



(22) Date de dépôt/Filing Date: 2004/05/05

(41) Mise à la disp. pub./Open to Public Insp.: 2004/12/13

(45) Date de délivrance/Issue Date: 2014/12/09

(62) Demande originale/Original Application: 2 466 458

(30) Priorité/Priority: 2003/06/13 (US10/460,949)

(51) Cl.Int./Int.Cl. *H04N 21/234* (2011.01),
H04N 19/177 (2014.01), *H04N 19/61* (2014.01),
H04N 21/236 (2011.01), *H04N 21/647* (2011.01)

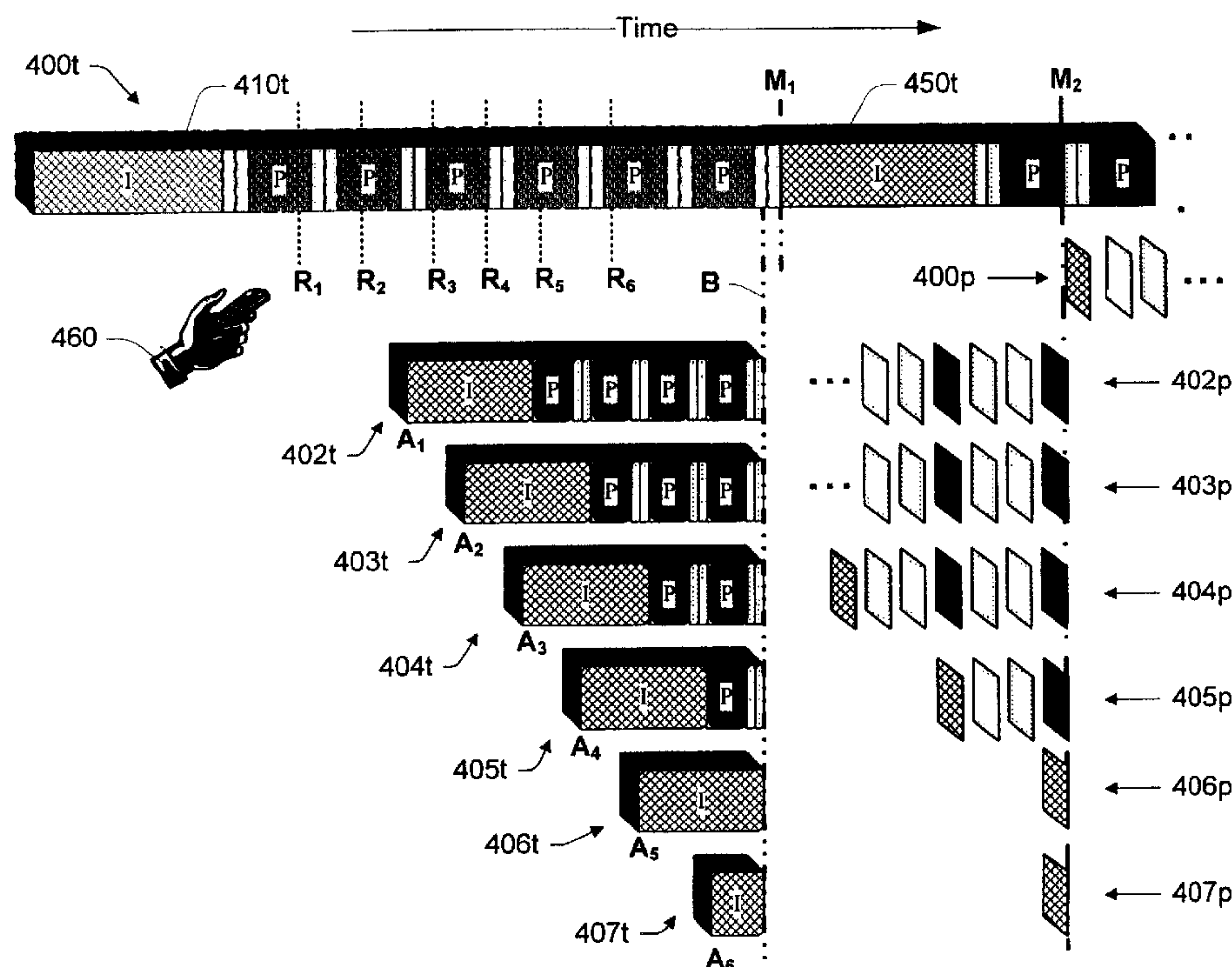
(72) Inventeurs/Inventors:
BALDWIN, JAMES ARMAND, US;
GREEN, DUSTIN L., US;
GROSSMAN, JOHN H., IV, US

(73) Propriétaire/Owner:
MICROSOFT CORPORATION, US

(74) Agent: SMART & BIGGAR

(54) Titre : SYSTÈME DE DEMARRAGE RAPIDE POUR TRAINS DE SEQUENCES VIDEO NUMERIQUE

(54) Title: FAST START-UP FOR DIGITAL VIDEO STREAMS



(57) Abrégé/Abstract:

Described herein is a technology facilitating the presentation of digital video streams. An implementation, described herein, reduces the effective start-up delay in the presentation of the first frames of the video content that occurs when a system tunes into a video stream by streaming at least an additional stream parallel with the main multicast stream. The alternative parallel stream having offset start points from the main multicast stream to more quickly provide a picture in response to a tuning command.



51331-688

Abstract

Described herein is a technology facilitating the presentation of digital video streams. An implementation, described herein, reduces the effective start-up delay in the presentation of the first frames of the video content that occurs when a system tunes into a video stream by streaming at least an additional stream parallel with the main multicast stream. The alternative parallel stream having offset start points from the main multicast stream to more quickly provide a picture in response to a tuning command.

51331-688D

FAST START-UP FOR DIGITAL VIDEO STREAMS

This is a divisional of Canadian Patent Application No. 2,466,458 filed May 5, 2004.

TECHNICAL FIELD

This invention generally relates to a technology for digital video streaming.

BACKGROUND

With advent of digital video streaming technology (such as video-on-demand (VOD) systems), users are able to see and hear digital videos, more or less, as the data is being received from a video server.

When video is streamed, the incoming video stream is typically buffered on the user's receiving device (e.g., computer or set-top box) while data is downloaded into it. At some defined point (generally, when the buffer is full), the video contents are presented to the user. As the video content plays, the receiving device empties the data stored in the buffer. However, while the receiving device is playing the stored video, more data is being downloaded to re-fill the buffer. As long as the data can be downloaded at least as fast as it is being played back, the file will play smoothly.

MPEG

The predominant digital video compression and transmission formats are from a family called MPEG (Moving Picture Experts Group). It is the name of family of standards used for coding audio-visual information (e.g., movies, video, music, and such) in a digital compressed format.

For the convenience of explanation of video streaming, the MPEG-family video stream is generally discussed and described herein. However, those who are

1 skilled in the art understand and appreciate that other such digital video
2 compression and transmission formats exist and may be used.

3 Of course, there are other digital video compression and transmission
4 formats, such as the H.264 codec. Those of ordinary skill in the art will
5 understand how the concepts discussed herein with relationship to MPEG apply to
6 other formats.

7 8 GOP and Frames

9 A MPEG video stream is typically defined by a series of segments called
10 Groups of Pictures (GOP). Typically, a GOP consists of a set of pictures intended
11 to be displayed in sequence over a short duration (e.g., ½ second) when displayed
12 at their intended speed.

13 A GOP typically includes three types of frames:

- 14 • an intra frame (**I**-frame);
- 15 • predictive frames (**P**-frames); and
- 16 • bi-directionally predictive frames (**B**-frames).

17 There is no specific limit to the number of frames which may be in a GOP,
18 nor is there a requirement for an equal number of pictures in all GOPs in a video
19 sequence.

20 The **I**-frame is an encoded still image. It is not dependent upon any other
21 frame that the decoder has already received. Each GOP typically has only one **I**-
22 frame. It is sometimes called a random access point (or "RAP") since it is an entry
23 point for accessing its associated GOP.

24 From the point of view of a video-stream decoder, the **P**-frames are
25 predicted from the most recently reconstructed **I**- or **P**-frame. A **P**-frame (such

1 as frame 120p) requires data from a previously decompressed anchor frames (e.g.,
2 **I**-frames or **P**-frames) to enable its decompression.

3 Switching to the point of view of video stream encoder and transmitter, the
4 **B**-frames are predicted from the closest two **I**- or **P**-frames—one frame in the
5 past and one frame in the future. A **B**-frame (such as frame 132p) requires data
6 from both preceding and succeeding anchor frames (e.g., **I**-frames or **P**-frames)
7 to decode its image. It is bi-directionally dependent.

8 Of course, other digital video compression and transmission formats (such
9 as H.264 codec) may employ other labels, some different types, and different
10 relationships between frames. For example, in H.264, the frame types, frame
11 dependence relationships, and frame ordering are much more decoupled than they
12 are in MPEG. In H.264, the **I**-frames are independently decodable and are
13 random access points. Also, frames have defined presentation order (like MPEG
14 does). However, the other frames relate differently than do the MPEG **P**-frames
15 and **B**-frames.

16 So, those of ordinary skill in the art will understand how the concepts
17 discussed herein with relationship to MPEG apply to other formats.

18 Transmission and Presentation Timelines

19
20 Fig. 1 illustrates two manifestations of the same MPEG video stream. The
21 first is the transmission timeline 100t and the other is the presentation timeline
22 100p.

23 The transmission timeline 100t illustrates a video stream from the
24 perspective of its transmission by a video-stream encoder and transmitter.

1 Alternatively, it may be viewed from the perspective of the receiver of the
2 transmission of the video stream.

3 As shown in Fig. 1, the **I**-frames (e.g., 110t and 150t) are typically
4 temporally longer than the other frames in the transmission timeline. Since it
5 doesn't utilize data from any other frame, it contains all of the data necessary to
6 produce one complete image for presentation. Consequently, an **I**-frame includes
7 more data than any of the other frames. Since the **I**-frame has more data than
8 others, it follows that it typically requires greater time for transmission (and, of
9 course, reception) than the other frame types.

10 Fig. 1 also shows **P**-frames (such as 120t) and **B**-frames (such as 130t and
11 132t) of the transmission timeline 100t. Relative to the **B**-frames, the **P**-frames
12 are temporally longer in the transmission timeline because they typically include
13 more data than the **B**-frames. However, **P**-frames are temporally shorter than **I**-
14 frames because they include less data than **I**-frames. Since the **B**-frames rely on
15 data from at least two other frames, they typically do not need as much data of
16 their own to decode their image as do **P**-frames (which rely on one other frame).

17 Fig. 1 also illustrates the presentation timeline 100p of the video stream
18 from the perspective of its presentation by the video decoder and presenter. In
19 contrast to their transmission duration, the presentation duration of each frame—
20 regardless of type—is exactly the same. In other words, it displays at a fixed
21 frequency.

22 The incoming frames of the video stream are decoded, buffered, and then
23 presented at a fixed frequency (e.g., 24 frames per second (fps)) to produce a
24 relatively smooth motion picture presentation to the user. In MPEG 2 used to
25 convey NTSC video, the field rate is fixed, and each MPEG 2 picture may produce

1, 2, or 3 fields. Field pictures are required to produce 1 field, and frame pictures may produce 2 or 3 fields. Thus, the frame picture presentation rate may not be fixed, but it is not dictated by the transmission rate of the frame pictures.

Fig. 1 also illustrates a typical decoded GOP 105 of MPEG in its presentation timeline. This GOP example includes an **I**-frame 110p; six **P**-frames (e.g., 120p); and 14 **B**-frames (e.g., 130p and 132p). Typically, each GOP includes a series of consecutively presented decoded frames that begin with an **I**-frame (such as frame 110p).

Order of Transmission and Presentation

As shown in Fig. 1, the order in which the frames are presented typically does not directly match the order in which the frames are transmitted. The arrows shown in Fig. 1 between the frames of the transmission timeline 100t and the presentation timeline 100p illustrate a typical way that frames are reordered between reception and presentation. The tail of each arrow has a bullet (i.e., circle) anchor at the end of a transmitted frame. The head of each arrow has an arrowhead pointing to its corresponding presentation frame.

For example, the transmission **I**-frame 110t corresponds to the presentation **I**-frame 110p. In reality these are the same frames, but their timeline representations indicate their different manifestations.

Returning to the explanation of this example, the transmission **P**-frame 120t corresponds to the presentation **P**-frame 120p. The transmission **B**-frames 130t and 132t corresponds to the presentation **B**-frames 130p and 132p. As shown in Fig. 1, these **B**-frames 130t and 132t are encoded, transmitted, received, and decoded *after* their **P**-frame 120t in the transmission timeline 100t, but their

1 corresponding presentation B-frames 130p and 132p are presented *before* their
 2 P-frame 120p in the presentation timeline 100t. Note that the encoder typically
 3 receives the frames in non-compressed form in the same order that the frames are
 4 eventually displayed, and the encoder typically performs the frame re-ordering
 5 before compressing the frames.

6 Furthermore, the next GOP to be transmitted starts with I-frame 150t, but
 7 two B-frames 134t and 136t typically come along after this new GOP has begun.
 8 As illustrated in Fig. 1, the straggling B-frames 134p and 136p are presented in
 9 sequence and before the presentation of the I-frame 150p of the new GOP.

10 GOP Presentation Delay

11
 12 Fig. 1 shows that the I-frame 110t of an example GOP is first received
 13 beginning at point T_1 in time; however, it is not first presented until point T_2 . The
 14 time gap between the two points is called herein the "GOP presentation delay" and
 15 is labeled 170 in Fig. 1. It represents the delay from when the receiver first begins
 16 receiving the first frame of a GOP (which is typically the I-frame) until the device
 17 first presents the first frame of the GOP.

18 There are many reasons for this delay. Some are a natural consequence of
 19 the video streaming technology and others are imposed into the process to address
 20 known technical issues. Some of reasons for the GOP presentation delay include:

- 21 • contrast between the time required to receive a frame transmission
 22 and the time required to display a frame;
- 23 • the time required to decode a frame (especially considering inter-
 24 frame dependencies for decoding); and

- built-in delay to facilitate smooth presentation of frames without needed to wait for frame transmission or decoding.

The details of these reasons and the knowledge of other reasons are known to those of ordinary skill in the art.

Video-Stream Presentation Start-up Delay

To tune channels in a video-streaming environment (such as digital cable), a receiver receives a video stream and waits for an access point into the stream. A channel change cannot occur until an access point is received. From the perspective of the user, this can lead to lengthy channel change times.

Fig. 2 illustrates an example of a video-stream presentation start-up delay at 280. The start-up delay is the effective delay experienced by a user. It includes a delay between when a particular video stream is requested and the actual presentation of the first frame of a GOP from the particular video stream. As shown in Fig. 2, the start-up delay 280 includes the GOP presentation delay 270 (discussed above).

Referring to Fig. 2, this example is explained. A GOP, starting with I-frame 210t, is being transmitted. This is shown in the transmission timeline 200t. The receiver tunes into this video stream at request point **R**. This selection is illustrated as a user selecting a video-stream channel using a remote control 260.

Again, this is an example illustration for explanatory purpose. This point **R** could be at any moment in time after the beginning (i.e., after the beginning of its I-frame 210t) of a GOP.

The receiver must wait for a random access point (or RAP) in order to access the video stream. In this example, each GOP has one RAP. An I-frame is

1 an example of a typical RAP. Therefore, each GOP has one I-frame. So, the
 2 receiver must wait for the next I-frame (at the beginning of the next GOP) before
 3 it can access the video-stream transmission as shown by transmission timeline
 4 200t.

5 Once the receiver has an I-frame in its buffer, it may refer back to it for
 6 dependency decoding of P- and B-frames. Consequently, a conventional system
 7 must wait for a RAP before it can start buffering frames (that are useful).

8 In Fig. 2, the receiver starts buffering the next GOP at point M_1 with I-
 9 frame 250t. Thus, the first frame that may be eventually presented to the user is I-
 10 frame 250t, because it is the first RAP in the stream after the point at which the
 11 receiver joined the stream. Because of the GOP presentation delay (discussed
 12 above), it actually starts presenting the GOP (with I-frame 250p of presentation
 13 timeline 200p) at point M_2 —which is also the presentation start-up point S of the
 14 start-up delay 280.

15 As demonstrated by the screens 262-266, the start-up delay is the effective
 16 delay experienced by a user. The user selects a video-stream channel at request
 17 point R (using, for example, a remote 260) and sees a blank screen, as shown by
 18 screen 262. Of course, there may be information presented here (such as
 19 electronic programming information), but since it is not yet the desired video-
 20 stream content it is effectively blank.

21 Screen 264 shows that screen remains blank even after the next GOP is
 22 currently being received. Screen 266 shows that the first image of frame 250p is
 23 finally presented to the user.

24 The average length of this start-up delay is directly proportional to the
 25 average GOP length. Some video-stream providers employ relatively long average

51331-688D

GOP lengths. For these instances, this delay is even more acute because the user is waiting longer for the next GOP to come round after she has changed channels.

In short, this start-up delay is very annoying to the typical users and tries their patience.

5 **SUMMARY**

Described herein is a technology facilitating the presentation of digital video streams. An implementation, described herein, reduces the effective start-up delay in the presentation of the first frames of the video content that occurs when a system tunes into a video stream.

10 In one aspect, there is provided a system comprising: a selector configured to select a lead-in video stream, where the lead-in stream corresponds to a particular main video-stream transmission; a receiver configured to receive the selected lead-in video-stream transmission; a presentation device configured to present content of the selected lead-in video stream transmission; a switcher configured to switch reception from the lead-in to the main
15 video-stream transmission and to switch presentation from the content of the lead in to the content of the main video-stream transmission; an inquisitor configured to query a video-stream provider asking the provider to identify which one of multiple lead-in video-stream transmissions is the first transmission with an available access point, wherein the identified lead-in transmission is the one selected, wherein a first one of the multiple lead-in video
20 stream transmissions includes data for a first number of frames, a second one of the multiple lead-in video-stream transmissions includes data for a second number of frames, the first number of frames being different from the second number of frames, the first number of frames being determined at least in part based on a first time and a next main video-stream access point transmission time, the first time being associated with a first beginning
25 transmission time for the first one of the multiple lead-in video-stream transmissions, the second number of frames being determined at least in part based on a second time and the next main video-stream access point transmission time, the second time being associated with

51331-688D

a second beginning transmission time for the second one of the multiple lead-in video-stream transmissions.

In another aspect, there is provided a system comprising: a selector configured to choose a particular main video-stream transmission for reception and presentation; a
 5 receiver configured to temporarily receive, before the transmission of the next access point transmitted in the main video-stream transmission, a lead in video stream that corresponds to the particular main video-stream transmission; a switcher configured to switch reception from the lead-in to the main video stream transmission; wherein the lead-in video-stream
 10 transmission is scheduled for transmission during transmission of one group-of-pictures (GOP) of the main video-stream transmission, wherein a GOP has only one access point, and wherein the lead-in video-stream transmission includes data for a number of frames, the number of frames being determined at least in part based on the scheduled beginning time for transmission of the lead-in video-stream transmission and a next main video-stream access point transmission time.

15 This summary itself is not intended to limit the scope of this patent. Moreover, the title of this patent is not intended to limit the scope of this patent. For a better understanding of the present invention, please see the following detailed description and appending claims, taken in conjunction with the accompanying drawings. The scope of the present invention is pointed out in the appending claims.

20 **BRIEF DESCRIPTION OF THE DRAWINGS**

The same numbers are used throughout the drawings to reference like elements and features.

Fig. 1 is diagram illustrating a typical video stream transmission timeline and its corresponding presentation timeline.

25 Fig. 2 is diagram illustrating the presentation start-up delay using a typical video stream transmission timeline and its corresponding presentation timeline.

1 Fig. 3 is diagram illustrating the new presentation start-up delay in
2 accordance with an implementation, described herein, that employs a single
3 alternative video-stream.

4 Fig. 4 is diagram illustrating the new presentation start-up delays in
5 accordance with another implementation, described herein, that employs multiple
6 alternative video-streams.

7 Fig. 5 is a flow diagram showing a methodological implementation
8 described herein.

9 Fig. 6 is a flow diagram showing a methodological implementation
10 described herein.

11 Fig. 7 illustrates exemplary environment in which an implementation
12 described herein may be employed.

13 Fig. 8 illustrates of an example presentation device, a television, and
14 various input devices that interact with the presentation device.

15 Fig. 9 is a block diagram that illustrates components of the example
16 presentation device(s) shown in Figs. 7 and 8.

17 **DETAILED DESCRIPTION**

18
19 In the following description, for purposes of explanation, specific numbers,
20 materials and configurations are set forth in order to provide a thorough
21 understanding of the present invention. However, it will be apparent to one skilled
22 in the art that the present invention may be practiced without the specific
23 exemplary details. In other instances, well-known features are omitted or
24 simplified to clarify the description of the exemplary implementations of the
25 present invention, and thereby, to better explain the present invention.

1 Furthermore, for ease of understanding, certain method steps are delineated as
2 separate steps; however, these separately delineated steps should not be construed
3 as necessarily order dependent in their performance.

4 The following description sets forth one or more exemplary
5 implementations of a Fast Start-up for Digital Video Streams that incorporate
6 elements recited in the appended claims. These implementations are described
7 with specificity in order to meet statutory written description, enablement, and
8 best-mode requirements. However, the description itself is not intended to limit
9 the scope of this patent.

10 The inventors intend these exemplary implementations to be examples. The
11 inventors do not intend these exemplary implementations to limit the scope of the
12 claimed present invention; rather, the inventors have contemplated that the
13 claimed present invention might also be embodied and implemented in other ways,
14 in conjunction with other present or future technologies.

15 An example of an embodiment of a Fast Start-up for Digital Video Streams
16 may be referred to as an "exemplary fast start-up system."

17 For the convenience of explanation, digital video streams are discussed and
18 described herein in terms of the MPEG-family standard format. However, those
19 who are skilled in the art understand and appreciate that other such digital video
20 compression and transmission formats exist.

21 Introduction

22
23 The one or more exemplary implementations, described herein, of the
24 present claimed invention may be implemented (in whole or in part) by a
25

1331-688

1 presentation device 708 (of Figs. 7-9) and/or as part of a computing environment
2 like that shown in Fig. 7.

3 To tune channels in a digital video multicast (e.g., IP multicast)
4 environment, a receiver receives a multicast video data stream and waits for an
5 access point into the stream. This is sometimes called a random access point
6 (RAP). A channel change cannot occur until a RAP is received. Thus may lead to
7 lengthy channel change times.

8 A multicast environment is an example of one type of environment that
9 employs video streaming. Such an environment may utilize Internet Protocol
10 multicasting (i.e., IP multicasting). Those of ordinary skill in the art are familiar
11 with multicast and its use in a multicast environment.

12 In many IP multicast scenarios, there is a limited bit-rate available to the
13 client device (i.e., a receiver). The bit-rate used to the client device typically
14 depends on which IP multicasts the client is currently listening to among all
15 available IP multicasts.

16 Because the total number of ongoing IP multicasts is not limited by the
17 available bit-rate to any particular client, the exemplary fast start-up system is
18 particularly applicable to such a scenario. One may multicast multiple join-in
19 streams (e.g., alternative transmissions 402t-407t) and a main stream at all times.
20 Furthermore, individual client device may determine for themselves which stream
21 to be listening to at any given point. Also, the bit-rate of the main stream may be
22 as high as the connection to the client device can support—so, a relatively high
23 apparent steady-state video quality may be achieved.

24 As shown in Fig. 3, the exemplary fast start-up system employs a main
25 multicast video stream transmission 300t and one or more alternative multicast

1 lead-in video stream transmissions (such as transmission 302t) to achieve a faster
2 tuning time. These alternative streams include the same original content as the
3 main stream, but they may have a lower bit-rate than the main stream.

4 The alternative streams may be low bitrate so that they may be transmitted
5 in less time than is required to transmit the main stream. This is graphically
6 illustrated by alternative video-stream transmission 302t being shorter than the
7 corresponding frames in the main video-stream transmission 300t. Herein, the
8 concept of "low bitrate" is in terms of number of bits per picture (i.e., frame).
9 These may be a normal bitrate in terms of the number of bits per second.

10 With the exemplary fast start-up system, the video stream content is
11 presented more quickly (than conventional approaches) in response to a tuning
12 command—even if the initial picture is of lesser quality.

13 While much of the discussion of the exemplary fast start-up system is
14 framed within terms of the MPEG family of digital video compression and
15 transmission formats, those of ordinary skill in the art will understand how the
16 concepts discussed herein with relationship to MPEG apply to other formats, such
17 as H.264.

18 **Exemplary Fast Start-up System with One Alternative Stream**

19
20 Fig. 3 illustrates the exemplary operation of the exemplary fast start-up
21 system with only one alternative video-stream. Fig. 3 shows the main multicast
22 video stream transmission 300t and the alternative multicast lead-in video stream
23 transmission 302t. Although this discussion refers to the use of only one
24 alternative stream, the bulk of the concepts illustrated in this example apply the
25 use of multiple alternative video-streams.

Fig. 3 also illustrates how this operation reduces the apparent video-stream presentation start-up delay—which is the effective delay experienced by the user. This start-up delay is the delay between the time a particular video stream is requested and the actual presentation of the first frame of a GOP.

Referring to Fig. 3, this example is explained. A GOP starting with I-frame 310t is transmitted in the main stream 300t. The receiver tunes into this video stream at request point **R**. This selection is illustrated as a user selecting a video-stream channel using a remote control 360. (However, in a multicast environment this stream 300t is requested, but since the receiver is not yet tuned in, the stream may not actually be transmitted to the receiver until some point after **R**.)

Again, this is an example illustration for explanatory purpose. This point **R** could be at any moment in time within a GOP after its beginning (i.e., after the beginning of its I-frame 310t).

The receiver typically waits for a random access point (or RAP) in order to access a video stream. (In this example, each GOP is assumed to have one RAP.) An I-frame is an example of a typical RAP. So, the receiver must wait for the next I-frame (at the beginning of the next GOP transmission) before it can access a video-stream transmission and present the video.

With conventional approaches, the user would start seeing the video presentation (as shown in presentation timeline 300p) upon the presentation of the first I-frame 350p at point **M₂**. Therefore, the conventional presentation start-up delay (**D**) would be the difference between the request point (**R**) and point **M₂**. Written as an equation that is

$$D = M_2 - R \quad [1]$$

Conventional Presentation Start-up Delay

However, instead of waiting for the next RAP in the main stream transmission 300t, the exemplary fast start-up system tunes into the alternative video-stream transmission 302t. It starts receiving this GOP transmission at point A. The RAP of alternative transmission 302t is I-frame 312t; therefore, it can begin presenting the alternative presentation timeline 302p with presentation I-frame 312p at point S'.

While alternative video-stream is presented (as shown in presentation timeline 302p), the exemplary fast start-up system requests that the multicast router switch the receiver to the main stream multicast transmission 300t at the next RAP (e.g., I-frame 350t) of the main stream.

While alternative video-stream is presented (as shown in presentation timeline 302p), the exemplary fast start-up system starts receiving main video-stream transmission 300t starting with the first frame (e.g., I-frame 350t) of the next GOP of the main stream. In Fig. 3, the first frame (e.g., I-frame 350t) of the next GOP starts at point M₁. Point B on the timeline represents the end of the reception of the last frame of the alternative stream transmission 302t.

The switch-over from the alternative stream transmission 302t back to the main stream 300t occurs during or around the gap between points B and M₁ in Fig. 3. An exaggerated and visible gap between these points is shown in Fig. 3 to illustrate that there is a clear opportunity for the exemplary fast start-up system to join the main stream. In reality, points B and M₁ may occur very nearly

1 concurrently and may indeed be concurrent. It is possible in some embodiments to
2 have an small overlap so that point **B** occurs shortly after point M_1 .

3 At point M_2 , the exemplary fast start-up system starts presenting main
4 video-stream presentation 300p starting with the first frame (e.g., **I**-frame 350p)
5 of the next GOP of the main stream. This starts immediately after (or very nearly
6 so) the presentation of the last frame (e.g., frame 322p) of the alternative stream
7 presentation 302p. This presentation switch-over point is designated in Fig. 3 at
8 point M_2 .

9 To accomplish switch-over, it may be desirable for the main stream be
10 tagged for splicing. Alternatively, the RAP locations within the main stream may
11 be conveyed to the multicast router via some other mechanism. Furthermore, it
12 may be desirable for the router have an extension to normal IP multicast so that it
13 starts delivering the main stream multicast transmission when the next RAP in the
14 main stream multicast arrives.

15 Although not necessarily required, it is desirable for the presentation of the
16 alternative stream 302p to be timed so that the last frame presented (e.g., frame
17 322p) is presented immediately before the presentation of the first frame (e.g.,
18 frame 350p) of the main stream presentation 300p. Doing so enhances the smooth
19 presentation of the video content—in particular, it smoothes the presentation of the
20 switch-over from the alternative to the main stream presentations at point M_2 .

21 As demonstrated by the screens 362-366, the start-up delay is the effective
22 delay experienced by a user. The user selects a video-stream channel at request
23 point **R** (using, for example, a remote 360) and sees a blank screen, as shown by
24 screen 362. Of course, there may be information presented here (such as
25

1 electronic programming information), but since it is not yet the desired video-
2 stream content it is effectively blank.

3 Screen 364 shows that screen remains blank even as the RAP of the
4 alternative stream 302t is being transmitted and received. However, screen 366
5 shows that the first image of frame 312p is presented to the user.

6 With the exemplary fast start-up system (as shown in Fig. 3), the user first
7 experiences the presentation of the video content (as shown in alternative
8 presentation timeline 302p) upon the presentation of the first I-frame 312p at
9 point S'. Therefore, the new presentation start-up delay (D') would be the
10 difference between the request point (R) and presentation of the alternative stream
11 at point S'. Expressed as an equation, that is

$$12 \quad D' = S' - R \quad [2]$$

13
14 *New Presentation Start-up Delay*

15
16 The time-savings effected by the exemplary fast start-up system is the
17 difference between the D' and D. Using equations 1 and 2, that time-savings may
18 be expressed as this equation:

$$19 \quad D - D' = M_2 - S' \quad [3]$$

20
21 *Time-savings*

22
23 With the exemplary fast start-up system, the user experience is improved
24 because the new start-up delay (D') is less than the conventional start-up delay (D)
25 (i.e., $D' < D$). The exemplary fast start-up system improves the user's experience

1 by decreasing the effective start-up delay experienced by the user when compared
2 to the delay experienced using conventional approaches.

3 4 **Exemplary Fast Start-up System with Multiple Alternative Streams**

5 To further minimize the start-up delay, the exemplary fast start-up system
6 may account for the randomness at which a user tunes into a video-stream channel
7 by employing multiple alternative, RAP phase-staggered video-stream
8 transmissions.

9 Fig. 4 shows the main multicast video stream transmission 400t and phase-
10 staggered multiple alternative lead-in video-stream transmissions 402t-407t. Each
11 alternative transmission may be sent using all of the available bandwidth—thereby
12 minimizing transmission time for each stream. Also, the transmission schedule of
13 the streams are phase-staggered so that the completed reception of each stream's
14 RAP is staggered. Since the RAP of each of the streams illustrated in Fig. 4 is at
15 the beginning of the transmissions, the beginning of each of the transmissions
16 402t-407t is staggered. This staggered start of the alternative transmissions is
17 illustrated in Fig. 4.

18 Furthermore, the time range of phase-staggering of the alternative streams
19 402t-407t is between RAPs of the main stream transmission 400t. In Fig. 4, this is
20 illustrated by the phase-staggering between I-frames 410t and 450t.

21 By sending multiple different streams, tuning time is improved because the
22 receiver may select one of the lead-in streams to play. The one selected will
23 typically be the one which will be ready to be presented the quickest after the time
24 at which the user tunes.

1 These alternative transmissions need not be sent concurrently to a particular
2 receiver within a multicast environment. Rather, each one is prepared for
3 transmission, but a particular multicast stream is sent to a particular receiver only
4 when requested by that receiver. Fig. 4 shows six alternative request points (R_1
5 through R_6) where each one corresponds to a particular alternative video-stream
6 transmission (streams 402t-407t, respectively).

7 For example, when the exemplary fast start-up system wishes to tune to a
8 channel (examples are indicated by points R_1 through R_6), it queries the multicast
9 server (such as content server 712 of Fig. 7) in order to determine which lead-in
10 alternative stream is the first lead-in that has not started yet, and the receiver joins
11 that alternative multicast transmission. Then, the exemplary fast start-up system
12 requests that the router switch the receiver back to the main stream multicast
13 transmission 400t just before the next RAP (e.g., frame 450t) of the main stream.

14 Since the alternative stream transmission(s) serves as a "bridge" until the
15 receiver can start receiving the next RAP of the main stream 400t, all of these
16 alternative streams (402t-407t) are shown in Fig. 4 ending at point **B**.

17 In Fig. 4, the first frame (e.g., I-frame 450t) of the next GOP starts at point
18 M_1 . Point **B** on the timeline represents the end of the reception of the last frame of
19 each of the alternative streams (402t-407t)

20 The switch-over from each of the alternative streams (402t-407t) back to
21 the main stream 400t occurs during or around the gap between points **B** and M_1 in
22 Fig. 4. An exaggerated and visible gap between these points is shown in Fig. 4 to
23 illustrate that there is a clear opportunity for the exemplary fast start-up system to
24 join the main stream. In reality, points **B** and M_1 may occur very nearly
25

1331-688

1 concurrently and may indeed be concurrent. It is possible in some embodiments to
2 have a small overlap so that point **B** occurs shortly after point M_1 .

3 Fig. 4 shows the alternative video-stream presentations 402p-407p that
4 correspond to the alternative video-stream transmissions 402t-407t, respectively.
5 Although not necessarily required, it is desirable for the alternative video-stream
6 presentations 402p-407p to be timed so that the last frame presented is presented
7 immediately before the presentation of the first frame of the main stream
8 presentation 400p. Doing so enhances the smooth presentation of the video
9 content—in particular, it smoothes the presentation of the switch-over from the
10 alternative to the main stream presentations at point M_2 .

11 Operation of Exemplary Fast Start-up System

12
13 Fig. 5 shows a methodological implementation of the exemplary fast start-
14 up system. This methodological implementation may be performed in software,
15 hardware, or a combination thereof.

16 At 510 of Fig. 5, the user tunes into a specific main video-stream multicast
17 transmission (e.g., 300t or 400t). This example point is designated at points **R** in
18 Fig. 3 and R_1 through R_6 in Fig. 4.

19 At 512, the exemplary fast start-up system queries a multicast server (such
20 as content server 712 of Fig. 7) in order to determine which of the alternative
21 RAP-phase-staggered lead-in alternative video streams is the first lead-in that has
22 not started yet. Of course, if there is only one alternative transmission, this query
23 may be viewed simply as a request.

24 At 514, the exemplary fast start-up system joins the alternative multicast
25 transmission identified by the query. It receives and buffers this alternative

transmission. This occurs, for example, at designated points **A** in Fig. 3 and **A₁** through **A₆** in Fig. 4.

At 516, it presents the video stream of the identified and buffered alternative transmission. This occurs, for example, at designated point **S** in Fig. 3.

At 518, the exemplary fast start-up system switches back to receiving and buffering the main stream multicast transmission (e.g., 300t and 400t) exactly at the next RAP (e.g., frame 350t and 450t) of the main stream. It may do this by requesting that the router switch the receiver back to the main stream multicast transmission exactly at the next RAP of the main stream. This occurs, for example, between or around designated points **B** and **M₁** in Figs. 3 and 4. This may also be described as occurring "on or about" such designated points.

At 520, it presents the video stream of the main stream multicast. This occurs, for example, at designated point **M₂** in Figs. 3 and 4.

Operation of Exemplary Fast Start-up Multicast System

Fig. 6 shows a methodological implementation of the exemplary fast start-up multicast system, which may be embodied by a content provider 702 and/or a content distribution system 706 of Fig. 7). This methodological implementation may be performed in software, hardware, or a combination thereof.

At 610 of Fig. 6, the exemplary fast start-up multicast system concurrently encodes one or more alternative video streams for transmission. The server system encodes each alternative stream so that the RAP of each is phase-staggered relative to the other streams.

Furthermore, each stream is encoded so that each ends at the same point, which is at or near when the next RAP is available in the main video stream. This designated point, for example, is point M_1 in Figs. 3 and 4.

At 612, the multicast system receives a query to which of the RAP-phase-staggered lead-in alternative video streams is the first lead-in that has not started yet. Of course, if there is only one alternative transmission, this query may be viewed simply as a request.

At 614, it transmits the alternative multicast transmission identified by the query to the receiver that requested it.

At 616, the multicast system receives a request for the router switch the receiver back to the main stream multicast transmission just before the next RAP of the main stream. At 618, it does so in response to such a request. This occurs, for example, between or around designated points B and M_1 in Figs. 3 and 4.. This may also be described as occurring "on or about" such designated points.

Exemplary Environment

Fig. 7 illustrates an exemplary environment 700 in which the techniques, systems, and other aspects described herein may be implemented (partially or wholly). Exemplary environment 700 is a television entertainment system that facilitates distribution of multi-media.

The environment 700 includes one or more multimedia content providers 702, a content distribution system 706, and one or more presentation devices 708(1), 708(2), ..., 708(N) coupled to the content distribution system 706 via a multicast-capable network 710.

1 Multimedia content provider 702 includes a content server 712 and stored
2 content 714, such as movies, television programs, commercials, music, and similar
3 audio and/or video content. Content server 712 controls distribution of the stored
4 content 714 from content provider 702 to the content distribution system 706.
5 Additionally, content server 702 controls distribution of live content (e.g., content
6 that was not previously stored, such as live feeds) and/or content stored at other
7 locations to the content distribution system 706.

8 Content distribution system 706 may be coupled to a network 720, such as
9 an intranet or the Internet. The content distribution system 706 includes a multicast
10 transmitter 728, and one or more content processors 730. Multicast transmitter
11 728 multicasts signals across multicast-capable network 710.

12 Content distribution system 706 is representative of a headend service that
13 provides multimedia content to multiple subscribers.

14 Multicast-capable network 710 can include a cable television network, RF,
15 microwave, satellite, and/or data network, such as the Internet, and may also
16 include wired or wireless media using any multicast format or multicast protocol.
17 Additionally, multicast-capable network 710 may be any type of network, using
18 any type of network topology and any network communication protocol, and may
19 be represented or otherwise implemented as a combination of two or more
20 networks.

21 Content processor 730 processes the content received from content provider
22 702 prior to transmitting the content across multicast-capable network 708. A
23 particular content processor 730 may encode, or otherwise process, the received
24 content into a format that is understood by the multiple presentation devices
25 708(1), 708(2), ..., 708(N) coupled to multicast-capable network 710.

1 Presentation devices 708 may be implemented in a number of ways. For
2 example, a presentation device 708(1) receives content multicasted from a
3 satellite-based transmitter via a satellite dish 734. Presentation device 708(1) is
4 also referred to as a set-top box or a satellite receiving device. Presentation device
5 708(1) is coupled to a television 736(1) for presenting the content received by the
6 presentation device (e.g., audio data and video data), as well as a graphical user
7 interface. A particular presentation device 708 may be coupled to any number of
8 televisions 736 and/or similar devices that may be implemented to display or
9 otherwise render content. Similarly, any number of presentation devices 708 may
10 be coupled to a single television 736.

11 Presentation device 708(2) is also coupled to receive content from
12 multicast-capable network 710 and provide the received content to associated
13 television 736(2). Presentation device 708(N) is an example of a combination
14 television 738 and integrated set-top box 740. In this example, the various
15 components and functionality of the set-top box are incorporated into the
16 television, rather than using two separate devices. The set-top box incorporated
17 into the television may receive multicast signals via a satellite dish or wireless
18 antenna (such as dish 734) and/or via multicast-capable network 710. In alternate
19 implementations, presentation devices 706 may receive content via the Internet or
20 any other multicast medium.

21 The exemplary environment 700 also includes live or stored pay-per-view
22 (PPV) content 742, such as PPV movie content. The stored or live content is
23 typically multicast on a schedule. When a device joins a PPV multicast channel,
24 the PPV content may be viewed with a presentation device 708.
25

Exemplary Presentation Device

Fig. 8 illustrates an exemplary implementation 800 of a presentation device 708 shown as a standalone unit that connects to a television 736. Presentation device 708 may be implemented in any number of embodiments, including as a set-top box, a satellite receiver, a TV recorder with a hard disk, a game console, an information appliance, a DVD player, personal video recorder, a personal computer, a home media center, a modem, and so forth.

Presentation device 708 includes a wireless receiving port 802, such as an infrared (IR) or Bluetooth wireless port, for receiving wireless communications from a remote control device 804, a handheld input device 806, or any other wireless device, such as a wireless keyboard. Handheld input device 806 may be a personal digital assistant (PDA), handheld computer, wireless phone, or the like. Additionally, a wired keyboard 808 is coupled to communicate with the presentation device 708. In alternate embodiments, remote control device 804, handheld device 806, and/or keyboard 808 may use an RF communication link or other mode of transmission to communicate with presentation device 708.

Presentation device 708 may have a storage medium reader 809 for reading content storage media, such as DVD disks. A standalone or non-standalone presentation device 708 may include the storage medium reader 809.

Presentation device 708 may receive one or more multicast signals 810 from one or more multicast sources, such from a multicast network.

Presentation device 708 also includes hardware and/or software for providing the user with a graphical user interface by which the user can, for

1 example, access various network services, configure the presentation device 708,
2 and perform other functions.

3 Presentation device 708 may be capable of communicating with other
4 devices via one or more connections including a conventional telephone link 812,
5 an ISDN link 814, a cable link 816, an Ethernet link 818, a DSL link 820, and the
6 like. Presentation device 708 may use any one or more of the various
7 communication links 812-820 at a particular instant to communicate with any
8 number of other devices. The multicast signals may also be received via the
9 various communication links 812-820.

10 Presentation device 708 generates video signal(s) 820 and audio signal(s)
11 822, both of which are communicated to television 736. Alternatively, video and
12 audio signal may be communicated to other audio/visual equipment, such as
13 speakers, a video monitor, a home theater system, an audio system, and the like.

14 Although not shown in Fig. 8, presentation device 708 may include one or
15 more lights or other indicators identifying the current status of the device.
16 Additionally, the presentation device may include one or more control buttons,
17 switches, or other selectable controls for controlling operation of the device.

18 Fig. 9 illustrates selected components of presentation device 708 shown in
19 Figs. 7 and 8. Presentation device 708 includes a first tuner 900 and an optional
20 second tuner 902. The tuners 900 and 902 are representative of one or more in-
21 band tuners that tune to various frequencies or channels to receive television
22 signals, as well as an out-of-band tuner or receiver or network interface card that
23 tunes to or receives the multicast communications channel over which other
24 content may be multicast to presentation device 708.
25

1 The tuners 900 and 902 may be digital tuners, analog tuners, or any
2 combination of analog and digital components used to get digital data into the
3 client device 708.

4 Presentation device 708 also includes one or more processors 304 and one
5 or more memory components. Examples of possible memory components include
6 a random access memory (RAM) 906, a disk drive 908, a mass storage component
7 910, and a non-volatile memory 912 (e.g., ROM, Flash, EPROM, EEPROM, etc.).

8 Alternative implementations of presentation device 708 can include a range
9 of processing and memory capabilities, and may include more or fewer types of
10 memory components than those illustrated in Fig. 9.

11 Processor(s) 904 process various instructions to control the operation of
12 presentation device 708 and to communicate with other electronic and computing
13 devices. The memory components (e.g., RAM 906, disk drive 908, storage media
14 910, and non-volatile memory 912) store various information and/or data such as
15 multimedia content, electronic program data, web content data, configuration
16 information for presentation device 708, and/or graphical user interface
17 information. The device may cache data into any one of these many memory
18 components.

19 An operating system 914 and one or more application programs 916 may be
20 stored in non-volatile memory 912 and executed on processor 904 to provide a
21 runtime environment. A runtime environment facilitates extensibility of
22 presentation device 708 by allowing various interfaces to be defined that, in turn,
23 allow application programs 916 to interact with presentation device 708.
24
25

1 The application programs 916 that may be implemented on the presentation
2 device 708 may include an electronic program guide (EPG), an email program to
3 facilitate electronic mail, and so on.

4 Presentation device 708 can also include other components pertaining to a
5 television entertainment system which are not illustrated in this example for
6 simplicity purposes. For instance, presentation device 708 can include a user
7 interface application and user interface lights, buttons, controls, etc. to facilitate
8 viewer interaction with the device.

9 Network interface 924 and serial and/or parallel interface 926 allows
10 presentation device 708 to interact and communicate with other electronic and
11 computing devices via various communication links. Although not shown,
12 presentation device 708 may also include other types of data communication
13 interfaces to communicate with other devices.

14 It may include a modem 928 or other communications device that
15 facilitates communication with other electronic and computing devices via a
16 conventional telephone line or other communications mediums.

17 The presentation device 708 has the ability to receive multicast digital data
18 and it may receive it using the tuners 900 or 902, the network interface 924, the
19 modem 928, or other communications device.

20 Presentation device 708 also includes an audio/video output 930 that
21 provides signals to a television or other device that processes and/or presents or
22 otherwise renders the audio and video data. This output may be called the display.

23 Presentation device 708 also includes a fast start-up module 940 that
24 partially or wholly implements the exemplary fast start-up system. It may be
25 application program or a hardware component.

1 Although shown separately, some of the components of presentation device
2 708 may be implemented in an application specific integrated circuit (ASIC).
3 Additionally, a system bus (not shown) typically connects the various components
4 within presentation device 708.

5 A system bus may be implemented as one or more of any of several types
6 of bus structures, including a memory bus or memory controller, a peripheral bus,
7 an accelerated graphics port, or a local bus using any of a variety of bus
8 architectures. By way of example, such architectures can include a CardBus,
9 Personal Computer Memory Card International Association (PCMCIA),
10 Accelerated Graphics Port (AGP), Small Computer System Interface (SCSI),
11 Universal Serial Bus (USB), IEEE 1394, a Video Electronics Standards
12 Association (VESA) local bus, and a Peripheral Component Interconnects (PCI)
13 bus also known as a Mezzanine bus.

14 Computer-Executable Instructions

15
16 An implementation of an exemplary fast start-up system may be described
17 in the general context of computer-executable instructions, such as program
18 modules, executed by one or more computers or other devices. Generally,
19 program modules include routines, programs, objects, components, data structures,
20 etc. that perform particular tasks or implement particular abstract data types.
21 Typically, the functionality of the program modules may be combined or
22 distributed as desired in various embodiments.

Computer Readable Media

An implementation of an exemplary fast start-up system may be stored on or transmitted across some form of computer readable media. Computer readable media may be any available media that may be accessed by a computer. By way of example, and not limitation, computer readable media may comprise “computer storage media” and “communications media.”

“Computer storage media” include volatile and non-volatile, removable and non-removable media implemented in any method or technology for storage of information such as computer readable instructions, data structures, program modules, or other data. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which may be used to store the desired information and which may be accessed by a computer.

“Communication media” typically embodies computer readable instructions, data structures, program modules, or other data in a modulated data signal, such as carrier wave or other transport mechanism. Communication media also includes any information delivery media.

The term “modulated data signal” means a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example, and not limitation, communication media includes wired media such as a wired network or direct-wired connection, and wireless

1 media such as acoustic, RF, infrared, and other wireless media. Combinations of
2 any of the above are also included within the scope of computer readable media.

3 4 Conclusion

5 Although the invention has been described in language specific to structural
6 features and/or methodological steps, it is to be understood that the invention
7 defined in the appended claims is not necessarily limited to the specific features or
8 steps described. Rather, the specific features and steps are disclosed as preferred
9 forms of implementing the claimed invention.
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

51331-688D

CLAIMS:

1. A system comprising:

a selector configured to select a lead-in video stream, where the lead-in stream corresponds to a particular main video-stream transmission;

5 a receiver configured to receive the selected lead-in video-stream transmission;

a presentation device configured to present content of the selected lead-in video stream transmission;

a switcher configured to switch reception from the lead-in to the main video-stream transmission and to switch presentation from the content of the lead in to the content of
10 the main video-stream transmission;

an inquisitor configured to query a video-stream provider asking the provider to identify which one of multiple lead-in video-stream transmissions is the first transmission with an available access point, wherein the identified lead-in transmission is the one selected, wherein a first one of the multiple lead-in video stream transmissions includes data for a first
15 number of frames, a second one of the multiple lead-in video-stream transmissions includes data for a second number of frames, the first number of frames being different from the second number of frames, the first number of frames being determined at least in part based on a first time and a next main video-stream access point transmission time, the first time being associated with a first beginning transmission time for the first one of the multiple lead-
20 in video-stream transmissions, the second number of frames being determined at least in part based on a second time and the next main video-stream access point transmission time, the second time being associated with a second beginning transmission time for the second one of the multiple lead-in video-stream transmissions.

2. The system as recited in claim 1, wherein the selector is further configured to
25 select the selected lead-in video stream from one of multiple lead-in video streams.

51331-688D

3. The system as recited in claim 1, wherein the switcher is configured to switch on or about the occurrence of an access point transmitted in the main video stream transmission.

4. The system as recited in claim 1, wherein the switcher is configured to switch just before the occurrence of an access point transmitted in the main video-stream transmission.

5. The system as recited in claim 1, wherein the available access point is an anchor frame.

6. The system as recited in claim 1, wherein the access point of each of the multiple lead-in video-stream transmissions are scheduled for phase staggered transmission relative to each other.

7. The system as recited in claim 1, wherein each of the multiple lead-in video-stream transmissions are encoded using a lower bitrate than that used by the main video-stream transmission.

8. A system comprising:

a selector configured to choose a particular main video-stream transmission for reception and presentation;

a receiver configured to temporarily receive, before the transmission of the next access point transmitted in the main video-stream transmission, a lead in video stream that corresponds to the particular main video-stream transmission;

a switcher configured to switch reception from the lead-in to the main video stream transmission;

wherein the lead-in video-stream transmission is scheduled for transmission during transmission of one group-of-pictures (GOP) of the main video-stream transmission,

wherein a GOP has only one access point, and wherein the lead-in video-stream transmission

51331-688D

includes data for a number of frames, the number of frames being determined at least in part based on the scheduled beginning time for transmission of the lead-in video-stream transmission and a next main video-stream access point transmission time.

9. The system as recited in claim 8 further comprising a presenter configured to
5 present content of the lead-in video stream transmission and configured to, after switching reception to the main video-stream, present the content of the main video stream transmission.

10. The system as recited in claim 8, wherein the switcher is further configured to switch on or about the transmission of an access point in the main video-stream transmission.

11. The system as recited in claim 10 wherein the video-stream transmissions are
10 multicast.

12. The system as recited in claim 8, wherein the switcher is further configured to switch on or about the transmission of the next access point to occur during the main video-stream transmission.

13. The system as recited in claim 8, wherein each of the multiple lead-in video-
15 stream transmissions are encoded using a lower bitrate than that used by the main video-stream transmission.

14. The system as recited in claim 8, wherein the video-stream transmissions are multicast.

15. The system as recited in claim 8, wherein the access points are anchor frames.

20 16. The system as recited in claim 8 further comprising:

the selector is further configured to select one of multiple lead-in video-stream transmissions for reception, where each lead-in stream corresponds to a particular main video-stream transmission, wherein the one lead-in transmission selected is the lead-in transmission that is temporarily received;

51331-688D

a presenter configured to present content of the selected lead-in video stream transmission;

a switcher further configured to switch reception from the lead-in to the main video-stream transmission and doing so on or about the occurrence of an access point

- 5 transmitted in the main video-stream transmission and switch presentation from the content of the lead-in to the content of the main video stream transmission.

17. The system as recited in claim 16, wherein the access point of each of the multiple lead-in video-stream transmissions are scheduled for phase staggered transmission relative to each other.

- 10 18. The system as recited in claim 16, wherein each of the multiple lead-in video-stream transmissions are encoded using a lower bitrate than that used by the main video-stream transmission.

19. The system as recited in claim 16, wherein the access points are anchor frames.

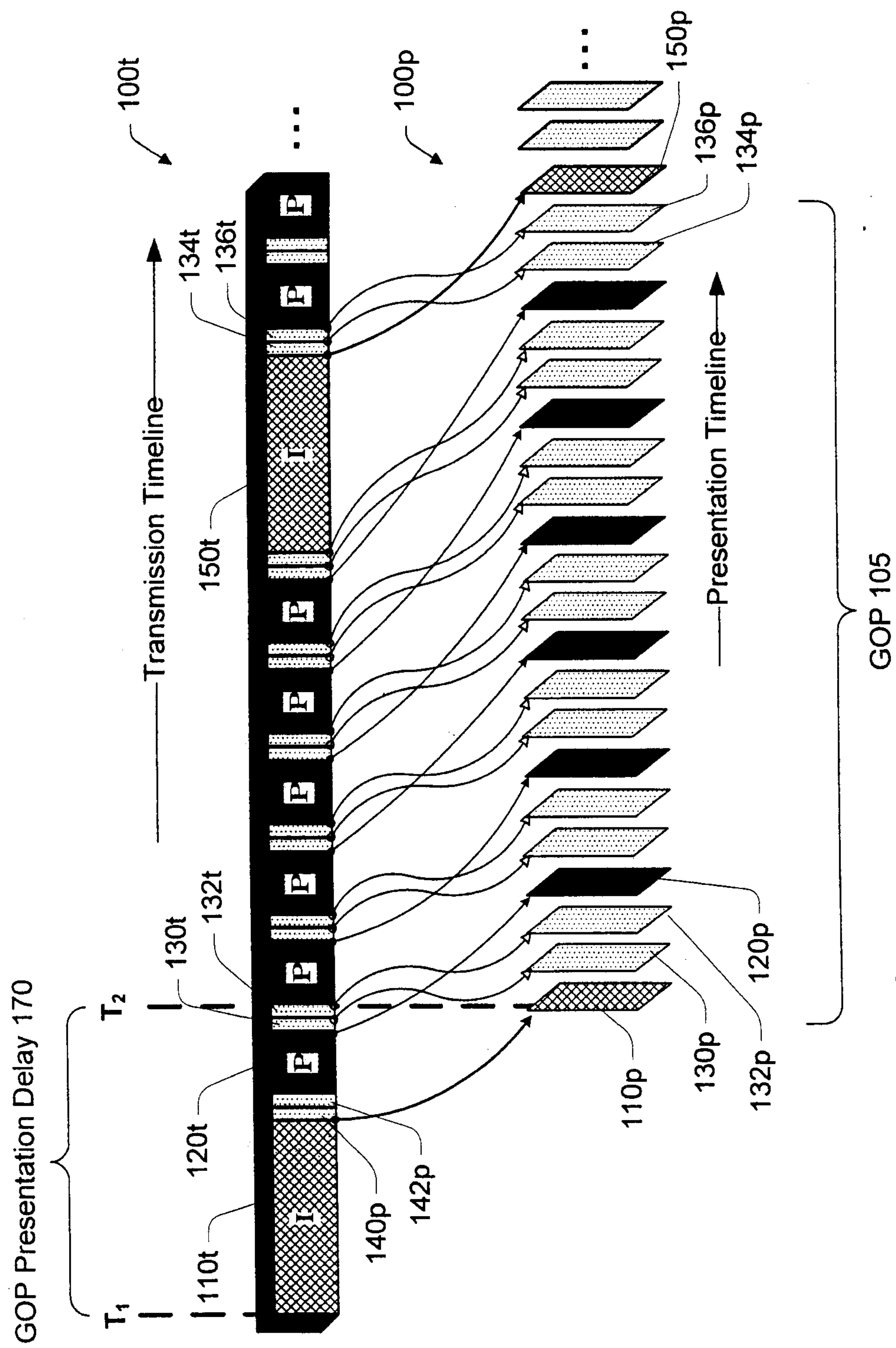
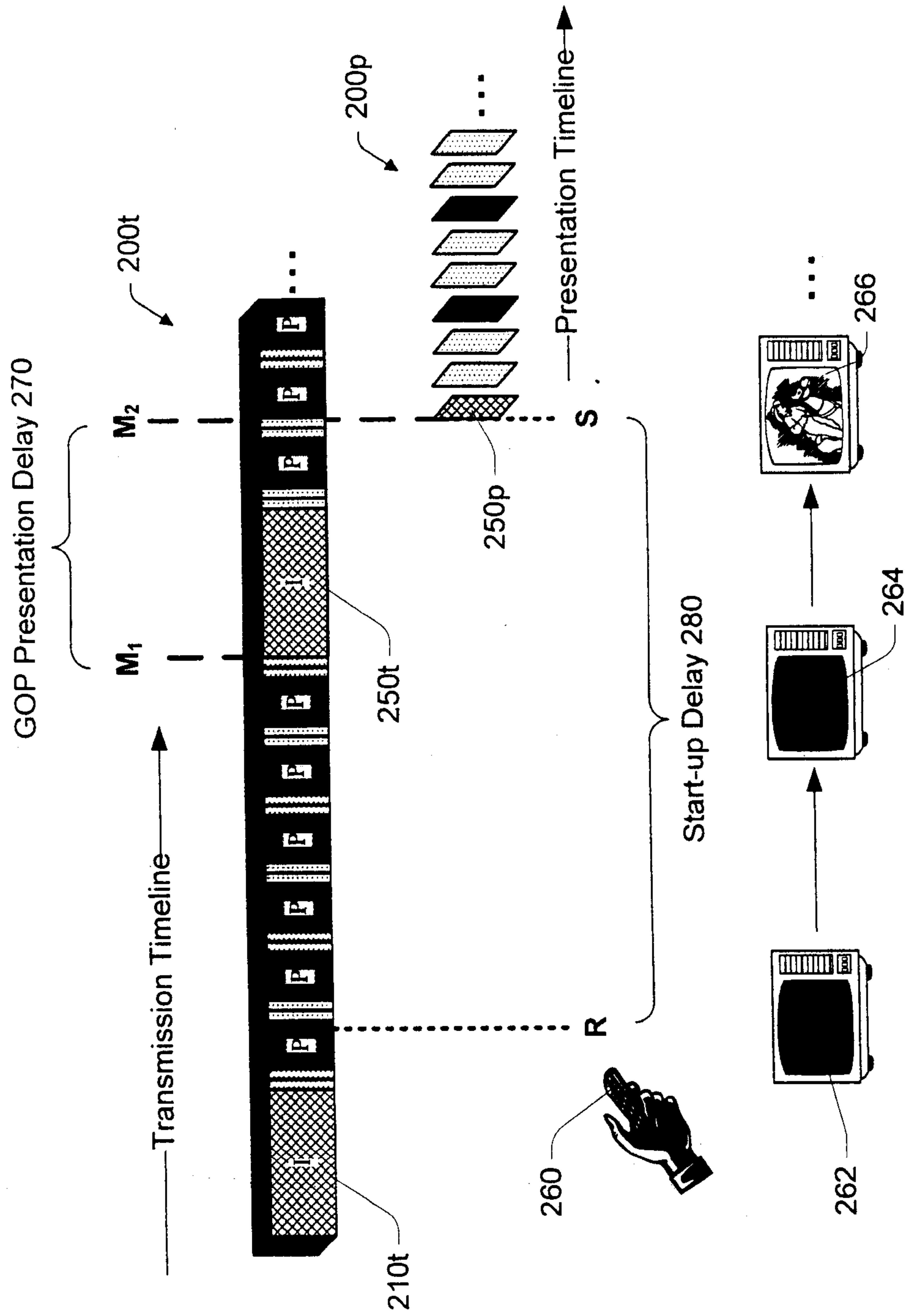


Fig. 1 (background)



$$S = M_2;$$

$$\text{Start-up Delay (D)} = M_2 - R$$

Fig. 2 (background)

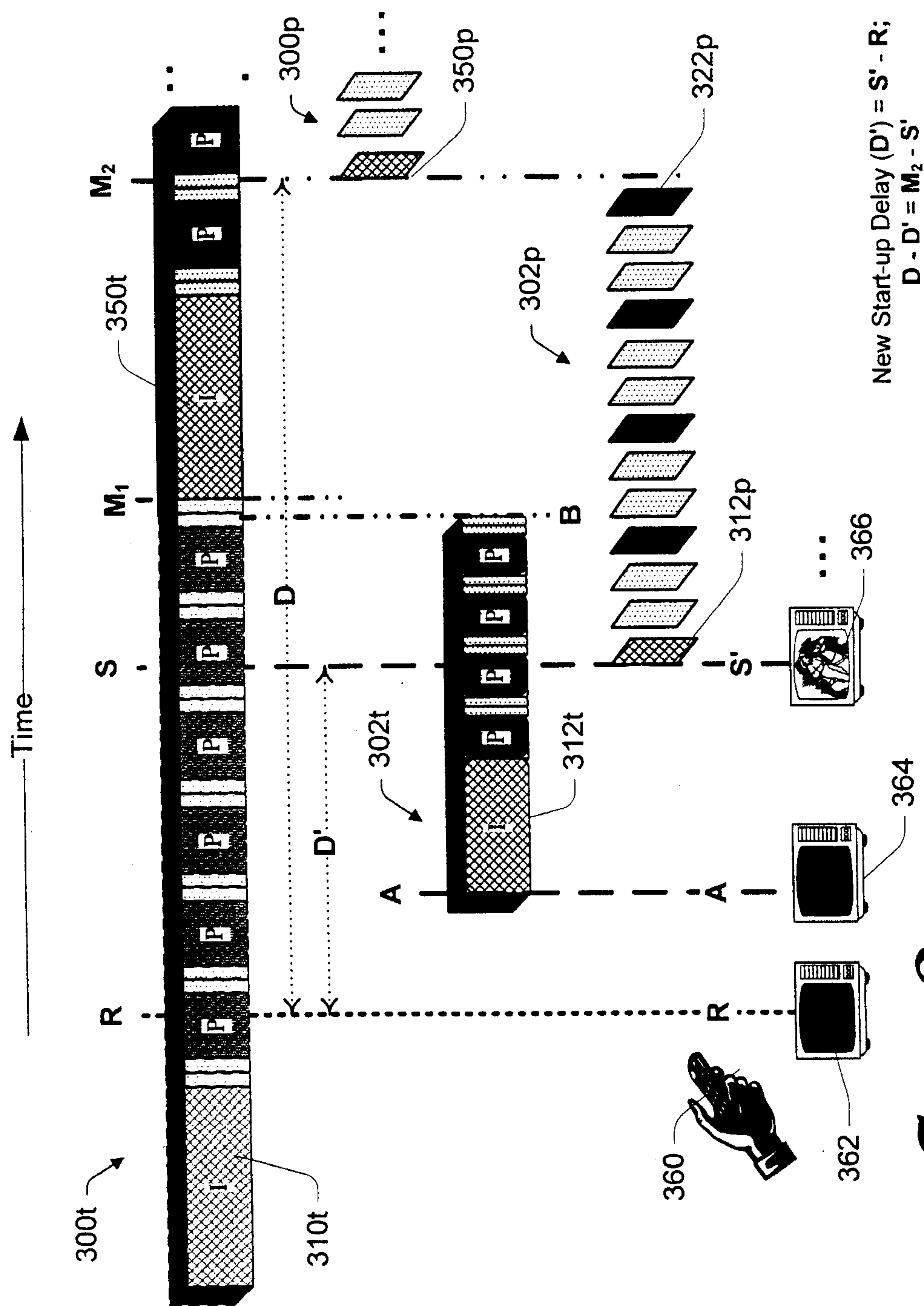


Fig. 3

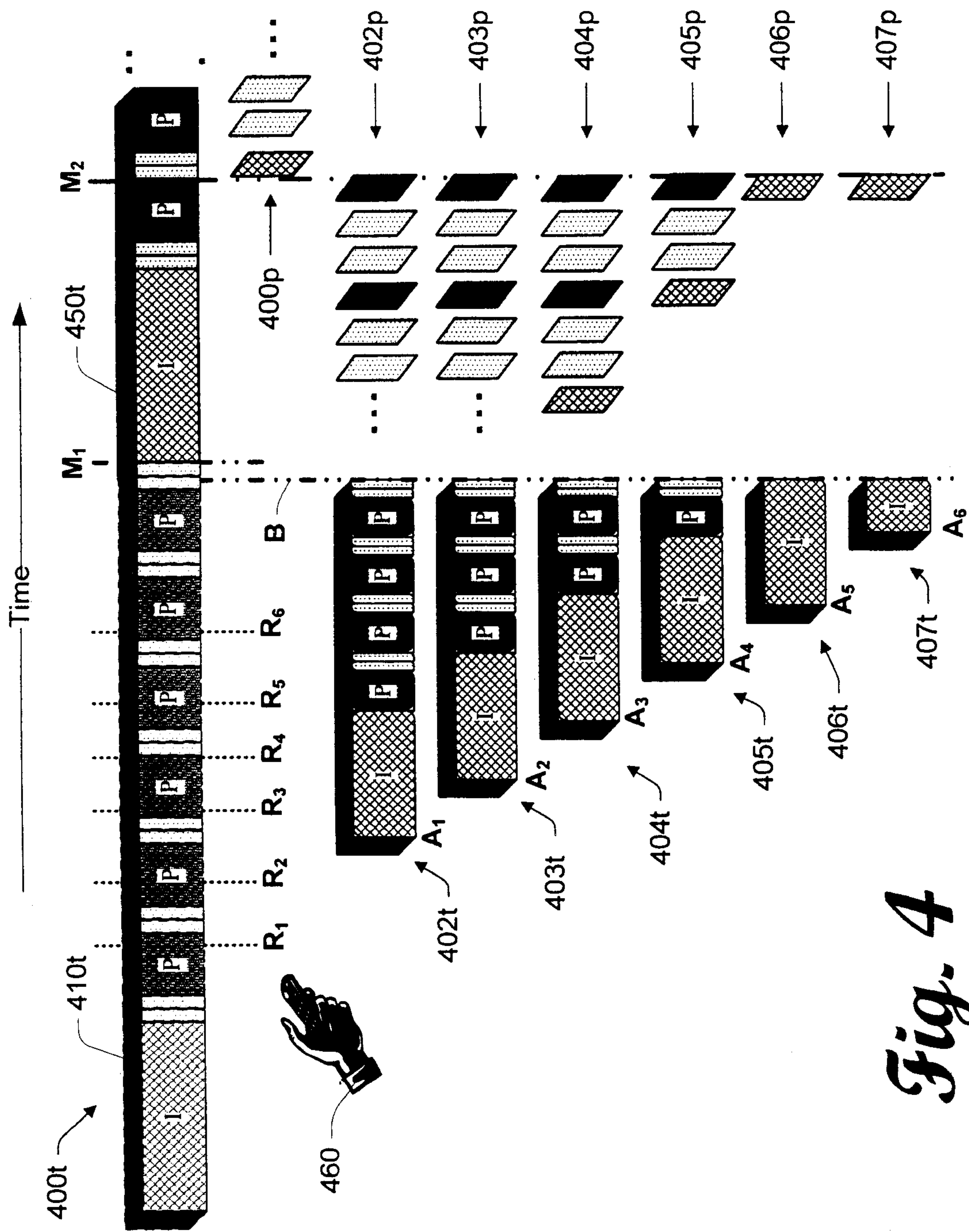
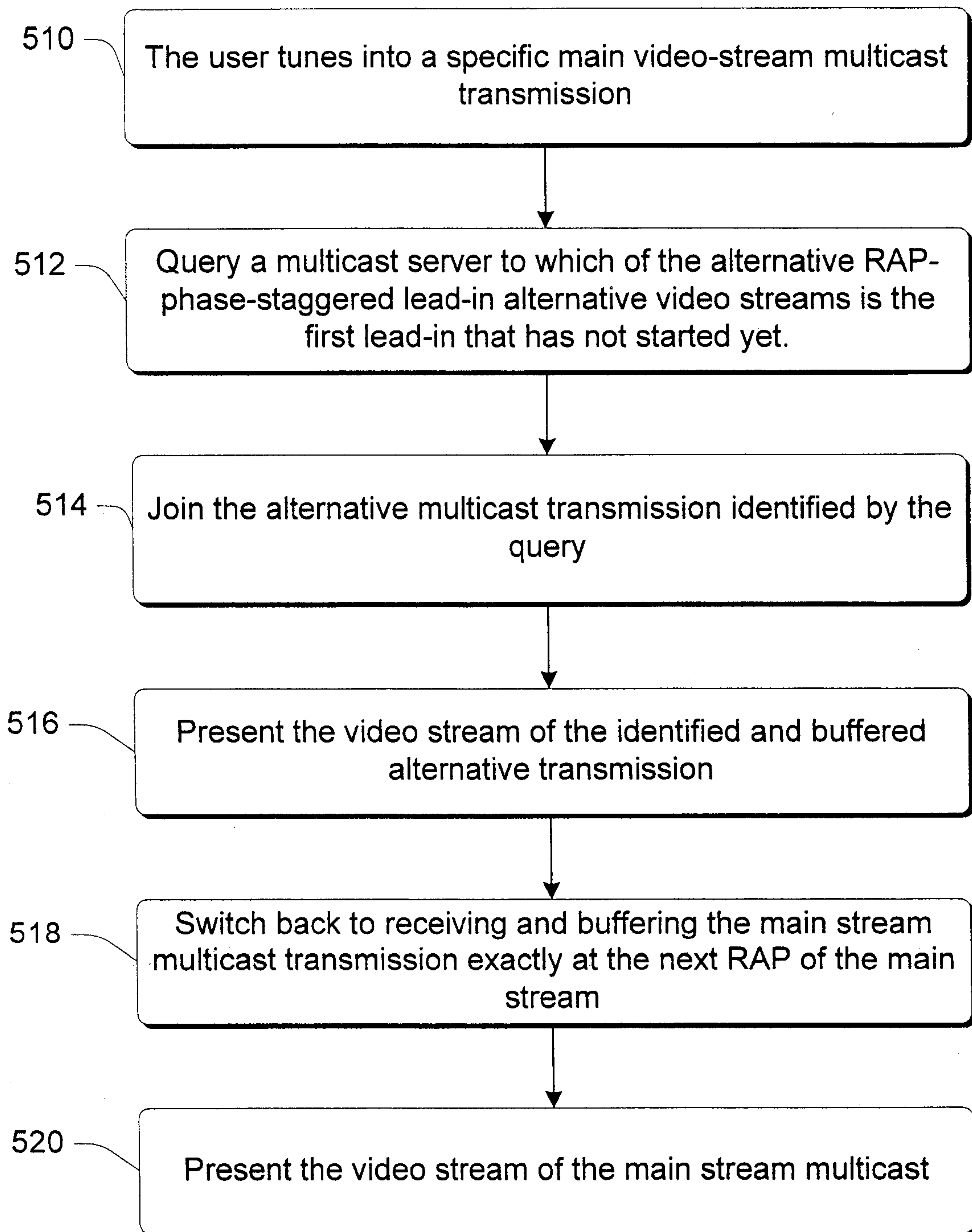
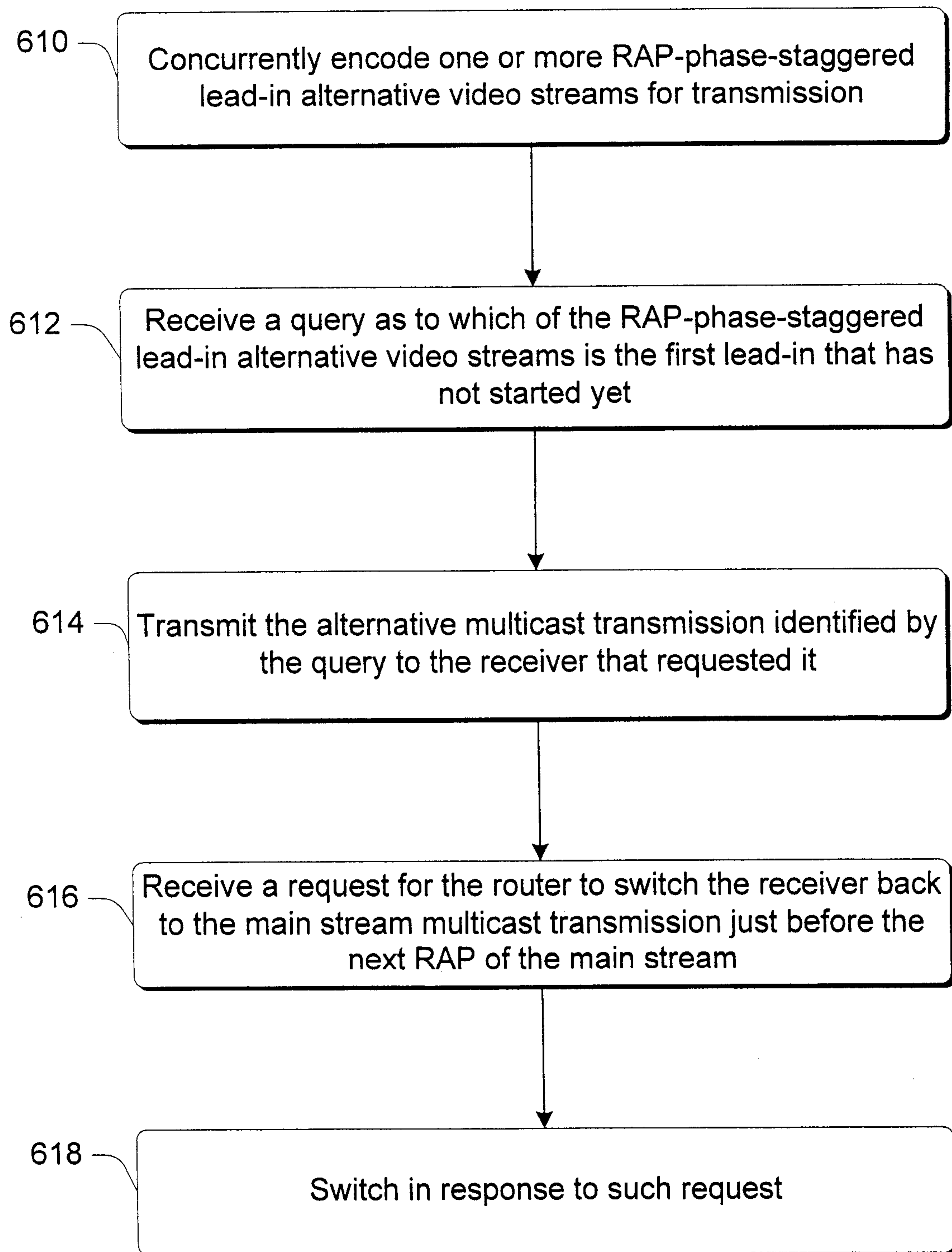
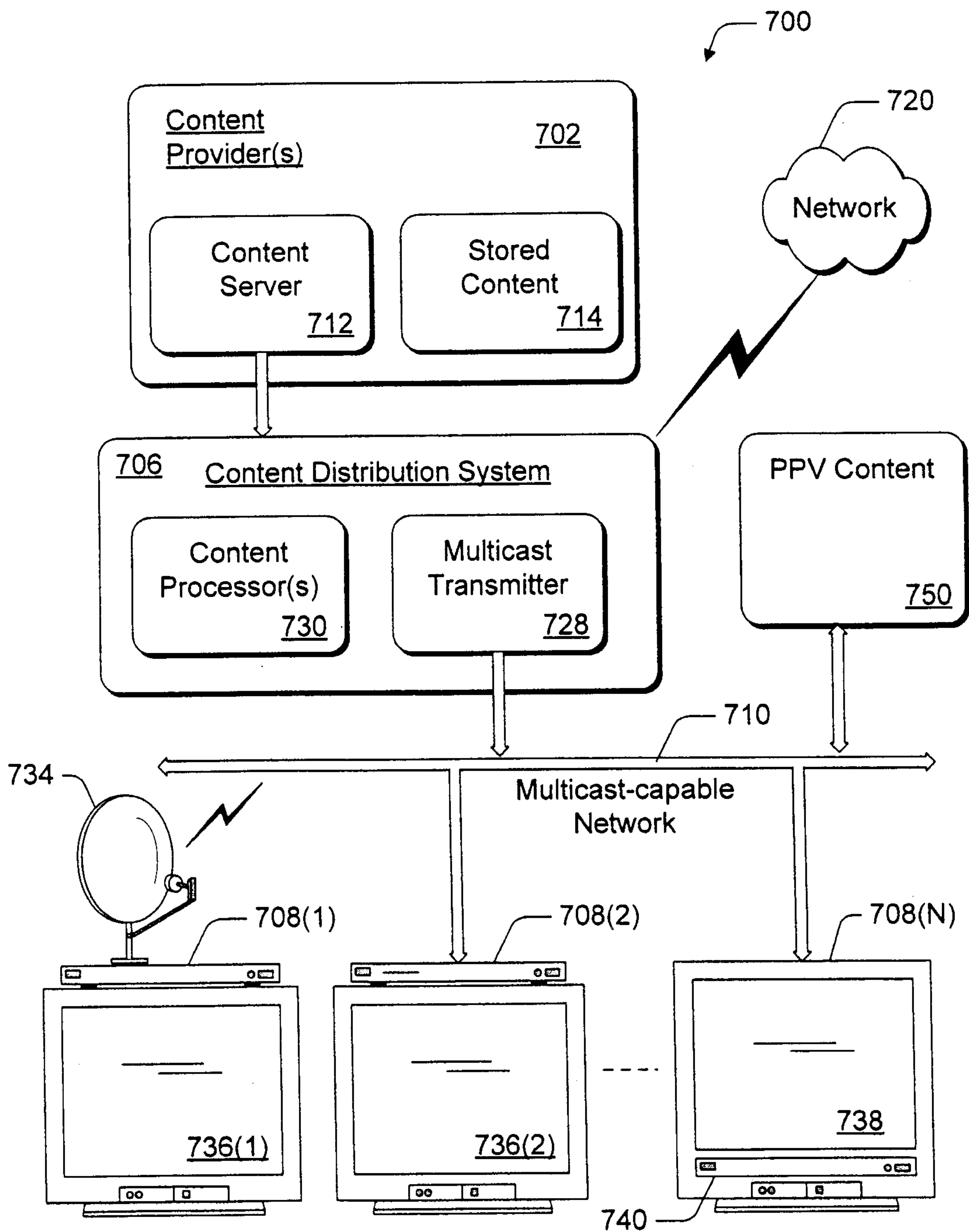


Fig. 4

*Fig. 5*

*Fig. 6*

**Fig. 7**

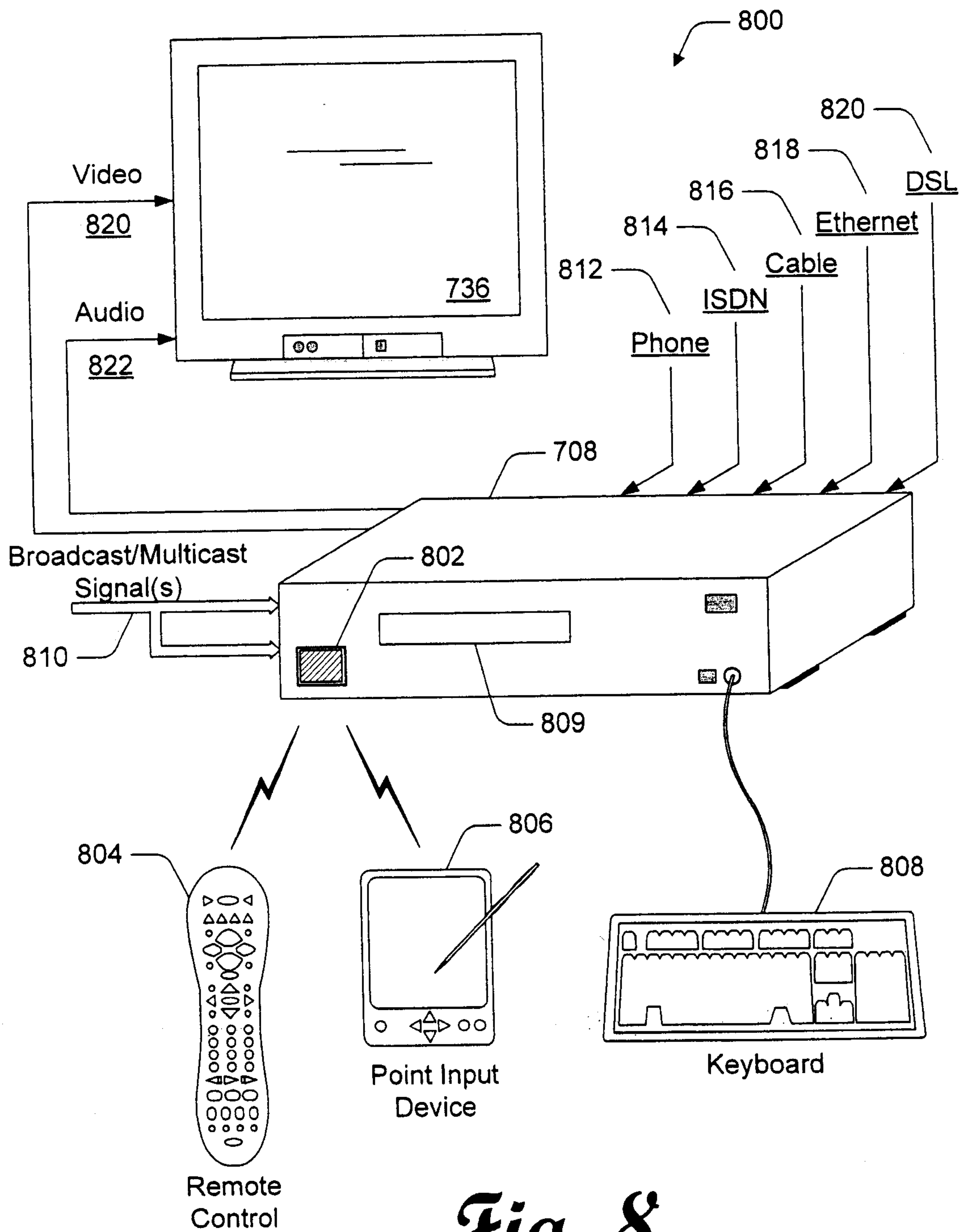


Fig. 8

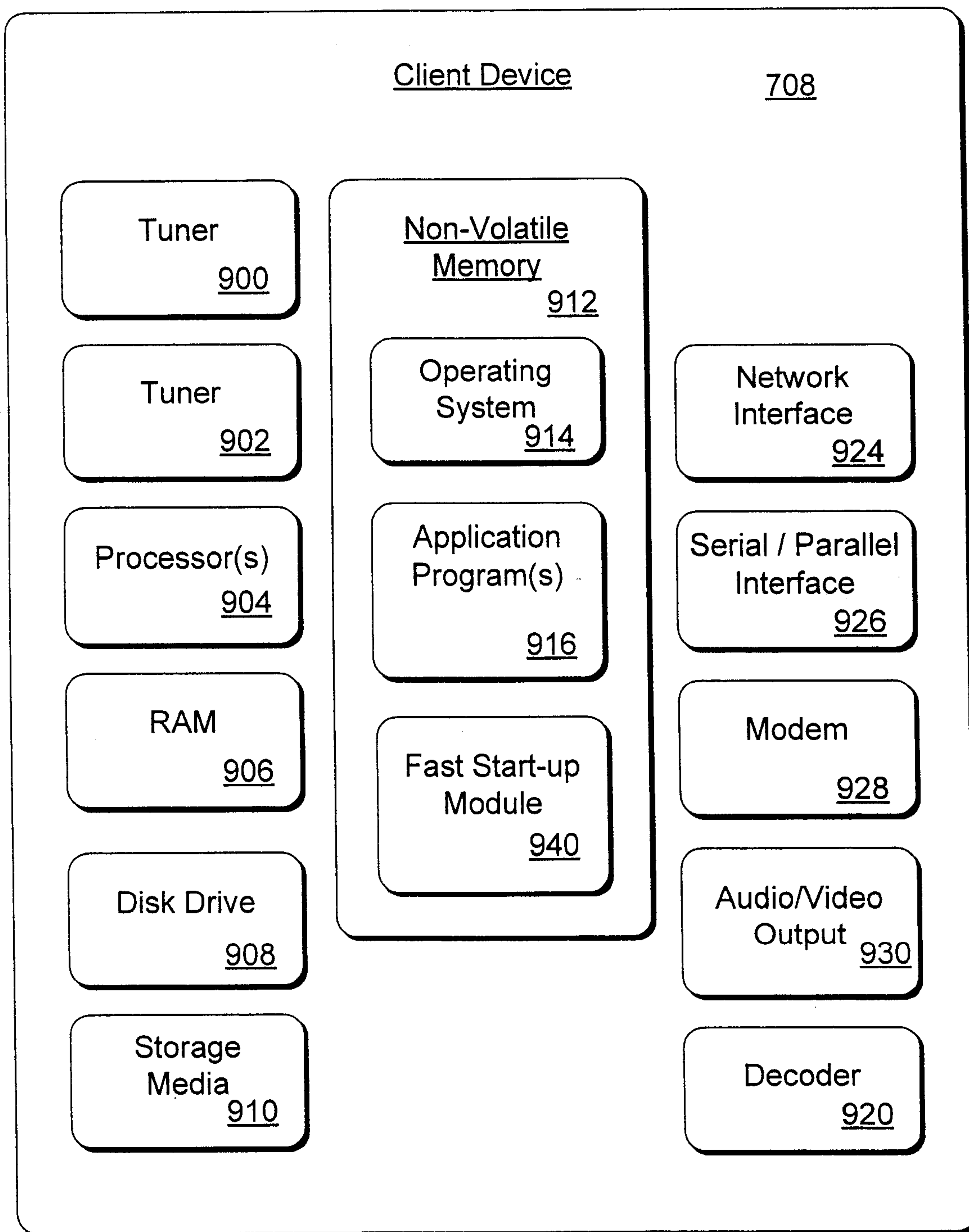


Fig. 9

