



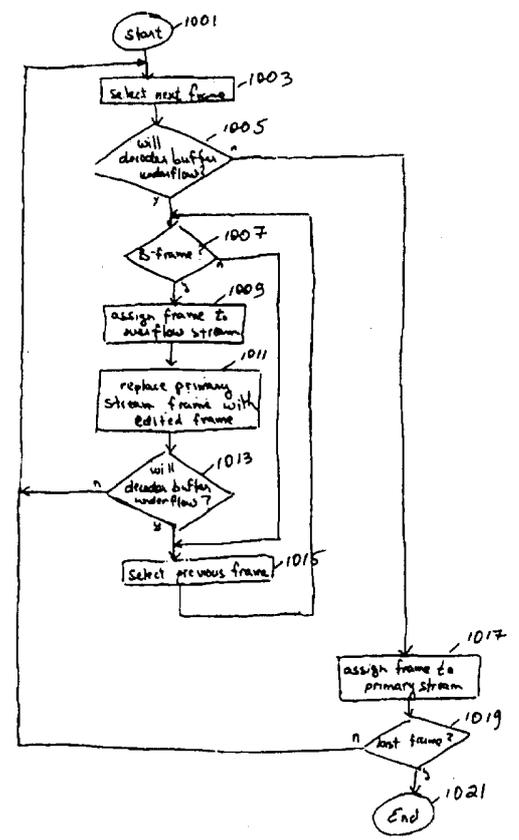
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification <sup>6</sup> : H04N 7/26, 7/58, 7/46</p>	<p>A1</p>	<p>(11) International Publication Number: <b>WO 97/19559</b> (43) International Publication Date: 29 May 1997 (29.05.97)</p>
<p>(21) International Application Number: PCT/US96/18091 (22) International Filing Date: 8 November 1996 (08.11.96) (30) Priority Data: 08/561,296 21 November 1995 (21.11.95) US (71) Applicant: IMEDIA CORPORATION [US/US]; Suite 2850, 425 Market Street, San Francisco, CA 94105 (US). (72) Inventors: SHEN, Paul; Apartment 2B, 1945 Pacific Avenue, San Francisco, CA 94109 (US). ARAZI, Efraim; 2780 Broadway, San Francisco, CA 94115 (US). KRAUSE, Edward, A.; 8318 Terrace Drive, El Cerrito, CA 94530 (US). TOM, Adam, S.; Apartment 3, 1717 Jones Street, San Francisco, CA 94109 (US). (74) Agents: HARTOGS, Michael, D. et al.; McCutchen, Doyle, Brown &amp; Enersen, Three Embarcadero Center, San Francisco, CA 94111 (US).</p>		<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b> <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>

(54) Title: METHOD AND APPARATUS FOR MODIFYING ENCODED DIGITAL VIDEO FOR IMPROVED CHANNEL UTILIZATION

(57) Abstract

A method for increasing channel utilization for a data channel transmitting a multiplex of a set of one or more encoded program streams, each program stream being decodeable by a corresponding decoder, each corresponding decoder including a corresponding decoder buffer, the decoder buffers having a maximum allowable size, the method comprising selecting encoded pictures to be modified, said selecting according to a criterion which includes preventing any underflow of any decoder buffer, modifying each said selected encoded picture to form a corresponding modified encoded picture, said modified encoded picture having less data than said selected encoded picture, and transmitting the corresponding modified encoded pictures through the channel in place of the selected encoded pictures.



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<b>AM</b>	Armenia	<b>GB</b>	United Kingdom	<b>MW</b>	Malawi
<b>AT</b>	Austria	<b>GE</b>	Georgia	<b>MX</b>	Mexico
<b>AU</b>	Australia	<b>GN</b>	Guinea	<b>NE</b>	Niger
<b>BB</b>	Barbados	<b>GR</b>	Greece	<b>NL</b>	Netherlands
<b>BE</b>	Belgium	<b>HU</b>	Hungary	<b>NO</b>	Norway
<b>BF</b>	Burkina Faso	<b>IE</b>	Ireland	<b>NZ</b>	New Zealand
<b>BG</b>	Bulgaria	<b>IT</b>	Italy	<b>PL</b>	Poland
<b>BJ</b>	Benin	<b>JP</b>	Japan	<b>PT</b>	Portugal
<b>BR</b>	Brazil	<b>KE</b>	Kenya	<b>RO</b>	Romania
<b>BY</b>	Belarus	<b>KG</b>	Kyrgystan	<b>RU</b>	Russian Federation
<b>CA</b>	Canada	<b>KP</b>	Democratic People's Republic of Korea	<b>SD</b>	Sudan
<b>CF</b>	Central African Republic	<b>KR</b>	Republic of Korea	<b>SE</b>	Sweden
<b>CG</b>	Congo	<b>KZ</b>	Kazakhstan	<b>SG</b>	Singapore
<b>CH</b>	Switzerland	<b>LI</b>	Liechtenstein	<b>SI</b>	Slovenia
<b>CI</b>	Côte d'Ivoire	<b>LK</b>	Sri Lanka	<b>SK</b>	Slovakia
<b>CM</b>	Cameroon	<b>LR</b>	Liberia	<b>SN</b>	Senegal
<b>CN</b>	China	<b>LT</b>	Lithuania	<b>SZ</b>	Swaziland
<b>CS</b>	Czechoslovakia	<b>LU</b>	Luxembourg	<b>TD</b>	Chad
<b>CZ</b>	Czech Republic	<b>LV</b>	Latvia	<b>TG</b>	Togo
<b>DE</b>	Germany	<b>MC</b>	Monaco	<b>TJ</b>	Tajikistan
<b>DK</b>	Denmark	<b>MD</b>	Republic of Moldova	<b>TT</b>	Trinidad and Tobago
<b>EE</b>	Estonia	<b>MG</b>	Madagascar	<b>UA</b>	Ukraine
<b>ES</b>	Spain	<b>ML</b>	Mali	<b>UG</b>	Uganda
<b>FI</b>	Finland	<b>MN</b>	Mongolia	<b>US</b>	United States of America
<b>FR</b>	France	<b>MR</b>	Mauritania	<b>UZ</b>	Uzbekistan
<b>GA</b>	Gabon			<b>VN</b>	Viet Nam

## METHOD AND APPARATUS FOR MODIFYING ENCODED DIGITAL VIDEO FOR IMPROVED CHANNEL UTILIZATION

### I. Background of the Invention

#### A. Field of the Invention

5 The field of the present invention is video compression.

#### B. Background

The present invention relates to the encoding, formatting, and distribution of compressed video programs. It is particularly, but not necessarily, applicable to a particular video compression technique known  
10 as *variable bit-rate(VBR) encoding*. VBR encoding can be used to overcome the well-known problem of most video compression encoders where the perceived image quality tends to vary as a function of image complexity. Typically, a video program will contain a variety of scenes. Many of these scenes are lacking in motion or detail and are therefore  
15 easily compressed, while many other scenes contain complex details which are generally more difficult to compress, particularly when moving in a complex or random manner. Therefore, unless the available bandwidth is very high, the perceived quality of the decompressed and reconstructed images will tend to vary from one scene to the next. This problem becomes  
20 more serious as the available bandwidth is reduced until, eventually, the video quality becomes unacceptable, often because of just a few problem scenes.

VBR encoding overcomes this problem by allocating more bits to those scenes which are difficult to compress and fewer bits to those scenes which are more easily compressed. In this way, the decompressed and reconstructed images can be made to appear consistently uniform, and  
5 therefore superior to the reconstructed images derived from the constant bit-rate encoder adjusted for the same average rate of compression. As a result, it is possible to compress a video program more efficiently when using the VBR encoding technique. This not only increases the number and variety of programs or program streams that can be delivered over a  
10 fixed-bandwidth communication channel, but also reduces the storage capacity requirements at the head end or other site where the program library is maintained.

The disadvantage of the VBR encoding technique is that it presents certain problems when manipulating or editing the compressed bit-streams.  
15 In particular, it becomes difficult to efficiently utilize a fixed-bandwidth communication channel since the variable bit-rate stream may at times exceed the capacity of the channel, while at other times, it may utilize only a fraction of the available channel capacity. As would be clear to one in the art, the possibility of exceeding the capacity of the channel is not only a  
20 problem with VBR encoded data. The present invention is applicable also to non-VBR encoded data.

One known technique that is used to alleviate this problem is to buffer the compressed bit-stream at the transmission end of the communication channel in order to convert the variable rate stream to a  
25 constant rate stream. In such a case, it is also necessary to buffer the signal

received at the other end of the channel in order to recover the variable rate stream that is necessary for proper timing of the reconstructed video images. Unfortunately, the required amount of buffering would be prohibitively expensive and would introduce long delays into the distribution system. Moreover, existing video compression standards such as the MPEG standards, a set of International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) standards, specify limits on the amount of buffering required for conforming decoders. Therefore, it is important that the received bit-streams be decodable without exceeding these limits. That is, it is important to find methods for reducing the decoder memory requirements.

Another prior art technique that can be used to reduce the inefficiency of transmitting VBR encoded (and non-VBR encoded) programs over a fixed-bandwidth channel combines a plurality of program streams into a single multiplex. Although each additional program stream will increase the overall data rate of the multiplex, the variance of the average per stream data rate of this multiplex will tend to decrease in approximate proportion to the number of program streams, assuming approximate statistical independence of them. Therefore, if the nominal rate that can be accommodated by the channel is significantly greater than the average rate of a single program stream, then the channel utilization can be significantly improved. This technique is known in the art as *statistical multiplexing*.

When using buffering, one prior art method of assuring that buffer overflow does not occur is to add feedback between the encoder buffer and

the encoder. When the buffer approaches a full state, a buffer fullness signal from the buffer informs the encoder to reduce the bit-rate so that the buffer does not overflow. When the buffer has more room, the feedback signal, from the buffer to the encoder, enables the encoder to increase the bit-rate in order to maintain image quality. Such feedback is particularly effective when combined with statistical multiplexing. A single encoder buffer may be used, or several buffers. In the case of a single buffer, it would be at the output of the multiplexer, and feedback from that buffer would be to all the encoders of the programs being multiplexed. Because the multiplexing reduces effective per stream data rate variance, in general, less memory would be required to convert the multiplex to a fixed bit-rate stream than if each individual encoder included a VBR-to-constant bit-rate buffer.

Statistical multiplexing with feedback for conveying encoded, particularly VBR encoded program streams over fixed-bandwidth channels, can be effective, but generally may not be sufficient when uniform perceived image quality is important.

In addition, there are often situations where no feedback is possible between the output buffer of the multiplexer and the individual encoders. One such situation occurs when multiplexing *already encoded* data streams. Another occurs when the encoders are located in an area physically remote from the multiplexer. Both these situations are referred to herein as *remote encoding*, indicating that encoding and multiplexing are remote in time, location, or otherwise, so that no feedback is possible from the multiplexer to the encoders of the program streams to be multiplexed.

In addition, one of the factors that limits the efficiency of transmitting the multiplexed bit stream is the finite overall memory available for buffering at the decoder/demultiplexer. While statistical multiplexing usually provides sufficient benefit due to the small probability  
5 that all program streams will have complex scenes at the same time, there is still a finite chance of such data rate peaks occurring. To accommodate these peaks, long delays, and therefore large buffers, must be designed into the decoders in order to prevent the decoder buffers from underflowing when the capacity of the channel is insufficient to deliver the data in time  
10 for decoding. The opposite extreme occurs when all of the programs contain scenes which are easily compressed. In this case, data can be delivered too quickly and the decoder buffers may overflow unless fill packets, containing stuffing bytes, are inserted into the bit stream. This reduces channel utilization (channel efficiency) since the fill packets are not  
15 useful for decoding and are discarded at the receivers.

Our co-pending U.S. patent application entitled "Method and Apparatus for Multiplexing Video Programs For Improved Channel Utilization," filed concurrently with the present application, incorporated herein by reference, and hereinafter referred to as "Our Multiplexing  
20 Invention," discloses a multiplexing method for determining the order of forming the multiplex that assures that no overflow occurs in any decoder buffer, thereby increasing the efficiency of transmitting the multiplex. Even when using the method of Our Multiplexing Invention, there is still a finite chance that all program streams have scenes at the same time of such  
25 complexity and that peaks in the decoder buffer memory requirement are such that no more decoder buffer space is available. A method for reducing

such data rate fluctuations is disclosed in our co-pending U.S. patent application entitled "Method and Apparatus for Increasing Channel Utilization for Digital Video Transmission," filed concurrently with the present application, incorporated herein by reference, and hereinafter referred to as "Our Channel Utilization Invention." The method of Our Channel Utilization Invention comprises varying the formation of the multiplex of the different program streams to reduce the memory requirement of the decoder buffers, varying including adjusting the relative alignment of the program streams. Different embodiments of adjusting include inserting program material such as a pause at selected points in time, causing decoder buffer underflow at selected points in time, and varying the rate of decoding and display at selected points in time.

The methods of Our Multiplexing Invention and Our Channel Utilization Invention do not reduce the quality of the pictures transmitted. There may still be situations in which these methods are not sufficient or cannot be used. In such situations, it may be necessary to reduce the amount of data at selected points in time in order to avoid exceeding the capacity of the channel.

Thus, there is a need in the art for technology that can be used to modify encoded video programs, in particular VBR-encoded programs, in order to reduce the data rate fluctuations in such programs without significantly affecting the image quality.

Thus, there also is a need for technology for modifying encoded video that reduces data rate fluctuations when transmitting a stream of a

multiplex of several encoded video programs, applicable to both real time encoding and remote encoding situations.

Thus, there also is a need for technology for modifying encoded video that eliminates the chance of a buffer overflow condition occurring  
5 at the decoder.

Thus, there also is a need for technology for modifying encoded video that maximizes channel efficiency.

Thus, there also is a need for technology for delivering the information that is removed from the encoded video by modifying to  
10 certain receivers capable of extracting and utilizing the information to deliver improved image quality.

## II. Summary of the Invention

### A. Objects of the Invention

Thus, an object of the invention is to provide a method (and apparatus) for modifying encoded video programs, in particular, VBR-  
5 encoded programs, to reduce the data rate fluctuations in such programs without significantly effecting the quality of programs.

Another object of the invention is to provide a method (and apparatus) for modifying encoded video that reduces data rate fluctuations when transmitting a stream of a multiplex of several encoded video  
10 programs, applicable to both real time encoding and remote encoding situations.

Another object of the invention is to provide a method (and apparatus) for modifying encoded video that reduces the chance of a buffer overflow condition occurring at the decoder.

15 Another object of the invention is to provide a method (and apparatus) for modifying encoded video that reduces the chance of a buffer underflow condition occurring at the decoder.

Another object of the invention is to provide a method (and apparatus) for modifying encoded video that maximizes channel efficiency.

Another object of the invention is to provide a method for sending the information removed by modifying in such a way that it can be recovered by certain receivers equipped with additional storage capacity and utilized to deliver improved image quality.

5 Another object of the invention is to provide a method for sending the information removed by modifying in such a way that it can be recovered by certain receivers with access to a second data channel and utilized to deliver improved image quality.

### **B. Overview of the Invention**

10 The above and other objects of the invention are provided for in a method (and apparatus) for increasing channel utilization for a data channel transmitting a multiplex of a set of one or more encoded program streams, each program stream in being decodeable by a corresponding decoder, each corresponding decoder including a corresponding decoder buffer, the  
15 decoder buffers having a maximum allowable size, the method comprising selecting encoded pictures to be modified, said selecting according to a criterion which includes preventing any underflow of any decoder buffer, modifying each said selected encoded picture to form a corresponding modified encoded picture, said modified encoded picture having less data  
20 than said selected encoded picture, and transmitting the corresponding modified encoded pictures through the channel in place of the selected encoded pictures. In one embodiment, modifying deletes each selected encoded picture. In another embodiment where the encoded program streams include predictively encoded pictures, selecting selects predictively

encoded pictures that are not anchor pictures, and modifying deleted the prediction error data from each said selected encoded picture.

In a further embodiment of the present invention, one or more additional data channels are used to send augmentation information. The augmentation information can be used by specially equipped receivers to correct the impairments that would normally occur when decoding the modified signal received from the data channel. In yet another embodiment of the present invention, augmentation information is sent using the same data channel that is used to transmit the modified pictures. In this case, the information that is removed by modifying is transmitted before it is needed for decoding and at a time when the data channel is not fully utilized. Certain receivers equipped with sufficient storage can receive and store the augmentation information until it is needed. Alternatively, if the additional storage is used to insert additional delay between the time that data is received and the time that data is decoded, then the augmentation information can be sent after it would be needed by a conventional receiver.

### C. Related Inventions

Co-pending with the present application are our U.S. patent applications entitled "Method and Apparatus for Multiplexing Video Programs for Improved Channel Utilization" and "Method and Apparatus for Increasing Channel Utilization for Digital Video Transmission," both filed concurrently with the present application, and incorporated herein by reference.

### III. Brief Description of the Drawings

Figure 1 shows an example of an encoding system.

Figure 2 shows a system for combining multiple compressed program streams into a single multiplex.

5 Figure 3 shows the details of the video program stream, the  $n$  different segments, and the multiplexer output.

Figure 4(a) shows a B-picture before modifying.

10 Figure 4(b) shows a typical B-picture after modifying the B-picture of Figure 4(a) according to the method of the invention. Modifying discards DCT coefficients, so that the macroblocks of the slices in the B-picture of Figure 4(a) are replaced with shorter macroblocks.

Figure 5(a) shows: (i) a typical MPEG-2 macroblock before modifying and (ii) after macroblock modifying, including the removal of the DCT coefficients according to the method of the present invention.

15 Figure 5(b) shows an MPEG-2 macroblock before and after modifying according to the method of the present invention in the case that the macroblock was intra-coded and the assumed value of `quantizer_scale` was erroneous so that a new value of `quantizer_scale` needed to be inserted.

20 Figure 6 shows an embodiment of an encoding system for implementing the B-picture modifying process of this invention.

Figure 7 shows an embodiment of an enhanced receiver, capable of receiving data from both the primary channel and the overflow channel of the system of Figure 6.

5 Figure 8 shows an alternate encoding system for implementing the B-picture modifying process of this invention.

Figure 9 shows an embodiment of an enhanced receiver, capable of receiving data from both the primary channel and the overflow channel of the system of Figure 8.

10 Figure 10 shows a flow chart for an embodiment of the method for analyzing a program stream to determine when to modify B-pictures, according to the present invention.

## IV. Description of the Preferred Embodiment

### A. Introduction

The preferred embodiment of the invention utilizes the presently known technique of variable bit-rate ("VBR") encoding and decoding of video data, including soundtrack or audio data, and uses a compression method to reduce the total amount of data required to represent a video stream. The present invention, however, is applicable also to non-VBR encoded data, as would be clear to one in the art. Although the preferred embodiment uses the MPEG coding standard, the invention is applicable to many other coding methods, as also would be apparent to one in the art.

The following detailed description is made with reference to Figures 1-10, in which like reference numerals indicate identical elements throughout all of the figures.

For convenience, several terms used in describing the invention are defined in the Appendix entitled "An Introduction to Video Coding and MPEG." The Appendix also explains many of the concepts of coding and MPEG that may be needed to understand the presented invention. It should be noted that some of the terms in the terminology section of the Appendix may have more than one meaning in the art. The meaning applicable in this description would be clear from the context to one in the art.

### B. Overall Architecture of the Preferred embodiment

The invention deals with *complete* coded streams, including both video and audio. The preferred embodiment deals with MPEG streams of a

single program. That is, MPEG encoded video, audio, and other data combined together as the MPEG stream of a single program. Such an MPEG stream is defined herein as a program stream.

Sometimes, details are provided herein only for dealing with the video stream component of the program stream, as it is the determining or limiting part. In any such case, how to process coded audio to be consistent with the method of the present invention would be straightforward to one in the art. Also, in all cases, realizing the present invention for streams coded by methods other than MPEG would be straightforward to one in the art.

An example of a video encoding system is shown in Figure 1. Program source 101, such as a video tape recorder (VTR) or other program source, is used to supply uncompressed video 103 to an encoder 105. Encoder 105 receives uncompressed video 103 and applies a video compression process. In the preferred embodiment, the video compression is that specified by the MPEG-2 standard. It is preferred, but not necessary for the invention, that the video compression technique provides for a VBR mode that can be used to produce video streams which can be decoded to obtain images of substantially uniformly consistent quality. Many compression methods and standards, including MPEG, have such a VBR mode. Typically, the compressed output 107 of encoder 105 would be written to a digital storage medium (DSM) such as a disk or tape storage media 109 for playback at a later time, or transmitted directly to a data channel for broadcast or other transfer. That is, in the nomenclature of the MPEG standard, 109 is a DSM or channel.

In the preferred embodiment of the present invention, multiple compressed program streams are combined into a single multiplex as shown in Figures 2(a) and 2(b). The multiplexer 201 combines each of the program streams 203 retrieved from storage 109 or obtained from any other source, such as an encoder 209 and/or satellite receiver 211, into a single multiplex 205 that can either be broadcast immediately over the distribution system or returned to storage for broadcasting at a later time. One advantage of using the multiplexing process and apparatus, also called the combining process and apparatus, is to reduce the statistical variance in the video as compared to that of the individual VBR program streams 203. Additionally, the multiplexing method has the advantage of providing the viewer with a choice of retrieving or viewing any of the program streams contained in multiplex 205.

The multiplexing in the preferred embodiment is for a plurality of compressed program streams or for the case of transmitting several segments of a single data stream, such as a VBR or otherwise compressed program. In the single data stream case, one first partitions the compressed program into a plurality of segments using the splitter 207 shown in Figure 2(a). For example, some or all of program streams 203 combined by the multiplexer 201 could correspond to different segments of the same video program. By continuously transmitting the same multiplex of program segments over and over, it becomes possible for the receiver to eventually reconstruct the entire program by selecting and decoding a different program segment during each repetition of the multiplex. This technique is described in more detail in our co-pending U.S. patent application, Application Number 08/326,511, filed October 19, 1994 entitled "Method

and Apparatus for Encoding and Formatting Data Representing a Video Program to Provide Multiple Overlapping Presentations of the Video Program,” hereinafter referred to as “Our Video-On-Demand Invention” and incorporated herein by reference. One application of Our Video-On-  
5 Demand Invention is video-on-demand, where the user may choose to begin watching a program during any of a wide range of time intervals.

The splitting and multiplexing are shown in more detail in Figure 3 for the case where program streams 203 correspond to different segments of the same video program and multiplexed output 205 is a multiplex of  
10 such different segments of the same video program. In the preferred embodiment of the present invention for the single original stream case, compressed program 107 is first split into a number,  $n$ , of segments 303 where each segment 303 corresponds to a fixed time interval 307 of  
15 length  $T$  seconds. In the example illustrated in Figure 3,  $T$ , the time interval 307, is 300 seconds (5 minutes). When segments 303 are compressed using VBR techniques, the compression ratio is allowed to vary over time, and different segments may contain differing amounts of compressed data.

As shown in Figure 3, in the segmented single-stream case, each of  
20 the  $n$  different segments 303 becomes an independent program stream, and each of these  $n$  independent program streams subsequently are aligned to start at the same time, herein referred to as “stacking up.” The stacked-up program streams 203 are multiplexed together using multiplexer 201.

Making the necessary changes for the case of multiplexing several  
25 video programs rather than segmenting a single stream before multiplexing

would be clear to one in the art. Similarly, changing for the hybrid case of multiplexing together segments of one or more programs with one or more independent video programs would also be clear to one in the art.

In the preferred embodiment, the multiplexing method used is that of  
5 Our Multiplexing Invention. The method assumes that all receivers that are  
“tuned” to a particular program stream use the same hypothetical decoder.  
Each such decoder has a single buffer to receive data from a single selected  
stream. Thus, when sending data from a particular program stream, the  
multiplexing method in the preferred embodiment effectively assumes the  
10 data will be sent along a particular communications channel including the  
tuned hypothetical decoder. The method of Our Multiplexing Invention  
uses this assumption, for example, to estimate the fullness of the buffer of  
the hypothetical decoder corresponding to any stream to which the receiver  
containing the decoder is tuned. The word “receiver” as used herein means  
15 a receiving device which includes a decoder. The term “channel buffer”  
refers to this decoder’s buffer. The assumption about such hypothetical  
decoders does not necessarily restrict the type of decoders or receivers  
with which the method of the present invention works. Rather, the  
assumption is used by the multiplexer to make certain decisions. Different  
20 levels of modeling the decoders are possible within the scope of the  
invention.

In Our Multiplexing Invention, each of the program streams 203 is  
assumed to have been subdivided into packets prior to multiplexing. Each  
time a packet is sent, the multiplexer selects the next packet by determining  
25 which of the decoder buffers corresponding to the  $n$  different channels will

become empty first. The complexity of the determining step depends on how complete a model of decoders is maintained at the multiplexer. The more complete the model, the less computation is involved in determining which channel to send to. The method of Our Multiplexing Invention  
5 determines the order of sending data packets of program streams into the data channel. Whenever data is to be sent, the method determines which packet to send next by selecting a next channel that has a decoder buffer that is not approaching a full state. The method further includes selecting a channel that substantially balances the decoder buffers of the decoders with  
10 which the program streams are decodeable. In a further feature of that embodiment, balancing the decoder buffers includes keeping approximately the same time duration stored in each decoder buffer.

Although the method of Our Multiplexing Invention is used in the preferred embodiment, those in the art will appreciate that the present  
15 invention is applicable also to all other multiplexing methods and also to sending single (non-multiplexed) program streams and other digital video streams.

In many situations, the data rate of the resulting multiplex may exceed the capacity of the transmission channel even when optimal  
20 multiplexing techniques are used. The method of the present invention modifies the resulting multiplex to fit into such data channels by introducing slight impairments to the image quality. The impairments are of a type and nature that they are not likely to be detected under normal viewing conditions. Thus the method of this invention maintains constant  
25 perceived image quality.

In a further embodiment of the present invention, one or more additional data channels are used to send augmentation information. The augmentation information can be used by specially equipped receivers to correct the impairments that would normally occur when decoding the modified signal received from the first data channel. For example, in  
5 satellite transmission, the method of the present invention can be used to transmit several program streams by satellite. Each satellite provides several transponders in both the uplink and downlink paths. Using the method of the present invention, the modified signal is transmitted using a  
10 first transponder, and the information removed by modifying is transmitted using a second transponder having spare capacity.

In yet another embodiment of the present invention, augmentation information is sent using the same data channel that is used to transmit the modified signal. In this case, the information that is removed by modifying  
15 is transmitted *before* it is needed for decoding and at a time when the data channel is not fully utilized. Certain receivers equipped with sufficient storage can receive and store the augmentation information until it is needed. Alternatively, if the additional storage is used to insert additional delay between the time that data is received and the time that data is  
20 decoded, then the augmentation information can be sent *after* it would be needed by a conventional receiver. The increased delay in the path of the enhanced receivers would allow sufficient time to receive the augmentation information and apply it to the decoding process at the appropriate time.

### C. Detailed Description

As stated earlier, even with statistical multiplexing and the method of Our Multiplexing Invention, there is still a finite chance that all program streams have scenes at the same time of such complexity that the capacity of the transmission channel is exceeded and the data cannot be delivered in time to prevent the decoder buffers from underflowing. Often, in such cases, the method of Our Channel Utilization Invention can be used to prevent such large data rate fluctuations while preserving the full quality of the encoded video.

The method of the present invention includes the modification of the signal that is transmitted by reducing the amount of data whenever necessary to avoid exceeding the capacity of the channel. The method can be used in conjunction with the method of Our Channel Utilization Invention, or alone. Reducing the amount of data is done, according to the present invention, by selectively increasing the compression ratio during certain complex scenes, thus affecting image quality. Depending on the particular compression method to which this is applied, the method of the present invention chooses particular portions of the program to decompress and recompress. It is known in the art that the human eye is less sensitive to artifacts, such as compression artifacts, in parts of a picture that have complex motion or that are fast moving. That is, in the case of using motion prediction, in those parts of a picture where the motion prediction method is not particularly accurate in prediction. This is known in the art as the "masking affect." According to the method of the present invention,

the choosing of particular portions of the program to decompress and recompress uses the masking effect.

In the preferred embodiment of the invention, MPEG is used for compression. MPEG uses motion prediction, and as described in the Appendix, pictures are classified in MPEG into three types. I-pictures use only spatial techniques for coding. P-pictures use forward prediction from a previous anchor picture (a P- or I-picture), followed by spatial coding to compress the differences between the picture and the predicted picture, and B-pictures use bidirectional prediction from a pair of anchor pictures followed by spatial coding. In general, P-pictures can be compressed more efficiently than I-pictures, since the prediction errors tend to contain less information than the original images. B-pictures are usually the most efficiently-compressed images because they use predictions based on information contained in a previous picture, a future picture, or both. Since the bidirectionally-derived predictions are usually more accurate than unidirectionally-derived predictions, greater compression ratios are possible in the spatial compression step with B-pictures. Sometimes, however, during detailed scenes with complex motion, the advantages of bidirectional prediction are reduced and the B-pictures become more difficult to compress.

Another distinguishing characteristic of B-pictures is that they are never used to predict other frames of the sequence. Therefore, if a B-picture is deleted or altered, the effect will be confined to that picture alone and will not cause any alteration of other pictures in the video sequence yet to be decoded. Therefore, in accordance with the method of the present

invention, post-processing techniques aimed at reducing peak data rates are limited only to data that does not affect the image quality of other video frames. In the case of MPEG, such post-processing is limited in the preferred embodiment to the B-pictures only. In this way, any impairments or artifacts introduced into a sequence will be limited to very short durations corresponding to a single frame time interval. As a result, such impairments will become difficult to detect visually, particularly if they are strategically confined to regions characterized by rapid, random, or complex movements (the masking effect). These are the regions which are difficult to reproduce using previous and/or future frame predictors. Sometimes, accurate reconstruction in these regions is accomplished by sending intra-coded or spatially-compressed information that is not based on prediction errors.

Thus, according to the method of the present invention, lowering peak data rates is accomplished by reducing or eliminating prediction-error information in selected B-pictures. In one embodiment, the visual masking effect is increased by only altering pictures that are separated by a certain minimum time-interval. For example, the number of altered pictures may be limited to no more than one out of every group of twelve consecutive pictures. Recalling that multiplex 205 may contain many program streams 203, the method of the invention may individually alter each of the streams during periods when the nominal data rate is being exceeded. An additional advantage of the method is that it is relatively straightforward to implement in real time. Additionally, it does not require reconstruction of the pixel-data.

In another embodiment, for any program stream, for B-pictures that are separated enough, the prediction-error information is first identified by parsing the compressed bit-stream. A B-picture is selected for processing if the energy in all the DCT coefficients is sufficiently large. This  
5 corresponds to either a B-frame of a complex scene, or a B-frame in which there is a large prediction error. As discussed before, these situations are when the masking effect is strongest. That is, errors are not easily perceived in the display. Thus, selecting such B-pictures for modifying still retains the property of constant perceived quality in the modified bit  
10 stream

Once B-pictures are selected, in one embodiment of the present invention, the spatially-coded prediction-error data (the DCT coefficients) sent with such selected B-pictures is discarded. As will be clear to those in the art, additional code words may need to be adjusted in order to indicate  
15 the absence of this prediction-error data.

In an alternate embodiment, the B-pictures selected are simply deleted.

The process of modifying MPEG B-pictures according to the preferred embodiment will now be described in detail with the aid of  
20 Figures 4 and 5. Figure 4(a) shows a B-picture before modifying. The picture consists of a picture header 403, and several slices, that number denoted by  $N$ . Each slice, for example, the first slice 405 and the last slice 407, has a slice header (slice header 409 in the case of the first slice) followed by a number,  $M$ , of macroblocks, each macroblock denoted by  
25  $MB_{ij}$  for the  $j$ 'th macroblock in the  $i$ 'th slice. For example, in the first

slice, slice header 409 is followed by the first macroblock 411, denoted by  $MB_{11}$ , followed by the next macroblock, *etc.*, through to the last macroblock 413, denoted by  $MB_{1M}$ . Similarly, in the last slice, slice header 415 is followed by the first macroblock 417, denoted by  $MB_{N1}$ , followed by the next macroblock, *etc.*, through to the last macroblock 419, denoted by  $MB_{NM}$ . Modifying according to the method of the invention, discards the DCT coefficients that are present in some of the macroblocks. Figure 4(b) shows the B-picture of Figure 4(a) after modifying. Each macroblock, denoted by  $M \bar{B}_{ij}$  for the  $j$ 'th macroblock in the  $i$ 'th slice, is a shorter version of corresponding macroblock  $MB_{ij}$  of the picture in Figure 4(a), except that some of the original macroblocks may not have contained DCT coefficients, and some of the original macroblocks may have been skipped. Skipped macroblocks are not affected by the modifying method of the present invention.

How each macroblock is modified, according to the preferred embodiment, is described with the aid of Figure 5(a), which shows a typical MPEG-2 macroblock 501 before modifying and the modified macroblock 503. MPEG-2 macroblock 501 contains several components defined by the macroblock layer syntax of the standard. The `macroblock_address_increment` is an integer, represented by a variable length code, which indicates how many macroblocks have been skipped to arrive at the present macroblock. The `macroblock_escape` component is used together with the `macroblock_address_increment` to determine the total when a large number of blocks has been skipped. The `macroblock_type` indicates, using a variable length code, the type of macroblock. Different types are modified in different ways according to

the preferred embodiment, as described below and as shown in Table I, where, for convenience, the variable length codes are represented by integers. The motion\_type, for example frame\_motion\_type or field\_motion\_type in MPEG-2, is a code indicating the type of motion prediction. The DCT\_type is a code indicating whether the DCT is field coded or frame coded. The quantizer\_scale is an integer used to scale the reconstruction level of the retrieved DCT coefficients. The motion vector or vectors are used for motion compensation and may be of a different type depending on the macroblock\_type, motion\_type, and other parameters, and typically will include vertical and horizontal vector components for both forward and backward prediction. The coded\_block\_pattern defines, for this macroblock, to which blocks the DCT coefficients belong. For example, the coded\_block\_pattern defines which luminance and chrominance blocks, within the macroblock, contain non-zero DCT coefficients. The DCT coefficients, for these blocks specified by coded\_block\_pattern, follow next in the bit stream and, according to the method of the present invention, are subsequently deleted.

Depending on the macroblock\_type and other codes from other layers in an MPEG-2 stream, some of this data may not be present in all macroblocks. In particular, macroblock\_escape, motion\_type, DCT\_type, quantizer\_scale, the motion vector or vectors, and or coded\_block\_pattern are not always present.

Table I indicates some properties of macroblocks having different macroblock\_type codeword values. For convenience, the macroblock\_type codeword is shown as a number rather than a variable length codeword.

Table I shows the new macroblock\_type codeword to insert according to the preferred embodiment of the method of the present invention. Table I also shows different coding features according to the macroblock\_type. These features are macroblock\_quant (shown by Quant and indicating that a new quantizer\_scale is being sent), macroblock\_motion\_forward (shown by Motion\_forward, and indicating that forward prediction is used), macroblock\_motion\_backward (shown by Motion\_backward, and indicating that backward prediction is used), macroblock\_pattern (shown by Pattern and indicating that coded\_block\_pattern is being sent), and macroblock\_intra (shown by Intra and indicating that the data is intracoded).

Figure 5(a) shows the macroblock 503 which results from modifying the macroblock 501 in the case where DCT coefficients are discarded. The macroblock\_escape and macroblock\_address\_increment are unchanged, the macroblock\_type is modified according to Table I, the motion\_type remains unchanged, the DCT\_type and the quantizer\_scale are discarded since no DCT coefficients are sent, the motion vectors remain unchanged, and the coded\_block\_pattern is discarded, as are the DCT coefficients. A variation of this preferred embodiment is to delete the motion vectors in addition to the DCT coefficients. This would further reduce the amount of data to be transmitted but would degrade the accuracy of the prediction, and therefore the image quality, in most regions of most images.

In the preferred embodiment of the present invention, the DCT coefficients are not deleted during modifying for intra-coded macroblocks, that is, for macroblocks with macroblock\_type 7 or 11. Instead, the intra-

coded macroblock is sent unchanged in all cases except in the case that the quantizer\_scale has changed since the last intra-coded macroblock was sent, and the current quantizer\_scale does not reflect such a change. This might happen, for example, if the intra-coded macroblock is of macroblock\_type 5 7 and the changed values of quantizer\_scale were to be sent in a previous macroblock which was modified according to the method of the present invention, such modifying removing the quantizer\_scale. In such a case, as shown in Figure 5(b), the macroblock 501 of macroblock\_type 7 is replaced with a macroblock 503 of macroblock\_type 11, and the new value of quantizer\_scale is inserted into macroblock 503.

TABLE I

Type	Quant	Motion_ forward	Motion_ backward	Pattern	Intra	New type
1		x	x			1
2		x	x	x		1
3			x			3
4			x	x		3
5		x				5
6		x		x		5
7					x	7 or 11*
8	x	x	x	x		1
9	x	x		x		5
10	x		x	x		3
11	x				x	11

One variation of the preferred embodiment is to convert intra-coded macroblocks to predictive-coded macroblocks. A simple method for performing the conversion is to substitute the macroblock\_type from 7 or 11 to 1. As shown in Table I, macroblock\_type 1 corresponds to a  
5 bidirectionally-predicted macroblock where the motion vectors are assumed to have a magnitude of zero. The prediction may not be very accurate in this case, but the resulting artifacts are likely to be difficult to detect in such regions where the original encoder determined that intra-coding was more effective than predictive-coding. No additional motion  
10 vectors, DCT coefficients, or other codewords are expected when this type of macroblock is specified.

A preferred system for implementing the B-picture modifying process is shown in Figure 6. A program analyzer and modifier 603 receives one or more encoded program streams, depending on whether the  
15 method is to be applied to a single program or to an entire multiplex. In this example, the program analyzer and modifier 603 selects the B-pictures to be modified, performs the modifying using the preferred method described previously, assigns the modified B-picture packetized data and the packetized data of all other unmodified pictures to a primary output  
20 stream 605, and assigns the original unmodified packets corresponding to each selected B-picture to an overflow output stream 607. In this example, the primary and overflow stream data are modulated by the primary modulator 609 and overflow channel demodulator 611, respectively, onto different RF channels 615 and 617, respectively, and combined by  
25 combiner 619 for distribution. In larger systems, additional flexibility and

efficiency can be gained by sharing one or more overflow channels among multiple primary channels.

The output 617 of Figure 6 is compatible with conventional receiver systems capable of receiving only one RF channel. Such receivers would  
5 only decode the data in the primary channel, and therefore would experience some, albeit imperceptible or almost imperceptible degradation in image quality when the modified B-pictures are encountered.

An example of an enhanced receiver, capable of receiving data from both the primary channel and the overflow channel, is shown in Figure 7.  
10 A first tuner/demodulator 703 demodulates the data from the primary channel and a second tuner/demodulator 705 demodulates the data from the overflow channel. The output of each tuner/demodulator is provided to a program DMUX which extracts the desired program stream from each of the respective multiplexes. In Figure 7, the primary channel and overflow  
15 channel program DMUX's are labeled 707 and 709, respectively. The output of each program DMUX is provided to a time stamp comparator 711 which instructs a packet MUX 713 to select a next packet from the primary stream if the associated time stamp is less than the time stamp corresponding to the next packet from the overflow stream. In such cases,  
20 the packet from the overflow stream is saved in a small buffer 715 and then compared once again with the following packet in the primary stream. Eventually, when the time stamps of the corresponding next packets are identical, the packet MUX 713 is instructed to select the next packet from the overflow stream and the corresponding primary stream packet is  
25 discarded. In this way, the full quality of the original encoded signal is

preserved, since the decoder in the enhanced receiver only receives packets corresponding to unmodified pictures.

A slightly different encoding system is shown in Figure 8. Like the system of Figure 6, this uses a program analyzer and modifier 603 which  
5 outputs the primary and overflow packet stream outputs, 605 and 607, respectively. In this case, outputs 605 and 607 from the program analyzer and modifier 603 are connected to respective packet storage systems 809 and 811, respectively. These storage systems are arranged as first-in-first-out (FIFO) storage devices. Once a sufficient amount of data has collected  
10 in the overflow packet storage device 811, the packets from the respective storage devices can begin to be multiplexed. The packet MUX 817 in Figure 8 normally selects packets 815 from the primary packet storage device 811 and only inserts packets 813 from the overflow storage device 809 when sufficient channel capacity is available. In this way, the overflow  
15 packets 815 are inserted and used in the same way as fill packets which are no longer needed unless the overflow packet storage device 811 becomes empty. Thus this systems avoids whenever possible sending fill packets, and thus increases channel utilization.

As before, the output produced by the encoding system in Figure 8 is  
20 compatible with conventional receivers, which ignore the overflow data packets embedded in the bit stream. However, an enhanced receiver, which includes sufficient local storage, can utilize both the primary and the overflow packets to reconstruct the original encoded bit stream as it existed before the B-pictures were modified. An example of such an enhanced  
25 receiver is shown in Figure 9. In this case, after demodulation by a channel

tuner/demodulator 903, a first program DMUX 905 extracts all primary packets 909 while a second program DMUX 907 extracts all overflow packets 911 associated with the same program. The overflow packets 911 are transferred to a FIFO storage device 913 where they are delayed until  
5 needed for decoding. As in the system of Figure 7, a time stamp comparator 711 compares the time stamp of the next primary packet with the time stamp of the next overflow packet from the FIFO storage device 913 and selects the one with the smallest associated time stamp. If both packets have the same time stamp, then the overflow packet from the  
10 storage device 913 is selected and the primary packet is deleted. In this way, the full quality of the original encoded signal is preserved, since the decoder 717 in the enhanced receiver only receives packets corresponding to unmodified pictures.

A particular embodiment of the operation of the program analyzer  
15 and modifier 603 is now described with reference to Figure 10. To simplify the explanation, a one-to-one correspondence between packets and pictures will be assumed. Those skilled in the art will recognize that packets generally are not aligned with picture boundaries and may be either larger or smaller than the size of the average picture. Those skilled in the  
20 art will also realize that all such packets which include even a small portion of an modified B-picture can be included in the overflow output stream, and also duplicated in the primary stream with the B-picture component of the packet subjected to the modifying process.

Referring to Figure 10, the method starts with step 1003, in which  
25 the next picture is selected. In the case of a single program stream, Several

methods can be used to select the next picture. In the preferred embodiment, the method of Our Multiplexing Invention is used for step 1003. The reader is referred to Figure 4(a) of the patent application for Our Multiplexing Invention for a flow chart of such a method, noting that  
5 the flow chart of that Figure 4(a) is for packets rather than video frames.

After step 1003, step 1005 determine if the decoder buffer corresponding to the selected next picture will underflow. Step 1005 is done using time stamps in a manner analogous to the method described in the co-pending patent application for Our Channel Utilization Invention.  
10 The reader in particular is referred to the flow chart of Figure 7 of that patent application. If it is determined in step 1005 that the corresponding decoder buffer will not underflow, the picture is assigned in step 1017 to the primary stream, and a determination is made, in step 1019, whether or not that was the last video frame. If that was the last picture, the method  
15 ends. If not, the method goes back to step 1003 to select the next picture.

If at step 1005, it is determined that the next picture would normally underflow the buffer, then, in the preferred embodiment of the method of the present invention, Step 1007 determines if the next frame selected in step 1003 is a B-picture. If step 1007 determined that the next picture is not  
20 a B-frame, then the pervious picture is selected in step 1015, and that picture is checked if it is a B-picture. When a B-picture is thus found, that picture is assigned to the overflow stream in step 1009, and n step 1011, the B-picture is modified, and the modified picture is assigned to the primary stream, replacing the B-picture assigned to the overflow stream.  
25 In step 1013, a determination is made if the decoder buffer will underflow

with the modified primary stream, and, if not, the method returns to step 1003 to select the next picture. If there is an underflow the pervious picture is selected, and the method returns to step 1007 to find the next previous B-picture to modify.

5           In an alternate implementation of the invention, rather than, as in step 1007, all possible B-pictures being candidates for modification, B-frames are selected for modification only if they contain a relatively large amount of data. That is, only if there is a high prediction error. Selecting such B-pictures for modification ensures, according to the masking effect,  
10 that errors would not be easily perceptible.

Thus, in an alternate embodiment, groups of pictures rather than just pictures are considered. In the preferred embodiment of this alternate, each group consists of 12 pictures, that is, 0.5 seconds of a 24 frames per second movie. Using methods or simple modification of methods that have been  
15 disclosed herein or in our co-pending patent applications of Our Multiplexing Invention and Our Channel Utilization Invention, the process knows the state of all decoder buffers at any point in time. Also, the process knows the next group of pictures for any program stream at any point in time. Thus, at any point in time, the process determines the effect  
20 n the state of the decoder buffers of transmitting all these next groups of pictures for all programs streams. In particular, the program can determine if an underflow situation would occur. If no overflow condition would occur in any of the decoder buffers, the method sends this next group of pictures for all program streams and then looks at the next group  
25 of pictures for all channels.

If an overflow condition would have occurred, the method chooses from all pictures in the next groups of pictures the B-frame which contains the most data—a good frame to modify with respect to the masking effect. That B-picture is modified, the unmodified B-picture is 1) assigned to the  
5 overflow stream, and 2) replaced in the primary stream with the modified picture. The method next checks if all groups of pictures can now be sent with no underflow condition occurring. If yes, they are sent, else, the B-picture which contains the next most data is selected for modification. This is repeated until a sufficient number of B-pictures have been modified to  
10 avoid the underflow condition.

An alternate embodiment, additional criteria may be added, such as not modifying B-frames that are too close together within a single program stream.

How to simplify either the embodiment described in the flow chart  
15 of Figure 10, or the method of the alternate embodiment to deal with the case of there not being an overflow stream would be clear to those in the art.

Referring to the flow chart of Figure 10, in a real-time  
implementation, it may not be practical to repeatedly select and modify  
20 previous pictures until the data rate is sufficiently reduced to match the capacity of the channel. In such real-time implementations, it may be necessary to halt the backward stepping process when necessary to maintain proper timing, and resort to alternative data rate reduction measures such as the controlled underflow technique described in Our Channel Utilization  
25 Invention.

Selectively modifying B-pictures can be done during formation of the complete multiplex. Alternatively, this can be generalized to selectively modifying B-pictures independently for each  $n$  program stream. Since modifying can be done independently for each of the program streams 203,  
5 the method of the present invention is applicable also to the case of modifying selected B-pictures in a single encoded stream when there is a single encoded stream to transmit to a decoder via a data channel. How to simplify either the embodiment described in the flow chart of Figure 10, or the method of the alternate embodiment to deal with the case of there  
10 being only a single program stream would be clear to those in the art.

The co-pending patent applications for Our Multiplexing Invention (in Figure 5) and Our Channel Utilization Invention (in Figure 15) disclosed a server system capable of generating encoded multiplexes 205. Referring to Figure 15 in the application for Our Channel Utilization  
15 Invention, Editing of the compressed data streams, according to the present invention, can be performed in such a server system controller 1509, which in the preferred embodiment, is a. For example, if the compression ratio corresponding to a portion of a particular program is to be increased, then the steps for decompressing the selecting portion, re-compressing at a  
20 higher reduction ratio, and inserting the results back into the data stream can all be performed. Alternatively, complete or partially compressed programs to be used for data stream editing can be entered into the system via the encoder 1511 or archive library 1505.

In the case that unedited versions of B-frames are also sent in an  
25 "overflow" channel when there is B-frame editing according to the

invention, a separate server module can store all edited B-frames of a program. When a multiplex is being formed using a MUX module 1513, a full B-frame is sent from the server module holding the program when the channel can accommodate the full B-frame. Otherwise, an edited B-frame  
5 is taken from the server module holding edited B-frames.

Optionally, whenever an edited B-frame is sent, a separate MUX module forms a separate multiplex that includes a stream containing the full versions of edited B-frames.

Thus, a method and apparatus has been disclosed for modifying  
10 encoded video for improved channel utilization for multiplexed digital video transmission.

The foregoing description of preferred and alternate embodiments of the invention has been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to  
15 the precise forms disclosed. Obvious modifications or variations are possible in light of the above teachings. The embodiments are chosen and described to provide the best illustration of the principles of the invention and its practical application in order to enable one of ordinary skill in the art to utilize the invention in various embodiments and with various  
20 modifications, as are suited to the particular use contemplated. All such modifications and variations are within the scope of the invention as determined by the appended claims when interpreted in accordance with the breadth to which they are fairly, legally, and equitably entitled.

## Appendix: An Introduction to Video Coding and MPEG

### A. Terminology

For convenience, several terms used in describing the invention are now defined. The concepts behind the terms listed in this terminology section are explained in the sections that follow. It should be noted that some of these terms may have more than one meaning in the art. The meaning applicable to the description section would be clear from the context to one in the art.

*MPEG:* The MPEG standard, a set of International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) standards. Two variations of MPEG have been agreed to; MPEG-1, documented in ISO/IEC publications 11172 (“Coding of Moving Pictures and Associated Audio for Digital Storage Media”), and MPEG-2, documented in ISO/IEC publications 13818 (“Generalized Coding of Moving Pictures and Associated Audio Information”). As used herein, MPEG refers to either MPEG-1 or MPEG-2 without distinction.

*Remote coding:* Encoding programs independently of encoding of any other programs and independently of any multiplexing of the coded streams, either because of remote physical location, or for other reasons.

*Frame:* The basic unit of audio coding. MPEG audio bit-streams are broken down into frames. The term “frame” is also commonly used

for the unit of a video sequence. A video frame often is referred to as a "picture" herein.

*Picture:* A video frame, the unit of a video sequence.

5 *Pixel:* A "picture element" of a picture. In the arts of digital video and video coding, a picture may be thought of as a discretized two-dimensional array of pixels.

*Pel:* A picture element. The same as a pixel.

10 *Field:* In video, a subset of the pixels of a picture transmitted separately. In interlacing, a picture is split into two or more fields. In NTSC and PAL, each field consists of the pixels in alternating horizontal lines of the picture.

*Interlacing:* The splitting of a picture into two or more fields.

*Image coding:* The art of image compression.

15 *Color quantization:* Reducing the number of bits in a pixel required to represent that pixel. *Color quantization* also means digitizing the numbers representing the colors of a pixel.

*Intra-coding:* Image coding which takes into account spatial redundancies in a picture, including color quantization, but which does not use temporal variations from picture-to-picture.

20 *Inter-coding:* Image coding which takes advantage of temporal redundancies from picture-to-picture in a video sequence.

- Intraframe coding:* The same as intra-coding.
- Interframe coding:* The same as inter-coding.
- Motion estimation.:* Computing the spatial displacement due to motion from picture-to-picture in a video sequence.
- 5 *Motion compensation:* Using motion estimation to compensate for motion from picture-to-picture in a video sequence.
- I-picture:* In MPEG, a picture that uses purely intra-coding, and thus, is processed independently of other pictures
- I-frame:* Same as an I-picture.
- 10 *Intraframe coded picture:* Same as an I-picture.
- P-picture:* In MPEG, a picture that is coded using a previous I-picture or a previous P-picture. The compression of P-pictures relies on temporal prediction from previous I- or P-pictures using forward motion estimation and compensation.
- 15 *P-frame:* Same as a P-picture.
- Predictively coded picture:* Same as a P-picture.
- B-picture:* In MPEG, a picture coded by a bidirectional motion estimation and compensation from a previous I- or P-picture and a future I- or P-picture.
- 20 *B-frame:* Same as a B-picture.

*Bidirectionally coded picture:* Same as a B-picture.

*Anchor picture* In MPEG, for a P-picture, the I-or P-picture used to predict the P-picture. For a B-picture, either of the adjacent I- or P-pictures, one forward, the other backward in time from the B-picture, used to predict that B-picture backwards and forwards in time, respectively.

*Reference pictures:* The same as anchor pictures.

*Reference frames:* The same as anchor pictures.

*Variable bit-rate(VBR) encoding:* Image coding which the output bit-rate varies over time to account for the different complexities of the original video scenes from time to time. VBR encoding often is associated with encoding to achieve a uniform picture quality.

*MPEG stream:* In MPEG, a bit-stream typically containing one or more video streams multiplexed with one or more audio streams and other data, such as timing information.

*Program stream:* A coded stream for a single video program, usually comprising all video, audio, and other data streams associated with that program. An example is an MPEG stream for a program.

*Elementary stream:* In MPEG, a video stream, audio stream, or private data stream.

*Video stream:* The bit-stream of coded video information.

*Audio stream:* The bit-stream of coded audio information.

- Packet:* In MPEG, the basic unit of an elementary stream. Packets include a packet header and packet data.
- Packet header:* In MPEG, a header in each packet. The packet header includes a stream identification code and may include one or more  
5 time-stamps.
- Pack:* In MPEG, the basic unit of MPEG streams. Each pack includes a pack header and packets from one or more elementary streams.
- Pack header:* In MPEG, the header of a pack, possibly including time stamps.
- 10 *Time-stamps:* In MPEG, information in packs or packets which provide means for synchronization during decoding and presentation to the viewer.
- Multiplex-wide operators:* In MPEG, operations performed in decoding on packs, including the coordination of retrieval off the  
15 DSM or channel, the adjustment of clocks, and the management of buffers.
- Stream-specific operations:* In MPEG, operations performed in decoding on one elementary stream, including demultiplexing and playback synchronization.
- 20 *DSM:* A general term for Digital Storage Medium, used to denote a digital storage or transmission device or system. In the case of encoding, coded data is sent to a DSM or to a data channel, and, in

the case of decoding, coded data is obtained from a DSM or from a data channel.

*Source clock reference (SCR):* In MPEG, data which might be present in a pack header used to adjust local clocks to a master time base.

- 5 *Decoding time-stamp (DTS):* In MPEG, data that may be present in a packet header which indicates the time that the first unit in the packet is to be decoded (audio frame or video picture).

10 *Presentation time-stamp (PTS):* In MPEG, data that may be present in a packet header which indicates the time that the first unit (audio frame or video picture) in the packet is to be presented to the viewer.

15 *System time clock (STC):* A common time base used to create time stamps. The use of a common time base to unify the measurement of the timing of coded data (SCR) and the timing of the presentation of data (the PTS and DTS fields), ensures correct synchronization and buffer management.

20 *System target decoder (STD):* In MPEG, a hypothetical model of a decoder used in encoding to create MPEG streams. Real decoders need not be implemented with the same architecture or timing structure as the STD.

*Receiver:* A device including a decoder. In general, a receiver includes a decoder, tuner, and demodulator.

*Reference clock:* A local clock in the decoder.

*Channel:* In the context of the present invention, a particular program stream. The word channel also has other meanings, for example, a communications channel or data channel, which is a digital medium that stores or transports a digital data.

5 *Channel buffer:* The buffer of a decoder tuned to a particular program stream.

*Block:* In the multiplexing method of the present invention, any arbitrary grouping of data corresponding to a fixed time duration of presentation. In MPEG, a *macroblock* is a block which is the unit of motion compensation and adaptive quantization, thus representing a fixed amount of spatial data. A macroblock in MPEG further comprises one or more luminance blocks and one or more color difference blocks.

10

*Reorder buffer:* In MPEG, a buffer used in a video decoder to store I- or P-pictures while the sequence of pictures is reordered before presentation. This is necessitated often in the presence of B-frames.

15

*Stuffing bits:* In MPEG, extra stuffing code-bits that are inserted into a compressed bit-stream for the purpose of increasing the bit-rate of the stream. Stuffing bits are discarded during decoding.

20 *Stuffing bytes:* Eight stuffing bits.

*MPEG Profile:* In MPEG-2, the image/picture resolution. The most common profile for broadcast quality is the main profile.

*Main Profile (MP):* The MPEG-2 profile for CCIR-601 video.

*MPEG Level:* In MPEG-2, a set of minimum specifications and a subset of MPEG-2 features.

*Medium Level (ML):* A common MPEG-2 level.

*MP@ML:* In MPEG-2, MP and ML compliance. An MP@ML compliant  
5 decoder has at least 1.8M bits of decoder buffer storage.

## **B. Digital Video Coding**

A video program typically is organized as an ordered sequence of *frames*, each frame a scene at a particular point in time. It should be noted that the word "frame" is also used in the art of audio-compression and in  
10 the MPEG standard to refer to a basic unit of compressed audio. For that reason, the word "picture" is also used herein to indicate a video frame. Each video frame may be thought of as discretized into a two-dimensional array of picture elements (called *pixels* or *pels*), and may be split into two or more *fields*, each field consisting of a subset of the pixels in the picture.  
15 Such splitting is called *interlacing*. In the NTSC (U.S.A and Japan), PAL and SECAM (elsewhere) broadcast television standards, for example, each video frame consists of two fields, and each field consists of the pixels in alternating horizontal lines of the picture.

Each pixel contains information to reproduce the color (hue and  
20 saturation) and brightness, and this data may be expressed in various ways. One way is as red, green, and blue ("RGB") values corresponding to the relative amounts of red, green, and blue primaries needed to reproduce the color and brightness. Cathode ray tubes common in television displays have guns that produce red, green, and blue on the screen in response to voltages

in the guns. The data can also be expressed at each pixel as intensity data  
(*e.g.*, luminance) and color data, where the color data is expressed in  
chrominance coordinates, in hue and saturation coordinates, or some other  
way. Techniques are known for converting between the various ways of  
5 representing pixel information.

The information in each pixel can be represented digitally. Thus, a  
video program can be converted into a digital data stream which consists of  
an ordered sequence of bits describing the pixel values in each video frame.  
Similarly, audio associated with the program can be converted into digital  
10 data, and can be identified with the video on a picture-by-picture basis.

It is well known in the art that video data can be compressed. First,  
the number of bits required to represent each individual pixel may be  
reduced on a pixel-by-pixel basis. That is, the data of each pixel is  
processed (reduced) without reference to the data of any other pixel.  
15 Second, the number of bits required to represent a field or a picture can be  
reduced by taking advantage of spatial redundancies in the field or picture.  
For example, regions of uniform, slowly, or smoothly varying color or  
brightness may be represented by fewer bits (*i.e.* less data) than regions of  
many changes and edges. Third, the number of bits required to represent a  
20 sequence of pictures may be reduced by taking advantage of temporal  
redundancies— areas in the pictures of a sequence that vary slowly or  
hardly at all in time. The art of image compression also is called *image  
coding*. The first method above commonly is called *color quantization*, the  
second is called *intraframe coding* or *intra-coding* and typically includes  
25 color quantization, and the third is called *interframe coding* or *inter-coding*  
and typically, but not necessarily, includes intra-coding.

### C. The MPEG Video Coding Standard

The preferred embodiment of the present invention is for video streams which have been compressed in accordance with the MPEG standard. As will be understood to one in the art, the present invention may also be used with video streams that have been compressed using other compression schemes.

The video part of the MPEG standard uses motion compensated predictive coding, the discrete cosine transform (DCT) and adaptive quantization, each of these operating on images on a block-by-block basis.

10 *Motion estimation* refers to the process of computing the spatial displacement (in the form of *motion vectors*) from picture to picture, and is usually done by measuring the motion of blocks of pixels from picture to picture. *Motion compensation* uses the resultant motion vectors to compensate for the motion. MPEG uses both forward motion estimation (in which the estimation is of the future referenced to the past), and backward motion estimation (in which the estimation is of the past referenced to the future). Forward and backward motion estimation are also combined to produce bidirectional motion estimation.

Video frames to be compressed typically consist of an array of luminance pixels, where luminance is denoted by  $Y$ , and two arrays of chrominance pixels, where the chrominance signals are denoted by  $C_r$  and  $C_b$ , respectively.

According to the MPEG compression method, video frames (pictures) are classified into one of three types: *I-frames*, also called *I-*

*pictures* and *intraframe coded pictures*; *P-frames* or *P-pictures*, also called *predictively coded pictures*; and *B-frames* or *B-pictures*, also called *Bidirectionally coded pictures*. The three types of video frames differ in their use of motion estimation.

5 I-pictures use purely spatial compression and are processed independently of other pictures. Thus, I-pictures are processed entirely by intraframe operations and only a moderate level of compression is achieved. On the other hand, a complete picture can be generated from an I-picture alone.

10 P-pictures and B-pictures also use predictive coding, so can be thought of as dependent pictures.

P-pictures are coded using previous I-pictures or previous P-pictures. The compression of P-pictures relies on temporal prediction from previous I- or P-pictures using motion compensation. Only forward motion  
15 estimation/compensation is used in the temporal prediction. The I- or P-picture from which a P-picture is temporally predicted is called the *anchor picture* to the P-picture. It is also called the *reference picture* or *reference frame*.

B-pictures are coded by a bidirectional motion compensated  
20 predictive encoder using the next "future" I- or P-picture for backwards prediction and the previous I- or P-picture for forward prediction. The two adjacent I- or P-pictures from which a B-picture is temporally predicted are thus called *anchor pictures* of the B-picture. The B-picture temporal prediction uses motion compensation in forward and/or backward

directions. B-pictures are never used to predict other pictures. Due to the dependence of B-pictures on the two adjacent anchor pictures, B-pictures alone do not contain sufficient data from which to generate a recognizable picture.

5           In accordance with the MPEG standard, pictures are arranged in ordered groups. The standard allows the encoder to choose the frequency and location of I-pictures. A group of pictures is intended to assist random access into the sequence, thus the first coded picture in a group is normally an I-picture. A typical group then is a series of pictures which might  
10 include, as an example, an I-picture as the first picture, P-pictures distributed following every third picture, and a B-picture between every "I and P" and "P and P" sequence. Thus, a typical display order of picture types consistent with this example, might include an I-picture every  
15 fifteenth frame, each I-picture followed by two B-pictures, then a sequence of a P-picture followed by two B-pictures, until the next group starts with an I-picture. That is, the display order of I B B P B B P B B P B B P B B I  
B B P B B P B B ... A thirty picture-per-second environment which includes an I-picture every fifteenth frame, corresponds to having an independent picture every one half-second.

20           In accordance with the MPEG standard, after the video signals are digitized, the digital data is organized into macroblocks. A macroblock is the unit of motion compensation and adaptive quantization. A number of macroblocks comprise a picture. Each macroblock defines a predetermined spatial region in a picture, and contains luminance and chrominance  
25 information.

The MPEG proposal provides for the arrangement of macroblocks into slices. A slice is an integer number of consecutive macroblocks from a raster of macroblocks. A slice represents the boundary within which differential coding of macroblock parameters, *e.g.*, DC coefficients of a DCT, and motion vectors, is performed. Each slice has its own header information, can be independent of other slices, and each contains at least one macroblock. Slices do not overlap. The position of slices may change from picture to picture. The first slice starts with the first macroblock in the picture and the last slice ends with the last macroblock in the picture.

5 The first macroblock in a slice has its macroblock parameters, *e.g.*, DC coefficients of a DCT (if intraframe coded) and motion vectors, differentially coded from a constant value. Each subsequent macroblock in a slice has its macroblock parameters measured as an offset from the previous macroblock in the slice. Accordingly, the size of the slice is the

10 minimum size for which a piece of data can be recovered and correctly decoded. If part of a slice is lost, it may not be possible to decode the differences in motion vectors or the DC coefficients contained in the remaining part of the slice.

A macroblock comprises four 8 x 8 luminance blocks ( $Y_0, Y_1, Y_2, Y_3$ ) and two 8 x 8 color difference blocks ( $C_r$  and  $C_b$ ). The four luminance blocks ( $Y_0, Y_1, Y_2, Y_3$ ) and two color difference blocks ( $C_r, C_b$ ), which form a single macroblock, are used to encode a 16 x 16 picture element array covering the same spatial region in a picture. As described above, a macroblock serves as the smallest unit of motion compensation and

25 adaptive quantization.

In accordance with the MPEG standard, motion-compensated predictive coding is carried out by calculating motion vectors for every macroblock in a P-picture or B-picture. MPEG compression usually encodes motion vectors on a macroblock basis, but does not specify the technique for computing them. Thus, a variety of different motion estimation techniques can be implemented consistent with the MPEG standard. One technique, for example, is to compute motion vectors from the picture-to-picture correlation of blocks of pixels in the luminance signal, resulting in a motion vector for the luminance component of the macroblock.

The best mode for encoding each macroblock is selected. Within a given picture, each macroblock is coded in one of several different modes. The intraframe coding mode refers to macroblock coding in which only spatial information is used. Conversely, the interframe coding modes (forward motion, backward motion, and bidirectional motion) refer to macroblock coding in which information from pictures other than the current picture is used in the coding, typically for temporal prediction in motion-compensated predictive coding. For I-picture macroblocks, only intraframe coding mode is available.

In a typical encoder, not specified by the MPEG standard, P-picture macroblocks are first checked to determine if interframe coding without motion compensation is appropriate. This decision is made by computing the luminance energy of a forward prediction residual for the macroblock that results from an interframe coding without motion compensation, and comparing it to a threshold value. If the residual energy is below the

threshold, then the macroblock will be coded without motion compensation. Otherwise, the residual macroblock from interframe coding with forward motion compensation will be derived and used to determine if inter-coding or intra-coding is to be used, the last step in the coding mode selection.

B-picture macroblocks are similarly processed to determine whether interframe coding is appropriate. Since B-pictures may be bidirectionally coded, interframe coding can be either forward or backward, based on the preceding and following anchor (*i.e.*, I- or P-) pictures. It may also be based on the average of those macroblocks from the preceding and the following anchor pictures. In interframe coding using motion compensation, there are three possible modes: forward, backward, and bidirectional. The choice of coding mode for B-picture macroblocks may be determined on the basis of luminance prediction residual energy.

The final step in the coding mode selection for both P- and B-picture macroblocks is to choose between interframe coding and intraframe coding. Generally, P-pictures and B-pictures are encoded using interframe encoding. This selection could be made by comparing the luminance energy of the original macroblock to the energy of the luminance interframe (with or without motion compensation) prediction residual macroblock. If the original macroblock has less energy than the prediction residual macroblock, the intraframe coding mode is selected.

After the motion vectors have been calculated, each macroblock is transform encoded. Thus, the macroblocks are transformed from pixel domain to the DCT coefficient domain. The picture information in each

picture (*i.e.*, pixel values for I-pictures, and residual error after prediction for B- and P-pictures) is transformed using the DCT and then adaptively quantized. For the purpose of performing the DCT, a picture (a video frame) is divided, for example, into blocks of values (*i.e.*, arrays of DCT  
5 coefficients).

The DCT process generates blocks of DCT coefficients in a zigzag-scanned format (*i.e.*, the low-frequency coefficients are followed by the higher frequency coefficients). This zigzag scan arrangement facilitates the subsequent run-length coding process. The DCT coefficient for which the  
10 frequency is zero in both dimensions is called the DC coefficient.

Next, adaptive quantization is performed on each block of DCT coefficients. After adaptive quantization has been applied to the DCT coefficients, the coefficients undergo further compression involving such known techniques as differential coding, run-length coding, and variable  
15 length coding. As a result, the video compression encoder module produces encoded data in the form of variable length code words, which includes information concerning the selected mode of encoding and any motion vectors needed for decoding. Various headers are also included in the encoded bit-stream which provide information such as the picture size in  
20 terms of pixels per line and a pixel aspect ratio. The video compression encoder module also outputs information that states which picture the encoded data represents and which macroblock and slice the encoded data represents.

The code words are then further encoded with other MPEG-specific data needed to provide reliable delivery of the variable length encoded compressed video stream.

#### D. The MPEG Audio Coding Standard

5 MPEG also includes an audio coding standard. MPEG-1 audio supports mono, stereo, and a dual-mode with two separate channels, useful, for example, for bilingual programs, and a mode called joint stereo, in which interchannel (left-right) dependencies are also used for bit-rate reduction. In the first stage of MPEG-1 audio compression, an audio signal  
10 is converted into spectral subband components, then each subband is quantized. Psychoacoustic models can be used in the encoding process. MPEG-2 adds multichannel support as dual standards: MPEG-1 backwards-compatible coding and nonbackwards-compatible coding.

MPEG audio bit-streams are broken down into units called *frames*.  
15 An MPEG-1 audio stream consists of variable length frames, each frame starting with an MPEG-1 header, followed by the MPEG-1 audio signal, then followed by a variable-length field, which is reserved for ancillary data and ignored by MPEG-1 decoders.

The word "frame" in this context is not to be confused with a video  
20 frame. For that reason, the word "picture" also is used herein to indicate a video frame.

A backwards compatible MPEG-2 audio stream has frames which include the necessary MPEG-2 multichannel extension signals in the fields of MPEG-1 frames for ancillary data.

This invention deals with *complete* MPEG streams including both video and audio. Details are provided herein only for dealing with the video stream, because, for example, it is the determining or limiting part, of processing the audio to be consistent with the method of the present  
5 invention that would be straightforward to one in the art.

#### **E. Audio/Video Synchronization and MPEG Receivers**

A complete MPEG stream consists of the MPEG video stream, the MPEG audio stream, and other data, such as timing information, all combined and multiplexed together to form a single stream suitable for  
10 digital transmission or storage. The MPEG standard imposes syntactical and semantic rules to enable synchronized playback. The standard does not specify the exact architecture or implementation of encoders or decoders, but does specify bit-stream properties and performance requirements that need to be met, such as minimum clock tolerances.

15 The term *program stream* is used herein to indicate a coded stream, such as an MPEG stream, for a single program, including the video and audio. Thus, a program stream may consist of the MPEG stream of a digitized motion picture, a digitized terrestrial, cable or satellite television program, or different time segments of the same motion picture or  
20 television program.

The video stream, audio stream, *etc.* are called *elementary streams*. An MPEG stream thus consists of one or more elementary streams multiplexed together. Data from each elementary stream is stored in units called *packets*, which typically are serialized in that stream. A packet

consists of a *packet header* followed by packet data. The packet header begins with a start-code, and also identifies the stream to which the packet data belongs. The packet header may also include one or more *time-stamps* to indicate the timing of the event represented by the packet data. The  
5 packet data contains a variable number of contiguous bytes from one elementary stream, *e.g.*, audio, video, or other data.

Packets are organized together into *packs*. Packs may contain packets from more than one elementary stream, for example, from both the audio and the video stream, and so are used for *multiplex-wide operations* in  
10 decoding. This is in contrast to operations on packet data, called *stream-specific operations* which reflect the fact that the data in a packet is from one specific elementary stream. A pack commences with a *pack header*, and may be followed by one or more packets. The pack header may include a time-stamp called the *source clock reference (SCR)*, which specifies the  
15 exact time at which the SCR data byte (and thus each other byte) is to enter the decoder from the digital storage medium (the *DSM*) or the data channel. This target arrival schedule serves as a reference for clock correction and/or for managing buffers in the decoder. The first pack in any stream also may convey information on the maximum data rate and the  
20 number of channels in the multiplex.

Thus, multiplex-wide operations performed on packs include the coordination of data retrieval off the DSM or channel, the adjustment of clocks, and the management of buffers.

The two principal stream-specific operations are demultiplexing and  
25 synchronizing the playback of multiple elementary streams. Demultiplexing

reconstitutes elementary streams from the multiplexed streams using the stream identification codes contained in packet headers. Elementary streams in addition to audio and video streams may include private, reserved, and padding streams.

5           Synchronization among multiple streams is effected with time-stamps which may be present in packet headers. *Presentation time-stamps* (PTS) and *decoding time-stamp* (DTS) fields are used in packet headers for synchronization of audio and video. The PTS is the time at which the first unit (audio frame or video picture) in the packet data is to be presented to  
10 the viewer, and the DTS is the time that the first unit (audio frame or video picture) in the packet data is to be decoded. All time-stamps, including those in pack headers, are adjusted during encoding to a master time base called the system time-clock (STC). The use of a common time base to unify the measurement of the timing of coded data (SCR in the pack  
15 header), and the timing of the presentation of data (the PTS and DTS fields in packet headers) ensures correct synchronization and buffer management. SCR, PTS, and, if used, DTS fields are inserted during encoding at intervals not necessarily equal, but not to exceed some specified value—  
0.7 seconds in MPEG-1. Also, in MPEG-1, these time stamps are in units  
20 of 90 kHz.

A decoding system, including all of the synchronized decoders and the source of the coded data, must have exactly one independent time-master. This fact is a natural result of the requirement to avoid overflow and underflow in finite size buffers, while maintaining synchronization of  
25 the presentation of data. All other synchronized entities must slave the

timing of their operation to the time-master. If a decoder attempts to have more than one simultaneous time-master it may experience problems with buffer management or synchronization. Thus, playback of several streams is synchronized by adjusting the playback of all streams to a single master time base rather than by adjusting the playback of one stream to match that of another. The master time base may be one of the individual stream's decoders' clocks, for example, the video decoder or the audio decoder, it may be the DSM or channel clock, or it may be a separate STC such as some external clock. The time-master must communicate to the others the correct value of the STC. A time slave will typically maintain a local STC which is incremented nominally at 90 kHz (for MPEG-1) between updates or corrections. In this way, each entity has a continuously updated value of the STC which is nominally correct and which it uses to compare with the time-stamps.

15           The MPEG stream properties and semantics defined above are set out in the MPEG standard using a hypothetical reference model of decoding in the form of a hypothetical decoder called the *system target decoder (STD)*. The STD model is used by encoders to create MPEG streams, but real decoders need not be implemented with the same architecture or timing structure as the STD.

          The STD model consists of a demultiplexer followed by, for each elementary stream, a first-in first-out (FIFO) buffer followed by a decoder. In the model, data present in headers of the packs or packets, but not part of packet data (*e.g.*, SCR, DTS, PTS, packet lengths, *etc.*) is not delivered to any of the stream buffers, but may be used to control the

system. In one preferred embodiment of the present invention, such data is used to optimize the efficiency of transmitting VBR data.

In the STD model, the data output of the buffer is decoded instantaneously, and may be delayed in a *reorder buffer* before being presented to the viewer at the output of the STD. Reorder buffers are used only for video decoding in order to store I-pictures and P-pictures while the sequence of presentation pictures is reordered before presentation. This is often necessitated in the presence of B-frames.

As an example, consider a decoder for a video stream which includes an internal clock. The buffer at any time includes packets of the stream. The output of the buffer is connected to the decoder. The decoder needs to know when to decode the next packet. If a packet header has a DTS, then the decoder takes that packet when the DTS is less than or equal to the decoder clock. If a packet does not contain a DTS, then the decoder takes that packet at the next time instant that the decoder expects a packet, based on previous packets decoded.

Thus, MPEG streams containing elementary streams are formed (video and/or audio) and multiplexed together.

#### F. MPEG-2 Profiles and Levels

The MPEG-2 Standard defines classes of image resolution and sound quality (the "*profile*") and a minimum subset of specifications and features of the MPEG standard (the "*level*"). For example, *main profile* ("*MP*") covers the common CCIR-601 standards of a resolution of 704 pixels per line for 525 line/60 Hz systems (NTSC) and 625 line/50 Hz systems (PAL

and SECAM). *Low profile* ("LP") covers the SIF resolution which is half of CCIR-601 and is commonly used in video conferencing, while *high profile* ("HR") is usually reserved for high definition television (HDTV). There are three levels, low level (LL), medium level (ML), and high level (HL). A decoder meeting MPEG-2 compliance specifications must meet these minimum specifications and offer at least the subset of MPEG-2 features of that level. For example, the most common decoder is the MP@ML compliant decoder. It must have a decoder buffer that is at least 1.8M bits in size. An MP@HL compliant decoder must have a larger decoder buffer and must be able to implement many of the scaleable features of MPEG-2.

## VI. The Claims

What is claimed is:

1. A method for increasing channel utilization for a first channel transmitting a multiplex of a set of one or more encoded program streams, each program stream in said set being decodeable by a corresponding decoder, each corresponding decoder including a corresponding decoder buffer, said decoder buffers having a maximum allowable size, said method comprising
  - (a) selecting encoded pictures to be modified, said selecting according to a criterion;
  - (b) modifying each said selected encoded picture to form a corresponding modified encoded picture, said modified encoded picture having less data than said selected encoded picture; and
  - (c) transmitting the corresponding modified encoded pictures through said first channel in place of the selected encoded pictures, whereby said criterion includes preventing any underflow of any decoder buffer.
2. The method of claim 1, wherein said step of modifying deletes each said selected encoded picture, said modified encoded picture having no data.
3. The method of claim 1, wherein said encoded program streams include predictively encoded pictures, each said predictively

encoded picture having one or more corresponding anchor pictures, each said predictively encoded picture having prediction data and prediction error data, said step of selecting selects predictively encoded pictures that are not anchor pictures, and said step of modifying deleted the prediction  
5 error data from each said selected encoded picture.

4. The method of claim 3, wherein said encoded program streams are MPEG encoded program streams, and said step of selecting selects B-pictures.

5. The method of claim 3, wherein said step of selecting includes  
10 a second criterion.

6. The method of claim 5, wherein said second criterion includes maintaining essentially constant perceived image quality.

7. The method of claim 1, further comprising the step of:

(d) transmitting the selected encoded pictures.

15 The method of claim 7, wherein said said of transmitting transmits through a second channel

9. The method of claim 7, further comprising the steps of:

(e) determining whether it is possible to decode said selected encoded picture from said first channel; and

(f) whenever possible, decoding said selected encoded picture from rather than the corresponding modified encoded.

10. The method of claim 9, wherein said encoded program streams include predictively encoded pictures, each said predictively  
5 encoded picture having one or more corresponding anchor pictures, each said predictively encoded picture having prediction data and prediction error data, said step of selecting selects predictively encoded pictures that are not anchor pictures, and said step of modification deleted the prediction error data from each said selected encoded picture.

10 11. The method of claim 10, wherein said encoded program streams are MPEG encoded program streams, and said step of selecting selects B-pictures.

15 12. The method of claim 11, wherein said step of determining includes comparing one or more time stamps of the selected encoded pictures and one or more time stamps of the corresponding modified encoded pictures.

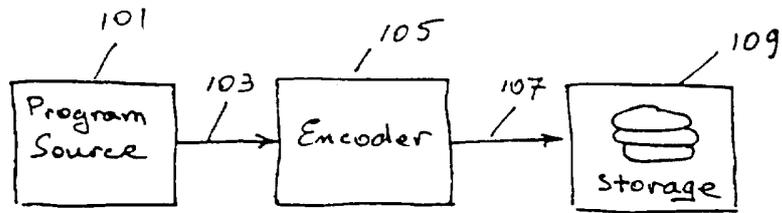


Figure 1

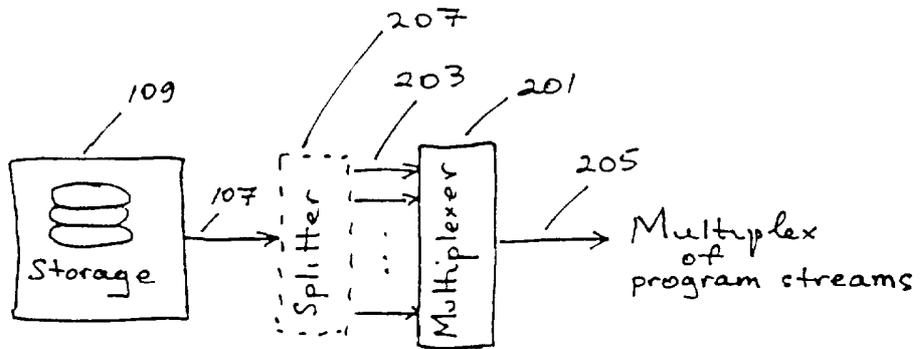


Figure 2(a)

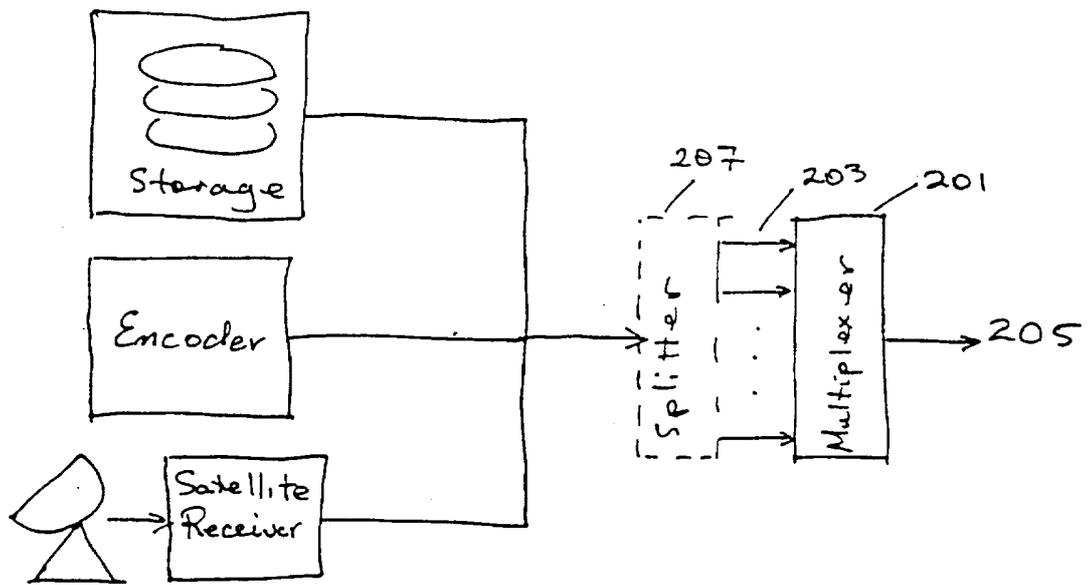


Figure 2(b)

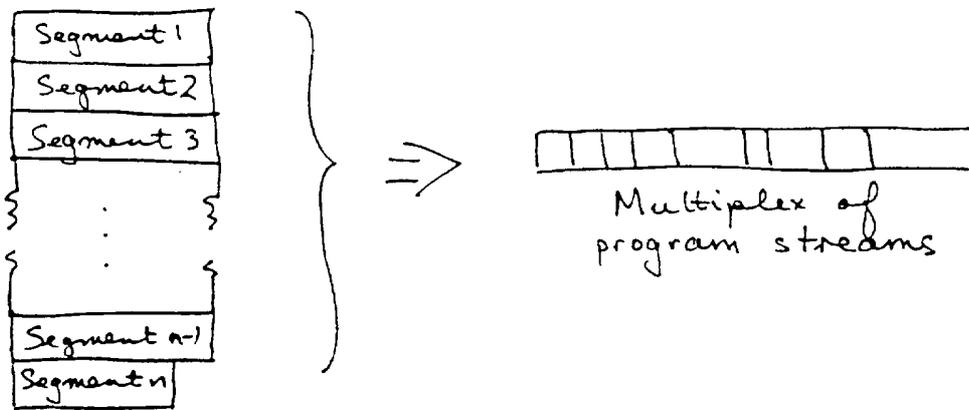
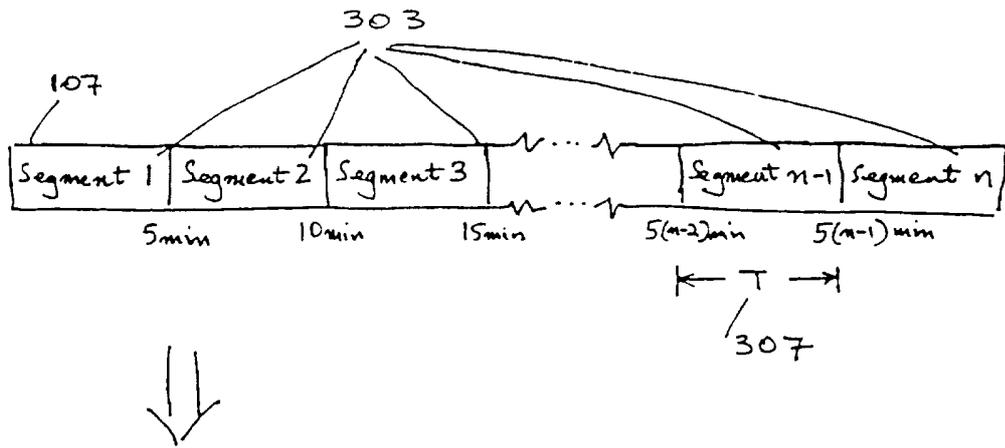


Figure 3

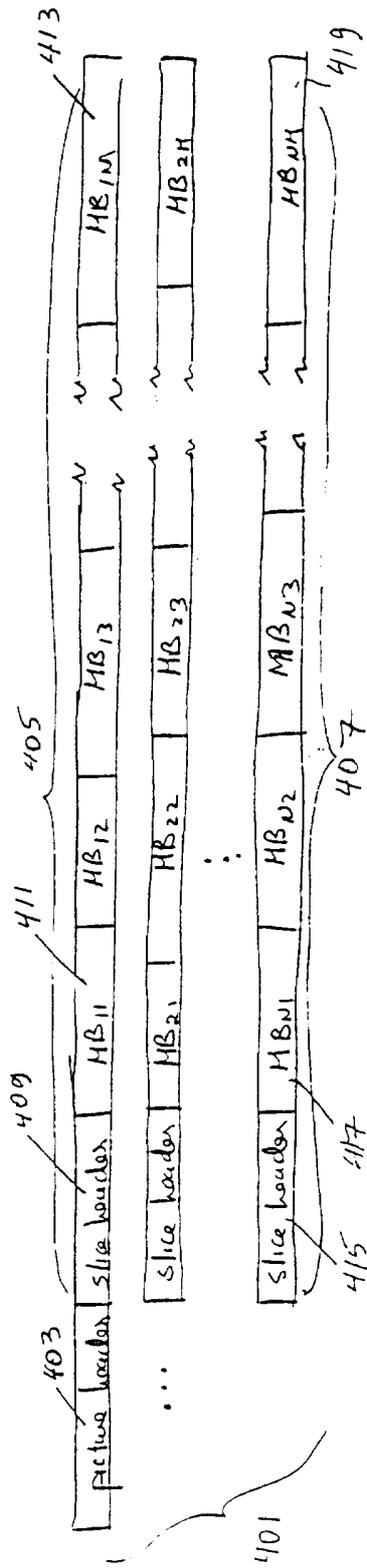


Figure 4(a)

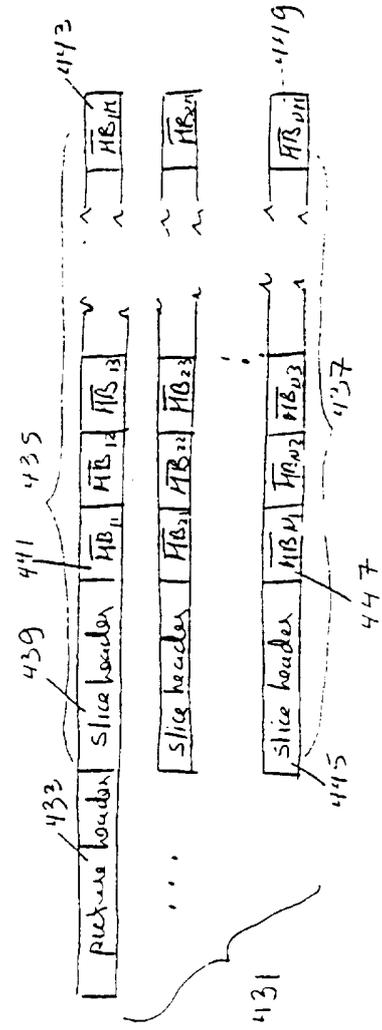


Figure 4(b)

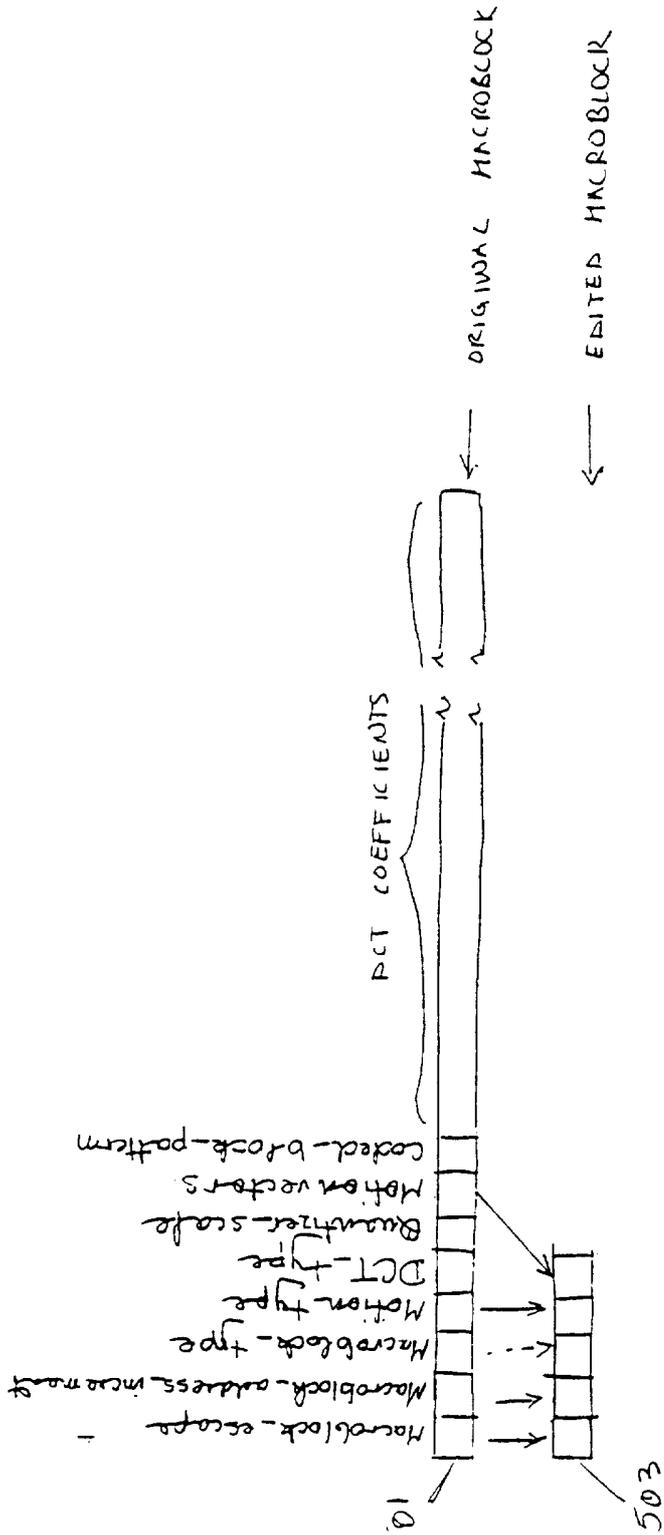


Figure 5(a)

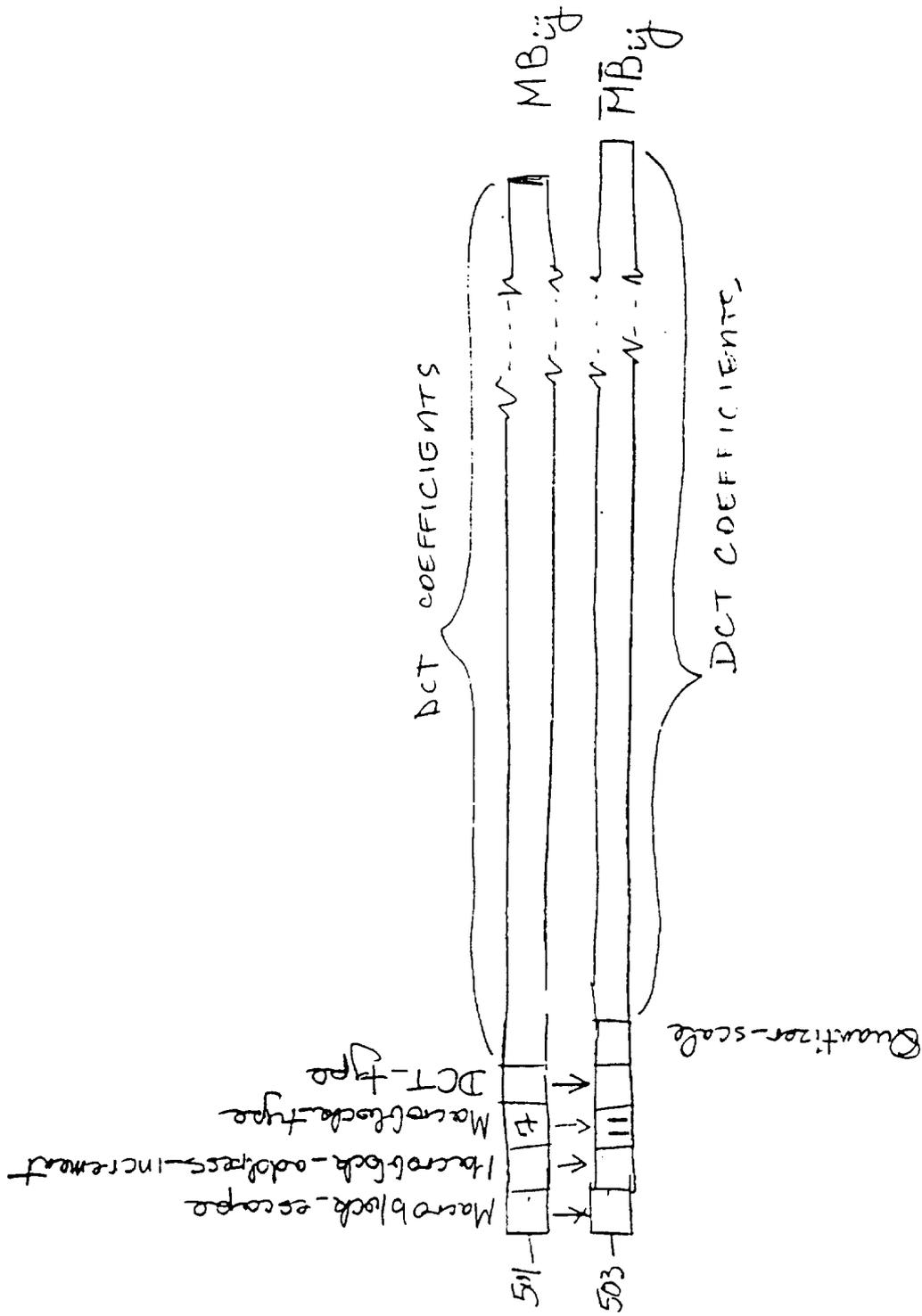


Figure 5(b)

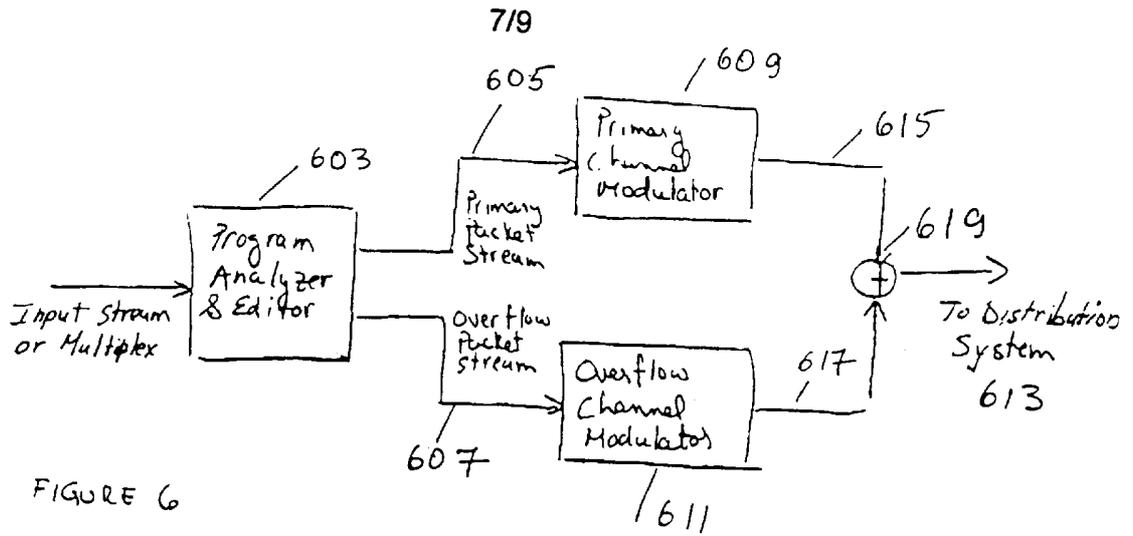


FIGURE 6

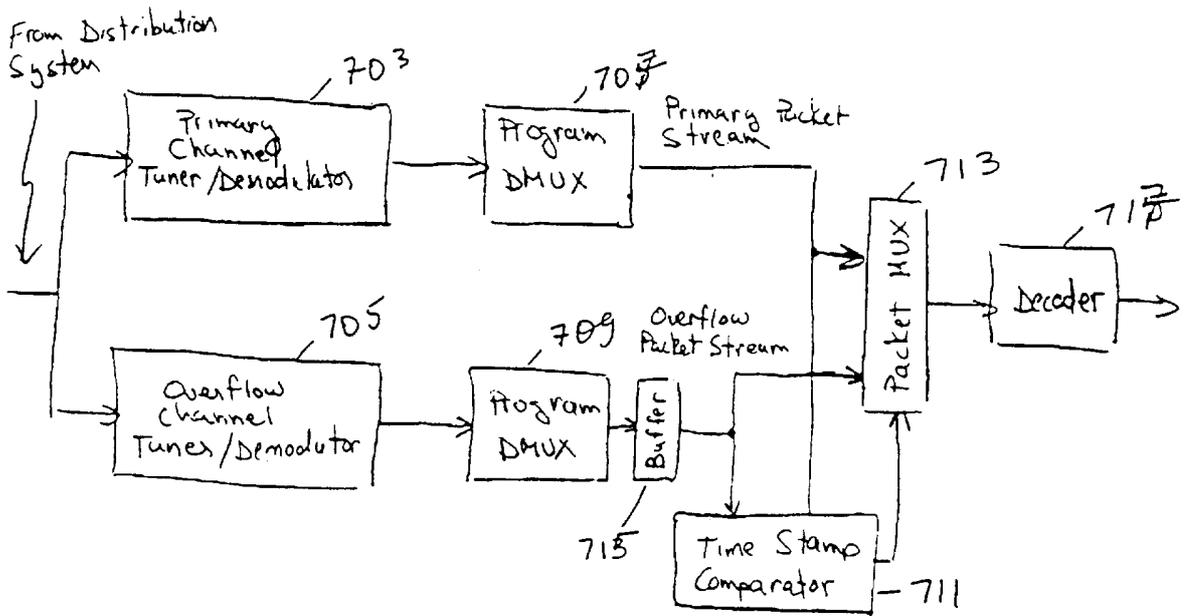


FIGURE 7 : Enhanced Multi-Channel Receiver

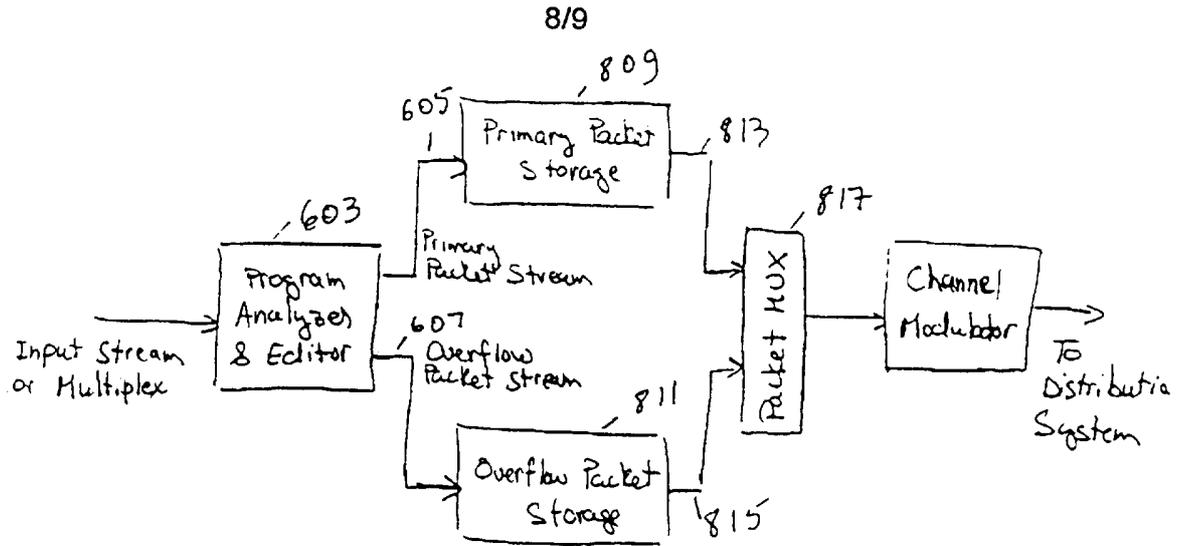


FIGURE 8

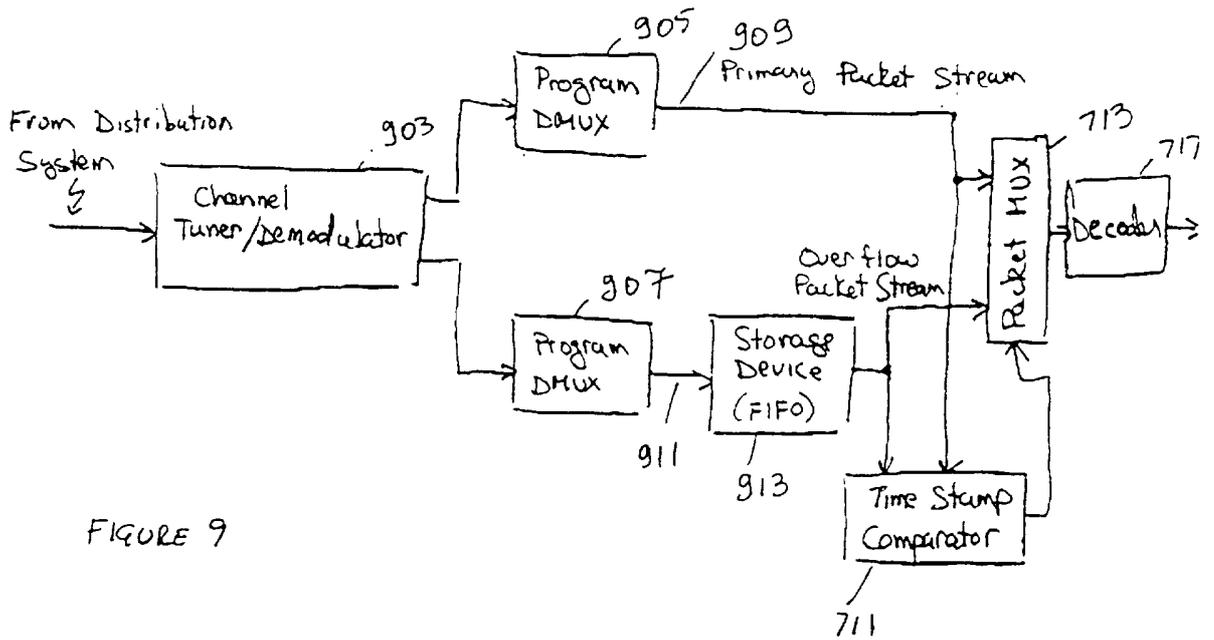


FIGURE 9

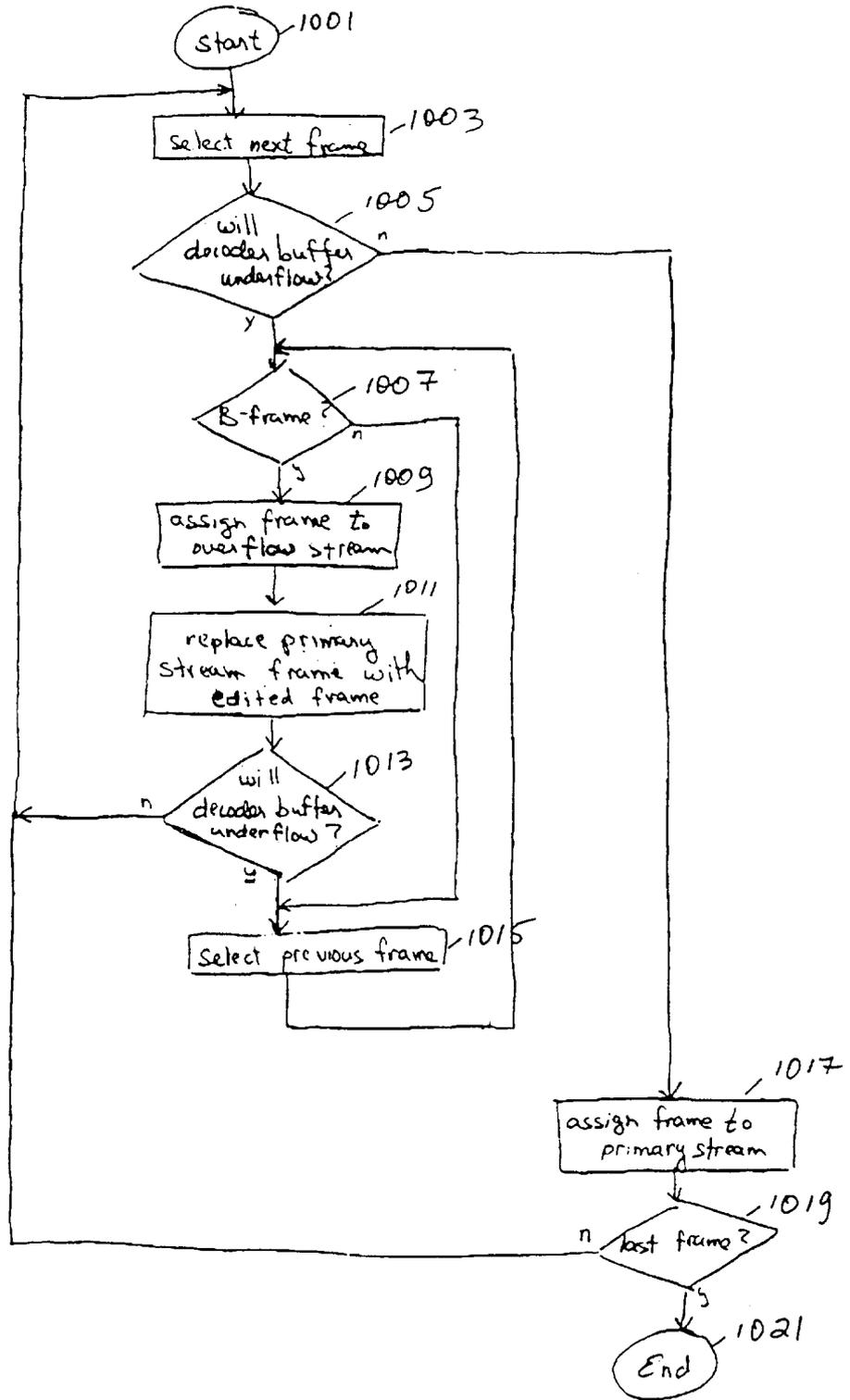


FIGURE 10

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 96/18091

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC 6 H04N7/26 H04N7/58 H04N7/46		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols) IPC 6 H04N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y A	EP 0 661 885 A (CANON KK) 5 July 1995 see the whole document ---	1-6 7-12
Y A	GB 2 289 194 A (SONY CORP.) 8 November 1995 see the whole document ---	1-6 7-12
A	ICIP 94, vol. 2, 13 November 1994, AUSTIN, TX, USA, pages 947-951, XP000522755 PURI ET AL.: "Temporal Resolution Scalable Video Coding" see the whole document ---	1,7-12
-/--		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C.		
<input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents :		
*A* document defining the general state of the art which is not considered to be of particular relevance	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention	
*E* earlier document but published on or after the international filing date	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone	
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.	
*O* document referring to an oral disclosure, use, exhibition or other means	*&* document member of the same patent family	
*P* document published prior to the international filing date but later than the priority date claimed		
Date of the actual completion of the international search  <div style="text-align: center; font-size: 1.2em;">19 March 1997</div>	Date of mailing of the international search report  <div style="text-align: center; font-size: 1.2em;">10.04.97</div>	
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. ( + 31-70) 340-2040, Tx. 31 651 epo nl, Fax: ( + 31-70) 340-3016	Authorized officer  <div style="text-align: center; font-size: 1.2em;">Foglia, P</div>	

# INTERNATIONAL SEARCH REPORT

Inter. Application No  
PCT/US 96/18091

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ICC 95, vol. 3, 18 June 1995, SEATTLE, WA, USA, pages 1742-1746, XP000535047 SUN ET AL.: "MPEG Video Coding with Temporal Scalability" see the whole document ---	1,7-12
A	US 5 144 425 A (JOSEPH KURIACOSE) 1 September 1992 see the whole document ---	1,7-12
P,A	EP 0 711 077 A (AT&T CORP.) 8 May 1996 ---	
A	EP 0 627 858 A (MATSUSHITA ELECTRIC INDUSTRIAL CO. LTD.) 7 December 1994 -----	

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No  
PCT/US 96/18091

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 661885 A	05-07-95	JP 7203429 A JP 7222146 A	04-08-95 18-08-95
-----			
GB 2289194 A	08-11-95	JP 5047108 A JP 5048565 A JP 5048566 A GB 2289195 A,B GB 2289196 A,B DE 4227451 A GB 2259229 A,B US 5291486 A	26-02-93 26-02-93 26-02-93 08-11-95 08-11-95 25-02-93 03-03-93 01-03-94
-----			
US 5144425 A	01-09-92	CN 1070531 A DE 4228220 A FR 2682846 A GB 2259216 A,B JP 5211645 A	31-03-93 04-03-93 23-04-93 03-03-93 20-08-93
-----			
EP 711077 A	08-05-96	CA 2159846 A JP 8242451 A	02-05-96 17-09-96
-----			
EP 627858 A	07-12-94	JP 6343156 A US 5570197 A	13-12-94 29-10-96
-----			