DVCAM Format Overview

DVCAM™

Table of contents

1

Introduction

The DVCAM™ format has been developed with the robustness and operability required for professional use while maintaining compatibility with the DV format.

In addition to its tape and cassette mechanics, the recorded data also provides full compatibility with DV recordings. This means that full upward/downward tape playback is guaranteed and that signal transfers are accomplished without manipulation to the originally recorded data by use of i.LINK or SDTI (QSDI) interfaces. These features have proven the DVCAM as the most suitable format for integrated use with the widely popular DV transports and DV-based NLE's.

The DVCAM format also takes into account the requirements in existing linear editing environments. The 15-micron track pitch assures frame accurate and stable editing at the tape edit point. The use of this track pitch also realizes full lip-sync audio and pre-read capabilities.

This document describes details on the DVCAM format as well as its associated interface technologies.

DSR-2000/P
Editing Recorder

2

Mechanical Specifications

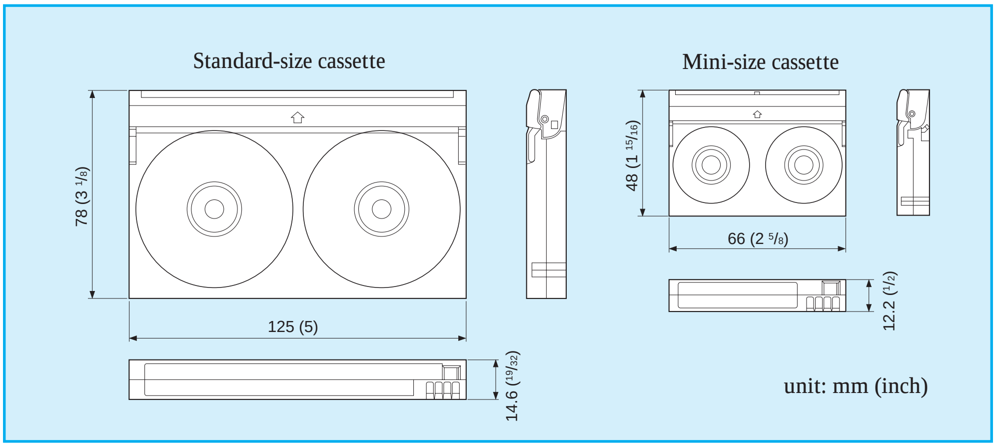
2-1 Tape Cassette

The DVCAM/DV formats use metal evaporated tape with a tape width of 6.350 ± 0.005 mm. The DV format specifies two types of tape thickness which are 7.0 µm and 5.3 µm including all tape coatings.
The DVCAM format uses only the tape thickness of 7.0 µm to achieve its professional robustness.

◆ Dimensions

There are two cassette sizes.
The dimensions of the two cassettes are in accordance with **Figure 1**.

Figure 1 - Appearance of cassette



The sizes of the two cassette types are identified as follows.

Standard-size cassette: approximate size: 125.0 mm x 78.0 mm x 14.6 mm
(5 x 3 $^{1}/_{8}$ x $^{19}/_{32}$ inches)
Mini-size cassette: approximate size: 66.0 mm x 48.0 mm x 12.2 mm
(2 $^{5}/_{8}$ x 1 $^{15}/_{16}$ x $^{1}/_{2}$ inches)

◆ Recording time

The maximum recording time is 184 minutes for a Standard-size cassette and 40 minutes for a Mini-size cassette.

◆ Cassette Identification and Cassette Memory

DVCAM/DV cassettes use either an ID board or a Cassette Memory (so-called 'MIC') for cassette identification.
Both cassettes have four electrode contacts which are used to communicate the cassette identification and other information to the VTR. The information each contact transfers to the VTR is described below with its characteristics.
For a cassette with an ID board, the cassette type is identified by the resister values of the contacts. On the other hand, for a cassette with a MIC, the identification information is stored in the memory.

The MIC is also used to store a variety of auxiliary information including a shooting log know as 'ClipLink™'. The simpler ID board is used only for cassette identification.

Cassette with ID board

- Contact number 1 indicates the tape thickness.
- Contact number 2 indicates the tape type.
- Contact number 3 indicates the tape grade.
- Contact number 4 is ground level.

The resistance value between contact number 1 to 3 and contact number 4 designate the cassette identification as specified in Table 1.

Cassette with memory (MIC)

- Contact number 1 is used for the memory power supply.
- Contact number 2 is used for data input/output.
- Contact number 3 is used for the clock signal.
- Contact number 4 is ground level.

The MIC data is transferred to the VTR through contact number 2. As mentioned above, the MIC includes cassette identification information, such as tape thickness, tape type and tape grade as well as auxiliary information.

In DVCAM, the auxiliary area is currently used for ClipLink and has provision for future applications.

A DVCAM VTR can detect cassette identification for both cassettes with an ID board or MIC.

Table 1 - Assignment of the four contacts

Contact number	Cassette with ID board			Cassette with memory (MIC)
	Assignment	Identification	Resistance value	Assignment
1	Tape thickness	7 µm	Open	VDD
		5.3 µm	1.80 kΩ ± 0.09 kΩ	
2	Tape type	ME	Open	SDA
		Reserved	6.80 kΩ ± 0.34 kΩ	
		Cleaning	1.80 kΩ ± 0.09 kΩ	
		MP	Short	
3	Tape grade	Consumer VCR	Open	SCK
		Non-consumer VCR	6.80 kΩ ± 0.34 kΩ	
		Reserved	1.80 kΩ ± 0.09 kΩ	
		Computer	Short	
4	GND	——		GND

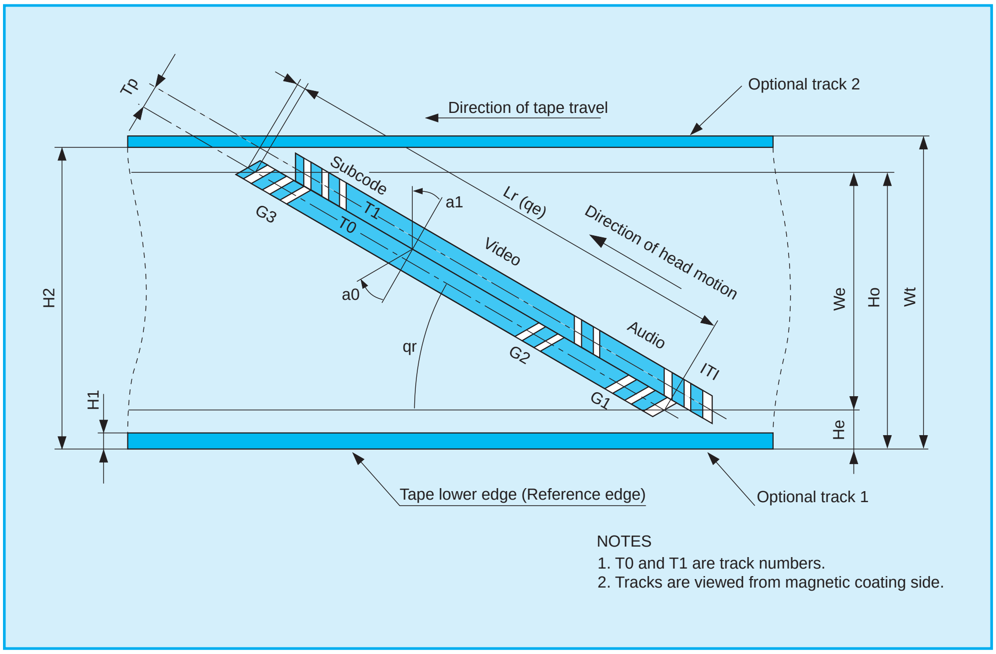
Where ME: Metal Evaporated
MP: Metal Particle

2-2 Helical Recordings

◆ Record location and dimensions

Record location and dimensions of each sector are as specified in **Figure 2** and **Table 2**. There are four sectors in a DVCAM helical track.
These are the ITI (Insert and Track Information), Audio, Video and Subcode sectors as shown in **Figure 3**. Their dimensions are specified as in **Table 3**.

Figure 2 – Record location and dimensions



NOTES
1. T0 and T1 are track numbers.
2. Tracks are viewed from magnetic coating side.

Table 2 – Track pattern parameters

Symbol	Description	Unit	525-60 system	625-50 system
Tp	Track pitch	µm	15	←
Ts	Tape speed	mm/sec	28.221/1.001	28.221
θr	Track angle	deg	9.1752	←
Lr	Length of track	mm	32.860	←
Wt	Tape width	mm	6.350	←
He	Effective area lower edge	mm	0.560	←
H0	Effective area upper edge	mm	5.800	←
We	Effective area width	mm	(5.240)	←
H1	Height of optional track 1 upper edge	mm	0.490	←
H2	Height of optional track 2 lower edge	mm	5.920	←
α0	Azimuth angle (track 0)	deg	-20	←
α1	Azimuth angle (track 1)	deg	+20	←

Figure 3 – Sector location

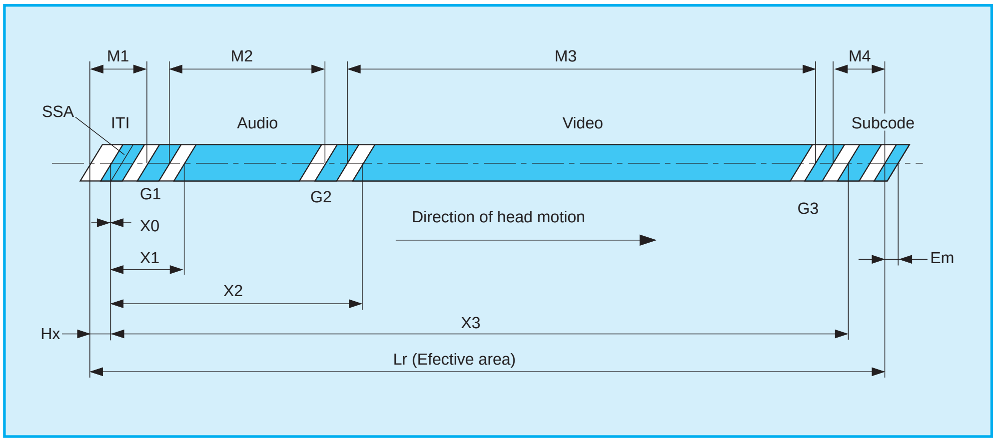


Table 3 – Sector location

Dimensions in millimeters

Dimensions		525-60 system	625-50 system
Hx	Length of ITI pre-amble	0.341	←
X0	Beginning of SSA	0	←
X1	Beginning of audio sync blocks	0.810	←
X2	Beginning of video sync blocks	3.792	3.796
X3	Beginning of subcode sync blocks	31.905	31.937
M1	Length of ITI sector	0.876	0.877
M2	Length of audio sector	2.812	2.815
M3	Length of video sector	27.565	27.593
M4	Length of subcode sector	0.907	0.908
Em	Length of overwrite margin	0.304	←

3

Video Signal Processing

◆ Sampling

The sampling raster of the DVCAM is the same as that of the ITU-R Rec.601.
Luminance video signals are sampled at 13.5 MHz, 720 pixels are transmitted per line for both 525-60 and 625-50 systems.
In the 525-60 system, each color difference signal (CR/CB) is sampled at 3.375 MHz and 180 pixels are transmitted per line (4:1:1). In the 625-50 system, each color difference signal is sampled line sequentially at 6.75MHz, i.e. 360 pixels of either color difference signal is transmitted per line (4:2:0). The decimation filter for the 625-50 system color difference signals is 3 taps FIR filter whose coefficients are 1-2-1 value. The interpolation filter is also a 3 taps FIR filter which has the same value of coefficients as the decimation filter.
The sampling start point of CR and CB signals is the same as the luminance signal in both systems.
Table 4 shows the number of active pixels/line in the 525-60 and 625-50 systems.

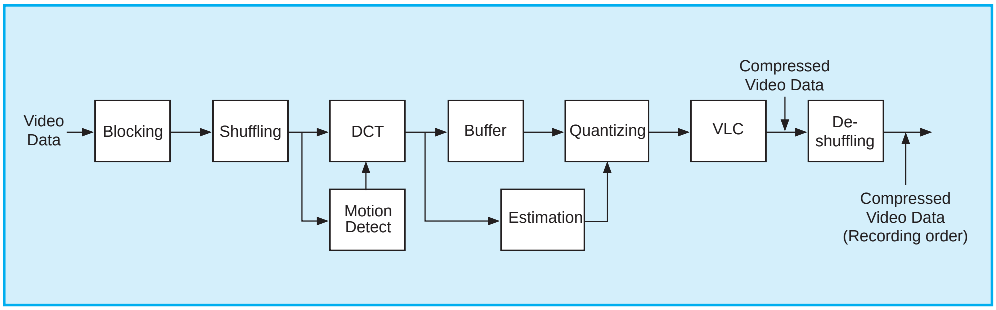
Table 4 – Construction of video signal processing

		525-60 system	625-50 system
Sampling frequency	Y	13.5 MHz	
	CR, CB	3.375 MHz	6.75 MHz
Number of active pixels per line	Y	720	
	CR, CB	180	360
Number of active lines per frame		480	576
Active line numbers	Field 1	23 to 262	23 to 310
	Field 2	285 to 524	335 to 622
Quantization		8 bits	
Correspondence between video signal level and quantized level	Scale	1 to 254	
	Y	220 quantization levels with the black level corresponding to level 16 and the peak white level corresponding to level 235	
	CR, CB	225 quantization levels in the center part of the quantization scale with zero signal corresponding to level 128	

Where Y: Luminance
CR, CB: Color difference

The sampled video data is reduced by a factor of 5:1 using bit rate reduction, resulting in a transfer rate of 25 Mb/s. Intra-frame coding which adopts DCT (Discreet Cosine Transform) and VLC (Variable Length Coding) is used.
To realize a good picture quality at the 25 Mb/s data rate, DV/DVCAM compression adopts a shuffling technique prior to the encoding process. This allows the video to be compressed with maximum efficiency and thus keeps a well-balanced picture quality for any type of images. Note that the data is shuffled only for the purpose of maximizing compression efficiency and thus de-shuffled before recorded to tape.
Figure 4 is a simplified block diagram of the video process.

Figure 4 – Video process block diagram

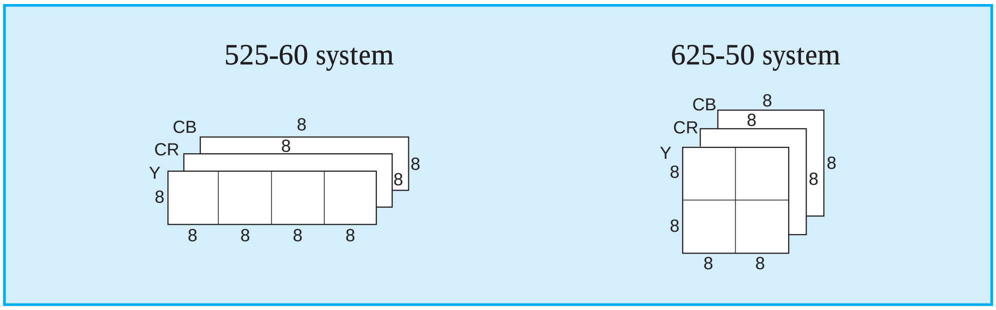


The following explains each process in the bit rate reduction.

◆ Blocking process

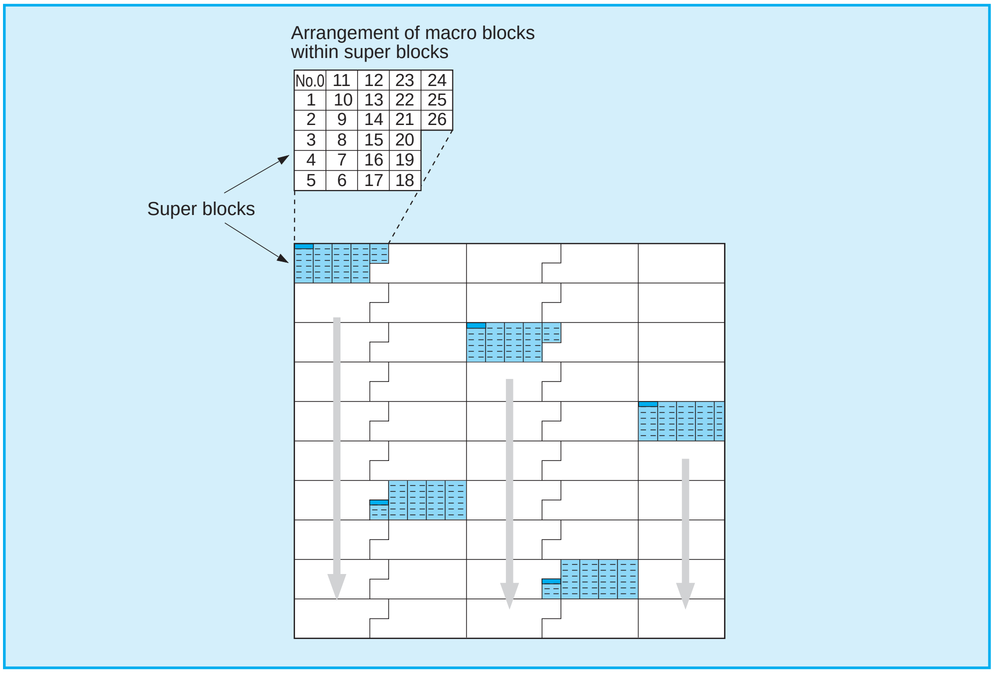
In the DVCAM/DV formats, the sampled video data is handled on a so-called ‘macro block’ basis. ‘Blocking’ is the process of preparing these macro block units.
First, the data of the vertical and horizontal blanking areas is discarded. The picture area of the video data is then divided into 8 x 8 pixel blocks, the size of the later mentioned DCT block. In the 525-60 system, a macro block is formed of four horizontally adjacent luminance pixel blocks and two chrominance pixel blocks, one each for the CR and CB component. Similarly, in the 625-50 system, a macro block is formed of four luminance pixel blocks neighboring in the horizontal and vertical directions, and two chrominance pixel blocks.
In either case, the macro block size was determined as the smallest unit to package one each of the 8 x 8 chrominace pixel blocks with their associated luminance pixel blocks. Since the 525 and 625 DVCAM systems use 4:1:1 and 4:2:0 processing respectively, there are four luminance pixel blocks associated with one each of the chrominace blocks, resulting in a macro block size of six 8 x 8 pixel blocks.
Figure 5 shows the arrangement of 525-60 and 625-50 macro blocks.

Figure 5 – Construction of macro block



27 neighboring macro blocks form a so-called ‘Super Block’ as shown in **Figure 6**.

Figure 6 – Super blocks (525-60 system)

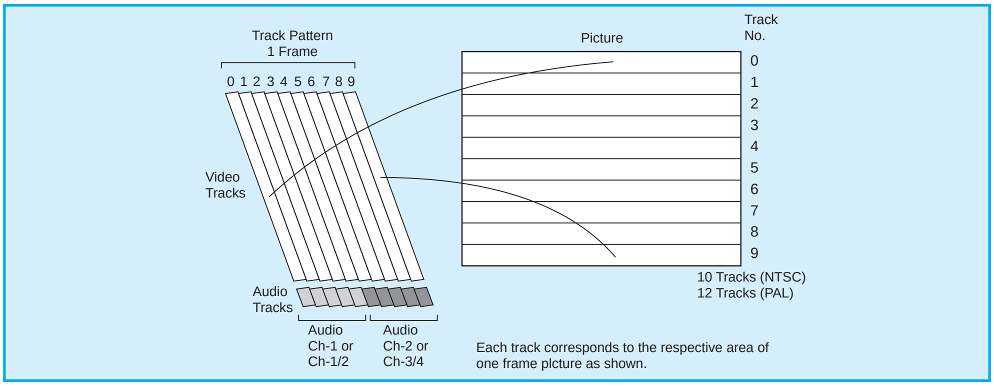


No.0	11	12	23	24
1	10	13	22	25
2	9	14	21	26
3	8	15	20	
4	7	16	19	
5	6	17	18	

Super blocks are used to average picture details in the screen to achieve efficient compression results. But before the details, it is important to understand how the compressed data is recorded to tape.

In a 525-60 system, one video frame is written on 10 tracks as shown in **Figure 7**. Similarly, in the 625-50 system, one video frame consists of 12 tracks. In 525-60, the video data from the top one tenth of the picture is written to the first track, the second one tenth to the second track and so on. In 625-50, the video data from the top twelfth of the picture is written to the first track, the second one twelfth to the second track and so on.

Figure 7 – Track pattern/picture to track allocation



The size of the super block was determined in relation to this screen-to-track data allocation.

◆ Height

The height of a super block corresponds to this screen-to-track data allocation. Thus, the number of vertical samples in a super block are equivalent to the number of scanning lines recorded on a track. In the 525-60 system, since there are 480 vertical samples from the top to bottom of the screen, the height of a super block is as follows.

$480 \div 10$ tracks = 48 pixels

Since there are eight pixels in the vertical direction per macro block, each super block is six macro blocks high. Likewise, in the 625-50 system, there are 576 pixels in the vertical direction meaning the super block height is as follows.

$576 \div 12$ tracks = 48 pixels

Since a macro block in the 625 system is 16 pixels high, each super block is three macro blocks high.

◆ Width

The super block width was determined as one fifth of the picture width since this size would be logical to achieve the end result of compressing the data from five to one.
There are 720 horizontal pixels in both the 525 and 625 systems. Hence, in the 525 system, there are 22.5 macro blocks in the screen's horizontal direction.

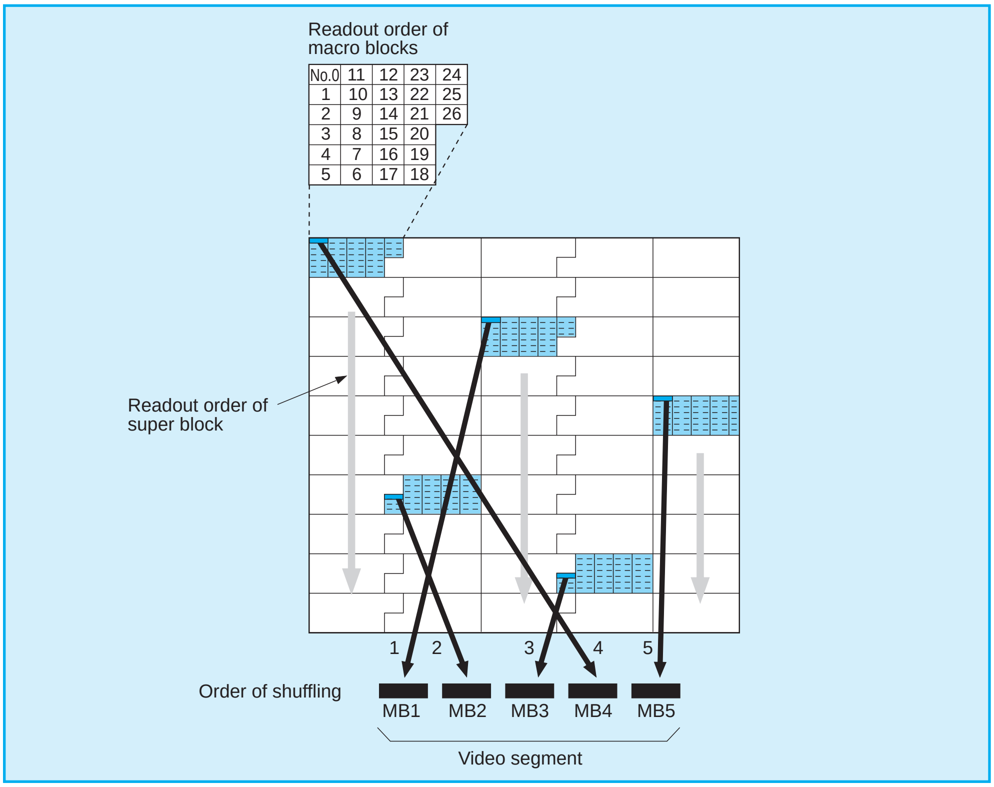
$720 \div 32$ pixels = 22.5 macro blocks
(as previously explained, one macro block is 32 pixels wide.)

Since the width of a super block is one fifth the screen, one super block is either four or five macro blocks wide.
Similarly in the 625 system, there are 45 macro blocks in the screen's horizontal direction.

$720 \div 16$ pixels = 45 macro blocks (one macro block is 16 pixels wide.)

Again, since the width of a super block is one fifth the screen, one super block is nine macro blocks wide.

Figure 8 – Shuffling method (525-60 system)



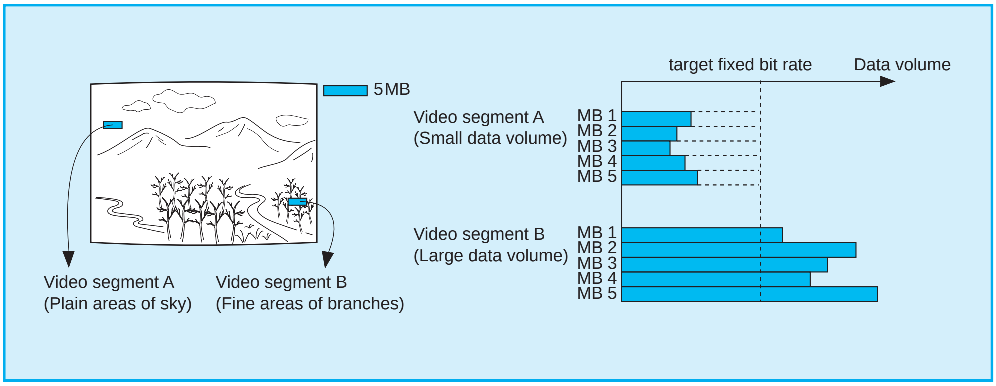
No.0	11	12	23	24
1	10	13	22	25
2	9	14	21	26
3	8	15	20	
4	7	16	19	
5	6	17	18	

◆ Shuffling

In the DVCAM/DV formats, the compression process is applied over five macro blocks gathered from five different super blocks. These five macro blocks form a so-called ‘Video segment’ as shown in Figure 8. Video segments are created by first selecting one super block each from the five super block columns and then gathering the five macro blocks that are located in the same position within their super blocks as shown in Figure 8. This process is called shuffling.

The use of shuffling greatly improves the compression efficiency. This is because in most pictures the amount of detail appears inconsistently across the screen, some areas have greater amount of information while other areas have less. It is also important that the center of the screen is not subjected to heavy compression since in most cases this is where the most important picture content appears. If the shuffling process was not used prior to the bit rate reduction, the amount of information to be compressed would vary for each picture area (or macro block) as shown in Figure 9. Since bit rate reduction is applied at a fixed bit rate, the distortion seen in each picture area (macro block) would vary as a result. For example, the flat picture areas would be less distorted as compared to those areas including greater picture detail. By adopting the shuffling process, the information of each picture area is averaged and kept uniform across the picture frame.

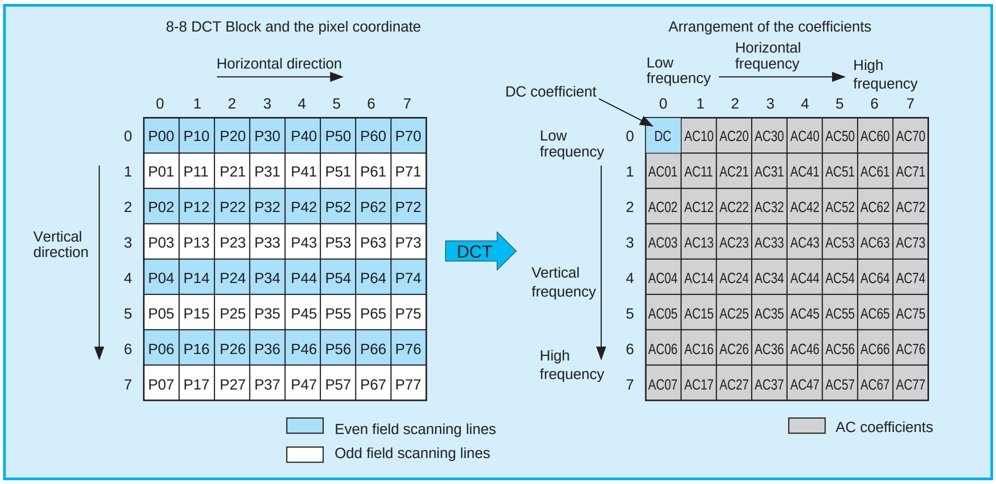
Figure 9 – Averaging data volume



◆ DCT process

After the blocking and shuffling process, each 8 x 8-pixel block is sent to the DCT encoder. The DCT encoder transforms the 8 x 8 base band pixel blocks from the spatial domain to the frequency domain. The result of this transform is an 8 x 8 DCT block with coefficients representing the energy of each frequency in the block as shown in Figure 10. In DCT blocks, the coefficient in the top left corner is called the DC coefficient, all others are called AC coefficients. The energy frequency that each coefficient represents becomes higher as the coefficient is positioned further from the DC coefficient.
The coefficients to the right of the DC coefficient represent higher horizontal frequencies than the left. The coefficients below the DC coefficient represent higher vertical frequencies than the ones above and the coefficients in the diagonal directions represent higher horizontal and vertical frequencies as they approach the coefficient on the right bottom.

Figure 10 – 8-8 DCT transform



8-8 DCT Block and the pixel coordinate

	0	1	2	3	4	5	6	7
0	P00	P10	P20	P30	P40	P50	P60	P70
1	P01	P11	P21	P31	P41	P51	P61	P71
2	P02	P12	P22	P32	P42	P52	P62	P72
3	P03	P13	P23	P33	P43	P53	P63	P73
4	P04	P14	P24	P34	P44	P54	P64	P74
5	P05	P15	P25	P35	P45	P55	P65	P75
6	P06	P16	P26	P36	P46	P56	P66	P76
7	P07	P17	P27	P37	P47	P57	P67	P77

Arrangement of the coefficients

	0	1	2	3	4	5	6	7
0	DC	AC10	AC20	AC30	AC40	AC50	AC60	AC70
1	AC01	AC11	AC21	AC31	AC41	AC51	AC61	AC71
2	AC02	AC12	AC22	AC32	AC42	AC52	AC62	AC72
3	AC03	AC13	AC23	AC33	AC43	AC53	AC63	AC73
4	AC04	AC14	AC24	AC34	AC44	AC54	AC64	AC74
5	AC05	AC15	AC25	AC35	AC45	AC55	AC65	AC75
6	AC06	AC16	AC26	AC36	AC46	AC56	AC66	AC76
7	AC07	AC17	AC27	AC37	AC47	AC57	AC67	AC77

It is important to know some basic facts about DCT.

- DCT does not compress anything. It arranges the video data in preparation to apply VLC in which the actual compression takes place.
- DCT transforms the video data from the spatial domain to the frequency domain. This means that the transformed DCT coefficients represent the ‘change’ of signal amplitude (pixel amplitude) within the DCT block instead of the amplitude itself.
- DCT is a ‘lossless’ process provided the coefficient accuracies are preserved.
- In the DVCAM/DV formats, the 8-bit pixels are transformed into 10-bit DCT coefficients.
- Most compression schemes including MPEG-2 use DCT.

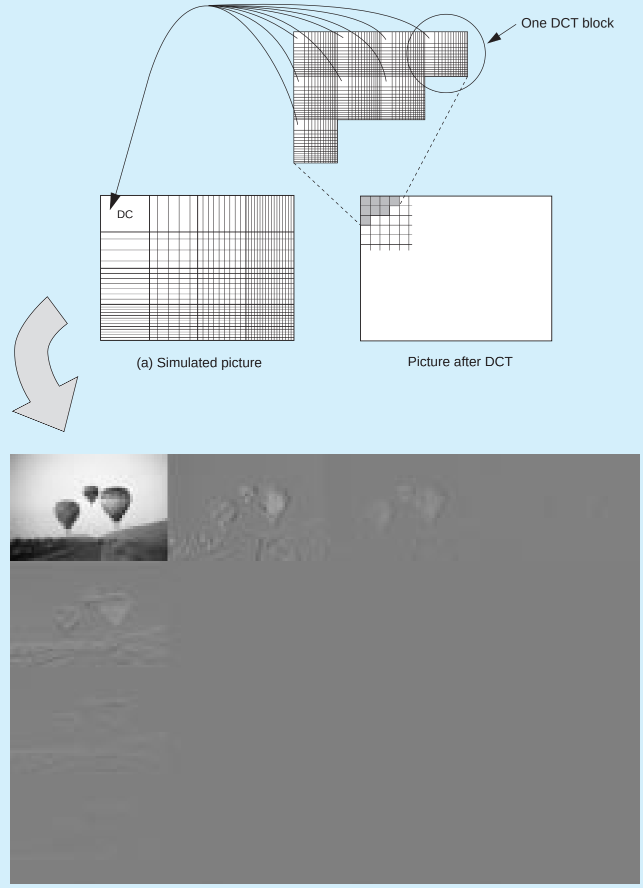
Of most importance is to know why the video must be transformed from the spatial domain to the frequency to be compressed.

Excluding extremely busy graphics, most video images contain a majority of the picture content in the low frequency areas and relatively very little content in the high frequency areas. This means that after the DCT transform, the coefficients near the DC coefficient have larger values than those near the bottom right corner of the DCT block.

Figure 11 is a visual representation of simulating the results of DCT transformation. For simplicity, a 4 x 4 block was used. The simulation was performed so that all coefficients (in each DCT block) with the same horizontal/vertical frequencies were gathered into one picture in the order of their positions within the entire video screen. This means that the upper left-hand picture would represent only the picture content of the DC coefficient. Notice that most of the video content is represented by this picture. It is also important to note that the other pictures show very little content.

It is easy to understand here that the low frequency DC coefficient is of most significance and that a significant number of bits should be allocated. On the contrary, allocating less bits to the high frequency coefficients will not effect the original picture much.

Figure 11 – DCT transform



(a) Simulated picture

Picture after DCT

(b) Simulated picture

Base band video does not take this into consideration and the video signal is merely sampled and encoded by its luminance and chrominance signal amplitudes. This results in the same number of bits being allocated for a flat picture with no information as well as an image that contains detailed picture content.

As mentioned earlier, using DCT in compression for converting the video signal into its frequency components allows the video signal to be encoded so that bit allocation can be applied depending on how much information each frequency represents in the entire video image. As explained earlier, most video signals contain a majority of the picture information in the low frequency areas. This means that low frequency components are most vital in the video signal and should be reproduced most accurately. Accordingly, a high bit resolution for low frequency areas is essential. In contrast, high frequency components represent less information in the picture and are less easily perceived, thus lower bit resolutions can be assigned.

◆ Two DCT modes

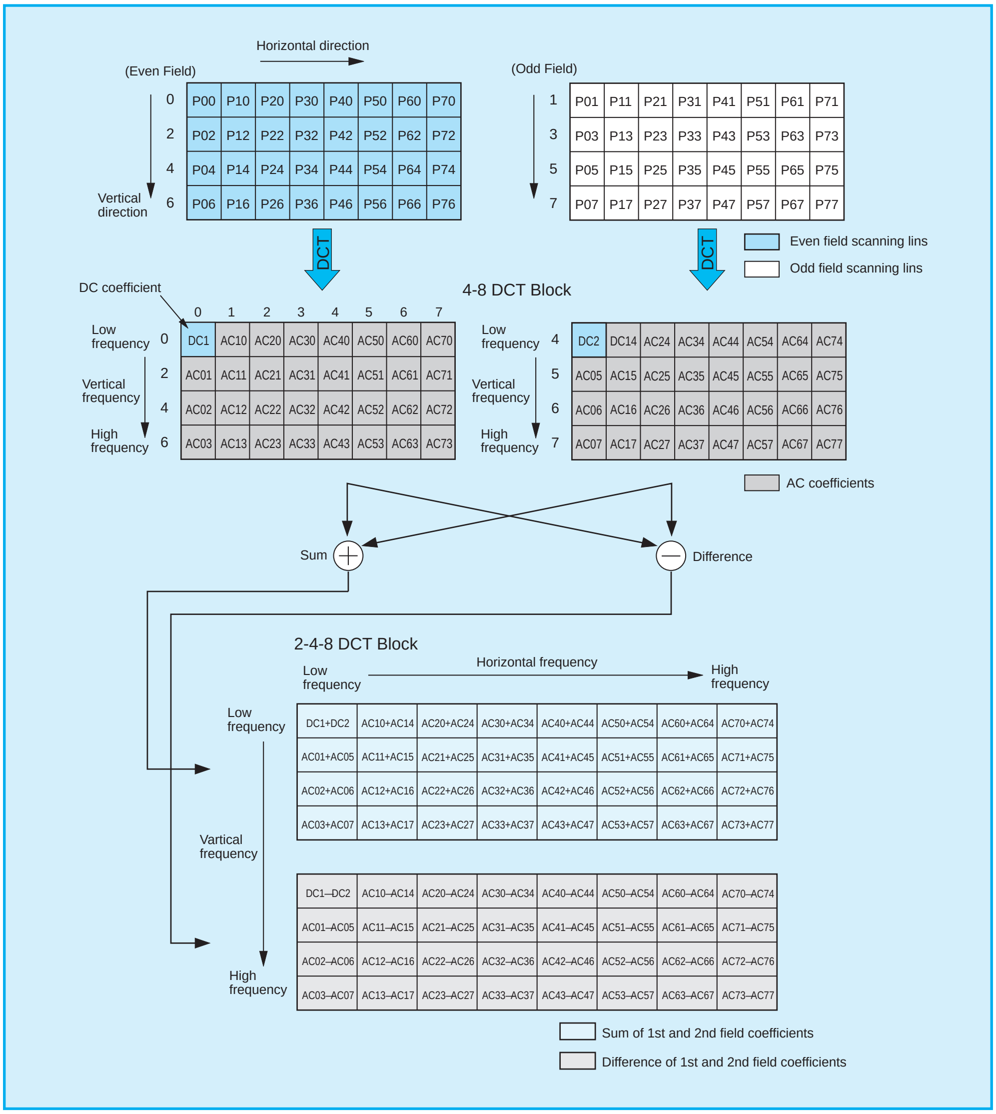
Since intra-frame based compression is used, there are two DCT modes in DVCAM/DV to cope with the amount of picture movement in a frame. The 8-8-DCT mode and the 2-4-8-DCT mode (Figure 12).

8-8-DCT mode is selected when there is no motion and the difference between the odd and even fields is small.

2-4-8-DCT mode is selected when there is motion and the difference between the two fields is significant.

This mode selection technique is important to keep good picture quality regardless of whether the picture is moving or static.

Figure 12 – 2-4-8 DCT mode



(Even Field)

0	P00	P10	P20	P30	P40	P50	P60	P70
2	P02	P12	P22	P32	P42	P52	P62	P72
4	P04	P14	P24	P34	P44	P54	P64	P74
6	P06	P16	P26	P36	P46	P56	P66	P76

(Odd Field)

1	P01	P11	P21	P31	P41	P51	P61	P71
3	P03	P13	P23	P33	P43	P53	P63	P73
5	P05	P15	P25	P35	P45	P55	P65	P75
7	P07	P17	P27	P37	P47	P57	P67	P77

	0	1	2	3	4	5	6	7
0	DC1	AC10	AC20	AC30	AC40	AC50	AC60	AC70
2	AC01	AC11	AC21	AC31	AC41	AC51	AC61	AC71
4	AC02	AC12	AC22	AC32	AC42	AC52	AC62	AC72
6	AC03	AC13	AC23	AC33	AC43	AC53	AC63	AC73

4	DC2	DC14	AC24	AC34	AC44	AC54	AC64	AC74
5	AC05	AC15	AC25	AC35	AC45	AC55	AC65	AC75
6	AC06	AC16	AC26	AC36	AC46	AC56	AC66	AC76
7	AC07	AC17	AC27	AC37	AC47	AC57	AC67	AC77

DC1+DC2	AC10+AC14	AC20+AC24	AC30+AC34	AC40+AC44	AC50+AC54	AC60+AC64	AC70+AC74
AC01+AC05	AC11+AC15	AC21+AC25	AC31+AC35	AC41+AC45	AC51+AC55	AC61+AC65	AC71+AC75
AC02+AC06	AC12+AC16	AC22+AC26	AC32+AC36	AC42+AC46	AC52+AC56	AC62+AC66	AC72+AC76
AC03+AC07	AC13+AC17	AC23+AC27	AC33+AC37	AC43+AC47	AC53+AC57	AC63+AC67	AC73+AC77

DC1–DC2	AC10–AC14	AC20–AC24	AC30–AC34	AC40–AC44	AC50–AC54	AC60–AC64	AC70–AC74
AC01–AC05	AC11–AC15	AC21–AC25	AC31–AC35	AC41–AC45	AC51–AC55	AC61–AC65	AC71–AC75
AC02–AC06	AC12–AC16	AC22–AC26	AC32–AC36	AC42–AC46	AC52–AC56	AC62–AC66	AC72–AC76
AC03–AC07	AC13–AC17	AC23–AC27	AC33–AC37	AC43–AC47	AC53–AC57	AC63–AC67	AC73–AC77

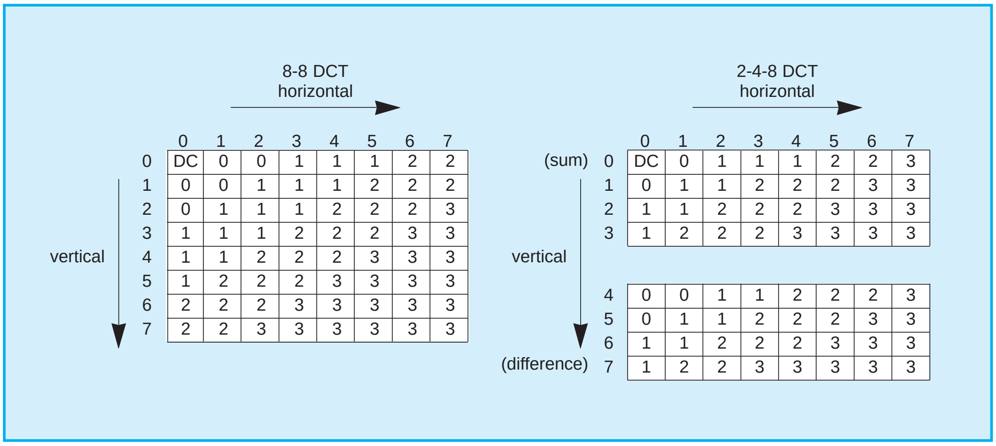
◆ Quantization process

In the quantization process, each transformed AC coefficient is divided by a certain number in order to limit the amount of data in a video segment. The number each AC coefficient is divided by is determined by a so-called quantizer which is an 8 x 8 table with coefficients equal to or larger than one.
It is important to note the following.

- Compression takes place at both the quantization process and the following VLC process. However, note that the quantization process is lossy whereas the VLC process is lossless.
- The quantization process also controls the ‘bit budget’ in order to achieve a 25 Mb/s transfer rate as the end result of compression – the transfer rates of the DVCAM/DV formats.

In the quantization process, the DC coefficient is not quantized since it contains the majority of the picture content.
The AC coefficients in the quantizer are determined by the class number and quantization number (QNO) as shown in **Figure 14** and **Table 5**. The class number is decided by estimating the activity (picture gradation) of each DCT block.
The quantization number, on the other hand, is determined so that the bit rate does not exceed the allowed rate with the selected class number.
Coefficients in the quantizer are determined by using **Table 5** and **Figure 13** as follows.
The coefficients of the quantizer table are represented by so-called ‘areas’ numbered from 0 to 3 (**Figure 13**). Areas with the same number (area number) have the same coefficients.

Figure 13 – Area numbers



	0	1	2	3	4	5	6	7
0	DC	0	0	1	1	1	2	2
1	0	0	1	1	1	2	2	2
2	0	1	1	1	2	2	2	3
3	1	1	1	2	2	2	3	3
4	1	1	2	2	2	3	3	3
5	1	2	2	2	3	3	3	3
6	2	2	2	3	3	3	3	3
7	2	2	3	3	3	3	3	3

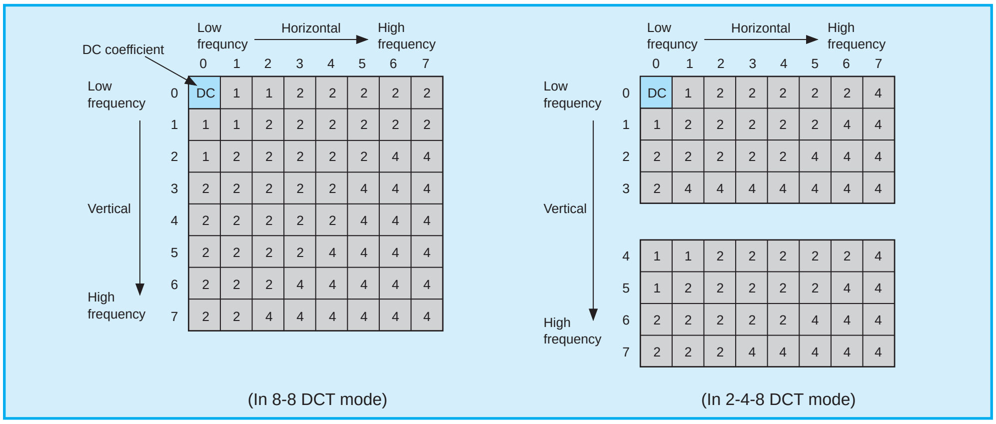
	0	1	2	3	4	5	6	7
0	DC	0	1	1	1	2	2	3
1	0	1	1	2	2	2	3	3
2	1	1	2	2	2	3	3	3
3	1	2	2	2	3	3	3	3
4	0	0	1	1	2	2	2	3
5	0	1	1	2	2	2	3	3
6	1	1	2	2	2	3	3	3
7	1	2	2	3	3	3	3	3

These area numbers correspond to the area numbers in **Table 5**.
For example, if the class number was ‘2’ and the quantization number was ‘10’, the coefficients in the quantizer table would be as shown in **Figure 14**.

It is important to note that if the 'activity' is large, a large class number is selected. This means that macro blocks with larger activity are divided by larger numbers, because they are less distorted even with coarse quantization. On the other hand, for areas with smaller activity such as smooth gradients, mild quantization is applied since distortion is more visible. In such cases, the quantizing uses a smaller class number.
It should be noted that when class number 3 is selected, one initial bit shift is applied before quantizing each AC coefficient, which is why class number 3 of Table 5 is shifted.

The quantization is applied so that the target bit rate after VLC is close to the transfer rate of the DVCAM/DV formats of 25 Mb/s.

Figure 14 – Example of using quantizer



	0	1	2	3	4	5	6	7
0	DC	1	1	2	2	2	2	2
1	1	1	2	2	2	2	2	2
2	1	2	2	2	2	2	4	4
3	2	2	2	2	2	4	4	4
4	2	2	2	2	2	4	4	4
5	2	2	2	2	4	4	4	4
6	2	2	2	4	4	4	4	4
7	2	2	4	4	4	4	4	4

(In 8-8 DCT mode)

	0	1	2	3	4	5	6	7
0	DC	1	2	2	2	2	2	4
1	1	2	2	2	2	2	4	4
2	2	2	2	2	2	4	4	4
3	2	4	4	4	4	4	4	4
4	1	1	2	2	2	2	2	4
5	1	2	2	2	2	2	4	4
6	2	2	2	2	2	4	4	4
7	2	2	2	4	4	4	4	4

(In 2-4-8 DCT mode)

Table 5 – Quantization step

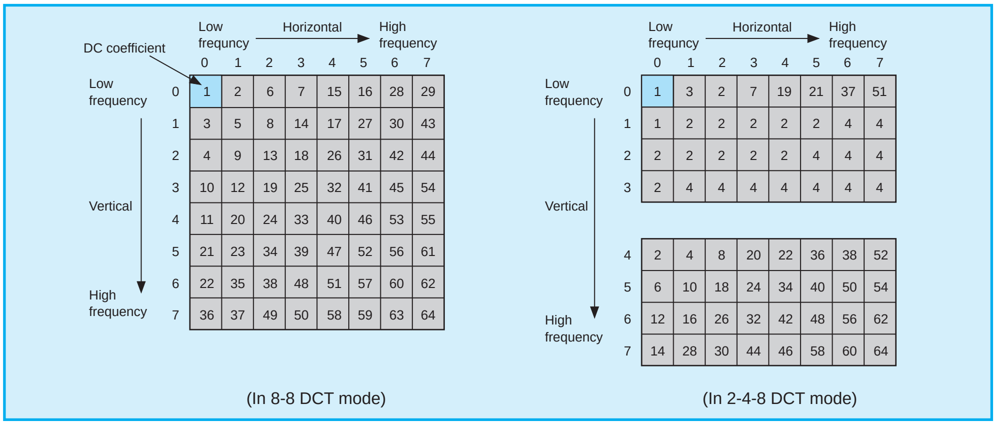
	Class number				Area number			
	0	1	2	3	0	1	2	3
Quantization number (QNO)	15				1	1	1	1
	14				1	1	1	1
	13				1	1	1	1
	12	15			1	1	1	1
	11	14			1	1	1	1
	10	13		15	1	1	1	1
	9	12	15	14	1	1	1	1
	8	11	14	13	1	1	1	2
	7	10	13	12	1	1	2	2
	6	9	12	11	1	1	2	2
	5	8	11	10	1	2	2	4
	4	7	10	9	1	2	2	4
	3	6	9	8	2	2	4	4
	2	5	8	7	2	2	4	4
	1	4	7	6	2	4	4	8
	0	3	6	5	2	4	4	8
		2	5	4	4	4	8	8
		1	4	3	4	4	8	8
		0	3	2	4	8	8	16
			2	1	4	8	8	16
			1	0	8	8	16	16
			0		8	8	16	16

◆ VLC

As mentioned earlier, the actual compression is applied at the VLC (Variable Length Coding) process.

The VLC transforms fixed length quantized AC coefficients to variable length code words. After the quantization process, the DC coefficient is output from the quantizer followed by the AC coefficients as shown in **Figure 15**. This method is called zigzag scan. The AC coefficients are output in the order of their frequencies. Note that the DC coefficients and AC coefficients are coded separately.

Figure 15 – Coefficient readout order (zigzag scan)



	0	1	2	3	4	5	6	7
0	1	2	6	7	15	16	28	29
1	3	5	8	14	17	27	30	43
2	4	9	13	18	26	31	42	44
3	10	12	19	25	32	41	45	54
4	11	20	24	33	40	46	53	55
5	21	23	34	39	47	52	56	61
6	22	35	38	48	51	57	60	62
7	36	37	49	50	58	59	63	64

(In 8-8 DCT mode)

	0	1	2	3	4	5	6	7
0	1	3	2	7	19	21	37	51
1	1	2	2	2	2	2	4	4
2	2	2	2	2	2	4	4	4
3	2	4	4	4	4	4	4	4
4	2	4	8	20	22	36	38	52
5	6	10	18	24	34	40	50	54
6	12	16	26	32	42	48	56	62
7	14	28	30	44	46	58	60	64

(In 2-4-8 DCT mode)

The zigzag scan results in the output AC coefficients forming a 'run' which is a stream of consecutive AC coefficients that have the same value. In most cases, AC coefficients near the lower left corner of the quantized block are '0' and a run of zeros is formed.

The VLC codes each 'zero run' of AC coefficients up to the next non-zero AC coefficient. Each codeword representing a run is determined by the length of the zero run and the amplitude of the non-zero AC coefficient that follows it.

For example, imagine a DCT block that does not have much activity and assume that some of the AC coefficients are zero. As highlighted in **Table 6**, if the length of the zero run is '5' and the amplitude of the non-zero AC coefficient following the run is '1', the codeword length is 7 bits. One bit is further added as a sign bit, forming the resultant 8-bit codeword. The sign bit is described with an 's'.

The actual codeword is determined as shown in **Table 7**. Note that this is only part of the entire codeword Table.

The highlighted column shows the codeword selected when the length of the zero run is '5' and the non-zero AC coefficient is '1'. Note that the resultant codeword is '1101000s' – an 8-bit word.

Since the original data before this process was 54 bits (9 bits x 6 AC coefficients), it is easy to understand how DCT blocks with little activity can be effectively compressed via the VLC process.

This method is called modified 2-dimensional Huffman coding.

Table 6 – Length of codewords

Run length	Amplitude																								
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	--------	255
0	11	2	3	4	4	5	5	6	6	7	7	7	8	8	8	8	8	8	9	9	9	9	9	15	--------	15
1	11	4	5	7	7	8	8	8	9	10	10	10	11	11	11	12	12	12								
2	12	5	7	8	9	9	10	12	12	12	12	12														
3	12	6	8	9	10	10	11	12																		
4	12	6	8	9	11	12																				
5	12	7	9	10																						
6	13	7	9	11																						
7	13	8	12	12																						
8	13	8	12	12																						
9	13	8	12																							
10	13	8	12																							
11	13	9																								
12	13	9																								
13	13	9																								
14	13	9																								
15	13																									
⋮	⋮																									
61	13																									

Note 1: Sign bit is not included. Note 2: The length of EOB = 4

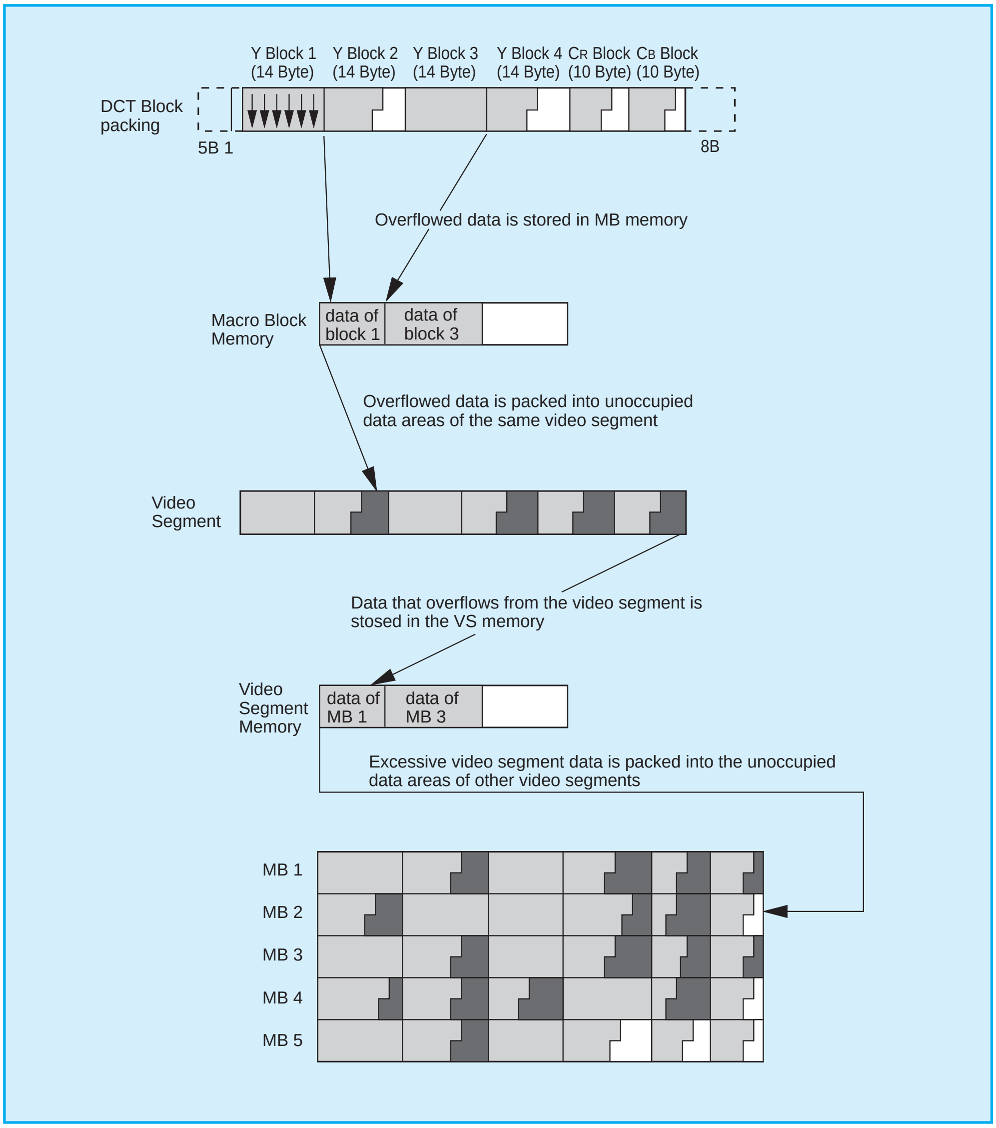
Table 7 – Modified 2-dimensional Huffman coding

(run, amp)		Code	Length	(run, amp)		Code	Length
0	1	0 0 s	2+1	5	1	1 1 0 1 0 0 0 s	7+1
0	2	0 1 0 s	3+1	6	1	1 1 0 1 0 0 1 s	
EOB		0 1 1 0	4	2	2	1 1 0 1 0 1 0 s	
1	1	0 1 1 1 s	4+1	1	3	1 1 0 1 0 1 1 s	
0	3	1 0 0 0 s		1	4	1 1 0 1 1 0 0 s	
0	4	1 0 0 1 s		0	9	1 1 0 1 1 0 1 s	
2	1	1 0 1 0 0 s	5+1	0	10	1 1 0 1 1 1 0 s	
1	2	1 0 1 0 1 s		0	11	1 1 0 1 1 1 1 s	
0	5	1 0 1 1 0 s		7	1	1 1 1 0 0 0 0 0 s	8+1
0	6	1 0 1 1 1 s		8	1	1 1 1 0 0 0 0 1 s	
3	1	1 1 0 0 0 0 s	6+1	9	1	1 1 1 0 0 0 1 0 s	
4	1	1 1 0 0 0 1 s		10	1	1 1 1 0 0 0 1 1 s	
0	7	1 1 0 0 1 0 s		⋮	⋮	⋮	⋮
0	8	1 1 0 0 1 1 s					

◆ Framing

The length (data rate) of the compressed data stream after zigzag scan and VLC is not the same for each quantized DCT. This is because the original DCT blocks each had different picture content. This also means that the length of the compressed data stream for each macro block is not the same. As mentioned earlier, the 25 Mb/s constant bit rate is achieved on the video segment level, not on the DCT block or macro block level.
On the other hand, by appropriately arranging the data in the compressed stream, a constant 25 Mb/s output can be achieved. This process is called framing.
The framing process is shown in **Figure 16.**

Figure 16 – Arrangement of a compressed macro block



As shown in Figure 16, the framing algorithm is comprised of three steps.
Note that this process takes place within a memory, prior to recording to tape.

Step 1: The data of each compressed stream representing a DCT block is packed into its associated memory areas (14 bytes) in the order of the Y1 – Y4, CR and CB data streams. As explained earlier, the data size (or length) of each stream varies – some do not fit in the designated 14-byte area (overflow) while others do not occupy it entirely.

Step 2: The data that overflowed from the 14-byte areas in Step 1 is stored in a macro block memory. This data is then packed in other 14-byte areas within the same macro block data stream that are not fully occupied.

Step 3: If the data still overflows after Step 2, this excessive data is stored in a video segment memory. The excessive data is then packed in other macro block data streams, which were not filled entirely.

Using these three steps, the data output from the VLC process is averaged across five macro blocks to achieve a constant 25 Mb/s data rate.

◆ Video product block

After the framing process, the de-shuffling process takes place and the macro block streams are re-arranged to their original positions – their locations at the ‘blocking process’. Although the original macro blocks were formed of 8 x 8 pixel blocks; after de-shuffle, they are compressed data streams which represent the picture content of each macro block.
Prior to recording to tape, the compressed macro block streams are read out to a memory to form a so-called ‘product block’. The macro block streams are read out to the product block memory in the order shown in Figure 17. The resultant product block is shown in Figure 18.
The product block in Figure 18 represents the data recorded to one video track.
For video data, there are 77 columns and 135 rows. The compressed data stream of one macro block (77 bytes) is stored across one row. There are 135 rows because there are 135 macro blocks in a track. In addition, there are three rows for Video auxiliary (VAUX) data. Each row is called a Sync block and includes a sync signal, an ID signal, the video data and the inner correction parities.

As the figure shows, each row starts with the sync signal, so correct synchronization can be maintained even under situations where large errors are seen, for example may occur with large burst errors.
The ID indicates information such as the frame sequence number, the track number, and the arrangement of the sync block within the track.
The inner correction parities are provided to correct errors which still remain even after the outer correction is performed.

Figure 17 – Readout order (525-60 system)

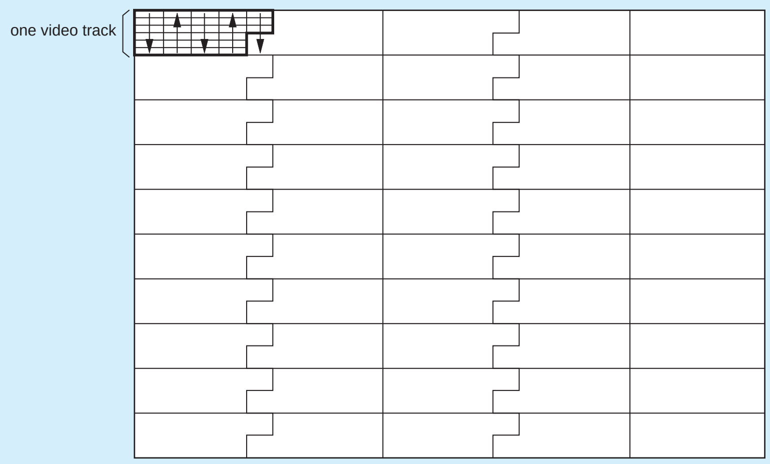
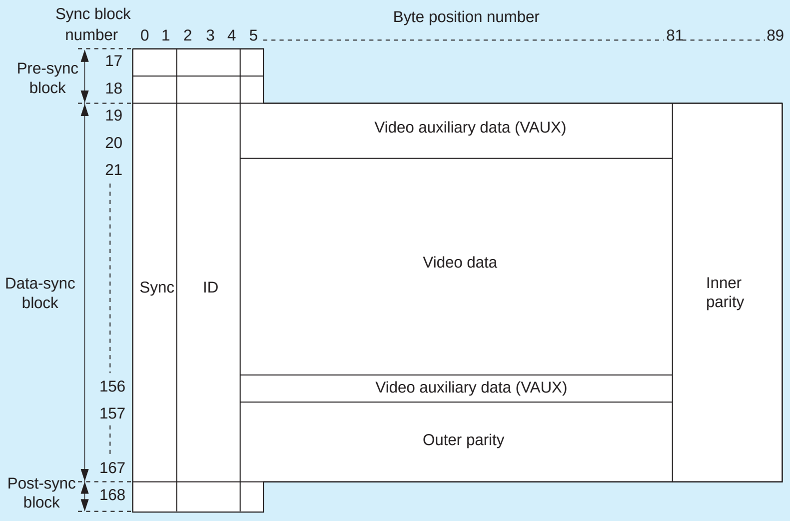


Figure 18 – Structure of video sector



◆ Video auxiliary data

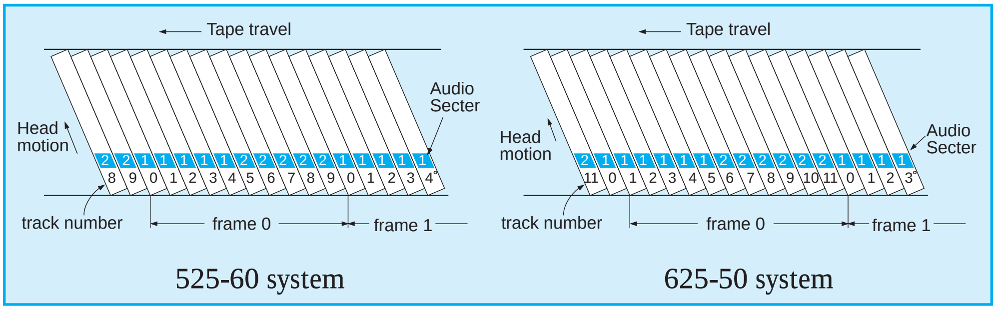
When the video product block is constructed, Video auxiliary data (VAUX) is multiplexed with the compressed video data, providing provision to write a variety of user data on the tape. Some examples of its applications are given later.

4

Audio Signal Processing

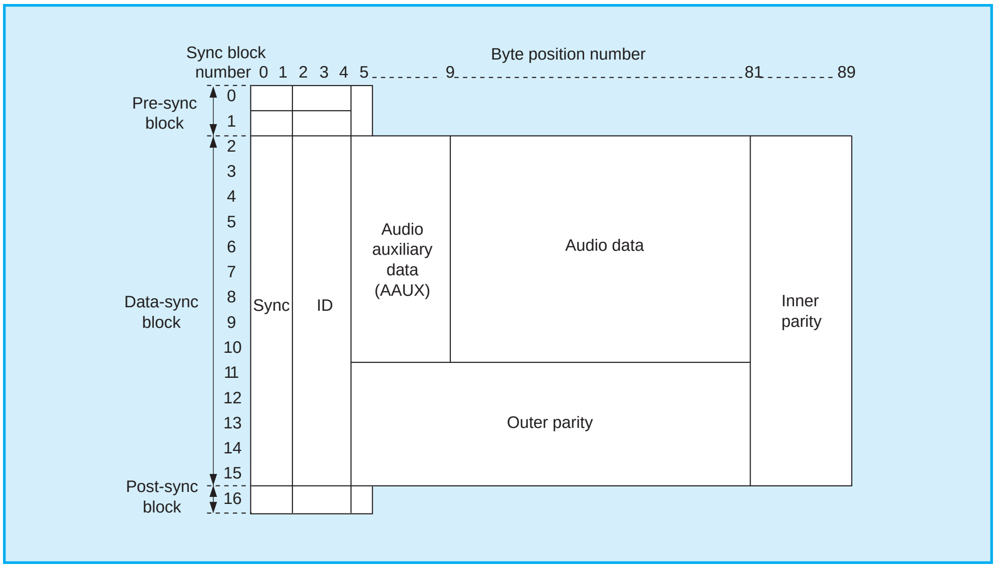
The audio signal is recorded on two audio blocks. Each audio block is processed independently and identically. The audio block is composed of five audio sectors in five consecutive tracks for the 525-60 system, six audio sectors in six consecutive tracks for the 625-50 system. **Figure 19** shows the audio track allocation.

Figure 19 – Audio track allocation



Each audio sector is processed in a product block with a dimension of 77 columns by nine rows. Notice that the length of one row in the product block (one sync block) is equivalent to that of video. The same sync block length was selected to simplify the processing circuitry. As with video, audio auxiliary data (AAUX) is multiplexed with each audio sector in the product block as shown in **Figure 20**. Since audio is sensitive to burst errors, the audio samples are shuffled within the audio block before the error correction data is added in the product block.
Inner error correction parities and outer error correction parities protect the audio data.

Figure 20 – Structure of audio sector



◆ Audio encoding mode

Audio encoding modes are defined in each audio block. They are classified by the sampling frequency, bit resolution and the number of channels in the audio block. This standard provides two types of audio encoding modes whose parameters are defined in Table 8.

Table 8 – Construction of audio block

Audio block		CH1	CH2
Track position	525-60 system	Track 0 to 4	Track 5 to 9
	625-50 system	Track 0 to 5	Track 6 to 11
Encoding mode	2-ch audio	48-kHz mode	48-kHz mode
	4-ch audio	32-kHz/2-ch mode	32-kHz/2-ch mode

In 48 K mode, one channel of audio is recorded in each of the two audio blocks, giving one-pair of stereo audio sampled at 48 kHz frequency. The encoded data is expressed by 2's complement representation with 16-bit linear resolution.

In 32-k/2-ch mode, two channels of audio signals are recorded in each of the two audio blocks, giving two-pairs of stereo audio sampled at 32 kHz frequency. The encoded data is expressed by 2's complement representation with 12-bit nonlinear quantization.

5 System Data

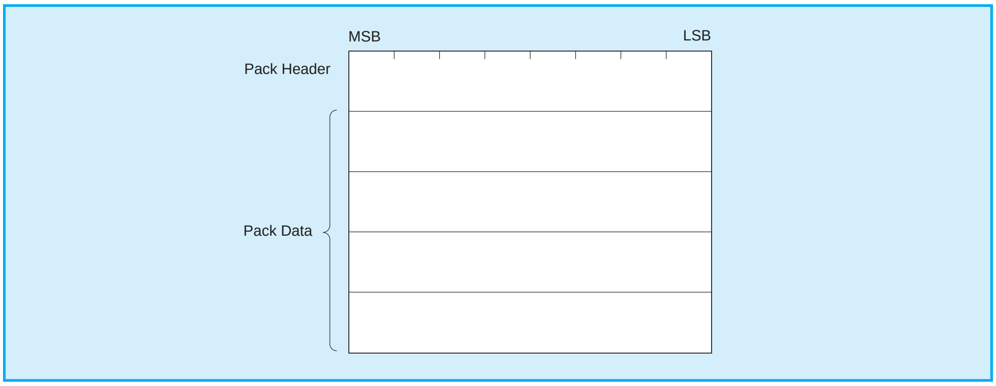
In the DVCAM format, system data is recorded in the Subcode, Video, Audio and ITI sectors of the helical tracks.

◆ System data pack structure

The DVCAM/DV formats adopt a fixed length pack structure to store the system data. The pack structure is optimized to reduce the hardware complexity for its storage and readout.

A pack consists of five bytes. The first byte is the pack header, the other four bytes are for the actual data related to the pack name as shown in **Figure 21**.

Figure 21 – Pack structure



Each pack has an eight bit-length header. The most significant four bits of the pack header are the upper header. A group of 16 packs can be designated for the upper header using the least significant four bits.

The pack adopts a layer structure. Using bit allocation, up to three levels of layers are permitted as shown in **Figure 22** and **Table 9**.

Figure 22 – Three levels of layer

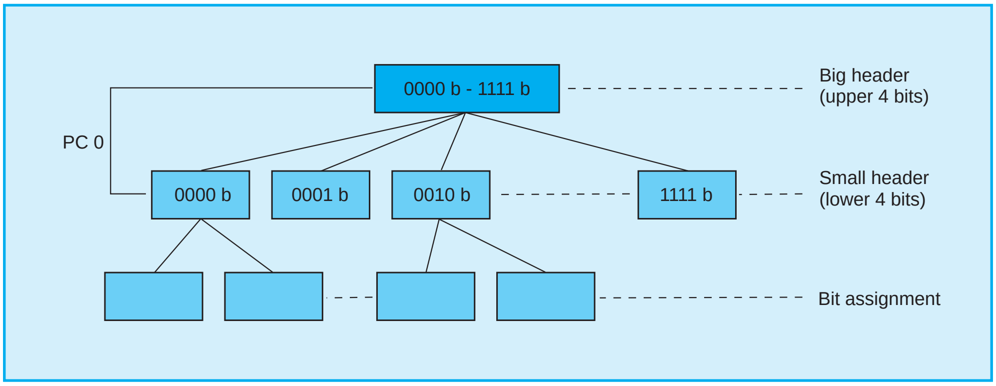


Table 9 – Pack group

Big header	Group name	Contents
0	Control	Pack related to video control
1	Title	Pack related to title
2	Chapter	Pack related to chapter
3	Part	Packrelated to part
4	Program	Pack related to program
5	AAUX	Pack related to Audio AUX
6	VAUX	Pack related to Video AUX
7	Camera	Pack related to camera
8	Line	Pack related to horizontal line
9-Eh	Reserved	Reserved
F	Soft mode	Pack related to option

The following is a summary of the DVCAM system data.

1. Subcode

Timecode and Binary group information is recorded in the Subcode sector.
The contents conform to SMPTE/EBU timecode standards.

2. AAUX

Audio source and audio source control information are recorded in the AAUX area of the Audio sector.
The contents of the audio source/source control information are as follows.

• Locked mode flag:	Locking condition of the audio sampling frequency and video signal
• Audio frame size:	The number of audio samples per frame
• Signal type information:	50 or 60 fields system
• Sampling frequency:	48 kHz or 32 kHz
• Quantization:	16 bits or 12 bits

3. VAUX

Video source and source control, Record date/time and Closed caption information are recorded in the VAUX area of the Video sector.
The contents of the video source/source control information are as follows.

• Signal type information:	50 or 60 fields system
• Display select mode information:	Aspect ratio

6

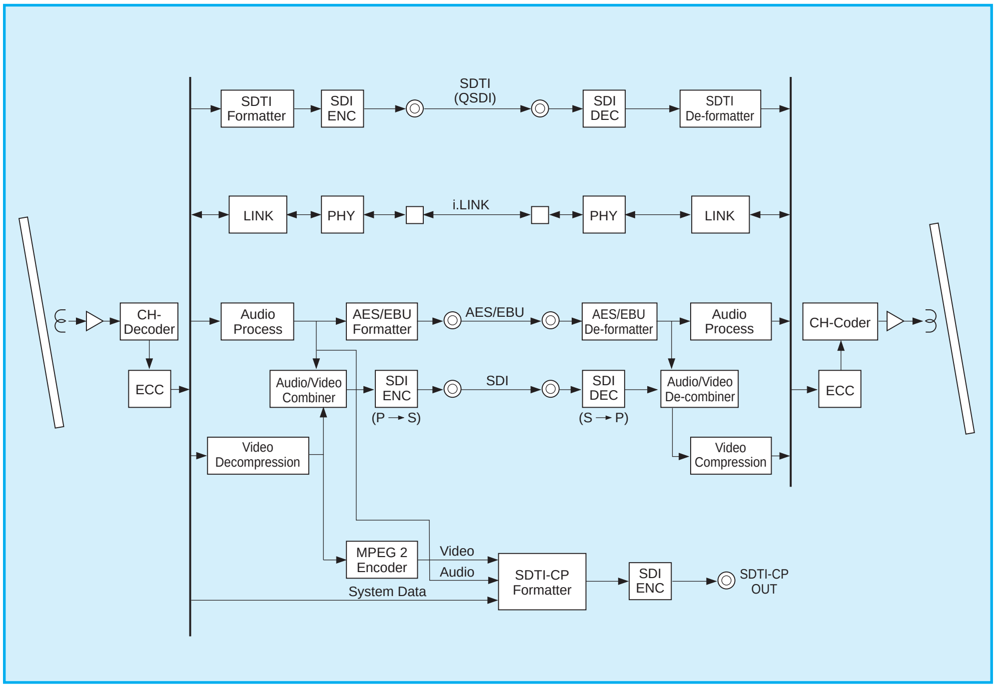
Interfaces

To satisfy a variety of system requirements, the DVCAM offers versatile digital interfaces as follows.

6-1 SDI

SDI is the standard digital interface to transfer uncompressed video and audio data in real time. DVCAM supports* 4:2:2 component digital video signals and four channels of digital audio signals.
Figure 23 shows a block diagram of the signal processing.
The play back video data is de-compressed to baseband 4:1:1 (525) / 4:2:0 (625) and then converted to 4:2:2 signals at the video de-compression block.
As mentioned before, the audio sampling structure of the DVCAM format is 48 kHz/16 bits/two channels or 32 kHz/12 bits/four channels. In the case of 32 kHz/12 bits/four channels, audio data is first converted to 48 kHz/16 bits/four channels then mapped to SDI.
*Note: Not for all models.

Figure 23 – Block diagram of signal processing



6-2 SDTI (QSDI)

The SDTI (QSDI) interface is for transferring compressed video data, uncompressed audio data and system data such as timecode, video and audio AUX data. SDTI (QSDI) is very useful for dubbing and connection with nonlinear editing systems, because the video data is transferred as compressed data with no quality degradation and reduced codec delay.

SDTI (QSDI) interface for DVCAM is standardized as SMPTE 322M.

6-3 SDTI-CP

Interconnectivity between systems with different compression formats is highly important in future operations.

Versions of DVCAM equipment support the SDTI-CP interface to feed MPEG-2 production systems.

In order to interface to MPEG systems, the DVCAM data is first transcoded to produce MPEG-2 Video Elementary Stream data which is then placed in the SDTI-CP interface together with audio and system data.

This interface not only has the capability to feed DVCAM sourced material but also becomes a bridge from the DV family 25 Mb/s format to the MPEG world.

6-4 i.LINK

i.LINK is a high-speed digital serial interface which carries video, audio, system data and control signals. i.LINK enables dubbing between two VTRs or editing operations via a single cable connection without the need of an RS-422A control cable.

Under the current specifications of its standards, transmission at up to 400 Mb/s is possible. 100 Mb/s is used for the DV and DVCAM interfaces.

The i.LINK interface for DVCAM is based on the following standards.

1. 1394-1995 IEEE Standard for a High Performance Serial Bus
2. AV/C Protocol
   2-1. IEC 61883-1 (1998-02) Ed. 1.0 Consumer audio/video equipment - Digital interface
      - Part 1: General
   2-2. IEC 61883-2 (1998-02) Ed. 1.0 Consumer audio/video equipment - Digital interface
      - Part 2: SD-DVCR data transmission
3. DV format documents
   3-1. IEC 61834-1 (1998-08) Ed. 1.0
   Recording - Helical-scan digital video cassette recording system using 6.35 mm magnetic tape for consumer use (525-60, 625-50, 1125-60 and 1250-50 systems)
      - Part 1: General specifications
   3-2. IEC 61834-2 (1998-08) Ed. 1.0
   Recording - Helical-scan digital video cassette recording system using 6.35 mm magnetic tape for consumer use (525-60, 625-50, 1125-60 and 1250-50 systems)
      - Part 2: SD format for 525-60 and 625-50 systems
4. AV/C Digital Interface Command Set
   4-1. AV/C Digital Interface Command Set General Specification Version 3.0
   TA Document 1998003
   4-2. AV/C Tape Recorder/Player Subunit Specification Version 2.1
   TA Document 1998012

SONY

Distributed by

V-11096
MK7507V1TC00JUN

Printed in Japan on recycled paper