



INTERNATIONAL TELECOMMUNICATION UNION

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

J.88

(09/99)

SERIES J: TRANSMISSION OF TELEVISION, SOUND
PROGRAMME AND OTHER MULTIMEDIA SIGNALS

Digital transmission of television signals

**Transmission of enhanced definition television
signals over digital links**

ITU-T Recommendation J.88

(Previously CCITT Recommendation)

ITU-T J-SERIES RECOMMENDATIONS

TRANSMISSION OF TELEVISION, SOUND PROGRAMME AND OTHER MULTIMEDIA SIGNALS

General Recommendations	J.1-J.9
General specifications for analogue sound-programme transmission	J.10-J.19
Performance characteristics of analogue sound-programme circuits	J.20-J.29
Equipment and lines used for analogue sound-programme circuits	J.30-J.39
Digital encoders for analogue sound-programme signals	J.40-J.49
Digital transmission of sound-programme signals	J.50-J.59
Circuits for analogue television transmission	J.60-J.69
Analogue television transmission over metallic lines and interconnection with radio-relay links	J.70-J.79
Digital transmission of television signals	J.80-J.89
Ancillary digital services for television transmission	J.90-J.99
Operational requirements and methods for television transmission	J.100-J.109
Interactive systems for digital television distribution	J.110-J.129
Transport of MPEG-2 signals on packetised networks	J.130-J.139
Measurement of the quality of service	J.140-J.149
Digital television distribution through local subscriber networks	J.150-J.159

For further details, please refer to ITU-T List of Recommendations.

ITU-T RECOMMENDATION J.88

TRANSMISSION OF ENHANCED DEFINITION TELEVISION SIGNALS OVER DIGITAL LINKS

Summary

This Recommendation considers a bit-rate reduction scheme to be applied for the digital transmission of EDTV-II composite signals compatible with NTSC television signals [1].

Digital transmission thereof involves direct coding of the composite signals.

Source

ITU-T Recommendation J.88 was prepared by ITU-T Study Group 9 (1997-2000) and was approved under the WTSC Resolution No. 1 procedure on 16 September 1999.

FOREWORD

ITU (International Telecommunication Union) is the United Nations Specialized Agency in the field of telecommunications. The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of the ITU. The ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Conference (WTSC), which meets every four years, establishes the topics for study by the ITU-T Study Groups which, in their turn, produce Recommendations on these topics.

The approval of Recommendations by the Members of the ITU-T is covered by the procedure laid down in WTSC Resolution No. 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

INTELLECTUAL PROPERTY RIGHTS

The ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. The ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, the ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementors are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database.

© ITU 2000

All rights reserved. No part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from the ITU.

CONTENTS

	<i>Page</i>
1 Scope	1
2 References	1
3 Definitions	1
4 Abbreviations.....	2
5 Transmission of enhanced definition television signals over digital links.....	3
5.1 Introduction.....	3
5.2 System description	3
Annex A – Specification of direct composite coding for EDTV-II and NTSC signals	3
A.1 Structure of coding bit stream	3
A.2 Coding process	10
A.3 Decoding process	25
Appendix I – Examples for adaptive scanning patterns	29
Appendix II – Examples for VLC of motion vectors	31
Appendix III – Examples for quantizer and mix mode parameters design	33
Appendix IV – Fast Hadamard transform	41
Appendix V – VLC patterns for WHT coefficients	44

Recommendation J.88

TRANSMISSION OF ENHANCED DEFINITION TELEVISION
SIGNALS OVER DIGITAL LINKS

(Geneva, 1999)

1 Scope

The introduction of an enhanced television system, namely EDTV-II (compatible with the NTSC system), has led to the need to carry this signal over digital links.

Due to the high cost expected for the enhanced television encoders, a specific coding system has been developed to allow transmission of the EDTV-II signals foreseen for primary distribution and also for contribution where no further post-processing is needed.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; all users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published.

- [1] Recommendation ITU-R BT.1298 (1997), *Enhanced wide-screen NTSC TV transmission system*.
- [2] ITU-T Recommendation H.222.0 (1995) | ISO/IEC 13818-1:1996, *Information Technology – Generic coding of moving pictures and associated audio informations: Systems*.

3 Definitions

This Recommendation defines the following terms:

- 3.1 adaptive quantizer:** Quantizer step size is controlled by a slice type, buffer occupancy and HVS (human visual system).
- 3.2 adaptive scanning:** Transmission orders of WHT coefficients. Four types of scanning are used to transmit composite TV signals efficiently by a limited number of bits.
- 3.3 bit rate:** The rate at which the compressed bit stream is delivered from the channel to the input of a decoder.
- 3.4 block:** A unit for direct execution of WHT on the composite EDTV-II signal under the 8 pixels × 8 lines size.
- 3.5 coding mode:** Mix A, Mix B and refresh (all intra) mode.
- 3.6 colour sub-carrier phase shift:** Phase shift of 3.58 MHz colour sub-carrier of the coding block from the motion compensated reference block in the previous frame.
- 3.7 composite motion compensation:** Motion Compensation (MC) is conducted on WHT domain not on pixel domain to solve colour sub-carrier phase shift.
- 3.8 criticality:** The reference for the difficulty of the picture judged in macro block units.
- 3.9 EDTV-II sampling:** 14.3 MHz with 8-bit precision.
- 3.10 EDTV-II signals:** NTSC-compatible components transmitted in the centre part of the picture and helper components located in the upper and lower parts of the picture.
- 3.11 Fukinuki hole:** Frequency regions around the colour sub-carrier of the compatible centre part where normal NTSC signals have less spectrum density.



Documents

Document Sharing Hub

Recommendation J.88 (09/99)

1

3.12 helper signals: Three types of spatio and temporal enhancement signals: Horizontal High frequency helper signal (HH), Vertical High frequency helper signal (VH) and Vertical Temporal helper signal (VT).

3.13 macro block: A size of 16 pixels $\times$ 16 lines and is composed of four 8×8 Walsh-Hadamard Transform.

3.14 mix mode: Mixing Inter and Intra mode in the same macro block to enhance coding efficiency.

3.15 PES packet: The data structure used to carry elementary stream data. A PES packet consists of a PES packet header followed by a number of continuous bytes from an elementary data stream. It is a layer in the system coding syntax described in 2.4.3.6 of ITU-T Rec. H.222.0 | ISO/IEC 13818-1 [2].

3.16 picture layer: A repeated structure composed of a head I-picture and a plurality of subsequent p-pictures.

3.17 sequence layer: Uppermost layer of a coding bit stream by which coding and decoding parameters are coordinated.

3.18 slice layer: Composed of 768 pixels $\times$ 16 lines active pixels, and comprises a horizontal arrangement of macro blocks. The slice is categorized into four types of ID and control signal area, boundary area, block area and active image area.

3.19 two-dimensional VLC: Huffman code having the combination of zero-run length and quantization output level as a symbol.

4 Abbreviations

This Recommendation uses the following abbreviations:

Bp	Maximum Buffer Capacity
BUFP	Buffer Pointer
CF	Colour Frame
EDTV	Enhanced Definition Television
FSW	Video Frame Synchronizing Word
GOP	Group of Pictures
HH	Horizontal High frequency component
NTSC	National Television System Committee
PC	Phase Compensation
SSW	Sequence Synchronizing Word
VH	Vertical High frequency component
VITS	Vertical Interval Test Signal
VLC	Variable Length Coding
VLD	Variable Length Decoding
VT	Vertical Temporal frequency component
WHT	Walsh-Hadamard Transform

5 Transmission of enhanced definition television signals over digital links

5.1 Introduction

This Recommendation considers a bit-reduction coding scheme for application to digital transmission of EDTV-II composite signals defined for NTSC-compatible enhanced television. The bit-reduction coding system described below is a composite coding system that does not require any decoding/re-encoding process for enhanced signals and colour component signals.

Consequently the system is free from picture quality loss brought about by composite-component conversion required in component coding schemes. The bit rate required for contribution and distribution purposes is approximately 20 Mbit/s.

5.2 System description

The EDTV-II signals consist of NTSC-compatible components transmitted in the centre part of the picture and the helper components located in the upper and lower parts of the picture. Three kinds of helper signals are employed: a horizontal high frequency component (HH) multiplexed in frequency regions around the colour sub-carrier (so-called Fukinuki hole) of the compatible centre part where normal NTSC signals have less spectrum density, the vertical high frequency component (VH), and the vertical temporal helper signal (VT) compressed in the spatial domain and multiplexed in the upper and lower parts of the picture. Thus, EDTV-II composite signals have a complex structure.

Although component bit-reduction coding can be applied to EDTV-II signals, it requires a decoding/re-encoding process for the colour and HH components as well as the VH/VT components at every stage of transmission chain. This decoding/re-encoding process inevitably causes severe damage to picture quality, especially in the case of tandem connections. This degradation is caused by the irreversible nature of the decoding/re-encoding process.

Hardware for the composite coding scheme was implemented based on an interframe motion-compensated Walsh-Hadamard Transform coding scheme. A preliminary picture quality evaluation test was conducted with the hardware and it showed that the system provides picture quality adequate for the contribution and primary distribution of the EDTV-II programmes at a transmission bit rate of approximately 20 Mbit/s. This evaluation test also showed that the system provides coding performance equivalent or superior to the MPEG-2 codec for NTSC composite signals at a transmission bit rate above 15 Mbit/s.

The above considerations and the test results mentioned show that the composite coding system is well suited for transmission of signals such as EDTV-II and NTSC when relatively high quality is required.

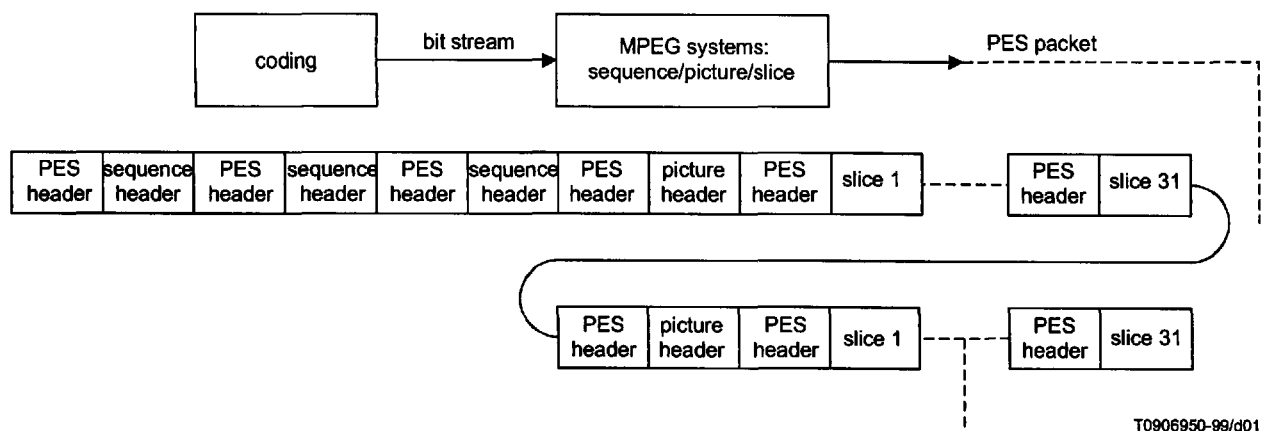
Detailed specification for a composite coding system suitable for digital transmission of EDTV-II as well as NTSC signals is given in Annex A.

Annex A

Specification of direct composite coding for EDTV-II and NTSC signals

A.1 Structure of coding bit stream

The coding bit stream is converted to the PES packet in order to be operable on ITU-T Rec. H.222.0 | ISO/IEC 13818-1 (MPEG systems) [2]. Therefore, the bit stream is arranged in a three-layer structure composed of a sequence layer, a picture layer, and a slice layer. The unit of the PES packet is basically a slice in view of the maximum code generation. The outline of the bit coding stream structure is shown in Figure A.1-1.



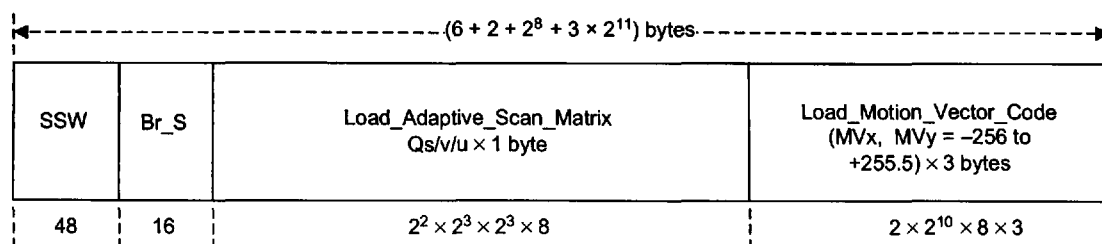
T0906950-99/d01

Figure A.1-1/J.88 – Outline of coding bit stream structure

A.1.1 Sequence layer

The sequence layer is the uppermost layer of a coding bit stream. The sequence layer has the purpose of coordinating coding and decoding parameters. In detail, the sequence header transmits required parameters from the coding side to the decoding side with the desired timing in order to coordinate automatically coding and decoding for channel hopping, switching of systems, and problems of systems. A sequence header may be interpolated as desired at any position as long as the position is a blank between pictures shown by the picture header in a picture layer as described later. There are three types of sequence headers from 1 to 3 as described below.

A.1.1.1 Sequence header 1



T0906960-99/d02

1) SSW: Sequence head word

SSW is a sequence synchronizing word. The pattern is composed of one bit "1" and 47 bits "0". SSW is transmitted once, at the timing when the receiving side becomes receivable, establishes the sequence synchronization and transmits parameters.

2) Br_S: Bit rate (sequence)

Under a premise of the change of parameters, bit rate change is executed. The multiple number of 90 kbit/s is represented by $90 \times Br_S$ kbit/s (0 to 5898.15 Mbit/s).

3) Load_Adaptive_Scan_Matrix

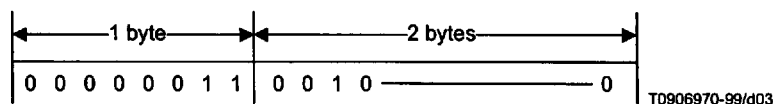
Four types of Adaptive_Scan parameters (represented by Qs 2 bit which will be described in A.2.6 and A.3.2 later) are transmitted.

The scanning pattern is transmitted in the order of Qs = 00, 01, 10, 11, and the scanning order of each coefficient for raster-scanning of 8×8 Hadamard transformation. As a result, the capacity of parameters is $8 \text{ bits} \times 64 \text{ coefficients} \times 4 \text{ patterns} = 2^{11} \text{ bits} = 2^8 \text{ bytes}$.

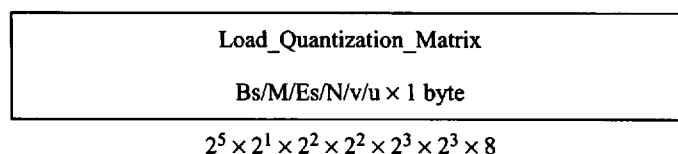
4) Load_Motion_Vector_Code

A coding pattern (Huffman code) which represents a horizontal motion vector (MVx) and vertical motion vector (MVy) is transmitted. MVx and MVy are composed of half-pixel/half-frame line units, and are retrievable up to the maximum -128.0 to +127.5 pixel/frame line. Therefore, the motion vector requires 1024 coding patterns of -256 to 255.5 pixel/frame line to transmit the differential value. This coding pattern is transmitted in the form as described below composed of 1-byte code length and subsequent 2-byte coding pattern in the order $MVx = -256.0 \geq 0 \geq 255.5$, $MVy = -256.0 \geq 0 \geq 255.5$ as Huffman code with the maximum code length of 16 bits or less. It is assumed that bits other than coding pattern is "0", and bits which have no coding pattern for defining the searching range is also "0".

Case that the coding pattern is "001" (code length of 3 bits).

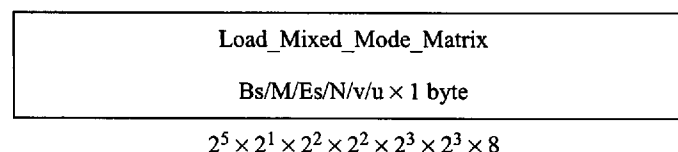


A.1.1.2 Sequence header 2



This sequence header 2 is data of fixed length 216 bytes subsequent to the sequence header 1. Load_Quantization_Matrix transmits a quantizing parameter described in A.2.6 and A.3.3. The transmission address order of transmission of the quantizing parameter contained in a quantizing parameter table g1 (16-bit input; equivalent to 1M) is represented by 1 byte of Delta (Bs, M, Es, N, v, u) under Bs/M/Es/N/v/u = 00 to FF. Therefore, the PES packet length is the maximum length.

A.1.1.3 Sequence header 3



This sequence header 3 is data of fixed length 216 subsequent to the sequence header 2. The Load_Mixed_Mode_Matrix transmits parameters for mixed mode coding described in A.2.5 and A.3.5. The transmission address order of transmission of the mixed mode parameter contained in a quantizing parameter table h1 (16-bit input; equivalent to 1M) is represented by 1 byte of A (Bs, M, Es, N, v, u) under Bs/M/Es/N/v/u = 00 to FF. Therefore, the PES packet is the maximum length.

A.1.2 Picture layer

The picture has a repeated structure composed of a head I-picture and a plurality of subsequent p-pictures. The head I-picture is directed by a refresh information (R). This R is included in the picture header shown in Figure A.1-2. The picture header constitutes one PES packet with one picture header.

(PES-packet-length = 000 000 0001 0010)

A.1.2.1 Picture header

The structure of a picture header is shown in Figure A.1-2, and the components are described in 1) to 8).

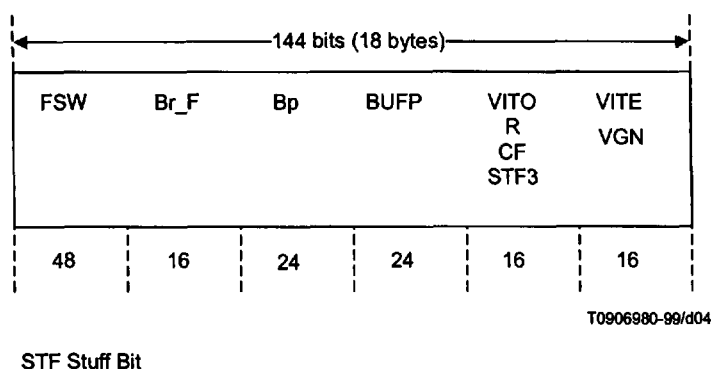


Figure A.1-2/J.88 – Structure of picture header

1) *FSW: Video frame head word*

FSW is a frame synchronizing word, and composed of 47 bits "0" and subsequent 1 bit "1". FSW is transmitted once, and at the timing when the receiving side becomes receivable, establishes the sequence synchronization and transmits parameters. Once synchronization is established and FSW has not been detected, it is regarded as a step out from the frame synchronization.

2) *Br_F: Bit rate (frame)*

Parameters are not changed, and bit rate is changed without any breaks of picture. The multiple number of 90 kbit/s is represented by $90 \times Br_F$ kbit/s (0 to 5898.15 Mbit/s).

3) *Bp: Maximum buffer capacity*

Transmission is implemented once at the head of each frame. Only when different patterns are received is the maximum buffer capacity change regarded as effective, and the change is implemented on the decoder side.

This is the parameter which represents the capacity of the buffer memory. One unit has the weight of 32 bits, maximum $32 \times 2^{24} = 512$ Mbits.

4) *BUFP: Buffer pointer*

The buffer pointer has 32-bit units and represents buffer occupancy quantity for each picture. It is possible to represent up to the maximum $32 \times 2^{24} = 512$ Mbits.

5) *VITO/VITE: VITS Position (ODD/EVEN)*

This shows whether the odd lines 10 to 20 (11 lines) and even lines 273 to 282 (10 lines) are selected by VITO-11 bits and VITE-10 bits as VITS line. For example, when the third bit of VITO is "1", 12-th line is selected as VITS line, and this line is addressed for transmission.

6) *R: Refresh information*

The existence of I-picture at the head of GOP is directed to a decoder with transmission of 1 bit. "1" represents the refresh mode and "0" represents the adaptive coding mode.

7) *CF: Colour frame*

Whether the picture has 1,2 field or 3,4 field is directed to a decoder with one transmission of 1 bit.

8) *VGN: Picture gain information*

The input signal is subjected to a signal compression in the range of 1/2 to 1 with 6 bits (1/64) precision.

A.1.2.2 EDTV-II composite picture format

Sampling frequency must be 4 times of colour sub-carrier (3.58 MHz). To obtain active pixel area to be coded out of these sampled data, the vertical blanking period signal (Figure A.1-3) and horizontal blanking period signal (Figure A.1-4) are removed.

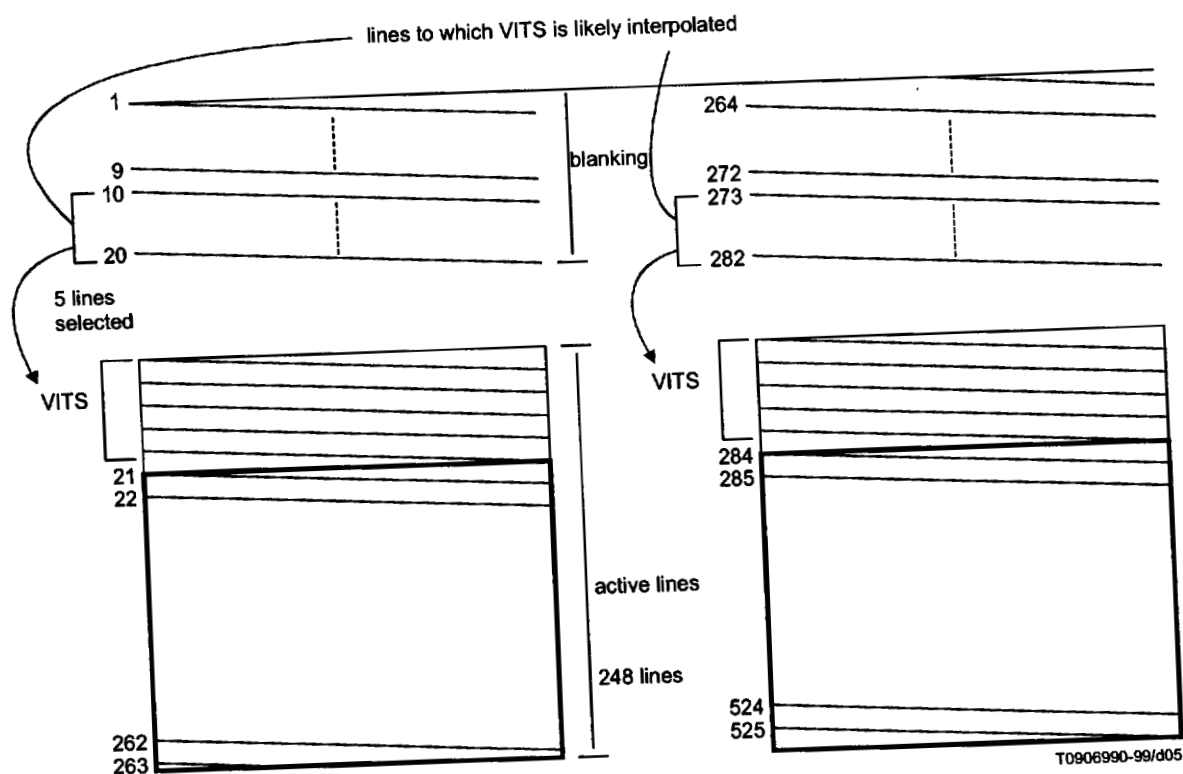


Figure A.1-3/J.88 – Removal of vertical blanking period signal

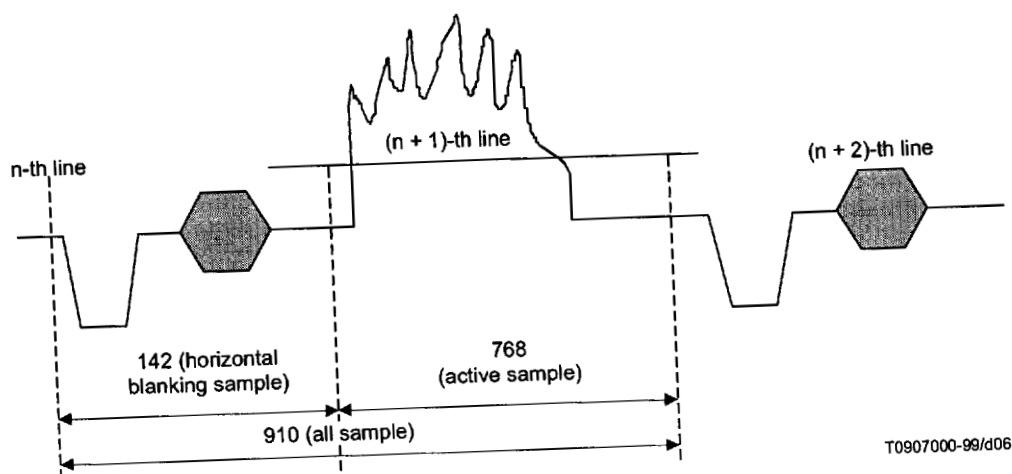


Figure A.1-4/J.88 – Removal of horizontal blanking period signal

A.1.3 Slice layer

A.1.3.1 Structure of slice

A slice is composed of $768 \text{ pixels} \times 16 \text{ lines}$ active pixels, and comprises a horizontal arrangement of macroblocks having a size of $16 \times 16 \text{ lines}$ which is the transmission unit of various control signals and motion vectors. Therefore, 1 slice is composed of 48 macroblocks taken in the horizontal direction from the active $768 \times 496 \text{ pixels}$ of the picture after having removed the vertical and horizontal blanking signals as described above, and 1 picture is composed of 31 slices.

The relationship between the slice and line in the EDTV-II signal is shown in Figure A.1-5, and the slice is categorized into four types (Es) of ID and control signal area, boundary area, black area, and active image area.

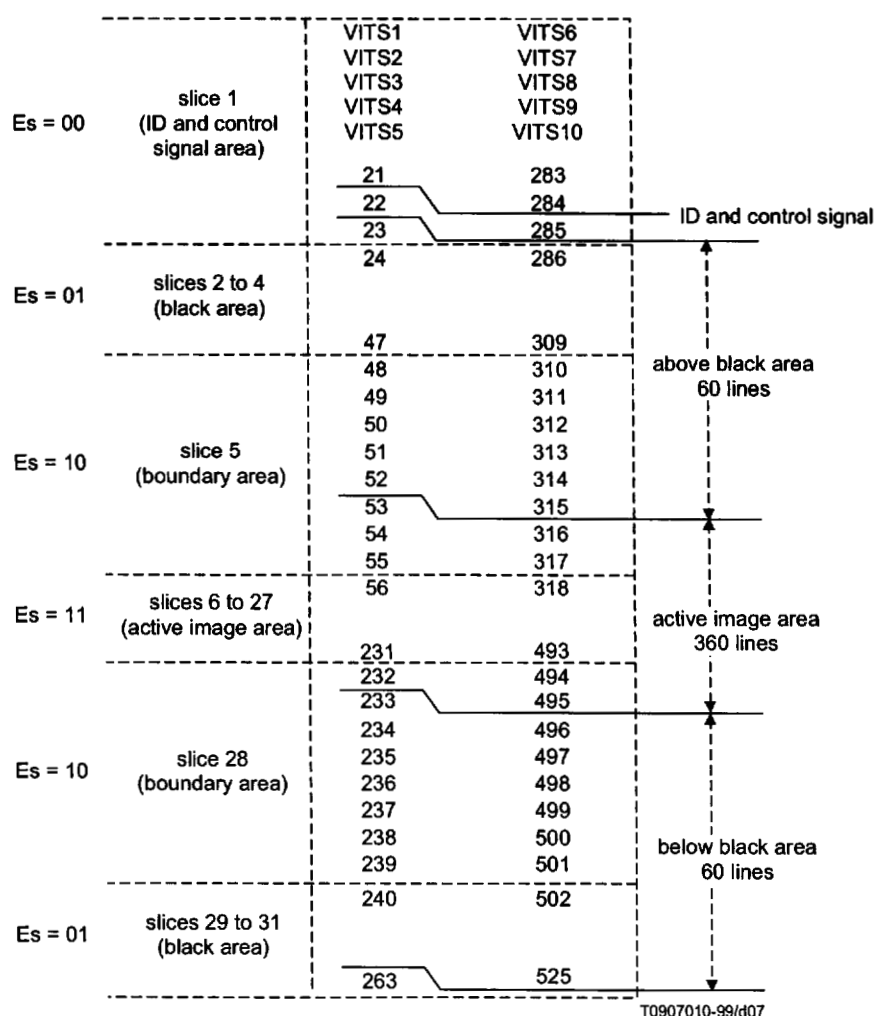


Figure A.1-5/J.88 – Four types of slices in EDTV-II signal

The structure of each slice is shown in Figure A.1-6 and described in 1) to 7) below:

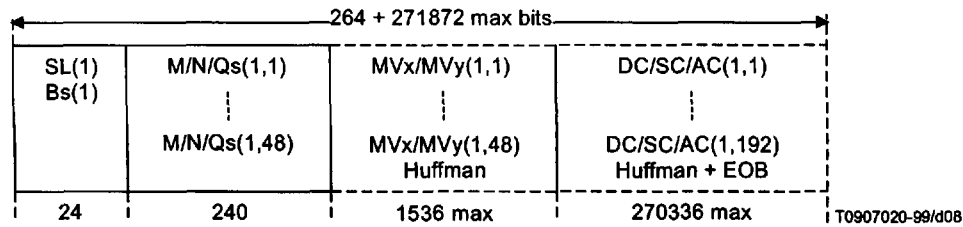


Figure A.1-6/J.88 – Internal structure of a slice

1) *SL(n)*

SL(n) positioned at the head of the *n*-th slice represents the quantity of information to be generated from the slice. This *SL* is represented using 19 bits. Because the maximum information quantity generated from variable length code of 1 slice is $272\,136\text{ bits} < 2^{19}\text{ bits} = 2^{16}\text{ bytes}$, which is the total of the assumable longest codes of all coefficients in a slice ($48 \times 64 \times 4$) and motion vectors (48), namely 22 bits and 32 bits respectively (described in A.2.7 and A.3.1 respectively), *SL* can express it and it can be contained in a PES packet. A counter decodes up to the bit length indicated by *SL(n)* and, when finished, starts to decode the next slice.

2) *Bs(n)*: Buffer information

Buffer occupancy quantity for coding the *n*-th slice is represented using 5 bits (32 steps).

3) *M(n, m)*: Coding mode selection information

The coding mode of the *m*-th macroblock ($m = 1$ to 48) of the *n*-th slice is represented using 1 bit. "0" is for inter (mixed) mode and "1" is for intra mode. The mode decision is described in A.2.5.

4) *N(n, m)*: Criticality information

The activity of the *m*-th macroblock ($m = 1$ to 48) of the *n*-th slice is represented using 2 bits (four types). "0" is for flat area, "1" is for gradient area, "2" is for edge area, and "3" is for detail area. The determination of the criticality information will be described later in A.2.4.

5) *Qs(n, m)*: Scanning selection information

The scanning selection for transmission order of Hadamard transform coefficient of the *m*-th macroblock ($m = 1$ to 48) of the *n*-th slice is represented using 2 bits (four types). The determination of the scanning selection information is described in A.3.6.

6) *MV(n, m)*: Motion vector information

The motion vector information of the *m*-th macroblock of the *n*-th slice is represented by Huffman code. The motion vector detection is described in A.2.3.

7) *DC/SC/AC(n, l)*: Transform coefficient data

The transform coefficient data is two-dimensionally Huffman-coded DS/SC/AC component of the 1st block ($l = 1$ to 192) of the *n*-th slice (described later).

In the above, slice data of 1) to 5) are in the form of fixed length, but 6) *MV* and 7) *DC/SC/AC* are in the form of variable length data. 2) *Bs* is control information with slice unit, and 3) to 6) are control information with macroblock unit.

A.1.3.2 Macroblock

The macroblock has a size of $16\text{ pixels} \times 16\text{ lines}$ always, and is composed of four 8×8 WHT (Hadamard transform). The counting of the macroblocks in a slice proceeds from left to right, namely raster scanning; the counting of blocks in a macroblock is carried out as shown in Figure A.1-7.

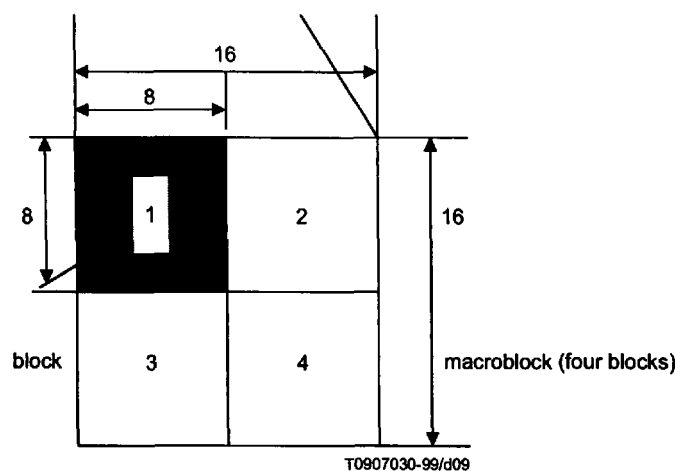


Figure A.1-7/J.88 – Structure of macroblock

Because the macroblock is addressed only on frame-based Hadamard transform coding, the macroblock is structured as shown in Figure A.1-8.

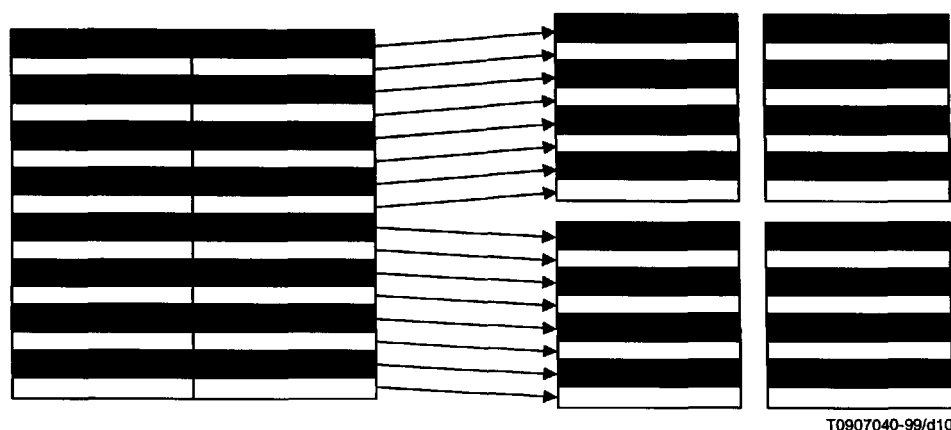


Figure A.1-8/J.88 – The relationship between frame and macroblock

A.1.3.3 Block

The block is a unit for direct execution of WHT on the composite EDTV-II signal under the 8 pixels $\times$ 8 lines size. It is assumed that Hadamard transform coding is performed under the frame base generated from field merging (Figure A.1-9).

A.2 Coding process

The whole structure of the coding process is shown in Figure A.2-1.

EDTV-II signal must be sampled at 14.3 MHz and with 8-bit precision. The coding process of this data involves basically the removal of time space redundant component by inter/intra composite motion compensated adaptive Hadamard transform coding. Elements associated with the coding process are described in the following.

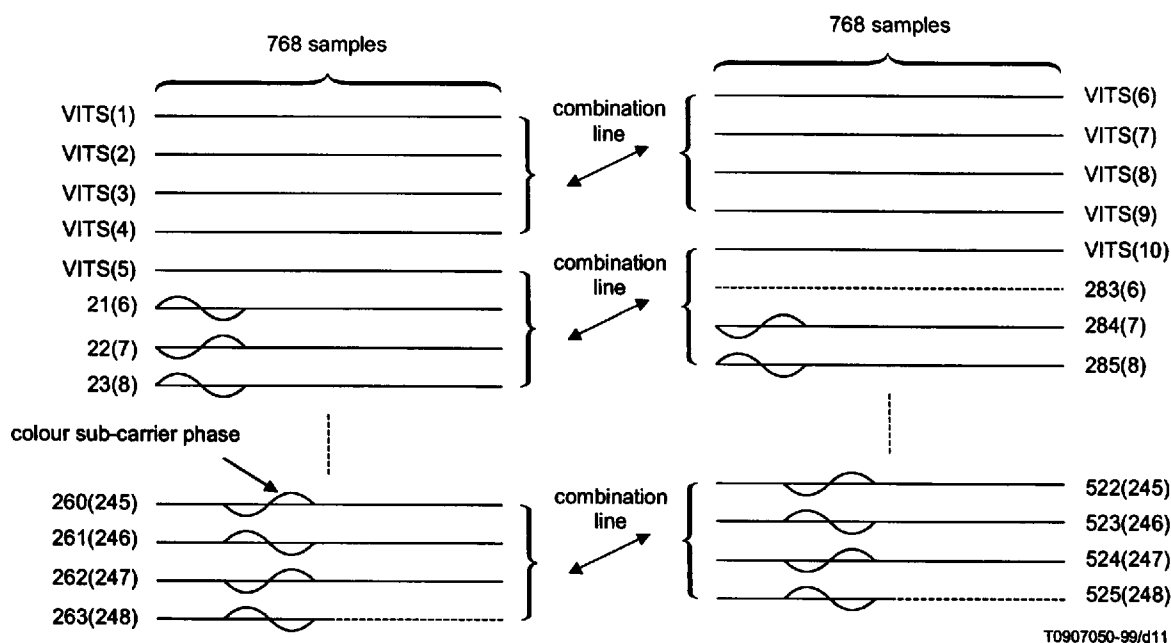


Figure A.1-9/J.88 – Frame-based WHT coding based on field merge

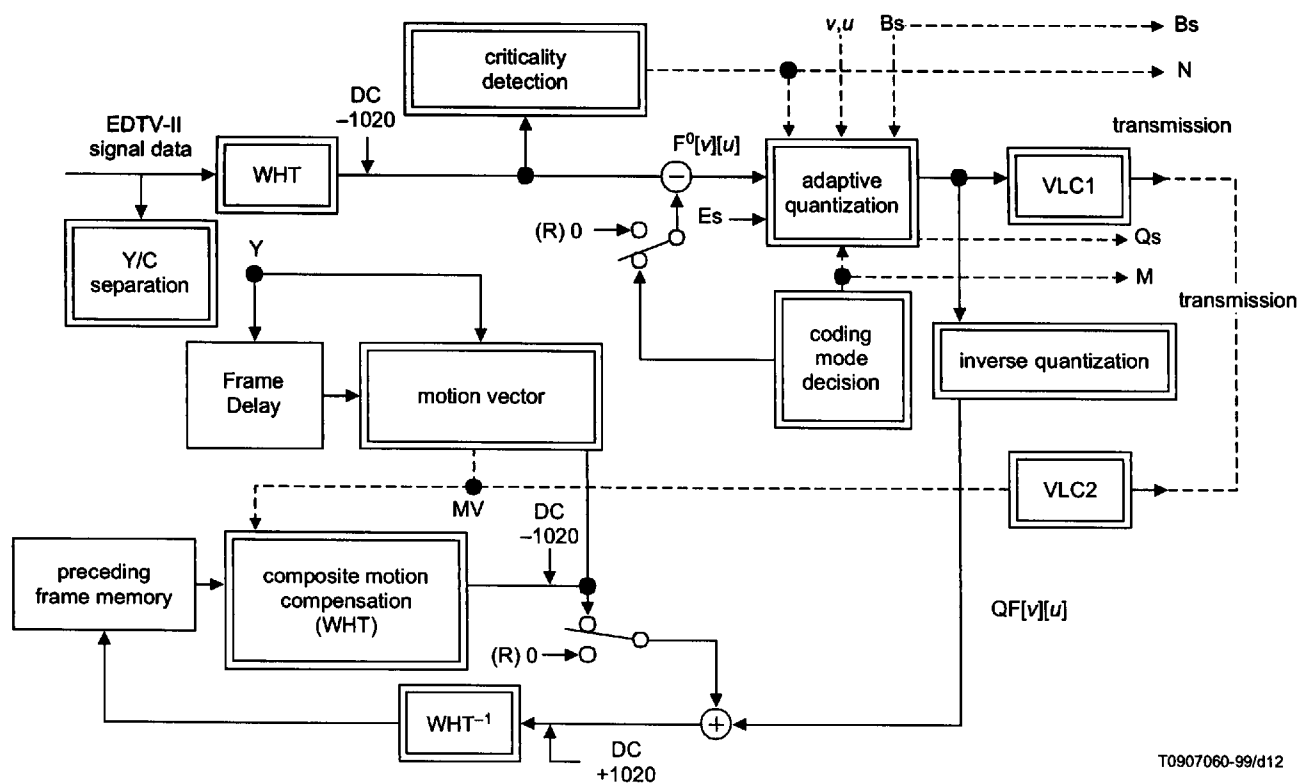


Figure A.2-1/J.88 – Whole structure of coding process