

(19) **DANMARK**

(10) **DK/EP 3503094 T3**



(12) **Oversættelse af
europæisk patentskrift**

Patent- og
Varemærkestyrelsen

-
- (51) Int.Cl.: **G 10 L 19/005 (2013.01)** **G 10 L 19/008 (2013.01)** **G 10 L 19/02 (2013.01)**
- (45) Oversættelsen bekendtgjort den: **2020-12-21**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2020-11-18**
- (86) Europæisk ansøgning nr.: **18215407.0**
- (86) Europæisk indleveringsdag: **2018-12-21**
- (87) Den europæiske ansøgnings publiceringsdag: **2019-06-26**
- (30) Prioritet: **2017-12-21 WO PCT/CN2017/117802** **2018-01-24 US 201862621176 P**
2018-02-16 EP 18157081
- (84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**
- (73) Patenthaver: **Dolby Laboratories Licensing Corporation, 1275 Market Street, San Francisco, CA 94103, USA**
- (72) Opfinder: **HUANG, Shen, Guandongdian Building 15-02, East 3rd Ring Rd., Beijing, 100022, Kina**
ECKERT, Michael, c/o Dolby Australia Pty Limited, Level 3, 35 Mitchell Street, McMahons Point, New South Wales NSW 2060, Australien
DICKINS, Glenn N., c/o Dolby Australia Pty Limited Level 3, 35, Mitchell Street, McMahons Point, NSW 2060, Australien
- (74) Fuldmægtig i Danmark: **Plougmann Vingtoft A/S, Strandvejen 70, 2900 Hellerup, Danmark**
- (54) Benævnelse: **SELEKTIV FREMADFEJLKORREKTION TIL RUMLIGE LYDCODECS**
- (56) Fremdragne publikationer:
US-A1- 2014 086 416
US-A1- 2016 148 618
E. C. MCCOY AND D. J. FURLONG: "Issues for Consideration in the Development of an Acoustical Design System Supporting Interactive Auralisation", AES, 60 EAST 42ND STREET, ROOM 2520 NEW YORK 10165-2520, USA, 11 November 1996 (1996-11-11), XP040370725,
HARDMAN V ET AL: "Reliable Audio for Use over the Internet", PROCEEDINGS INET, XX, XX, 1 June 1995 (1995-06-01), pages 1-8, XP002203997,

DESCRIPTION

TECHNICAL FIELD

[0001] Embodiments herein relate generally to audio signal processing, and more specifically to reducing complexity and bit rate overhead of forward error correction for multi-channel audio signals when applied after packet loss during transmission over a packet-switched network.

BACKGROUND ART

[0002] US 2016/0148618 A1 relates to concealment of artifacts that result from loss of spatial audio packets during audio transmission over a packet-switched network. US 2014/0086416 A1 discloses scene-based audio coding and decoding, which involves representing the sound field using coefficients of spherical harmonic basis functions. Scene-based audio is typically encoded using an Ambisonics format. Such encoding may include one or more lossy or lossless coding techniques for bandwidth compression, such as quantization (e.g., into one or more codebook indices), error correction coding, redundancy coding, etc. Additionally or alternatively, such encoding may include encoding audio channels (e.g., microphone outputs) into an Ambisonic format, such as B-format, G-format, or Higher-order Ambisonics (HOA). McCoy E C et al, "Issues for Consideration in the Development of an Acoustical Design System Supporting Interactive Auralisation", 101st AES Convention, 8-11 November 1996, page 1-4, XP040370725, inter alia relates to Forward Error Correction (FEC) mechanisms commonly used in used, which provide reliability without increased latency of redundant information along with the original information providing a means to reconstruct the original data in the event of packet loss. Hardmann V et al, "Reliable Audio for Use over the Internet", Proc. INET'95, pages 1-8, June 1995, XP002203997, proposes a method of repairing damaged speech using cheap redundancy within the packets sent from the transmitter. The redundancy is synthetic speech, which, when split into packets, only adds a very small amount of overhead, and therefore does not add to the congestion at the network level. The redundancy for any given packet of speech is piggybacked onto a later packet. This mechanism means that when the receiver suffers the loss of the primary speech information, it still has the possibility of substituting something sensible in the output stream of speech, provided that the redundancy can be received.

SUMMARY OF THE INVENTION

[0003] According to the present invention, there are provided a method for providing forward error correction for a multi-channel audio signal, as set forth in independent claim 1, a corresponding computer program product, as set forth in independent claim 9, another method for providing forward error correction for a multi-channel audio signal, as set forth in independent claim 10, and a system for providing forward error correction for a multi-channel

audio signal, as set forth in independent claim 15. As aforementioned, the invention is set forth in the independent claims. All following occurrences of the word "embodiment(s)", if referring to feature combinations different from those defined by the independent claims, refer to examples which were originally filed but which do not represent embodiments of the presently claimed invention; these examples are still shown for illustrative purposes only.

[0004] Systems and methods for providing forward error correction for a multi-channel audio signal are described. Blocks of a captured audio stream, which may include signals from one or more microphones, are buffered into a frame. For example, a three-microphone capture may result in a sequence of blocks that include three channels of audio, where each channel represents samples from one of the microphones. These blocks of audio are buffered into a frame, typically consisting of twenty milliseconds of audio for many voice applications. A transformation, such as a mixing matrix, can be applied to each block of audio samples that compacts the energy of each block into a plurality of transformed channels. The energy compaction transform may be implemented to compact the most energy of a block into the first transformed channel and to compact decreasing amounts of energy into each subsequent transformed channel. Examples of such a transformation used as a mixing matrix may be a Karhunen Loueve Transform ("KLT") coding or Singular Value Decomposition ("SVD"). Typically, this mixing matrix may be fixed for a frame of audio, but the mixing matrix may vary for each frame of audio in some embodiments. Each channel of audio can then be encoded into a frame using any suitable codec (e.g., EVS, Dolby Voice codec, Opus, etc.). The encoded frame, which may include all channels of encoded audio, may then be transmitted in packets over a network.

[0005] Improved forward error correction may then be provided by attaching a low bit rate encoding of only the first channel (highest energy channel) into a subsequent frame, which is then transmitted in a subsequent packet. To reconstruct a frame of audio corresponding to a lost packet, the low bit rate encoding of the highest energy channel for the lost packet may be combined with a packet loss concealment version of the other channels, constructed from a previously-received packet. The combination of the low bit rate copy of the first channel with the packet loss concealment from the other channels may sufficiently recover a lost frame with reasonably high quality. In some embodiments, spatial parameters for the energy compaction transform applied to each audio block are also included with both the transmitted packets for a frame and the low bit rate encoding of the highest energy channel in the subsequent packet, where these parameters may be used to reconstruct the block corresponding to the lost packet using forward error correction.

[0006] While the above embodiments use the energy compaction transform to aid in providing more bit-efficient forward correction, alternative processing may be used. In some embodiments, the energy compaction transform can be replaced with an alternative mixing matrix, where the mixing matrix converts the captured signals to an Ambisonics format (e.g., a three-channel or four-channel format). The block of audio is buffered into a frame and each channel can be encoded by a relevant codec such as EVS, Dolby Voice Codec, Opus, etc.. In this case, Forward Error Correction may be implemented by encoding a low bit rate version of

the first channel, the W channel for an Ambisonics representation, and including this low bit rate copy in a subsequent packet. The same approach as described above may be used to reconstruct a lost frame of audio, by utilizing the low bit rate copy of the first channel along with packet loss concealment techniques applied to the other channels to reconstruct the audio. A variation to this approach may be implemented by creating Forward Error Correction from more than the first channel, but fewer than the total number of channels. For example, a 3rd order Ambisonics stream, consisting of 16 channels of audio, can have Forward Error Correction applied to only the first four channels, or the first order Ambisonics stream.

[0007] In other embodiments, a transform is used that attempts to assign the energy of audio objects in the scene to individual channels (instead of the energy compaction transform). An example of such a transform is described in U.S. Patent number 9,460,728, entitled "Method and apparatus for encoding multi-channel HOA audio signals for noise reduction, and method and apparatus for decoding multi-channel HOA audio signals for noise reduction" and assigned to Dolby Laboratories Licensing Corporation, Dolby International Ab of the Netherlands. The transform, such as the adaptive Spherical Harmonic Transform described in US 9,460,728, may be applied after converting the received audio frame into a higher-order Ambisonics ("HOA") representation. A subset of channels less than n , including the greatest amount of energy, may then be encoded using a lower bit-rate and transmitted with subsequent packets. The number of channels n being encoded may depend on the desired quality of the reconstruction of the lost packet. The alternative processing methods described herein may produce a similar result as energy compaction approaches.

BRIEF DESCRIPTION OF THE FIGURES

[0008] This disclosure is illustrated by way of example and not limitation in the figures of the accompanying drawings, in which like references indicate similar elements, and in which:

FIG. 1 shows a simplified block diagram of a transmitted encoded audio stream and received audio stream that uses forward error correction, according to an embodiment.

FIG. 2 shows a flow diagram for a method for providing forward error correction for a multi-channel audio signal, according to an embodiment.

FIG. 3 shows a simplified block diagram of a system for providing forward error correction for a three-channel audio signal, according to an embodiment.

FIG. 4 shows a signal flow diagram for providing forward error correction for a three-channel audio signal, according to an embodiment.

FIG. 5 shows a simplified block diagram of a system for providing forward error correction for a four-channel audio signal, according to an embodiment.

FIG. 6 shows a signal flow diagram for providing forward error correction for a four-channel

audio signal, according to an embodiment.

FIG. 7 shows a flow diagram for a method for providing forward error correction for a multi-channel audio signal from a decoder perspective, according to an embodiment.

FIG. 8 shows a signal flow diagram for providing forward error correction for a multi-channel audio signal from a decoder perspective, according to an embodiment.

FIG. 9 is a block diagram of an exemplary system for concealing packet loss using multi-sinusoid detection, according to an embodiment.

DETAILED DESCRIPTION

[0009] In real-time transmission of internet protocol ("IP") packets over a network, degradation of speech quality can be observed when packet loss occurs. Such degradation may be proportional to a packet loss rate and a burst ratio of the data stream. Two well-known techniques of compensating for packet loss are packet loss concealment and low bit-rate redundant forward error correction (see Perkins et al, "RTP Payload for Redundant Audio Data," Request for Comments 2198, September, 1997, hereby incorporated by reference). In low bit-rate redundant forward error correction ("FEC"), a low bit-rate replication of a packet is attached to a later packet, thus allowing partial or full recovery if a packet is lost. The low bit-rate replication may be a low-bitrate version of every channel of the original block of the multi-channel transmitted audio stream, or may be a full bit rate version of every channel. FIG. 1 shows a simplified block diagram of a single-channel transmitted encoded audio stream 105 and received audio stream 110 that uses low bit-rate FEC, according to an embodiment. In the received audio stream 110, packet P3 115 from the transmitted stream 105 is lost. Using low bit-rate FEC, the low bit-rate copy of P3 120, contained in received packet P4 125, may be decoded and used for playback in place of packet P3 115. The recovered packet is generally of lower-quality than the original transmitted packet, and generally extra delay is needed for the decoder to recover the information for the lost packet from the subsequent packet. The temporary reduction in quality, caused by the reduced bit-rate, may often be unnoticeable by end listeners. Generally, FEC techniques provide significantly higher quality of recovery over packet loss concealment techniques, which generally extrapolate a lost packet from previous packets of the received audio stream.

[0010] A simple extension of conventional FEC would be to extend this single channel approach to multiple channels by applying a low bit rate FEC to all channels (see, e.g., Perkins). Applying low bit rate FEC to all channels may increase the bit rate in proportion to the number of channels, and can consume significant bandwidth. Accordingly, a bit rate efficient FEC technique is proposed which incorporates applying FEC to only a select subset of the channels. Specifically, an objective would be to reduce the number of FEC channels required for a multiple channel signal so as to obtain maximum quality with lower increased bit rate.

[0011] Systems and methods for providing FEC for a multi-channel audio signal having improved bit efficiency are described below. In various embodiments, a multi-microphone (e.g., three or four microphone) capture is mixed into a 1st-order Ambisonics representation. The Ambisonics capture may be further processed by applying a decorrelator and energy compaction transform, such as a Karhunen Loeve Transform ("KLT"), Singular Value Decomposition, or Principle Components Analysis. The resulting transformed channels can be encoded independently. The largest non-stationary energy variations in the sound field tend may to be packed into the highest energy components of the capture. Accordingly, significant quality benefits may be seen when these high energy components are reconstructed using a high-quality packet loss technique, such as FEC. By contrast, the lower energy components of the audio stream may tend to capture relatively stationary room ambiance, of which a high-quality reconstruction may be generated using packet loss concealment techniques. While the packet recovery process describe above can also be applied directly to the Ambisonics representation (without the energy compaction step), the quality of recovery from low-order FEC in practice may not be as high as when the additional step of energy compaction is applied.

[0012] FIG. 2 shows a flow diagram for a method 200 for providing forward error correction for a multi-channel audio signal, according to an embodiment. The method 200 will be explained with reference to a three-microphone array, as shown in FIGS. 3 and 4; however, as the description below shows, signals captured with more than three microphones may also benefit from the principles described in method 200. Additionally, the audio channels may not necessarily be constructed from a microphone capture and may represent any multi-channel audio signal. FIG. 3 shows a simplified block diagram of a system 300 for capturing a three-channel audio signal, according to an embodiment. FIG. 4 shows a signal flow diagram 400 for providing forward error correction for a captured three-channel audio signal (such as a signal captured by the system 300 of FIG. 3).

[0013] In method 200, blocks of a captured audio, which include signals from a plurality of microphones, are buffered at step 210 (by, for example, encoder 325) into a frame. System 300 illustrates an exemplary three-channel sound field microphone 315, which includes three microphones that capture sound along only a single horizontal axis. As shown in system 300, the three microphones may be oriented along a forward direction 350 such that one microphone captures a left channel, one microphone captures a right channel, and the third microphone captures a rear channel (also known as the surround channel, abbreviated by "S"). Table A illustrates an exemplary frame having three channels and m samples.

Table A: Exemplary frame having m samples

L1	L2	L3	L4	...	Lm	Channel 1
R1	R2	R3	R4	...	Rm	Channel 2
S1	S2	S3	S4	...	Sm	Channel 3

As shown in Table A, each channel includes time samples, which may be captured by a microphone, or retrieved from storage. The frame represents a time course of samples of length m and includes all channels of the audio stream. A block represents a single time instant

from each channel (e.g., {L1, R1, S1} forms a single block). After the three channels of audio are captured and buffered into a frame, encoder 325 may be used to encode the audio channel for transmission.

[0014] Returning to FIG. 2, a mixing matrix may be applied to each block of the captured audio stream at step 215, where the mixing matrix converts the captured signals to an Ambisonics format. Signal flow diagram 400 illustrates an exemplary embodiment of the processing steps applied by the encoder 325 to the captured audio stream. As is shown in FIG. 4, mixing matrix A1 410 may be applied to each captured (L,R,S) block of audio samples to create a 1st-order Ambisonics stream, represented by (W,X,Y). The term 'channel' is used to refer to the stream of audio samples, so the stream of microphone samples L represents a single channel. Likewise, the stream of audio samples W represents a single channel. Thus, method 200 may involve either a direct map of microphone samples into an energy compacted form or include an intermediate step of transforming into WXY. While there may not be a quality or energy compaction advantage provided by using the intermediate WXY step, the W channel may be easier to use in other processing by the encoder prior to transmission due to WXY being a standard representation that is independent of microphone configuration.

[0015] Returning to FIG. 2, at step 220, a transformation, which may take the form of a mixing matrix, may be applied to each block of audio samples of a frame. The transformation of step 220 compacts the energy of each block so that the first transformed channel contains the highest energy and each subsequent transformed channel has decreasing amounts of energy. An example of such an energy compaction transform is a Karhunen Loeve Transform, a Singular Value Decomposition, or Principle Components Analysis. In the example of FIG. 4, the 1st order Ambisonics representation (410) can be further transformed using an energy compaction mixing matrix (420), represented by (E1,E2,E3,k). The k may represent the parameters for the energy compaction matrix, noting that k typically varies with every frame but stays constant within a frame that includes multiple blocks of audio samples. A parametric representation for the KLT found in U.S. Patent number 9,502,046, entitled "Coding of a Sound Field Signal" and assigned to Dolby Laboratories Licensing Corporation, Dolby International Ab of the Netherlands (and hereinafter referred to as the "'046 Patent").

[0016] An example of the energy compaction transformation is illustrated in block 420 of FIG. 4, which shows a KLT matrix (as described in the '046 Patent). The transformation of block 420 can be applied to the (W,X,Y) signal from the mixing matrix 410 to achieve the encoded signal (E1',E2',E3',k) after encoding is performed at block 430. The parameter k shown during the transformation and encoding steps in FIG. 4 may be a set of spatial parameters that defines the frame-by-frame KLT matrix. As described in the '046 Patent, in an exemplary embodiment the spatial parameters may include decomposition parameters d , ϕ , and θ and determine a rotation used in the transformation.

[0017] At step 230 of method 200, the transformed frame may be encoded and transmitted via packet over a network. The encoding applied at block 430 may be done using any suitable

codec (e.g., EVS, Dolby Voice codec, AC-4, Opus, etc.) by encoding each channel in a frame. While generally encoding is applied to each channel for a frame of samples, in some embodiments the channels may be combined into a single encode. All encoded channels may be combined into a packet for transmission over the network. Packets generally include one frame of encoded data, but may include more than one frame in various embodiments.

[0018] Once the energy compaction transform mixing matrix for the block has been derived, a bit-efficient version of Forward Error Correction is applied by generating a low bit rate encoding of only the first channel from the energy compacted form for each block of the captured audio stream at step 240 of method 200. This low bit rate encoded channel is attached to packets that are subsequent to the transmitted packets for each encoded frame at step 250. In FIG. 4, this step may be observed as applying redundant FEC only to the encoded E1' channel 440 and allowing the E2' and E3' channels 450 (as well as the spatial parameter k) to be recovered using any suitable packet loss concealment technique. In an exemplary embodiment, the E2 and E3 are constructed from a packet received before the lost packet and the FEC parameters are found in a packet received after the lost packet was supposed to arrive. This technique ensures that PLC and k are predicted causally and thereby avoid pre-echo that is audible and annoying. By applying FEC to the lowest order basis function, greater bit efficiency may be obtained for transmission of the audio stream compared to conventional FEC (which is applied to every channel of the captured stream), while playback may still benefit from the improved accuracy provided by FEC (compared to merely using a PLC technique to regenerate a lost packet).

[0019] As stated above, the spatial encoding of the soundfield contains both the compressed media as well as side information describing the frame-by-frame KLT parameters ("k", as shown in FIG. 4). The KLT parameters may be necessary for the decoder to reconstruct an encoded block of the audio stream. The parameters, k, may be included into the FEC packet or predicted from a prior packet. In some embodiments, the k parameter is not included in the FEC packet because k varies slowly from frame to frame and thus prediction from a previous packet provides high-quality performance.

[0020] An extension of the embodiment described in FIGS. 3-4 is to implement method with a soundfield microphone with more than three microphones. FIG. 5 shows a simplified block diagram of a system 500 for providing forward error correction for a four-channel audio signal, according to an embodiment. FIG. 6 shows a signal flow diagram 600 for providing forward error correction for a captured four-channel audio signal (such as a signal captured by the system 500 of FIG. 5). While four channels are shown in FIGS. 5-6, the invention is not limited in this regard, as method 200 may be applied to any soundfield microphone having n microphones (where n is greater than 3), albeit with a greater degree of calculation complexity with respect to applying the mixing matrix and the transform. The four-channel soundfield microphone shown in FIG. 5 includes four microphones 505, 510, 520, and 525 for capturing sound channels Lf, Lb, Rf, Rb (corresponding to left-front, left-back, right-front, and right-back channels, respectively).

[0021] As previously described, a mixing matrix A1 610 may convert the captured signals to an Ambisonics format, creating the sound field (W,X,Y,Z). The transformation of block 620 can be applied to the (W,X,Y,Z) signal from the mixing matrix 610 to create the four basis functions (E1',E2',E3',E4') after encoding is performed at block 630. Once the basis functions for the block have been derived, redundant FEC may then be applied only to the encoded E1' channel 640 and allowing the E2', E3', and E4' channels 650 (as well as the spatial parameter k) to be recovered using any suitable packet loss concealment technique.

[0022] A further extension of this concept is a higher order Ambisonics capture, e.g., from an eigenmicrophone, in which case there are more basis functions and the application of FEC can be truncated at the E1' or extended to cover as many basis functions as needed to appropriately recover the signal. Applying FEC only to the 1st order Ambisonics representation (E1,E2,E3,E4) and packet loss concealment for higher order Ambisonics (E4...En) is often sufficient for any order Ambisonics representation when there is low levels of packet loss. This approach can also work for parametric spatial encoded stream, where the first channel, E1, is encoded as described above, and higher order channels, E2 and E3, are encoded parametrically. In this case, the same approach is used - apply FEC to the highest energy channel and use the appropriate packet loss concealment technique for other channels.

[0023] In addition to the foregoing, transforms may be used in place of the energy compaction transforms described above that assign the energy of an audio object or multiple objects within a captured audio stream to individual channels. The transform, such as the adaptive Spherical Harmonic Transform described in US 9,460,728, may be applied after converting the received audio frame into a higher-order Ambisonics ("HOA") representation. The adaptive Spherical Harmonic Transform may apply a rotation of the HOA representation of each frame to endeavor to focus basis functions such that individual audio objects correspond to individual channels that have greater energy than the other channels. This subset of channels, less than the total number of HOA channels, including the greatest amount of energy may then be encoded using a lower bit-rate and transmitted with subsequent packets. The number of channels n being encoded may depend on the desired quality of the reconstruction of the lost packet.

[0024] FEC is applied by the decoder after receiving packets of the audio stream over a network connection. FIG. 7 shows a flow diagram for a method 700 for providing forward error correction for a multi-channel audio signal from a decoder perspective, according to an embodiment. FIG. 8 shows a signal flow diagram 800 for providing forward error correction for a multi-channel audio signal from the decoder perspective, illustrating an embodiment of method 700 being applied to received packets.

[0025] At step 710, packets of a captured audio stream comprising signals from a plurality of microphones (e.g., over a network connection). As shown in diagram 800, The packets P1, P2, P4 may arrive in a jitter buffer at the receive end of a decoder. Each packet P1, P2, and P4 805 may include a plurality of channels for a block of the captured audio stream and a low bit rate-encoded form of a high energy channel of a past block of the captured audio stream. P1, being

the first packet, would not include any FEC data from a past frame; however, for example packet P4 includes the high energy channel of past block P3 of the captured audio stream. While the past block P3 is the block of the captured audio stream immediately preceding received packet P4, the invention is not limited in this regard. For example, due to latency, Forward Error Correction may be used on frames two, three, or any plurality of frames after the frame corresponding to the lost packet in some embodiments.

[0026] The received packets are decoded at step 720, wherein the decoded channels for the block of the captured audio stream are used to play back the block of the captured audio stream. This may be observed in diagram 800 at block 835, where the received packet is decoded to generate basis functions (E1,E2,E3) and spatial parameters k for the block corresponding to packet P2. Accordingly, in diagram 800, P1 is decoded into WXY and P2 is decoded into WXY. However, packet P3 is not available for playback.

[0027] Returning to FIG. 7, in response to a determination that the past block of the captured audio stream has been lost at step 730, steps 740, 750, and 770 may be applied. At step 740, the low bit rate-encoded first transformed channel of the past block, located in a subsequent packet, may be decoded. Using packet loss concealment, the other transformed channels of the past block may be extrapolated at step 750. And finally, based on the decoded low bit rate-encoded first transformed channel and the extrapolated other channels, the lost past block may be reconstructed for playback at step 770. These three steps are displayed in diagram 800, where, when P4 arrives in the jitter buffer, a reconstruction of P3 is made by decoding the redundant packet, p3 (which includes the highest energy channel (i.e. the first transformed channel) of the frame corresponding to lost packet P3), located in packet P4 at block 810. An estimate of channels E2, E3 for reconstructed packet P3 may be obtained by packet loss concealment extrapolation of E2 and E3 from past packet P2 at block 815. In embodiments where energy compaction is used, spatial parameters k are used to reconstruct the lost packet P3. In some embodiments, the spatial parameters for a packet are included in the same packet as the low bit rate-encoded first transformed channel. In other embodiments, as shown in diagram 800, spatial parameters for the lost past frame are copied from a most recent available frame of the encoded audio stream and are used to reconstruct the lost past frame of the captured audio stream. The most recent available frame, meaning the most proximate frame in the received encoded audio stream to the lost frame, may be a past frame, as is shown in diagram 800, where the spatial parameters k from P2 are used (as described above) to reconstruct the lost packet P3. In other embodiments, the most recent available frame may be a subsequent frame to the lost packet (e.g., copied from packet P4). Using the decoded high energy channel E1 (from redundant packet p3) and the extrapolated E2 and E3 basis functions, an inverse of the transform applied to generate the basis functions may be applied by block 825 to generate an Ambisonics representation WXY 830 of the block of the audio stream corresponding to P3. The Ambisonics representation 830 is then rendered for playout on headphones or speakers. Any suitable mixing matrices for rendering a WXY representation for audio to various speaker configurations may be utilized to render the Ambisonics representation 830.

[0028] The low bit rate FEC payload may often be encoded at reduced bandwidth. For example, the original signal may be encoded with a bandwidth of 32 Khz but the FEC payload for the redundant packet may be limited to 8 Khz bandwidth for bit efficiency purposes. This means that the reconstructed packet may have lower bandwidth than the surrounding packets. For isolated packet loss, this is usually not observable, but it may become noticeable as the packet loss rate increases. The perception of reduced bandwidth and discontinuities can be avoided if, as is done in various embodiments, a single ended blind Spectral Band Replication of the signal is performed as part of the FEC reconstruction. An example of this may be seen in U.S. Patent Number 9,653,085, entitled "Reconstructing an Audio Signal Having A Baseband And High Frequency Components Above The Baseband" and assigned to Dolby Laboratories Licensing Corporation, Dolby International Ab of the Netherlands.

[0029] The methods and modules described above may be implemented using hardware or software running on a computing system. FIG. 9 is a block diagram of an exemplary computing system for concealing packet loss using multi-sinusoid detection according to various embodiments of the present invention. With reference to FIG. 9, an exemplary system for implementing the subject matter disclosed herein, including the methods described above, includes a hardware device 900, including a processing unit 902, memory 904, storage 906, data entry module 908, display adapter 910, communication interface 912, and a bus 914 that couples elements 904-912 to the processing unit 902.

[0030] The bus 914 may comprise any type of bus architecture. Examples include a memory bus, a peripheral bus, a local bus, etc. The processing unit 902 is an instruction execution machine, apparatus, or device and may comprise a microprocessor, a digital signal processor, a graphics processing unit, an application specific integrated circuit (ASIC), a field programmable gate array (FPGA), etc. The processing unit 902 may be configured to execute program instructions stored in memory 904 and/or storage 906 and/or received via data entry module 908.

[0031] The memory 904 may include read only memory (ROM) 916 and random access memory (RAM) 918. Memory 904 may be configured to store program instructions and data during operation of device 900. In various embodiments, memory 904 may include any of a variety of memory technologies such as static random access memory (SRAM) or dynamic RAM (DRAM), including variants such as dual data rate synchronous DRAM (DDR SDRAM), error correcting code synchronous DRAM (ECC SDRAM), or RAMBUS DRAM (RDRAM), for example. Memory 904 may also include nonvolatile memory technologies such as nonvolatile flash RAM (NVRAM) or ROM. In some embodiments, it is contemplated that memory 904 may include a combination of technologies such as the foregoing, as well as other technologies not specifically mentioned. When the subject matter is implemented in a computer system, a basic input/output system (BIOS) 920, containing the basic routines that help to transfer information between elements within the computer system, such as during start-up, is stored in ROM 916.

[0032] The storage 906 may include a flash memory data storage device for reading from and writing to flash memory, a hard disk drive for reading from and writing to a hard disk, a

magnetic disk drive for reading from or writing to a removable magnetic disk, and/or an optical disk drive for reading from or writing to a removable optical disk such as a CD ROM, DVD or other optical media. The drives and their associated computer-readable media provide nonvolatile storage of computer readable instructions, data structures, program modules and other data for the hardware device 900.

[0033] It is noted that the methods described herein can be embodied in executable instructions stored in a non-transitory computer readable medium for use by or in connection with an instruction execution machine, apparatus, or device, such as a computer-based or processor-containing machine, apparatus, or device. It will be appreciated by those skilled in the art that for some embodiments, other types of computer readable media may be used which can store data that is accessible by a computer, such as magnetic cassettes, flash memory cards, digital video disks, Bernoulli cartridges, RAM, ROM, and the like may also be used in the exemplary operating environment. As used here, a "computer-readable medium" can include one or more of any suitable media for storing the executable instructions of a computer program in one or more of an electronic, magnetic, optical, and electromagnetic format, such that the instruction execution machine, system, apparatus, or device can read (or fetch) the instructions from the computer readable medium and execute the instructions for carrying out the described methods. A non-exhaustive list of conventional exemplary computer readable medium includes: a portable computer diskette; a RAM; a ROM; an erasable programmable read only memory (EPROM or flash memory); optical storage devices, including a portable compact disc (CD), a portable digital video disc (DVD), a high definition DVD (HD-DVD™), a BLU-RAY disc; and the like.

[0034] A number of program modules may be stored on the storage 906, ROM 916 or RAM 918, including an operating system 922, one or more applications programs 924, program data 926, and other program modules 928. A user may enter commands and information into the hardware device 900 through data entry module 908. Data entry module 908 may include mechanisms such as a keyboard, a touch screen, a pointing device, etc. Other external input devices (not shown) are connected to the hardware device 900 via external data entry interface 930. By way of example and not limitation, external input devices may include a microphone, joystick, game pad, satellite dish, scanner, or the like. In some embodiments, external input devices may include video or audio input devices such as a video camera, a still camera, etc. Data entry module 908 may be configured to receive input from one or more users of device 900 and to deliver such input to processing unit 902 and/or memory 904 via bus 914.

[0035] The hardware device 900 may operate in a networked environment using logical connections to one or more remote nodes (not shown) via communication interface 912. The remote node may be another computer, a server, a router, a peer device or other common network node, and typically includes many or all of the elements described above relative to the hardware device 900. The communication interface 912 may interface with a wireless network and/or a wired network. Examples of wireless networks include, for example, a BLUETOOTH network, a wireless personal area network, a wireless 802.11 local area network

(LAN), and/or wireless telephony network (e.g., a cellular, PCS, or GSM network). Examples of wired networks include, for example, a LAN, a fiber optic network, a wired personal area network, a telephony network, and/or a wide area network (WAN). Such networking environments are commonplace in intranets, the Internet, offices, enterprise-wide computer networks and the like. In some embodiments, communication interface 912 may include logic configured to support direct memory access (DMA) transfers between memory 904 and other devices.

[0036] In a networked environment, program modules depicted relative to the hardware device 900, or portions thereof, may be stored in a remote storage device, such as, for example, on a server. It will be appreciated that other hardware and/or software to establish a communications link between the hardware device 900 and other devices may be used.

[0037] It should be understood that the arrangement of hardware device 900 illustrated in FIG. 9 is but one possible implementation and that other arrangements are possible. It should also be understood that the various system components (and means) defined by the claims, described above, and illustrated in the various block diagrams represent logical components that are configured to perform the functionality described herein. For example, one or more of these system components (and means) can be realized, in whole or in part, by at least some of the components illustrated in the arrangement of hardware device 900. In addition, while at least one of these components are implemented at least partially as an electronic hardware component, and therefore constitutes a machine, the other components may be implemented in software, hardware, or a combination of software and hardware. More particularly, at least one component defined by the claims is implemented at least partially as an electronic hardware component, such as an instruction execution machine (e.g., a processor-based or processor-containing machine) and/or as specialized circuits or circuitry (e.g., discrete logic gates interconnected to perform a specialized function), such as those illustrated in FIG. 9. Other components may be implemented in software, hardware, or a combination of software and hardware. Moreover, some or all of these other components may be combined, some may be omitted altogether, and additional components can be added while still achieving the functionality described herein. Thus, the subject matter described herein can be embodied in many different variations.

[0038] In the description above, the subject matter may be described with reference to acts and symbolic representations of operations that are performed by one or more devices, unless indicated otherwise. As such, it will be understood that such acts and operations, which are at times referred to as being computer-executed, include the manipulation by the processing unit of data in a structured form. This manipulation transforms the data or maintains it at locations in the memory system of the computer, which reconfigures or otherwise alters the operation of the device in a manner well understood by those skilled in the art. The data structures where data is maintained are physical locations of the memory that have particular properties defined by the format of the data. However, while the subject matter is being described in the foregoing context, it is not meant to be limiting as those of skill in the art will appreciate that various of the acts and operation described hereinafter may also be implemented in hardware.

[0039] For purposes of the present description, the terms "component," "module," and "process," may be used interchangeably to refer to a processing unit that performs a particular function and that may be implemented through computer program code (software), digital or analog circuitry, computer firmware, or any combination thereof.

[0040] It should be noted that the various functions disclosed herein may be described using any number of combinations of hardware, firmware, and/or as data and/or instructions embodied in various machine-readable or computer-readable media, in terms of their behavioral, register transfer, logic component, and/or other characteristics. Computer-readable media in which such formatted data and/or instructions may be embodied include, but are not limited to, physical (non-transitory), non-volatile storage media in various forms, such as optical, magnetic or semiconductor storage media.

[0041] Unless the context clearly requires otherwise, throughout the description and the claims, the words "comprise," "comprising," and the like are to be construed in an inclusive sense as opposed to an exclusive or exhaustive sense; that is to say, in a sense of "including, but not limited to." Words using the singular or plural number also include the plural or singular number respectively. Additionally, the words "herein," "hereunder," "above," "below," and words of similar import refer to this application as a whole and not to any particular portions of this application. When the word "or" is used in reference to a list of two or more items, that word covers all of the following interpretations of the word: any of the items in the list, all of the items in the list and any combination of the items in the list.

[0042] In the description above and throughout, numerous specific details are set forth in order to provide a thorough understanding of the disclosure. It will be evident, however, to one of ordinary skill in the art, that the disclosure may be practiced without these specific details. In other instances, well-known structures and devices are shown in block diagram form to facilitate explanation. The description of the preferred an embodiment is not intended to limit the scope of the claims appended hereto. Further, in the methods disclosed herein, various steps are disclosed illustrating some of the functions of the disclosure. One will appreciate that these steps are merely exemplary and are not meant to be limiting in any way. Other steps and functions may be contemplated without departing from this disclosure.

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US20160148618A1 [0002]
- US20140086416A1 [0002]
- US9460728B [0007] [0007] [0023]
- US9502046B [0015]
- US9653085B [0028]

Non-patent literature cited in the description

- **MCCOY E C et al.**Issues for Consideration in the Development of an Acoustical Design System Supporting Interactive Auralisation101st AES Convention, 1996, 1-4 [0002]
- **HARDMANN V et al.**Reliable Audio for Use over the InternetProc. INET'95, 1995, 1-8 [0002]
- **PERKINS et al.**RTP Payload for Redundant Audio DataRequest for Comments 2198, 1997, [0009]

Patentkrav

1. Fremgangsmåde til tilvejebringelse af fremadfejlkorrektion for et flerkanalslydsignal, hvilken fremgangsmåde omfatter:

5 at buffe blokke af en lydstrøm i en lydramme, hvilken lydstrøm omfatter en flerhed af lydkanaler;

10 at påføre en transformation til hver blok af lydrammen, hvilken transformation komprimerer energien af hver blok til en flerhed af transformerede kanaler, hvor den første transformerede kanal for hver blok indeholder mest energi og efterfølgende transformerede kanaler indeholder

15 at kode den transformerede ramme;

at transmittere, over et netværk, den kodede ramme i en pakke;

at kode den første transformerede kanal af den transformerede ramme ved en lavere bitrate end kodningen anvendt til den transmitterede pakke; og

20 at transmittere, over netværket, den lavere bitrate-kodede kanal i en pakke der efterfølger den transmitterede pakke.

2. Fremgangsmåden ifølge krav 1, yderligere omfattende at påføre en blandingsmatrix til hver blok af den fangede lydstrøm forud for påføring af
25 transformationen, blandingsmatrixen konverterer de fangede signaler til et Ambisonics-format.

3. Fremgangsmåden ifølge krav 1 eller krav 2, hvor transformationen er en af en Karhunen Loeve-transformation, en enkelt værdi-nedbrydning (Singular Value
25 Decomposition), Principle Component-analyse, eller en adaptiv harmonisk sfærisk transformation.

4. Fremgangsmåden ifølge et hvilket som helst af kravene 1-3, hvor den efterfølgende pakke også inkluderer rumparametre for den transformerede ramme
30 der parameteriserer transformationen, hvor, når den lavere bitrate-kodede kanal anvendes til at rekonstruere en tabt pakke, de inkluderede rumparametre anvendes til at rekonstruere den tabte pakke.

5. Fremgangsmåden ifølge et hvilket som helst af kravene 1-3, hvor, når den lavere bitrate-kodede kanal anvendes til at rekonstruere den tabte pakke, den efterfølgende pakke inkluderer ikke rumparametre for den transformerede ramme, og et rumparameter af en efterfølgende transformeret ramme af lyd 5 inkluderet i den efterfølgende pakke anvendes i kombination med den lavere bitrate-kodede kanal til at rekonstruere den tabte pakke.

6. Computerprogramprodukt omfattende computerlæsbar programkode til at blive eksekveret af en eller flere processorer når hentet fra et ikke-midlertidigt 10 computerlæsbart medium, hvilken programkode inkluderer instruktioner der, når eksekveret, forårsager den ene eller flere processorer at udføre handlinger ifølge et hvilket som helst af kravene 1-5.

7. Fremgangsmåde til tilvejebringelse af fremadfejlkorrektion for et 15 flerkanalslydsignal, hvilken fremgangsmåde omfatter:

at modtage en pakke af en kodet lydstrøm omfattende en flerhed af transformerede lydkanaler, hvilken pakke omfatter en kodet ramme og en lavere bitrate-kodet transformeret kanal af en tidligere ramme af den kodede lydstrøm, hvor den lavere bitrate-kodede kanal kodes ved en lavere 20 bitrate end en kodning anvendt for den kodede ramme;

at afkode den kodede ramme, hvor flerheden af transformerede lydkanaler anvendes til playback; og

i reaktion på en bestemmelse at en pakke svarende til den tidligere ramme af den kodede lydstrøm er blevet tabt:

25 at afkode den lavere bitrate-kodede transformerede kanal af den tidligere ramme lokaliseret i den modtagne pakke;

at anvende pakketabsskjul til at ekstrapolere andre transformerede kanaler af den tidligere ramme; og

30 baseret på den lavere bitrate-kodede transformerede kanal og de ekstrapolerede transformerede kanaler, at rekonstruere den tabte tidligere ramme til playback.

- 8.** Fremgangsmåden ifølge krav 7, hvor den tabte tidligere ramme lokaliseres en flerhed af rammer forud for den kodede ramme i den kodede lydstrøm.
- 9.** Fremgangsmåden ifølge krav 7 eller krav 8, hvor den modtagne pakke
5 yderligere omfatter rumparametre for den tabte tidligere ramme, hvor rumparametrene anvendes til at rekonstruere den tabte tidligere ramme af den fangede lydstrøm.
- 10.** Fremgangsmåden ifølge et hvilket som helst af kravene 7-9, yderligere
10 omfattende, i reaktion på bestemmelsen at den tidligere ramme af den kodede lydstrøm er blevet tabt, at påføre enkelt-endet blind spektrumbåndreplikation som en del af rekonstruktionen af den tabte tidligere ramme til playback.
- 11.** Fremgangsmåden ifølge et hvilket som helst af kravene 7-10, hvor
15 rumparametre for den tabte tidligere ramme kopieres fra en nyeste tilgængelig ramme af den kodede lydstrøm og anvendes til at rekonstruere den tabte tidligere ramme af den fangede lydstrøm.
- 12.** System til tilvejebringelse af fremadfejlkorrektion for et flerkannelslydsignal,
20 hvilket system omfatter:
- en eller flere processorer; og
 - et ikke-midlertidigt computerlæsbart medium der lagrer instruktioner der, når eksekveret af den ene eller flere processorer, forårsager den ene eller flere processorer at udføre handlinger omfattende:
25
 - at buffe blokke af en lydstrøm i en lydramme hvilken lydstrøm omfatter en flerhed af lydkanaler;
 - at påføre en transformation til hver blok af lydrammen, hvor transformationen konverterer hver blok til et Ambisonics-format omfattende en flerhed af kanaler og inkluderer en W-kanal, hvor W-
30 kanalen er den første af flerheden af kanaler;
 - at kode den transformerede ramme;
 - at transmittere, over et netværk, den kodede ramme i en pakke;

at kode W-kanalen af den transformerede ramme ved en lavere
bitrate end

kodningen anvendt til den transmitterede pakke; og

at transmittere, over netværket, den lavere bitrate-kodede kanal i en

5 pakke der efterfølger den transmitterede pakke, hvor:

Ambisonics-formatet omfatter tre kanaler, fire kanaler, eller et tredje orden

Ambisonics-format omfattende 16 kanaler.

DRAWINGS

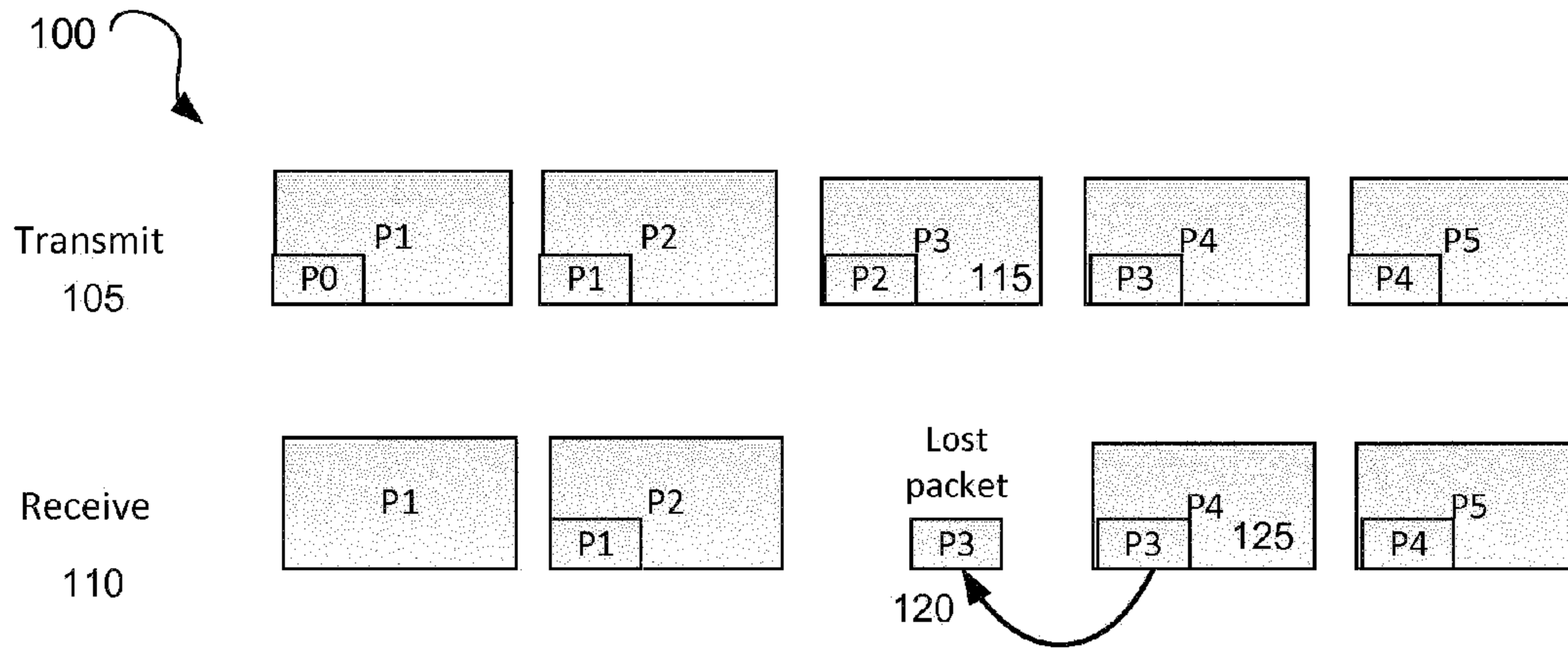
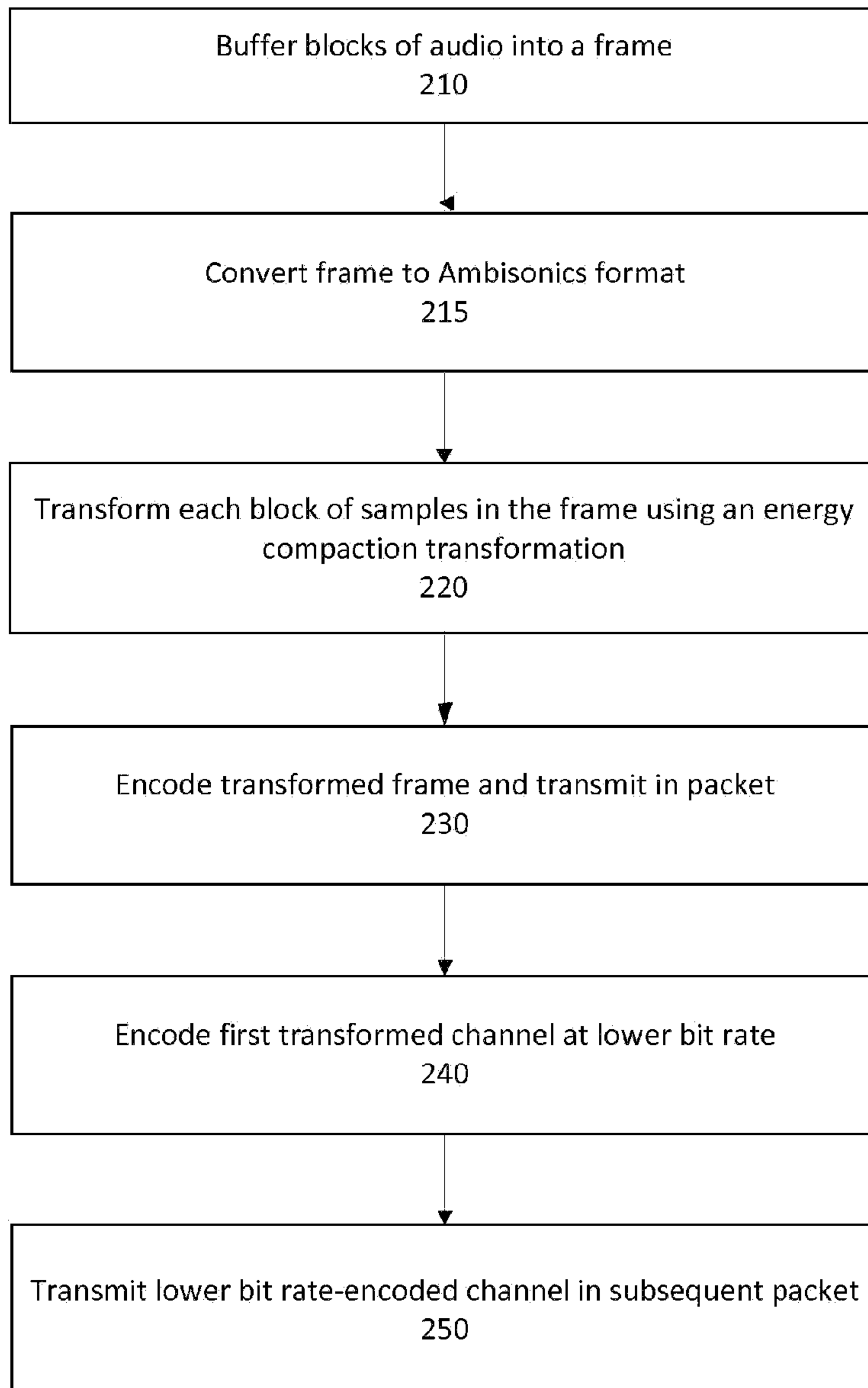


FIG. 1

200**FIG. 2**

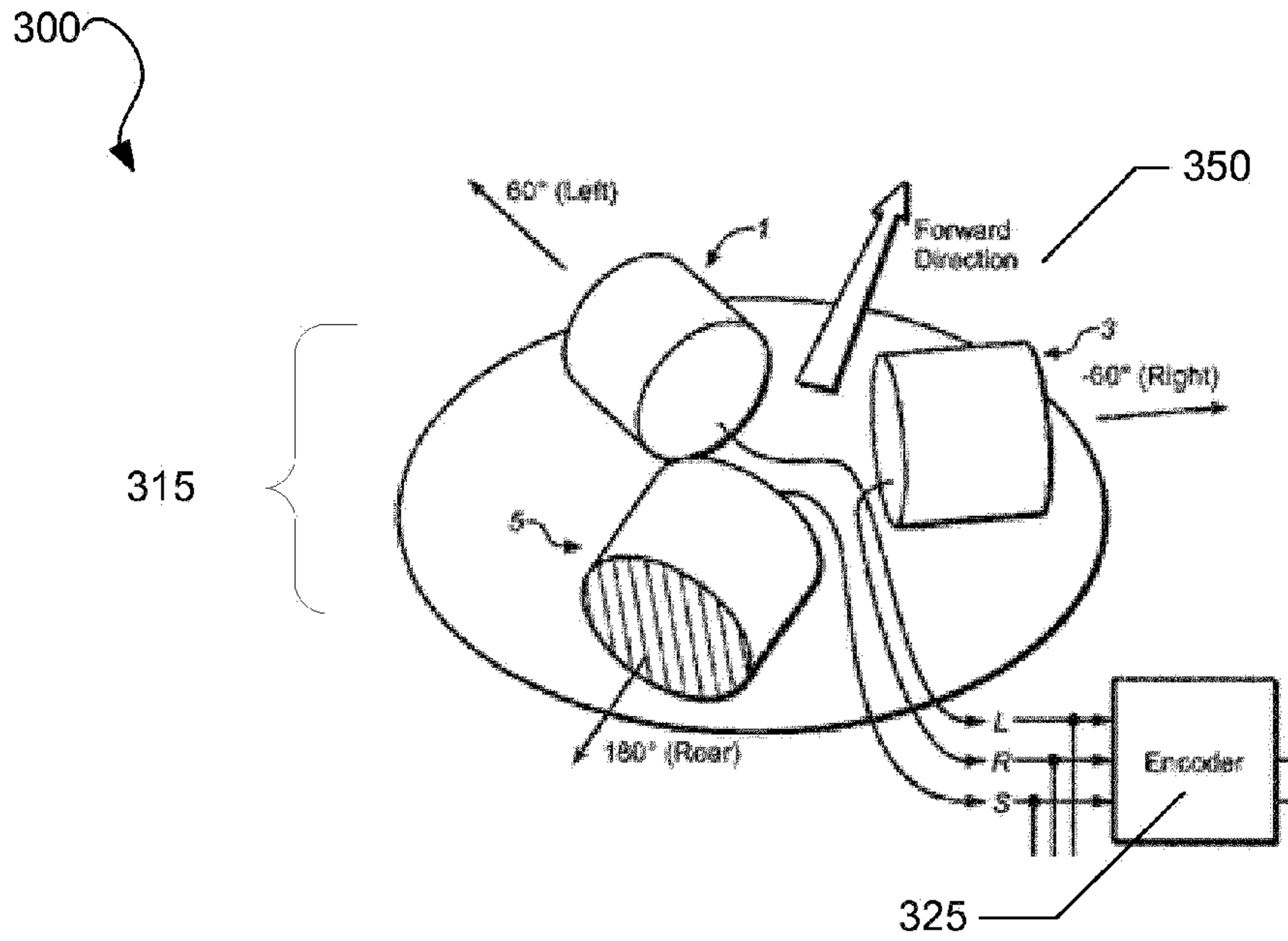


FIG. 3

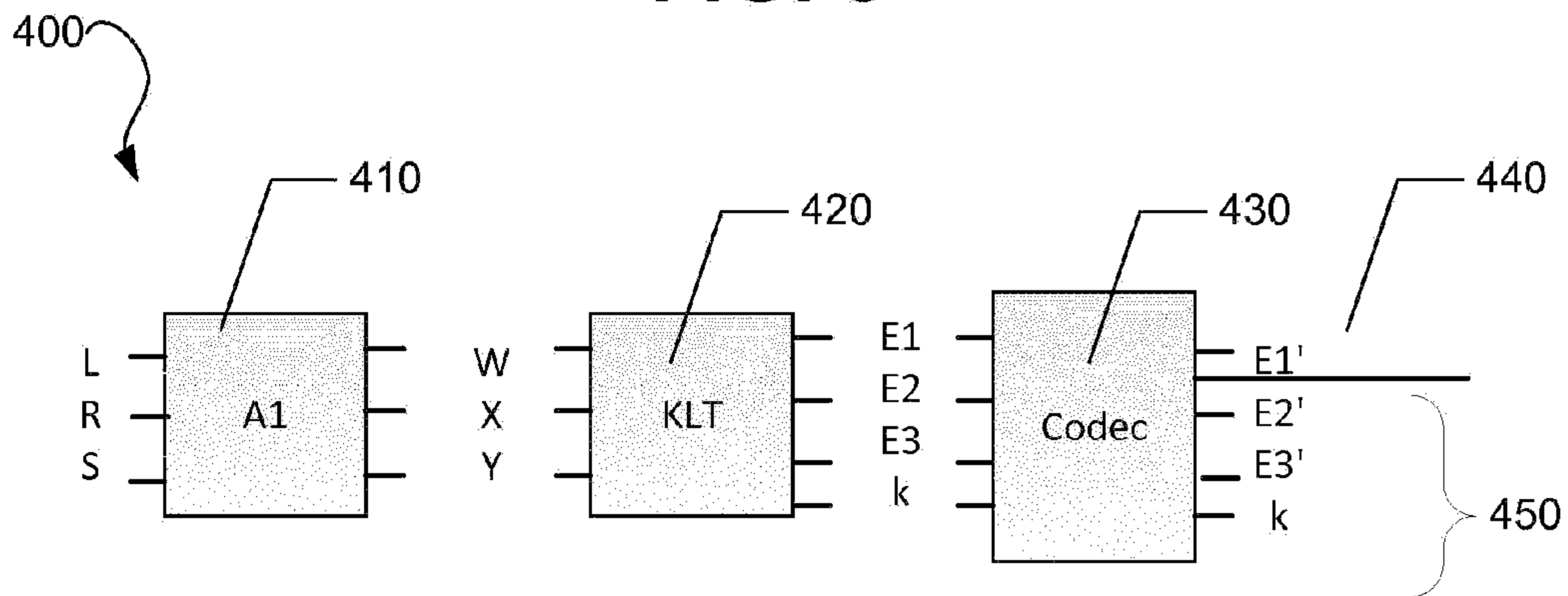


FIG. 4

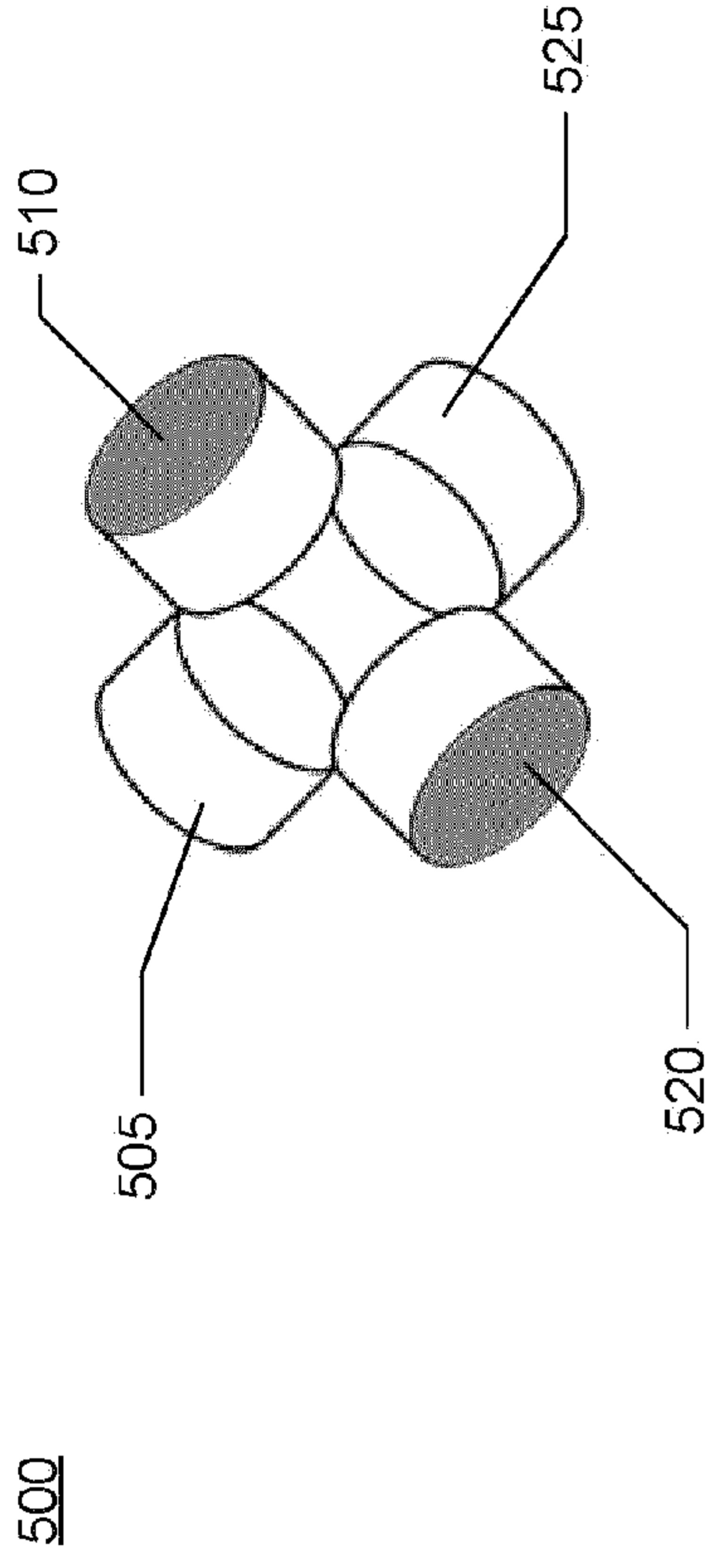


FIG. 5

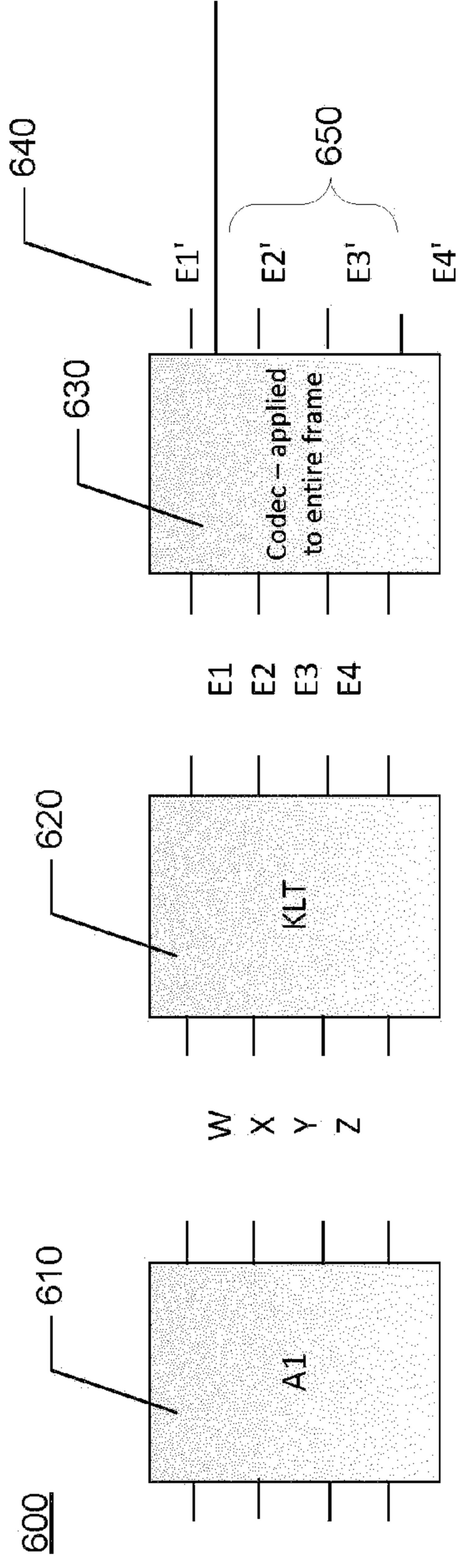
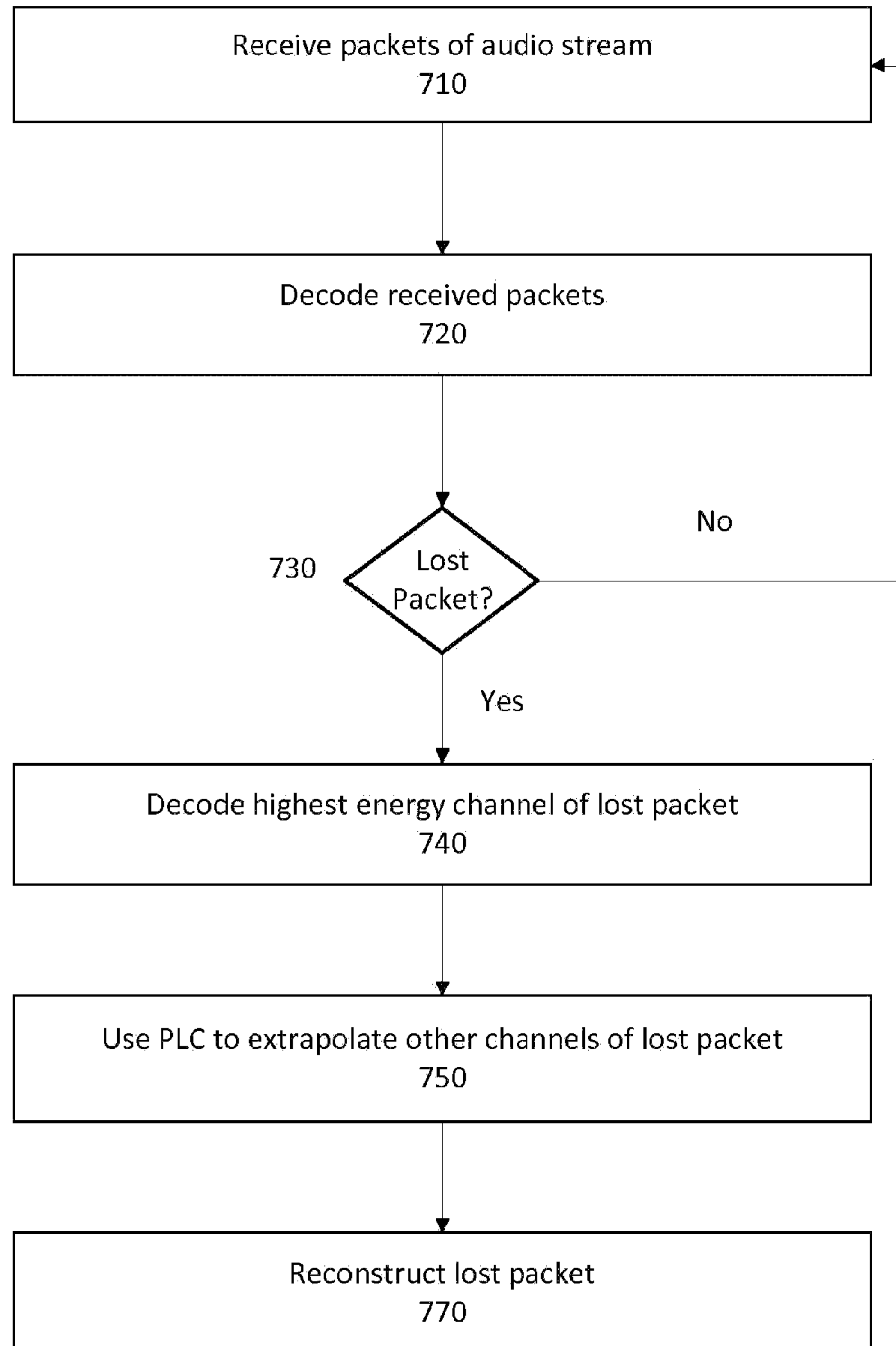


FIG. 6

700

**FIG. 7**

800

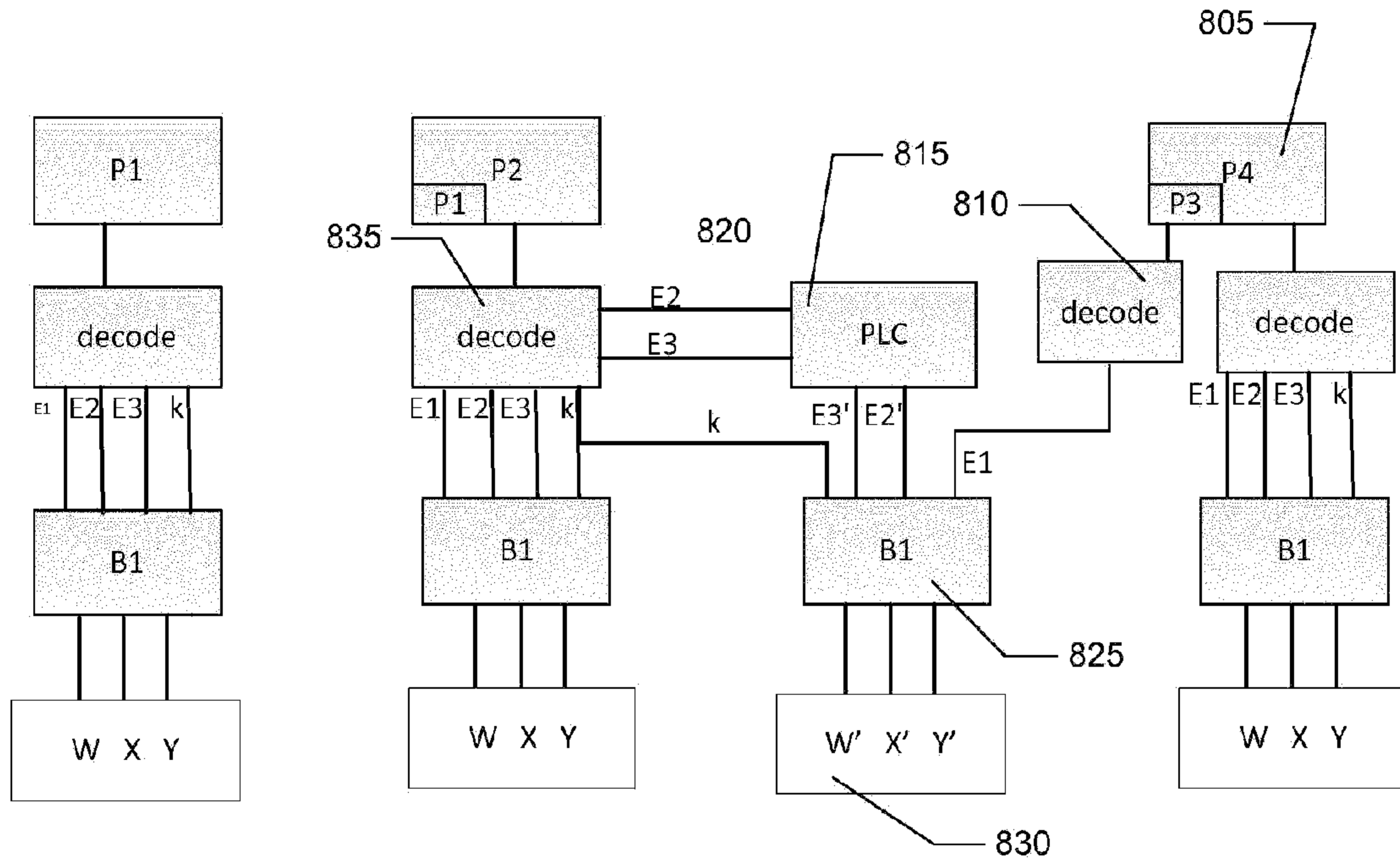


FIG. 8

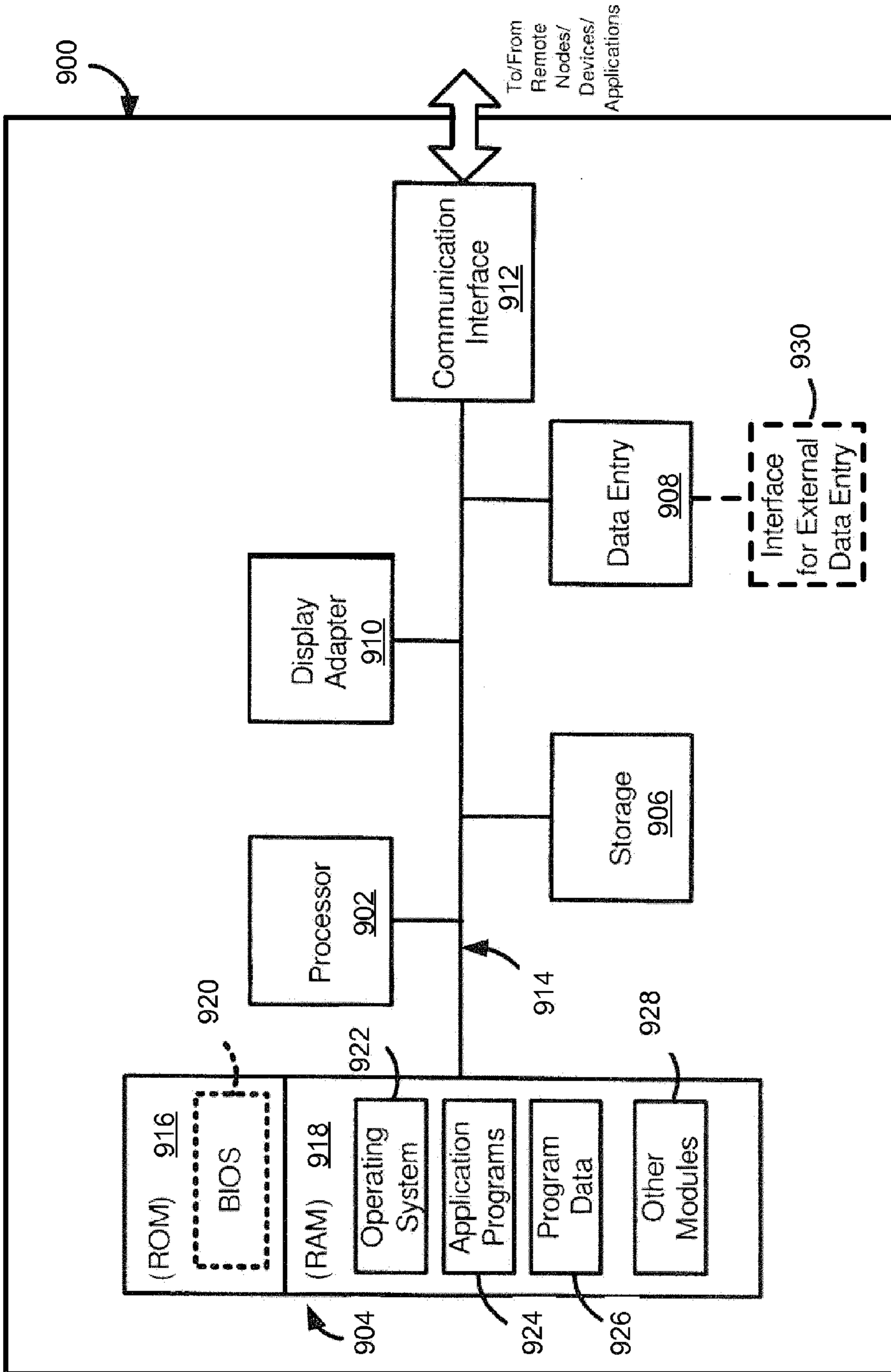


FIG. 9