such a way that the sum of the squared difference over all inputs falls into the minimum value, that is, the value of equation (10) is minimized.

$$\varepsilon'_Y = \sum_{\text{for all } RGB} \left\{\left(\frac{k'_{Y1}}{2^m} D'_R + \frac{k'_{Y2}}{2^m} D'_G + \frac{k'_{Y3}}{2^m} D'_B\right) - \left(\frac{r'_{Y1}}{2^m} D'_R + \frac{r'_{Y2}}{2^m} D'_G + \frac{r'_{Y3}}{2^m} D'_B\right)\right\}^2 \tag{10}$$

In addition to providing the minimum r.m.s. errors, this LSE optimization automatically minimizes the peak error that takes place at a particular input colour (a particular combination of input RGB signals), as well as the discrepancy between different signal processing sequences (analogue-matrixing and digital-matrixing).

The optimization procedure is as follows:

Step 1: For the initial value of each integer coefficient r'_{Yj} (j = 1, 2, 3), take the nearest integer to the real value of the coefficient r'_{Yj};

Step 2: With the initial integer coefficients, calculate the r.m.s. errors or the squared difference sum (equation (10)) over the input RGB signal range, e.g., 16 through 235 for an 8-bit system (a simple calculation method without using summation is described in § 3);

Step 3: Examine the r.m.s. errors when increasing/decreasing each integer coefficient by one. 27 ($= 3^3$) combinations must be evaluated in total, because each coefficient can take three values, i.e. increased, decreased and unchanged from the initial value.

Step 4: Select the combination of the coefficients that gives the minimum r.m.s. error. This combination is the resultant optimized one.

The same procedure is applied for the colour-difference equations, using equations (11) and (12).

$$\varepsilon'_{CB} = \sum_{\text{for all } RGB} \left\{ \left(\frac{k'_{CB1}}{2^m} D'_R + \frac{k'_{CB2}}{2^m} D'_G + \frac{k'_{CB3}}{2^m} D'_B + 2^{n-1} \right) - \left(\frac{r'_{CB1}}{2^m} D'_R + \frac{r'_{CB2}}{2^m} D'_G + \frac{r'_{CB3}}{2^m} D'_B + 2^{n-1} \right) \right\}^2 \tag{11}$$

$$\varepsilon'_{CR} = \sum_{\text{for all } RGB} \left\{ \left(\frac{k'_{CR1}}{2^m} D'_R + \frac{k'_{CR2}}{2^m} D'_G + \frac{k'_{CR3}}{2^m} D'_B + 2^{n-1} \right) - \left(\frac{r'_{CR1}}{2^m} D'_R + \frac{r'_{CR2}}{2^m} D'_G + \frac{r'_{CR3}}{2^m} D'_B + 2^{n-1} \right) \right\}^2 \tag{12}$$

3 Simple calculation method for squared difference sum

By expressing the difference between integer and real coefficients value as $\delta ij = k'ij - r'ij$, and the digital RGB signals as Xj, the sum of the squared differences of equations (10) to (12) can be written as the following:

$$\varepsilon'_i = \frac{1}{2^m} \sum_{X_1=L}^{H} \sum_{X_2=L}^{H} \sum_{X_3=L}^{H} \left(\delta_{i1} X_1 + \delta_{i2} X_2 + \delta_{i3} X_3 \right)^2 \tag{13}$$

where L and H denote the lower and upper boundaries of the input signal range, respectively, for which the integer coefficients are to be optimized.

As L and H are constant in the digital system under consideration, the summations for Xj are also constant. Then equation (13) can be expressed as a function only of δ_{ij}.

$$\varepsilon'_i = \frac{1}{2^m} \left\{ N_1 \left(\delta_{i1}^2 + \delta_{i2}^2 + \delta_{i3}^2 \right) + 2N_2 \left(\delta_{i1}\delta_{i2} + \delta_{i2}\delta_{i3} + \delta_{i3}\delta_{i1} \right) \right\} \tag{14}$$

where:

$$N_1 = \sum_{X_2=L}^{H} \sum_{X_3=L}^{H} \left(\sum_{X_1=L}^{H} X_1^2 \right) = \sum_{X_1=L}^{H} \sum_{X_3=L}^{H} \left(\sum_{X_2=L}^{H} X_1^2 \right) = \sum_{X_1=L}^{H} \sum_{X_2=L}^{H} \left(\sum_{X_3=L}^{H} X_1^2 \right)$$

$$= (H - L + 1)^2 \ \{H(H+1)\,(2H+1)/6 - (L-1)\,L(2L-1)/6\}$$

$$N_2 = \sum_{X_3=L}^{H} \left(\sum_{X_1=L}^{H} \sum_{X_2=L}^{H} X_1 X_2 \right) = \sum_{X_1=L}^{H} \left(\sum_{X_2=L}^{H} \sum_{X_3=L}^{H} X_2 X_3 \right) = \sum_{X_2=L}^{H} \left(\sum_{X_3=L}^{H} \sum_{X_1=L}^{H} X_3 X_1 \right)$$

$$= (H - L + 1)\,\{H(H+1)/2 - (L-1)\,L/2\}^2$$

Thus the calculation of r.m.s. errors or equations (10) to (12) can be simply performed by equation (14).