

(51) International Patent Classification:
H04N 7/24 (2006.01) *H04N 5/00* (2006.01)(21) International Application Number:
PCT/EP2009/007027(22) International Filing Date:
30 September 2009 (30.09.2009)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/113,671 12 November 2008 (12.11.2008) US(71) Applicant (for all designated States except US):
FRAUNHOFER-GESELLSCHAFT ZUR FÖRDERUNG DER ANGEWANDTEN FORSCHUNG E.V. [DE/DE]; Hansastrasse 27, 80686 München (DE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **FUCHS, Harald** [DE/DE]; Auf der Höhe 28, 91341 Röttenbach (DE). **DOEHLA, Stefan** [DE/DE]; Hartmannstrasse 47a, 91052 Erlangen (DE). **JENNEHAG, Ulf** [DE/DE]; Mittlere Schulstrasse 7, 91054 Erlangen (DE). **THOMA, Herbert** [DE/DE]; Geschwister-Scholl-Strasse 1, 91058 Erlangen (DE). **FAERBER, Nikolaus** [DE/DE]; Thomas-Dehler-Strasse 28, 91052 Erlangen (DE).(74) Agents: **ZIMMERMANN, Tankred** et al.; SCHOPPE, ZIMMERMANN, STÖCKELER & ZINKLER, P.O. Box 246, 82043 Pullach (DE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))

(54) Title: REDUCING A TUNE-IN DELAY INTO A SCALABLE ENCODED DATA STREAM

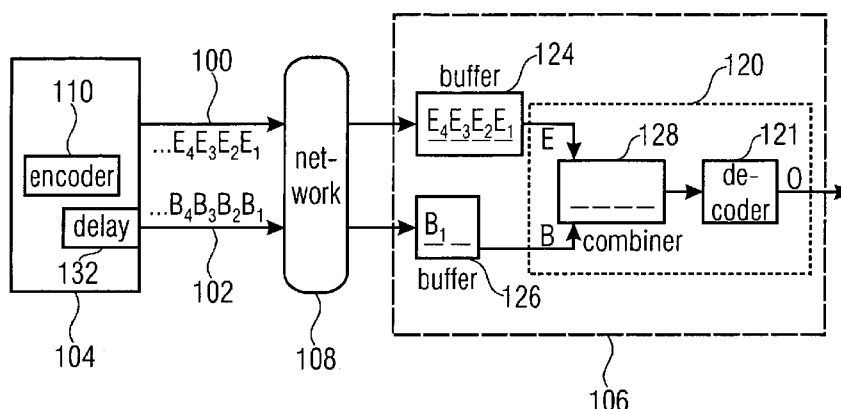


FIGURE 1A

(57) Abstract: A method for reducing a tune-in delay upon tuning into a scalable encoded data stream is described, wherein the scalable encoded data stream is provided by an encoder and comprises a base layer stream and at least one enhancement layer stream. The base layer stream and the enhancement layer stream are provided from the encoder to a decoder with substantially the same encoder-to-decoder delay.

REDUCING A TUNE-IN DELAY INTO A SCALABLE ENCODED DATA STREAM

5

Background of the Invention1. Technical Field of the Invention

Embodiments of the invention relate to the field of generating streams of data including a plurality of encoded data blocks, wherein such kind of streams are transmitted to a receiver and decoded for presentation of the data within the stream. More particularly, and not by way of any limitation, the invention is related to the field of media transmission, reception and playback, and embodiments of the invention concern a fast tune-in into a stream transmission over IP networks.

2. Description of the Related Art

In the field of media transmission over an IP network (for example IPTV systems) video data is transmitted in encoded and compressed form, and popular video compression standards, such as MPEG-2 and JVT/H.264/MPEG AVC, use intra-coding and inter-coding. For proper decoding a decoder decodes a compressed video sequence beginning with an intra-coded picture (e.g. an I-picture or I-frame) and then continues to decode the subsequent inter-coded pictures (e.g. the P-pictures or P-frames and/or the B-pictures or B-frames). A group of pictures (GOP) may include an I-picture and several subsequent P-pictures and/or B-pictures, wherein I-pictures require more bits to code than P-pictures or B-pictures for the same quality of the video. Upon receipt of the video stream on a particular channel, for example after changing to this channel or after turning on the receiver, decoding has to wait until the first I-picture is received. To minimize the delay in the coding of the video stream I-pictures are sent frequently, i.e. are included within the video stream at a fixed distance, for example every 0.5 seconds.

One problem of current IPTV systems is the so-called tune-in time into streams that are distributed over multicast IP. The delay between the initialization of tuning into a channel and rendering the content of this channel is due to several effects of which client pre-buffering time and acquisition time for random access points within the stream to be switched to are the dominant ones. Both effects are direct implications of the design of modern video codec schemes. In differential video coding schemes, like MPEG-2 video or MPEG-4 AVC/H.264, only a few pictures of a stream are self-contained, e.g. the above-mentioned I-pictures. These pictures include all information that is necessary to decode the complete picture. Most other pictures are differentially coded and depend on one or more previously transmitted and decoded pictures, e.g. the above-mentioned P-pictures or B-pictures. In other words, the P-pictures or B-pictures do not include all information that is necessary to decode a complete picture, rather additional information from preceding or following pictures is required.

To obtain the best coding efficiency at a given bit rate, the number of I-pictures should be low. On the other hand, the I-pictures serve as random access points (RAP) to the stream where decoding can be started. Hence, there is a delay when tuning into a new stream, since the client (receiver) has to wait for a random access point within the stream to arrive, before it can start decoding and displaying video.

In differential coding schemes, the encoded video bit rate is not necessarily constant but rather depends on the complexity of the video scene. Within a video stream, the variation in coded picture size can be large, for example the I-pictures can be many times as large as differentially encoded pictures, the P-pictures and B-pictures. Upon transmitting such a bit stream over a channel with constant channel bit rate, the client needs to pre-buffer incoming

picture data so that the video can be played with the same rate as it was sampled. This buffer needs to be large enough to avoid buffer overflow and shall only be emptied on reaching a certain buffer fullness for avoiding buffer
5 underrun during playout.

Whenever the buffer cannot be filled instantly to the point where the client can start emptying it, delay occurs before rendering can be started.

10

This functionality is disadvantageous as the receiver which begins receiving a program on a specific channel, for example following a channel change or turning on the receiver must wait until the random access point, for example an I-
15 picture is received, so that decoding can start. Thus, the distance of random access points within the main stream is one of the main causes for the tune-in delay.

One approach to reduce such delay in a multicast linear TV
20 scenario is to send a second stream in parallel to the main stream, wherein the second stream has a higher frequency of random access points. This second stream is for example called the "tune-in stream" or the "side stream".

25 Fig. 3 illustrates tuning into a main stream using a secondary or tune-in stream. Fig. 3 illustrates along the X-axis the time and along the Y-axis the quality level of the respective streams. In Fig. 3, the full quality Q_s of the main stream 100 is 100% and the side stream or tune-in
30 stream 102 has a lower quality Q_i , which is an intermediate quality level, which is lower than the quality level of the main stream 100. When the user initiates a channel change at time t_0 , tuning into the side stream 102 occurs. The side stream comprises more frequent random access points so
35 that the initial start-up delay ($t_r - t_0$) for decoding is reduced by using the tune-in stream 102 having more frequent I-pictures. A decoder within a receiver will obtain a first I-picture from the tune-in stream 102 for the new channel

earlier than the first I-picture of the main stream 100. However, as mentioned above, the quality of the tune-in stream 102 is lower than the quality of the main stream, e.g. the pictures are encoded at different quality levels, which is necessary to limit the additional bit rate that is necessary for the tune-in stream as same comprises more I-pictures which are many times larger than the other pictures. Therefore, the tune-in stream 102 is encoded at a lower intermediate quality level Q_i , for example, using a lower image resolution, for example, only a quarter resolution when compared to the full resolution of the main stream.

During the transition period ($t_T - t_R$) starting at t_R , the receiver or client decodes the pictures derived from the tune-in stream 102 until a full resolution I-picture arrives on the main stream at time t_T . Once this I-picture arrives, the low resolution stream is stopped and the full quality pictures of the main stream are decoded and rendered.

The above-described behavior of fast tune-in using a transition period of intermediate quality may also be achieved in combination with scalable video coding (SVC), for instance with the MPEG-4 SVC standard. It is noted that in this document the term "SVC" is used as a generic acronym for scalable coding or scalable video coding and does not only refer to MPEG-4 SVC.

In case of SVC, the base layer stream corresponds to the tune-in stream and the enhancement layer corresponds to the main stream. However, one difference to conventional approaches using streams of different quality is that in accordance with the SVC technique the base layer is always used, not only during the transition period, so that the client or receiver always receives both streams. As described above, also when using the SVC technique the shorter distance of random access points, I-pictures, in

the base layer stream reduces the delay contribution caused by waiting for the I-picture to arrive. However, as shall be described below in further detail, the client needs not only to wait for the first I-picture but also needs to pre-
5 buffer incoming data (independent of reception of the first I-picture) before starting to decode and playing out frames.

Fig. 4 is an example of a conventional system for transmitting a SVC stream from a transmitter 104 to a client or receiver 106. The transmitter 104 and the client 106 are connected via a network 108, for example the Internet. The transmitter 104 comprises an encoder 110 which receives at an input I an input signal 112 which is to be encoded in
10 accordance with the scalable coding technique. In case of the signal 112 being a video signal or including video contents, the encoder 110 is provided to apply the scalable video coding technique to provide at a first output E the enhancement layer stream or main stream 100. At a second
15 output B the encoder 110 provides the base layer stream or tune-in stream 102. The transmitter 104 further comprises an enhancement layer stream buffer 114 and a base layer stream buffer 116 receiving from the first and second outputs, respectively, the enhancement layer stream 100 and
20 the base layer stream 102, respectively. The buffers 114 and 116 are provided to allow for a constant transmission rate to be provided by the transmitter 104 for both the enhancement layer stream 100 and the base layer stream 102. The buffered streams 100 and 102 are transmitted via the
25 network 108 to the client 106.

The client 106 comprises a decoder 120 having an output O for providing a decoded signal 122 for further processing. The decoder 120 further comprises a first input E and a
35 second input B for receiving the enhancement layer stream and the base layer stream, respectively, for decoding thereof. The client 106 further comprises an enhancement layer stream buffer 124 and a base layer stream buffer 126.

The enhancement layer stream buffer 124 receives the enhancement layer stream transmitted by the network 108. Once a required buffer fill level for the buffer 124 is reached the enhancement layer stream data is output to the decoder 120 for decoding. The base layer stream buffer 126 receives the base layer stream which is transmitted by the encoder 104 via the network and, like the buffer 124, buffers a predetermined amount of data from the base layer stream before forwarding the base layer stream data to the decoder.

10

Reference is now made to Fig. 5 showing a detailed block diagram of the decoder 120 of the client 106 in Fig. 4. In the example described with regard to Fig. 3 the use of a main stream and a tune-in stream was described, wherein these two streams are independent from each other and two independent decoders are used to decode the main and tune-in streams. In the scalable coding technique this is different, as always all bit stream data of a frame or picture are to be decoded during one pass, which is also called "single loop decoding". Thus, to allow decoding the bit stream data must be provided in the correct order at the input of the SVC decoder, and not only in the correct temporal order but also the base layer data and the enhancement layer data for each frame need to be interleaved correctly. When transmitting the two layers independent from each other it is required to sort the data into the correct order which is especially important at the time of switching to the enhancement layer for obtaining an output signal with improved quality. However, it is noted that this re-ordering, i.e. providing the correct order of the data from the enhancement layer and the base layer to the decoder is also needed during the further operation.

As is shown in Fig. 5, for SVC the client 106 comprises the two buffers 124 and 126 and the decoder 120 which comprises a SVC decoder 121 and a combiner 128 coupled between the buffers 124, 126 and the SVC decoder 121. The combiner 128 is provided for re-generating the combined bit stream of

35

all layers so that the SVC decoder 121 can correctly decode the contents from the original input signal. However, to be in a position to correctly re-generate a combined bit stream of all layers the combiner 128 requires input from the enhancement layer stream and from the base layer stream, as it is provided by the respective buffers 124 and 126.

Turning now back to Fig. 4 the drawbacks of conventional designs when tuning into a stream will be discussed. As can be seen from Fig. 4 the buffers 114, 116, 124 and 126 are depicted with different dimensions to indicate that the buffer size of these respective buffers is different. For transmitting the base layer stream only a small amount of base layer stream data needs to be buffered, as it is indicated by the smaller size buffers 116 and 126, whereas a higher amount of data needs to be buffered for the enhancement layer stream, as it is indicated by the larger size buffers 114 and 124. Assuming now a situation that tuning into the stream generated from the input signal 112 just took place the encoder will start transmitting the enhancement layer stream 100 and the base layer stream 102 via the network 108 to the client 106. At the client the received streams are buffered however, as is indicated by the hatched areas 124a and 126a the base layer stream buffer 126 has reached its required buffer fill level while, at the same time, only a part of the necessary data to be buffered for the enhancement layer stream is received by buffer 124, i.e. the buffer fill level for the enhancement layer stream is not yet reached. Thus, despite the fact that information from the base layer stream is already available at the input B of the decoder 120 (as is indicated by the hatched area 130) decoding cannot yet start as the combiner 128 (see Fig. 5) is not in a position to re-generate the combined bit stream of all layers as the required information from the enhancement layer buffer 124 is still missing.

Thus, in accordance with the conventional system described above, the client 106 is not in a position to start decoding even in case the data 130 already includes an I-picture as the required buffer fill level for buffer 124 is not yet reached so that the necessary data from the enhancement layer stream cannot be presented to the combiner 128. Thus, besides the necessary time to wait for an I-picture also the time to fill up the buffer 124 to the required level adds to the overall tune-in delay when tuning to a scalable video encoded stream received at the client 106.

Thus, a need exists to provide an approach providing a fast tune-in into a scalable coded stream transmitted to a client.

SUMMARY OF THE INVENTION

An embodiment of the invention provides a method for reducing a tune-in delay upon tuning into a scalable encoded data stream which is provided by an encoder and comprises a base layer stream and at least one enhancement layer stream, wherein the base layer stream and the at least one enhancement layer stream are provided separate from the encoder to a decoder, wherein the base layer stream and the at least one enhancement layer stream are combined for decoding such that the combined stream comprises for each data block to be decoded data from the base layer stream and data from the at least one enhancement layer stream, and wherein the method comprises transmitting the base layer stream and the at least one enhancement layer stream time-shifted from the encoder to the decoder.

Another embodiment of the invention provides a method for providing a scalable encoded data stream for transmission to a decoder, wherein the scalable encoded data stream comprises a base layer stream and at least one enhancement layer stream, and wherein the base layer stream and the at

least one enhancement layer stream are combined for decoding such that the combined stream comprises for each data block to be decoded data from the base layer stream and data from the at least one enhancement layer stream, and
5 wherein the method comprises delaying the base layer stream with respect to the at least one enhancement layer stream.

Yet another embodiment of the invention provides a system for reducing a tune-in delay upon tuning into a scalable
10 encoded data stream which comprises a base layer stream and at least one enhancement layer stream, the system comprising an encoder which is configured to provide the base layer stream and the at least one enhancement layer stream of the scalable encoded data stream; a network which is
15 coupled to the encoder and is configured to transmit the base layer and the at least one enhancement layer stream; and a decoder which is coupled to the network and is configured to receive the base layer stream and the at least one enhancement layer stream and to decode a combined
20 stream comprising the base layer stream and the at least one enhancement layer stream, wherein the base layer stream and the at least one enhancement layer stream are combined for decoding such that the combined stream comprises for each data block to be decoded data from the base layer
25 stream and data from the at least one enhancement layer stream, wherein the system is configured to transmit the base layer stream and the at least one enhancement layer stream with a time-shift from the encoder to the decoder.

30 A further embodiment of the invention provides an encoder for providing a scalable data stream, comprising an encoder section which is configured to scalable encode an input signal to obtain a base layer stream and at least one enhancement layer stream, wherein the base layer stream and
35 the enhancement layer stream are combined for decoding such that the combined stream comprises for each data block to be decoded data from the base layer stream and data from the at least one enhancement layer stream; a first buffer

which is configured to buffer the base layer stream; and a second buffer which is configured to buffer the at least one enhancement layer stream; wherein the encoder is configured to transmit the base layer stream and the at least one enhancement layer stream with a time-shift.

Embodiments of the invention concern encoders, streaming servers, network components and clients, for example for IPTV systems, wherein a fast tuning into a stream or a fast channel change is achieved by using scalable coding, wherein the pre-buffering delay in a client device is reduced by a time shifted transmission of the base and enhancement layers which is obtained by an intentionally introduced additional delay for the base layer stream and, if present, optionally for lower level enhancement layers.

BRIEF DESCRIPTION OF THE DRAWINGS

20

The embodiments of the invention will be described in the following with reference to the accompanying drawings, wherein:

25 Fig. 1 is a schematic representation of a system for reducing the tune-in delay in accordance with an embodiment of the invention, wherein Fig. 1(a) represents a situation shortly after tuning into the stream with the decoder buffers not yet having reached a required fill level, and wherein
30 Fig. 1(b) shows the system with its decoder buffers filled so that decoding of the combined stream can start;

35 Fig. 2 is a block diagram of a system for reducing the tune-in delay in accordance with another embodiment of the invention;

Fig. 3 illustrates a conventional approach for tuning into a stream using a tune-in stream;

5 Fig. 4 is an example of a conventional system for transmitting a SVC stream, wherein a situation shortly after tuning into the stream is shown; and

10 Fig. 5 is a detailed diagram of the decoder of the client shown in Fig. 4.

DESCRIPTION OF EMBODIMENTS OF THE INVENTION

15 In the following, embodiments of the invention will be described. One embodiment of the invention concerns an approach for reducing a tune-in delay when tuning into a stream of data, for example a video stream or an audio stream, wherein the stream is encoded in accordance with
20 the scalable coding technique, for example in accordance with the scalable video coding approach so that a base layer stream and at least one enhancement layer stream are generated. Each of these streams comprises a plurality of encoded data blocks which comprise a plurality of self-
25 contained blocks including all information for decoding the block and a plurality of blocks including only partial information for decoding. In accordance with embodiments of the invention, the stream of data may be a video stream being encoded using intra- and inter-coding. The encoded
30 streams may comprise I-pictures as random access points and P-pictures and/or B-pictures, wherein the number and distance of I-pictures in the base layer stream and in the at least one enhancement layer stream are different from each other.

35

In accordance with embodiments of the invention a reduction of the pre-buffer fill delay described above is achieved by a time-shifted transmission between the base layer streams

and the enhancement layer stream. The base layer stream may be delayed at the server or transmitter so that the enhancement layer stream is virtually transmitted "ahead of time". In this way, the delay of the base layer stream compensates the buffer time that the client needs to pre-buffer the enhancement layer stream's data packets. This buffer time insures that the pre-buffering conditions of the enhancement layer stream and the total bit stream comprised of the base layer stream and the enhanced stream, respectively, are already fulfilled as soon as the first I-picture of the enhancement layer stream is needed for decoding and rendering.

Fig. 1 is a schematic representation of the system for reducing the tune-in delay in accordance with an embodiment of the invention. In the subsequent description of the embodiments of the invention given with regard to Figs. 1 and 2 elements, which were already described with reference to Figs. 4 and 5 have the same reference sign and are not again described in detail.

Fig. 1(a) shows a situation shortly after tuning into a stream which is provided by the provider or transmitter 104. As described above, the transmitter 104 comprises the encoder 110 and also the encoder buffers which are not shown in Fig. 1. The transmitter 104 outputs the base layer stream 102 and the enhancement layer stream 100 for transmission via the network 108 to the client 106. The transmitter 104 includes a delay element 132 which is provided to delay the base layer stream 102 with respect to the enhancement layer stream 100 in such a manner, that transmission of the base layer stream 102 is delayed when compared to transmission of the enhancement layer stream 100 so that the enhancement layer stream is virtually transmitted "ahead of time". Delaying the base layer stream 102 in accordance with the inventive approach is done such that the pre-buffer fill delay at the client 106 is reduced. For example, the delay 132 is selected such that the delay be-

tween the encoder 110 and the decoder 120 is substantially the same which results in a filling of the respective buffers 124 and 126 such that same reach the predefined or required buffer fill level at approximately the same instant, i.e. the buffers 124 and 126 are uniformly filled. In the situation shown in Fig. 1(a) it is assumed that the base layer stream comprises the base layer packets B_1 to B_4 and further base layer packets, whereas the enhancement layer stream comprises the packets E_1 to E_4 and additional packets. As is schematically indicated in Fig. 1(a) the buffers 124 and 126 of the client 106 are filled uniformly, which means that the buffers filled with the same rate and as is seen in Fig. 1(a) the buffers 124 and 126 did not yet reach their required buffer fill level. In the situation shown in Fig. 1(a) the combiner 128 did not yet receive packets for combining the streams 100 and 102 into a combined stream needed for decoding and rendering.

Fig. 1(b) shows a situation in which the buffers 124, 126 of the client 106 reached their required buffer fill level so that the combiner 128 already received packets from the two buffers for recombining same into the combined stream which can be used for decoding at SVC decoder 121. In view of the inventive approach of delaying the base layer stream the pre-buffer fill delay discussed above is reduced or even avoided as the buffers 124, 126, due to the delay of the base layer stream, reach the required fill level at the same time or instant or with a reduced delay so that the combiner 128 receives data from both buffers 124 and 126 for generating the combined stream.

To be able to align both time lines at the decoder/playout or rather at the input of the combiner 128, both the base layer stream 102 and the enhancement layer stream 100 need to have the same end-to-end delay so that the tune-in stream or base layer stream needs to be additionally delayed at the encoder by a delay that is the difference of the end-to-end delay of the total stream (base layer stream

+ enhancement layer stream) and the end-to-end delay of the base layer stream. Actually, the pre-buffering delay in the client device may be reduced to zero, in case of an ideal network without jitter. If the delay at the encoder is chosen to just compensate the pre-buffering delay in the decoder, no additional end-to-end delay is introduced. Additional network de-jitter buffer for the enhancement layer stream can be added without additional tune-in delay by increasing the time-shift between the base layer stream and the enhancement layer stream.

The delay 132 can be introduced into the base layer stream, for example by providing an additional delay element in the transmitter 104 or by modifying the base layer stream buffer in the encoder to set a desired buffer delay.

Following the situation shown in Fig. 1(b) decoding of the stream received at the client 106 can start. For example, once a tune-in request is issued by the client 106 the scalable encoded stream is received at the client 106 in form of the base layer and the enhancement layer streams. Once the required buffer-fill level for the enhancement layer stream is reached and once a first self-contained data block in the base layer, like an I-picture in the base layer, is received at the decoder 120 decoding of only the base layer starts thereby already providing an output -with reduced quality- despite the fact that the enhancement layer did not yet include the I-picture. Once the I-picture in the enhancement layer is received at the decoder 120 both the base layer stream and the enhancement layer stream will be decoded thereby changing the output from the lower quality to the higher quality.

Fig. 2 is a block diagram of a system for reducing the time-in delay in accordance with another embodiment of the invention. Fig. 2 is similar to the system shown in Fig. 4, however, in accordance with the invention the additional delay 132 is introduced into the base layer stream path of

the transmitter 104. The delay 132 is realized by providing an additional delay element, for example an additional buffer element in the just-mentioned base layer stream path so that the base layer stream 102 is delayed with respect to the enhancement layer stream 100. Instead of realizing the delay 132 by means of an additional buffer inside the transmitter 104 another kind of delay element may be used, like for example a buffer device in the network at any position between the transmitter and the client.

In addition, instead of providing an additional element 132 a buffer delay time of the existing buffer 116 may be increased, for example by using a buffer having a larger capacity.

In a conventional system as it is shown in Fig. 4, the end-to-end delays D_{100} and D_{102} of the enhancement layer stream and the base layer stream are as follows:

$$\begin{aligned} D_{100} &= d_{114} + d_{124} \\ D_{102} &= d_{116} + d_{126} \end{aligned}$$

wherein d indicates the delays of the respective buffers. As mentioned above, to be able to align both time lines at the decoder/playout both streams need to have the same end-to-end delay. Therefore, the tune-in stream or base layer stream needs to be additionally delayed at the decoder by d^* , that is the difference of the end-to-end delay of the total stream and the end-to-end delay of the base layer stream, namely

$$d_{114} + d_{124} = d_{116} + d^* + d_{126}$$

To allow for a short tune-in time, the base layer stream need to have a very low decoder buffer delay D_{126} so that the end-to-end delay of the base layer stream is much lower than the end-to-end delay of the enhancement layer stream, $D_{102} \ll D_{100}$. This can be achieved by a VBR-like encoding and

transmission of the base layer stream. This is possible even for bit rate constrained channels because the bit rate of the base layer stream is much lower than a total bit rate of the complete stream being the sum of the base layer stream and the enhancement layer stream. The enhancement layer stream is encoded in a way so that the total base and enhancement bit rate fulfills the buffer requirements of a CBR encoded stream for bandwidth-constrained transmission. This can be achieved by using an encoder rate control that handles the buffer management of all (base and enhancement layer) streams during the encoding process.

The inventive approach is not limited to the use of only a single enhancement layer stream, rather the above-described inventive approach can also be used with more than one enhancement layer. Fig. 2 shows in dotted lines an additional enhancement layer path 100' including at the transmitter site an additional buffer 114' and at the client side also an additional enhancement layer stream buffer 124'. Also, more than two enhancement layers may be used. Several enhancement layers may be used to achieve a smoothly increasing video quality during the transition period (see Fig. 3 described above). To enable this, every higher layer is using a longer I-picture distance so that after tune-in the client is starting with a base layer, and at the end of the first transition period additionally the first enhancement layer is used. Then, at the end of the second transition period the base layer, the first enhancement layer and the second enhancement layer are used and so on. The end of the complete transition period is reached once the I-picture in the top layer is received and ready for decoding. In such an environment it is noted that the buffers of the different enhancement layer streams may not require the same size as the buffer requirements may differ from enhancement layer to enhancement layer, i.e. the data required to be buffered increases with the level of the enhancement layer. In such a situation, also the enhancement layers may require additional delay, for example all enhancement layers

except for the top-level layer are associated with an increasing amount of delay so that the time-line of all streams are aligned.

- 5 While in the above-described embodiment the enhancement layer stream was buffered at the client 106 it is noted that the invention is not limited to such an environment. Rather, the inventive approach is also applicable to a receiver/server system (edge server). In such a system the
- 10 one or more enhancement layers are buffered in the edge server and the memory requirements for buffering at the client are reduced, because only a sub-part of the stream, not the full stream needs to be buffered. In addition, the edge server is enabled to push (=transmission faster than
- 15 real-time) only a sub-part of the stream, for example a lower level enhancement layer, to the client without exceeding the nominal bit rate of the full stream. This approach may be important in case the last mile link, for example the DSL line, does not enough additional bandwidth,
- 20 for example because other channels are transmitted in parallel. The edge server is able to adapt the fast push to the available bandwidth on the last mile link. If sufficient bandwidth is available, the full stream (all enhancement layers) will be pushed. If not, only a sub-stream (for
- 25 example one or more lower enhancement layers) will be pushed, without affecting the fast tune-in. In the latter case, only the second level of intermediate quality, for example a lower frame rate, has to be accepted.
- 30 The inventive approach of time-shifting the transmission of the respective streams is further advantageous as different I-picture distances in the base and enhancement layer streams also offer a higher error robustness during normal channel reception. The time-shifted distribution allows the
- 35 client to early detect losses and possibly request lost packets from a retransmission server, if available. In case of lost enhancement layer packets, the client can switch back to the base layer. Error distribution in the base

layer is restricted to a few frames, because of the small distances of I-pictures.

5 The above-described solution can also be adapted to temporal scalability, using for example MPEG-4AVC without the MPEG-4SVC extensions. In this case, the base layer corresponds to a low frame rate layer, the lowest hierarchy level, and the enhancement layer to, for example non-reference pictures that are inserted between the base layer
10 frames to reach the full frame rate, the highest hierarchy level. The inserted pictures may be divided into several enhancement layers, hierarchy levels, what is known as "hierarchical B-frames". In addition, the solution can also be adapted to spatial or SNR (signal to noise ratio) scalability.
15 ity.

The above embodiments were described in combination with video data, however, it is noted that the invention is not limited to the transmission of video data, rather the principals described in the embodiments above can be applied to
20 any kind of data which is to be scalable encoded in a data stream, wherein the term scalable encoded data stream means a data stream which is encoded in accordance with the principals of the scalable coding technique, for example in
25 case of a video contents same is encoded in accordance with the principals of the scalable video coding technique (SVC technique).

The embodiments described above may be used in the context
30 of a multi-channel transmission system in which the scalable encoded data stream is associated with a first channel of the multi-channel transmission system, for example an IPTV system. Tuning-in may either comprise switching into a TV channel for the first time or switching from a currently
35 viewed or displayed TV channel to another TV channel.

The invention is not limited to the embodiments described above. Rather, the invention, in general, is concerned with

improving the tune-in characteristics upon tuning into a scalable encoded stream which comprises a base layer stream and at least an enhancement layer stream as described in detail above, wherein the scalable encoded stream may be a
5 single stream which is provided to a user, e.g. over a network, like the Internet.

The stream containing e.g. a video contents may be provided by a service provider such that a user may tune into the
10 stream at any time. In such a situation, after receiving the tune-in request the stream including both the main and the secondary streams is received by the user, and the secondary stream is decoded until the self-contained block arrives on the main stream and the required main stream decoder buffer-fill level is reached.
15

In another embodiment of the invention, the stream is obtained by a user on the user's demand, e.g. from a service provider. The stream (e.g. video on demand) is received by
20 the user and when tuning into the stream decoding of the stream starts in accordance with the principles of embodiments of the invention.

In the description of the embodiments of the invention, the
25 self-contained blocks and the non-self-contained blocks of the streams were named as I-pictures and P- or B-pictures, respectively. It is noted, that the term "picture", in general, determines an encoded contents that includes data or information that is necessary to decode the contents of the
30 block. In case of I-pictures all data or information is included that is necessary to decode the complete contents of the block, whereas in case of P- or B-pictures not all information is included that is necessary to decode a complete picture, rather additional information from preceding
35 or following pictures is required. Alternatively, the I-, P- and B-pictures may be named I-, P- and B-frames.

Further, it is noted that the embodiments were described in combination with video data, however it is noted that the invention is not limited to the transmission of video data, rather the principles described in the embodiments above
5 can be applied to any kind of data which is to be encoded in a data stream. To be more specific, the above described principles also apply to audio data or other kind of timed multimedia data that uses the principle of differential encoding, utilizing the principle of different types of
10 transmitted data fragments within a data stream, like full information (that enables the client to decode the full presentation of the encoded multimedia data) and delta (or update) information (that contains only differential information that the client can only use for a full presentation
15 of the encoded multimedia data if preceding information was received). Examples of such multimedia data, besides video, are graphics data, vector graphics data, 3D graphics data in general, e.g. wireframe and texture data, or 2D or 3D scene representation data.

20

It should be understood that depending on the circumstances, the methods of embodiments of the invention may also be implemented in software. Implementation may occur on a digital storage medium, in particular a disc, a DVD or
25 a CD with electronically readable control signals which can interact with a programmable computer system such that the respective method is executed. Generally, the invention thus also consists in a computer program product with a program code stored on a machine-readable carrier for performing the inventive method, when the computer program
30 product runs on a PC and/or a microcontroller. In other words, the invention may thus be realized as a computer program with a program code for performing the method when, the computer program runs on a computer and/or a microcontroller.
35

It is noted that the above description illustrates the principles of the invention, but it will be appreciated

that those skilled in the art will be able to devise various arrangements that, although not explicitly described or shown here, embody the principles of the invention without departing from the spirit or scope of the invention. It will be appreciated by those skilled in the art that the block diagrams presented herein represent conceptual views of illustrative circuitry embodying the principles of the invention. Similarly, it will be appreciated that any flow-charts, flow-diagrams, transmission diagrams and the like represent various processes which may be substantially presented in a computer readable media and so executed by a computer or processor whether or not such a computer or processor is implicitly shown. The functions of the various elements shown in the figures may be provided through the use of dedicated hardware as well as hardware capable of executing software in association with the appropriate software. When provided by a processor, the functions may be provided by a single dedicated processor, by a single shared processor or by a plurality of individual processors, some of which may be shared. Moreover, explicit use of the term "processor" or "controller" should not be construed to refer exclusively to hardware capable of executing software, it may implicitly include, without limitation, a digital signal processor hardware, read-only memory for storing software, random access memory and non-volatile storage.

What is claimed is:

1. A method for reducing a tune-in delay upon tuning into a scalable encoded data stream which is provided by an encoder and comprises a base layer stream and at least one enhancement layer stream, wherein the base layer stream and the at least one enhancement layer stream are provided separate from the encoder to a decoder, wherein the base layer stream and the at least one enhancement layer stream are combined for decoding such that the combined stream comprises for each data block to be decoded data from the base layer stream and data from the at least one enhancement layer stream, and wherein the method comprises:
transmitting the base layer stream and the at least one enhancement layer stream time-shifted from the encoder to the decoder.
2. The method of claim 1, wherein the base layer stream and the at least one enhancement layer stream are time-shifted such that the base layer stream and the at least one enhancement layer have substantially the same encoder-to-decoder delay.
3. The method of claim 1, wherein time-shifted transmitting the base layer stream and the at least one enhancement layer stream comprises delaying the base layer stream at the encoder such that the base layer stream and the at least one enhancement layer have substantially the same encoder-to-decoder delay.
4. The method of claim 1, further comprising
buffering the base layer stream and the at least one enhancement layer stream prior to decoding, wherein a buffer delay of the base layer stream is shorter than

a buffer delay of the at least one enhancement layer stream,

5 wherein time-sifted transmitting the base layer stream and the at least one enhancement layer stream comprises delaying the base layer stream at the encoder such that the buffer delay of the at least one enhancement layer stream is compensated.

10 5. The method of claim 4, further comprising

buffering the base layer stream and the at least one enhancement layer stream at the decoder,

15 wherein delaying the base layer stream at the encoder comprises setting an encoder buffer delay for the base layer stream to a predefined value or introducing into the base layer stream at the encoder a further delay in addition to an encoder buffer delay for the base
20 layer stream.

6. The method of claim 5, further comprising:

25 in response to a tune-in request, receiving at the decoder both the base layer stream and the at least one enhancement layer stream;

30 in response to the arrival of a self-contained data block in the base layer stream and in response to reaching a required buffer fill level for the at least one enhancement layer stream, decoding the base layer stream only; and

35 in response to the arrival of a self-contained data block in the at least one enhancement layer stream, decoding both the base layer stream and the at least one enhancement layer stream.

7. The method of claim 6, wherein the scalable encoded stream comprises a plurality of enhancement layer streams of different enhancement levels, wherein the base layer stream and the plurality of enhancement layer streams are delayed at the encoder such that the substantially same encoder-to-decoder delay for all streams is obtained, and wherein at least a subset of the enhancement layer streams is successively used for tuning into the scalable encoded stream.
8. The method of claim 6, wherein the scalable encoded stream is associated with a channel of a multi-channel transmission system, and wherein the tune-in request indicates a change from a current channel of the multi-channel transmission system to a new channel of the multi-channel transmission system.
9. The method of claim 6, wherein the scalable encoded stream is associated with a stream which is obtained on demand by a user, and wherein the tune-in request initiates an initial tuning into the stream.
10. The method of claim 1, wherein
- the base layer stream and the at least one enhancement layer stream each comprise a plurality of self-contained blocks including all information for decoding the block, and a plurality of blocks including only partial information for decoding,
- the base layer stream is encoded at a quality different from a quality of the at least one enhancement layer stream, and
- wherein at least some of the blocks of the at least one enhancement layer stream including only partial information for decoding the block depend on informa-

tion from the base layer stream to allow decoding thereof.

11. The method of claim 1, wherein the scalable encoded
5 data stream represents video content, and wherein the
content is encoded using the scalable video coding
scheme yielding the at least one enhancement layer and
the base layer, wherein the I-picture distance in the
10 base layer is shorter than the I-picture distance in
the enhancement layer.
12. The method of claim 11, wherein the scalable video
coding scheme provides for a temporal, spatial or SNR
scalability.
13. The method of claim 1, wherein
15 the scalable encoded stream comprises a plurality of
enhancement layer streams of different enhancement
20 levels, and
wherein the base layer stream and the plurality of en-
hancement layer streams are provided from the encoder
to the decoder with substantially the same encoder-to-
25 decoder delay.
14. A method for providing a scalable encoded data stream
for transmission to a decoder, wherein the scalable
encoded data stream comprises a base layer stream and
30 at least one enhancement layer stream, and wherein the
base layer stream and the at least one enhancement
layer stream are combined for decoding such that the
combined stream comprises for each data block to be
decoded data from the base layer stream and data from
35 the at least one enhancement layer stream, and wherein
the method comprises:

delaying the base layer stream with respect to the at least one enhancement layer stream.

15. The method of claim 14, comprising:

5

scalable encoding an input signal to obtain the base layer stream and the at least one enhancement layer stream; and

10

buffering the base layer stream and the at least one enhancement layer stream,

15

wherein a buffer delay of the base layer stream is selected or a further delay in addition to an encoder buffer delay for the base layer stream is introduced such that the desired delay of the base layer stream with respect to the at least one enhancement layer stream is obtained.

20

16. A system for reducing a tune-in delay upon tuning into a scalable encoded data stream which comprises a base layer stream and at least one enhancement layer stream, the system comprising:

25

an encoder which is configured to provide the base layer stream and the at least one enhancement layer stream of the scalable encoded data stream;

30

a network which is coupled to the encoder and is configured to transmit the base layer and the at least one enhancement layer stream; and

35

a decoder which is coupled to the network and is configured to receive the base layer stream and the at least one enhancement layer stream and to decode a combined stream comprising the base layer stream and the at least one enhancement layer stream, wherein the base layer stream and the at least one enhancement

layer stream are combined for decoding such that the combined stream comprises for each data block to be decoded data from the base layer stream and data from the at least one enhancement layer stream,

5

wherein the system is configured to transmit the base layer stream and the at least one enhancement layer stream with a time-shift from the encoder to the decoder.

10

17. The system of claim 16, wherein

15

the encoder and the decoder each comprise respective buffers for buffering the base layer stream and the at least one enhancement layer stream, respectively, and

the encoder's base layer stream buffer is configured such that a buffer delay of the decoder's enhancement layer stream buffer is compensated.

20

18. The system of claim 17, wherein

25

the decoder comprises a base layer stream buffer, an enhancement layer stream buffer, a combining unit, and a decoding unit,

30

the decoder is configured to tune into the scalable encoded data stream in response to a tune-in request and to receive at the base layer stream buffer the delay base layer stream and at the enhancement layer stream buffer the enhancement layer stream,

35

the combining unit is configured to combine the base layer stream and the enhancement layer stream for decoding such that the combined stream comprises for each data block to be decoded data from the base layer stream and data from the at least one enhancement layer stream, and

the decoding unit is configured to decode the base layer stream only, in response to the arrival of a self-contained data block of the base layer stream in the combined stream, and to decode both the base layer stream and the enhancement layer stream in response to the arrival of a self-contained data block of the enhancement layer stream in the combined stream.

19. An encoder for providing a scalable data stream, comprising:

an encoder section which is configured to scalable encode an input signal to obtain a base layer stream and at least one enhancement layer stream, wherein the base layer stream and the enhancement layer stream are combined for decoding such that the combined stream comprises for each data block to be decoded data from the base layer stream and data from the at least one enhancement layer stream;

a first buffer which is configured to buffer the base layer stream; and

a second buffer which is configured to buffer the at least one enhancement layer stream;

wherein the encoder is configured to transmit the base layer stream and the at least one enhancement layer stream with a time-shift.

20. The encoder of claim 19, wherein the encoder is configured to delay the base layer stream with respect to the at least one enhancement layer stream.

21. The encoder of claim 20, wherein the first buffer is configured to delay the base layer stream so as to provide the base layer stream and the at least one en-

hancement layer stream with substantially the same encoder-to-decoder delay.

- 5 22. The encoder of claim 20, comprising a delay element provided in a base layer stream path of the encoder and being configured to delay the base layer stream so as to provide the base layer stream and the at least one enhancement layer stream with substantially the same encoder-to-decoder delay.
- 10 23. A computer readable medium for storing instructions which, when being executed by a computer, carry out a method for reducing a tune-in delay upon tuning into a scalable encoded data stream which is provided by an encoder and comprises a base layer stream and at least one enhancement layer stream, wherein the base layer stream and the at least one enhancement layer stream are provided separate from the encoder to a decoder, wherein the base layer stream and the at least one enhancement layer stream are combined for decoding such that the combined stream comprises for each data block to be decoded data from the base layer stream and data from the at least one enhancement layer stream, and wherein the method comprises:
- 15 20 25 transmitting the base layer stream and the at least one enhancement layer stream time-shifted from the encoder to the decoder.
- 30 24. A computer readable medium for storing instructions which, when being executed by a computer, carry out a method for providing a scalable encoded data stream for transmission to a decoder, wherein the scalable encoded data stream comprises a base layer stream and at least one enhancement layer stream, and wherein the base layer stream and the at least one enhancement layer stream are combined for decoding such that the combined stream comprises for each data block to be
- 35

decoded data from the base layer stream and data from the at least one enhancement layer stream, and wherein the method comprises:

- 5 delaying the base layer stream with respect to the at least one enhancement layer stream.

1/5

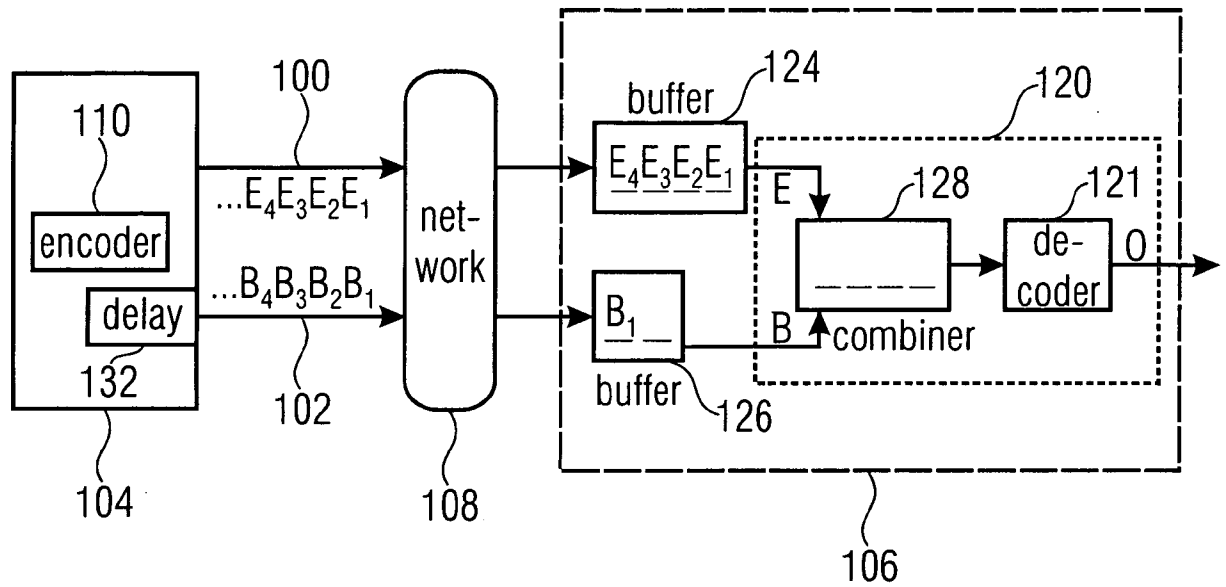


FIGURE 1A

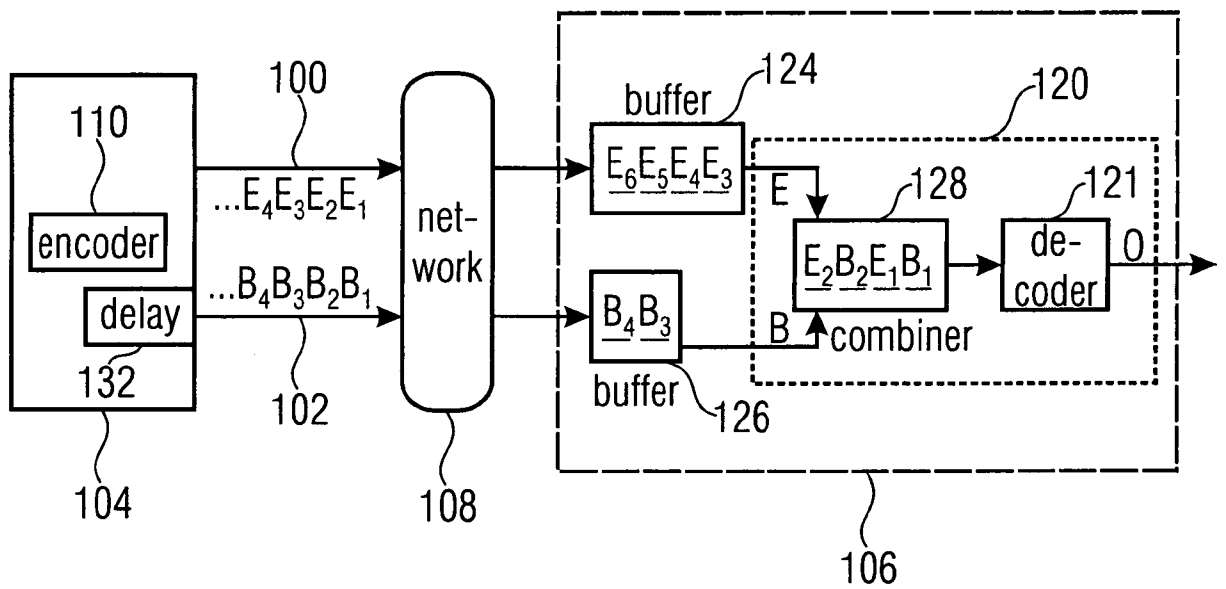


FIGURE 1B

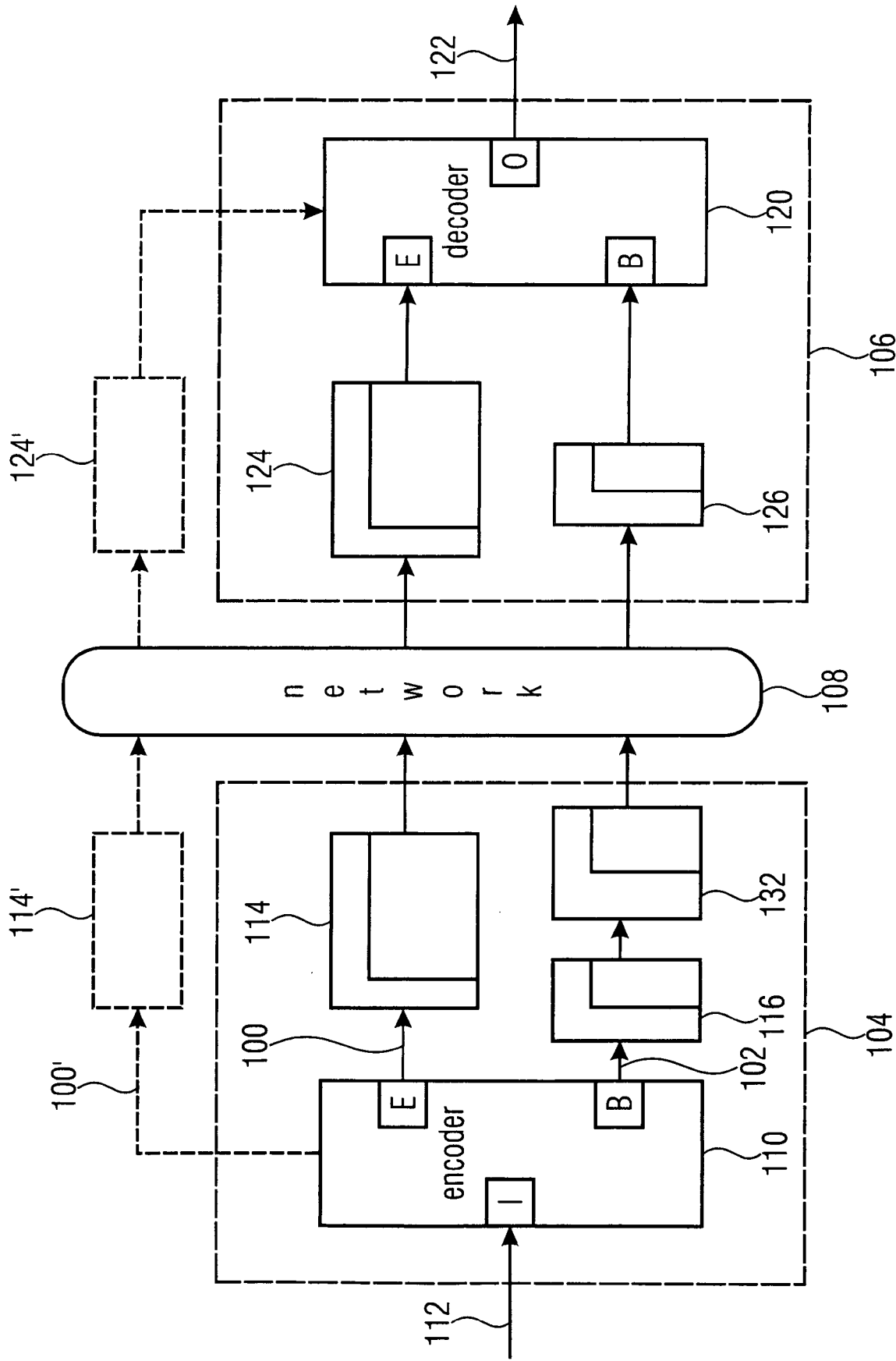


FIGURE 2

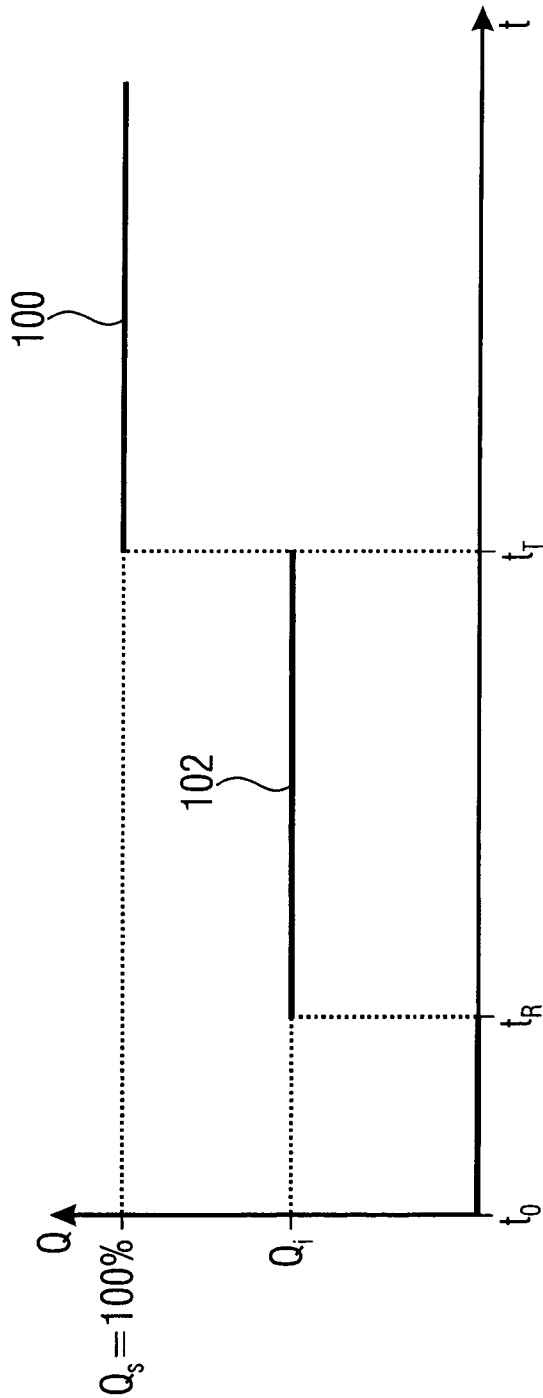


FIGURE 3

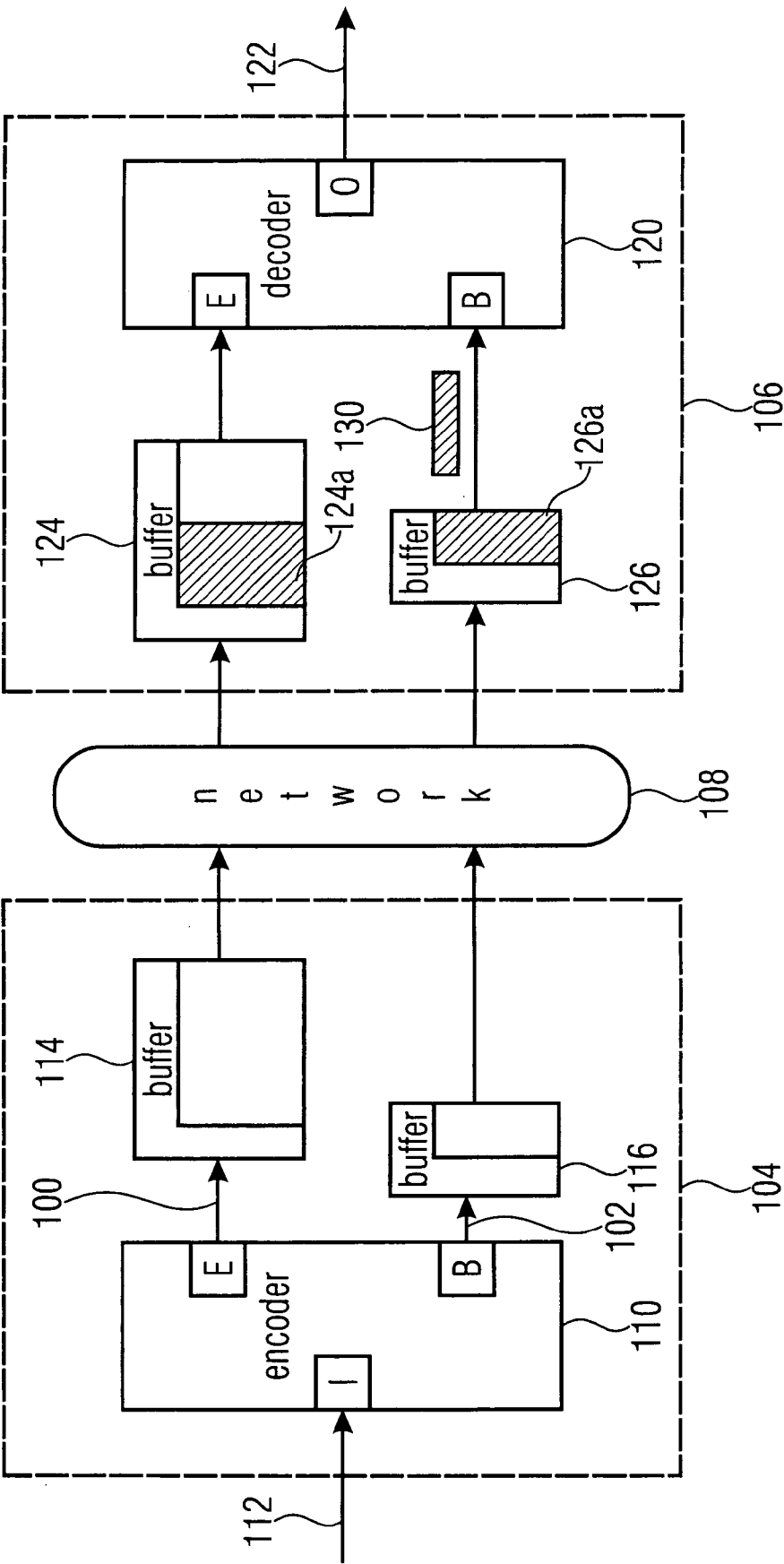


FIGURE 4

5/5

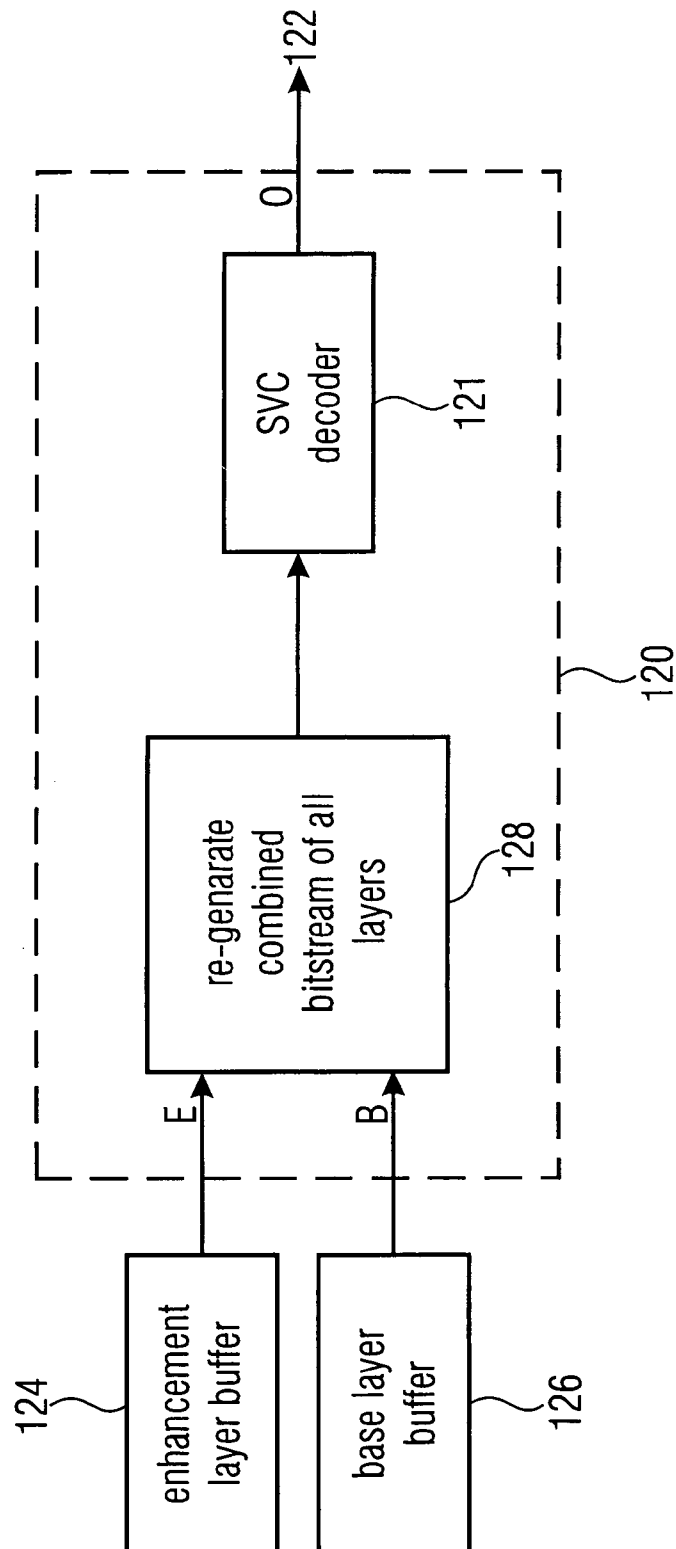


FIGURE 5

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2009/007027

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N7/24 H04N5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
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)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1 043 877 A1 (LUCENT TECHNOLOGIES INC [US]) 11 October 2000 (2000-10-11)	1-5, 10, 13-17, 19-24
Y	abstract paragraph [0010] - paragraph [0011] paragraph [0013] paragraph [0015] - paragraph [0017] paragraph [0019] figure 1 ----- -/--	6-9, 11-12, 18



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"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)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"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

"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

"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.

"&" document member of the same patent family

Date of the actual completion of the international search

5 January 2010

Date of mailing of the international search report

14/01/2010

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Hampson, Frances

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2009/007027

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	FUCHS H ET AL: "Optimizing channel change time in IPTV applications" BROADBAND MULTIMEDIA SYSTEMS AND BROADCASTING, 2008 IEEE INTERNATIONAL SYMPOSIUM ON, IEEE, PISCATAWAY, NJ, USA, 31 March 2008 (2008-03-31), pages 1-8, XP031268571 ISBN: 978-1-4244-1648-6 Section "Scalable coding" -----	6-9, 11-12,18
A	EP 1 926 322 A (HUAWEI TECH CO LTD [CN]) 28 May 2008 (2008-05-28) abstract paragraphs [0005], [0008] -----	1-24

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2009/007027

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1043877	A1	11-10-2000	CA 2300757 A1 08-10-2000
			DE 60000319 D1 19-09-2002
			DE 60000319 T2 13-03-2003
			JP 3705729 B2 12-10-2005
			JP 2000349825 A 15-12-2000
			US 6842724 B1 11-01-2005
EP 1926322	A	28-05-2008	CN 1960453 A 09-05-2007
			US 2008117336 A1 22-05-2008