

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 734 551**

51 Int. Cl.:

**H04N 19/169** (2014.01)

**H04N 19/70** (2014.01)

**H04N 19/46** (2014.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.05.2014 PCT/US2014/040318**

87 Fecha y número de publicación internacional: **04.12.2014 WO14194243**

96 Fecha de presentación y número de la solicitud europea: **30.05.2014 E 14732799 (3)**

97 Fecha y número de publicación de la concesión europea: **10.04.2019 EP 3005700**

54 Título: **Paquetes de una sola unidad de capa de abstracción de red con número de orden de decodificación para la codificación de vídeo**

30 Prioridad:

**31.05.2013 US 201361829950 P**  
**29.05.2014 US 201414290537**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**10.12.2019**

73 Titular/es:

**QUALCOMM INCORPORATED (100.0%)**  
**International IP Administration, 5775 Morehouse Drive**  
**San Diego, California 92121-1714, US**

72 Inventor/es:

**COBAN, MUHAMMED ZEYD y**  
**WANG, YE-KUI**

74 Agente/Representante:

**FORTEA LAGUNA, Juan José**

ES 2 734 551 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

## DESCRIPCIÓN

Paquetes de una sola unidad de capa de abstracción de red con número de orden de decodificación para la codificación de vídeo

5 [0001] Esta solicitud reivindica el beneficio de la solicitud provisional de patente estadounidense Nº 61/829.950, presentada el 31 de mayo de 2013.

## Campo técnico

10 [0002] Esta divulgación se refiere al procesamiento de datos de vídeo.

## Antecedentes

15 [0003] Las capacidades del vídeo digital pueden incorporarse a una amplia gama de dispositivos, que incluye televisores digitales, sistemas de difusión digital directa, sistemas de difusión inalámbrica, asistentes digitales personales (PDA), ordenadores portátiles o de escritorio, ordenadores de tableta, lectores de libros electrónicos, cámaras digitales, dispositivos de grabación digital, reproductores de medios digitales, dispositivos de videojuegos, consolas de videojuegos, teléfonos celulares o de radio por satélite, los denominados «teléfonos inteligentes», dispositivos de videoconferencia, dispositivos de transmisión continua de vídeo y similares. Los dispositivos de vídeo digital implementan técnicas de compresión de vídeo, tales como las descritas en las normas definidas por MPEG-2, MPEG-4, ITU-T H.263, ITU-T H.264/MPEG-4, parte 10, codificación de vídeo avanzada (AVC), la norma de codificación de vídeo de alta eficacia (HEVC) actualmente en desarrollo y las ampliaciones de dichas normas. Los dispositivos de vídeo pueden transmitir, recibir, codificar, decodificar y/o almacenar información de vídeo digital con más eficacia implementando dichas técnicas de compresión de vídeo.

20 [0004] Las técnicas de compresión de vídeo realizan predicción espacial (intra-imagen) y/o predicción temporal (entre imágenes) para reducir o eliminar la redundancia intrínseca a las secuencias de vídeo. Para la codificación de vídeo basada en bloques, un fragmento de vídeo (por ejemplo, una trama de vídeo o una parte de una trama de vídeo) puede dividirse en bloques de vídeo, que también pueden denominarse bloques arbolados, unidades de codificación (CU) y/o nodos de codificación. Los bloques de vídeo en un fragmento intracodificado (I) de una imagen se codifican usando predicción espacial con respecto a muestras de referencia en bloques vecinos en la misma imagen. Los bloques de vídeo en un fragmento intercodificado (P o B) de una imagen pueden usar la predicción espacial con respecto a muestras de referencia en bloques vecinos en la misma imagen o la predicción temporal con respecto a muestras de referencia en otras imágenes de referencia. Las imágenes pueden denominarse tramas y las imágenes de referencia pueden denominarse tramas de referencia.

30 [0005] La predicción espacial o temporal da como resultado un bloque predictivo para un bloque a codificar. Los datos residuales representan diferencias de píxeles entre el bloque original a codificar y el bloque predictivo. Un bloque intercodificado se codifica de acuerdo a un vector de movimiento que apunta a un bloque de muestras de referencia que forman el bloque predictivo, y los datos residuales que indican la diferencia entre el bloque codificado y el bloque predictivo. Un bloque intracodificado se codifica de acuerdo a una modalidad de intracodificación y a los datos residuales. Para una mayor compresión, los datos residuales pueden transformarse desde el dominio del píxel a un dominio de transformación, dando como resultado coeficientes de transformación residuales, que a continuación se pueden cuantizar. Los coeficientes de transformación cuantizados, dispuestos inicialmente en una formación bidimensional, pueden explorarse con el fin de generar un vector unidimensional de coeficientes de transformación, y puede aplicarse codificación por entropía para lograr aún más compresión.

35 [0006] Los datos de vídeo pueden ser transmitidos y recibidos usando uno o más protocolos. Cada protocolo puede especificar diversos requisitos de contenido y formato para la transmisión y/o recepción de los datos cuando se utiliza el protocolo. Por ejemplo, algunos protocolos pueden dividir un flujo o un conjunto de datos en trozos para su transporte a través de una o más redes. En algunos protocolos, este procedimiento de división puede denominarse paquetización o entramado.

55 El documento D1 - Wenger Independent Y-K Wang Huawei Technologies T Schierl Fraunhofer HHI A Eleftheriadis-Vidyo S: "RTP Payload Format for Scalable Video Coding; draft-ietf-avt-rtp-svc-27.txt [Formato de carga útil del RTP para codificación de vídeo ajustable a escala; draft-ietf-avt-rtp-svc-27.txt]", del 2 de febrero de 2011, revela una unidad de NAL de Información de ajustabilidad a escala de contenido de carga útil (PACSI), que contiene un campo DONC optativo para indicar una cruz. Número de orden de decodificación entre sesiones (CS-DON) de una unidad de NAL no de PACSI, transmitida posteriormente.

60 El documento D2 - Carminati L et al: "Towards Scalable Audiovisual Broadcasting Networks: Overview of the Toscane Project [Hacia redes de difusión audiovisual ajustables a escala: Descripción general del Proyecto Toscane]", 9 de septiembre de 2009, divulga el uso de una modalidad de transmisión de una sola unidad de NAL, en la que el orden de transmisión de paquetes de una sola unidad de NAL es el mismo que el orden de decodificación de la unidad de NAL.

**SUMARIO**

5 **[0007]** Las técnicas de la presente divulgación proporcionan procedimientos y aparatos para procesar datos de vídeo que se envían y reciben mediante un protocolo de red, como el Protocolo de transporte en tiempo real (RTP). Más específicamente, las técnicas descritas en el presente documento proporcionan un formato de paquete de una sola unidad de NAL, que se puede utilizar con varios parámetros de transmisión y varias modalidades. La invención se estipula en las reivindicaciones adjuntas.

10 **[0008]** En un ejemplo de la divulgación, un procedimiento para procesar datos de vídeo en una carga útil del protocolo de transporte en tiempo real (RTP) incluye encapsular datos de vídeo en un paquete de una sola unidad de capa de abstracción de red (NAL) para una sesión del RTP, donde el paquete de una sola unidad de NAL contiene una sola unidad de NAL, y encapsula la información del número de orden de decodificación en el paquete de una sola unidad de NAL, en función de al menos uno entre: la sesión del RTP está en una modalidad de transmisión de múltiples flujos (MST), o un número máximo de unidades de NAL, que pueden preceder la unidad de NAL en un almacén temporal de desempaquetado en orden de recepción, y seguir a la unidad de NAL en orden de decodificación, es mayor que 0.

20 **[0009]** En otro ejemplo de la divulgación, un procedimiento de procesamiento de datos de vídeo en una carga útil del protocolo de transporte en tiempo real (RTP) incluye la desencapsulación de datos de vídeo que están encapsulados en un paquete de una sola unidad de capa de abstracción de red (NAL) para una sesión del RTP, donde el paquete de solo una unidad de NAL contiene una sola unidad de NAL y la desencapsulación de información del número de orden de decodificación que está encapsulada en el paquete de una sola unidad de NAL, basándose en al menos uno entre: la sesión del RTP está en una modalidad de transmisión de múltiples flujos (MST), o un número máximo de unidades de NAL que pueden preceder a la unidad de NAL en un almacén temporal de desempaquetado en orden de recepción y seguir a la unidad de NAL en el orden de decodificación, es mayor que 0.

30 **[0010]** En otro ejemplo de la divulgación, un aparato configurado para procesar datos de vídeo en una carga útil del protocolo de transporte en tiempo real (RTP) incluye una memoria configurada para almacenar datos de vídeo y un procesador configurado para: encapsular, dentro de una carga útil del protocolo de transporte en tiempo real (RTP), datos de vídeo en un paquete de una sola unidad de capa de abstracción de red (NAL), para una sesión del RTP, donde el paquete de una sola unidad de NAL contiene una sola unidad de NAL, y encapsular la información del número de orden de decodificación en el paquete de una sola unidad de NAL, basándose en al menos uno entre: la sesión del RTP está en una modalidad de transmisión de múltiples flujos (MST), o un número máximo de unidades de NAL que pueden preceder a la unidad de NAL en un almacén temporal de desempaquetado en orden de recepción y seguir a la unidad de NAL en orden de decodificación, es mayor que 0.

40 **[0011]** En otro ejemplo de la divulgación, un aparato configurado para procesar datos de vídeo en una carga útil del protocolo de transporte en tiempo real (RTP) incluye medios para encapsular datos de vídeo en un paquete de una sola unidad de capa de abstracción de red (NAL) para una sesión del RTP, donde el paquete de una sola unidad de NAL contiene una sola unidad de NAL y medios para encapsular la información del número de orden de decodificación en el paquete de una sola unidad de NAL, basándose en al menos uno entre: la sesión del RTP está en una modalidad de transmisión de múltiples flujos (MST), o un número máximo de unidades de NAL que pueden preceder a la unidad de NAL en un almacén temporal de desempaquetado en el orden de recepción y seguir a la unidad de NAL en el orden de decodificación, es mayor que 0.

50 **[0012]** Los detalles de uno o más ejemplos se estipulan en los dibujos adjuntos y en la siguiente descripción. Otras características, objetos y ventajas resultarán evidentes a partir de la descripción y de los dibujos, y a partir de las reivindicaciones.

**BREVE DESCRIPCIÓN DE LOS DIBUJOS**

**[0013]**

55 La figura 1 es un diagrama conceptual que ilustra un sistema ejemplar de codificación y decodificación de vídeo que puede utilizar las técnicas descritas en la presente divulgación.

La figura 2 es un diagrama conceptual que muestra la estructura de una cabecera de unidad de capa de abstracción de red (NAL) de la HEVC.

60 La figura 3 es un diagrama conceptual que muestra la estructura de un formato de carga útil del protocolo de transporte en tiempo real (RTP) para un paquete de agrupación.

65 La figura 4 es un diagrama conceptual que muestra la estructura de la primera unidad de agrupación en un paquete de agrupación.

La figura 5 es un diagrama conceptual que muestra la estructura de un formato de carga útil del RTP para un paquete de una sola unidad de NAL.

5 La figura 6 es un diagrama conceptual que muestra un ejemplo de un formato de carga útil del RTP para un paquete de una sola unidad de NAL, de acuerdo a las técnicas de la divulgación.

La figura 7 es un diagrama conceptual que muestra otro ejemplo de un formato de carga útil del RTP para un paquete de una sola unidad de NAL, de acuerdo a las técnicas de la divulgación.

10 La figura 8 es un diagrama conceptual que ilustra un codificador de vídeo ejemplar que puede implementar las técnicas descritas en esta divulgación.

La figura 9 es un diagrama de bloques que ilustra un decodificador de vídeo ejemplar que puede implementar las técnicas descritas en esta divulgación.

15 La figura 10 es un diagrama de bloques que ilustra un conjunto ejemplar de dispositivos que forman parte de una red.

20 La figura 11 es un diagrama de flujo que ilustra operaciones ejemplares para encapsular datos de vídeo en un formato de carga útil del RTP, de acuerdo a las técnicas de la divulgación.

La figura 12 es un diagrama de flujo que ilustra operaciones ejemplares para desencapsular datos de vídeo encapsulados en un formato de carga útil del RTP, de acuerdo a las técnicas de la divulgación.

## 25 DESCRIPCIÓN DETALLADA

[0014] Esta divulgación introduce varias técnicas y dispositivos para empaquetar datos de vídeo. En uno o más ejemplos, esta divulgación propone diseños mejorados de un formato de carga útil del protocolo de transporte en tiempo real (RTP) para el transporte de datos de vídeo. En particular, esta divulgación propone técnicas para señalar un número de orden de decodificación (DON) para paquetes del RTP de una sola unidad de capa de abstracción de red (NAL). Las técnicas anteriores para transmitir paquetes de una sola unidad de NAL eran incompatibles con ciertas modalidades de transmisión y ciertos parámetros de transmisión. En cambio, las técnicas anteriores requerían que las unidades de una sola NAL se transmitieran en un paquete de agrupación, lo que incurría en un aumento de la sobrecarga y una disminución del caudal. Al incluir la información de número de orden de decodificación en un paquete flexible de una sola unidad de NAL, las técnicas descritas en el presente documento pueden admitir una transmisión más eficaz de unidades de NAL individuales y habilitar el uso de paquetes de una sola unidad de NAL con varias modalidades de transmisión y varios parámetros de transmisión.

[0015] La figura 1 es un diagrama de bloques que ilustra un sistema ejemplar de procesamiento de vídeo 10 que puede utilizarse junto con las técnicas descritas en esta divulgación. El sistema 10 puede, por ejemplo, generar, procesar y transmitir datos de vídeo usando las técnicas del RTP descritas en esta divulgación. Como se muestra en la figura 1, el sistema 10 incluye un dispositivo de origen 12 que genera datos de vídeo codificados, a decodificar más tarde por parte de un dispositivo de destino 14. Los datos de vídeo codificados pueden encaminarse desde el dispositivo de origen 12 hasta el dispositivo de destino 14 mediante un elemento de red consciente de los medios (MANE) 29. El dispositivo de origen 12 y el dispositivo de destino 14 pueden comprender cualquiera entre una amplia gama de dispositivos, incluidos ordenadores de escritorio, ordenadores plegables (es decir, portátiles), ordenadores de tableta, decodificadores, equipos telefónicos de mano tales como los denominados teléfonos «inteligentes», los denominados paneles «inteligentes», televisores, cámaras, dispositivos de visualización, reproductores de medios digitales, consolas de videojuegos, dispositivos de transmisión de vídeo en tiempo real o similares. En algunos casos, el dispositivo de origen 12 y el dispositivo de destino 14 pueden estar equipados para la comunicación inalámbrica.

[0016] El sistema 10 puede funcionar de acuerdo a diferentes normas de codificación de vídeo, una norma patentada o cualquier otra forma de codificación de vistas múltiples. Por ejemplo, el codificador de vídeo 20 y el decodificador de vídeo 30 pueden funcionar de acuerdo a una norma de compresión de vídeo, tal como ITU-T H.261, ISO/IEC MPEG-1 Visual, ITU-T H.262 o ISO/IEC MPEG-2 Visual, ITU-T H.263, ISO/IEC MPEG-4 Visual e ITU-T H.264 (también conocida como ISO/IEC MPEG-4 AVC), incluyendo sus extensiones de codificación de vídeo ajustable a escala (SVC) y de codificación de vídeo de múltiples vistas (MVC). El reciente borrador conjunto disponible al público de la extensión de la MVC se describe en "Advanced video coding for generic audiovisual services [Codificación de vídeo avanzada para servicios audiovisuales genéricos]", Recomendación ITU-T H.264, marzo de 2010. Un borrador conjunto más reciente disponible al público de la extensión de la MVC se describe en "Codificación de vídeo avanzada para servicios audiovisuales genéricos", Recomendación ITU-T H.264, junio de 2011. A partir de enero de 2012 se ha aprobado un borrador conjunto actual de la extensión de la MVC.

[0017] Además, hay una nueva norma de codificación de vídeo, concretamente, la Codificación de Vídeo de Alta Eficacia (HEVC), que ha sido desarrollada por el Equipo de Colaboración Conjunta en Codificación de Vídeo (JCT-VC) del Grupo de Expertos en Codificación de Vídeo (VCEG) de la ITU-T y del Grupo de Expertos en Imágenes en

Movimiento (MPEG) de ISO/IEC. Un borrador de la norma HEVC, denominado "HEVC Working Draft 10 [Borrador 10 de trabajo de la HEVC]" o "WD10", se describe en el documento JCTVC-L1003v34, de Bross et al., titulado "High Efficiency video Coding (HEVC) Text Specification Draft 10 [Memoria descriptiva textual de la Codificación de Vídeo de Alta Eficacia (HEVC), Borrador 10]", Equipo de Colaboración Conjunta en Codificación de Vídeo (JCT-VC) de ITU-T SG16 WP3 e ISO/IEC JTC1/SC29/WG11, 12ª conferencia: Ginebra, Suiza, 14 al 23 de enero de 2013, que, a partir del 30 de abril de 2014 puede descargarse desde [http://phenix.int-evry.fr/jct/doc\\_end\\_user/documents/12\\_Geneva/wg11/JCTVC-L1003-v34.zip](http://phenix.int-evry.fr/jct/doc_end_user/documents/12_Geneva/wg11/JCTVC-L1003-v34.zip).

**[0018]** Para los fines de la descripción, el codificador de vídeo 20 y el decodificador de vídeo 30 se describen en el contexto de la HEVC o la norma H.264 y las extensiones de dichas normas. Sin embargo, las técnicas de esta divulgación no están limitadas a ninguna norma de codificación particular. Otros ejemplos de normas de compresión de vídeo incluyen MPEG-2 e ITU-T H.263. Las técnicas de codificación patentadas, tales como las denominadas On2 VP6/VP7/VP8, también pueden implementar una o más de las técnicas descritas en el presente documento. Las técnicas de esta divulgación son potencialmente aplicables a varias normas de codificación de vídeo, incluyendo la HEVC y otras.

**[0019]** El dispositivo de destino 14 puede recibir los datos de vídeo codificados a decodificar, mediante un enlace 16. El enlace 16 puede comprender cualquier tipo de medio o dispositivo capaz de desplazar los datos de vídeo codificados desde el dispositivo de origen 12 hasta el dispositivo de destino 14. En un ejemplo, el enlace 16 puede comprender un medio de comunicación para permitir que el dispositivo de origen 12 transmita datos de vídeo codificados directamente al dispositivo de destino 14 en tiempo real. Los datos de vídeo codificados se pueden modular de acuerdo a una norma de comunicación, tal como un protocolo de comunicación inalámbrica, y transmitir al dispositivo de destino 14. El medio de comunicación puede comprender cualquier medio de comunicación, inalámbrica o cableada, tal como un espectro de radiofrecuencia (RF) o una o más líneas físicas de transmisión. El medio de comunicación puede formar parte de una red basada en paquetes, tal como una red de área local, una red de área amplia o una red global tal como Internet. El medio de comunicación puede incluir encaminadores, conmutadores, estaciones base o cualquier otro equipo que pueda ser útil para facilitar la comunicación desde el dispositivo de origen 12 hasta el dispositivo de destino 14. El enlace 16 puede incluir uno o más MANE, tales como el MANE 29, que encaminan los datos de vídeo desde el dispositivo de origen 12 al dispositivo de destino 14.

**[0020]** De forma alternativa, los datos codificados pueden emitirse desde la interfaz de salida 22 hasta un dispositivo de almacenamiento 27. De forma similar, una interfaz de entrada puede acceder a los datos codificados procedentes del dispositivo de almacenamiento 27. El dispositivo de almacenamiento 27 puede incluir cualquiera entre varios medios de almacenamiento de datos de acceso distribuido o local, tales como una unidad de disco duro, discos Blu-ray, DVD, CD-ROM, una memoria flash, una memoria volátil o no volátil, o cualquier otro medio de almacenamiento digital adecuado para almacenar datos de vídeo codificados. En otro ejemplo, el dispositivo de almacenamiento 27 puede corresponder a un servidor de ficheros o a otro dispositivo de almacenamiento intermedio que pueda contener el vídeo codificado generado por el dispositivo de origen 12.

**[0021]** El dispositivo de destino 14 puede acceder a datos de vídeo almacenados desde el dispositivo de almacenamiento 27, mediante transmisión en tiempo real o descarga. El servidor de ficheros puede ser cualquier tipo de servidor capaz de almacenar datos de vídeo codificados y transmitir esos datos de vídeo codificados al dispositivo de destino 14. Los servidores ejemplares de ficheros incluyen un servidor de la Red (por ejemplo, para una sede de la Red), un servidor del FTP, dispositivos de almacenamiento conectados en red (NAS) o una unidad de disco local. El dispositivo de destino 14 puede acceder a los datos de vídeo codificados mediante cualquier conexión de datos estándar, incluida una conexión a Internet. Esto puede incluir un canal inalámbrico (por ejemplo, una conexión de Wi-Fi), una conexión cableada (por ejemplo, DSL, módem de cable, etc.) o una combinación de ambas que sea adecuada para acceder a datos de vídeo codificados almacenados en un servidor de ficheros. La transmisión de datos de vídeo codificados desde el dispositivo de almacenamiento 27 puede ser una transmisión en tiempo real, una transmisión de descarga o una combinación de ambas. Los datos de vídeo recuperados desde el dispositivo de almacenamiento 27 pueden encaminarse al dispositivo de destino 14 utilizando uno o más MANE, tales como el MANE 29.

**[0022]** Las técnicas de esta divulgación no están limitadas necesariamente a aplicaciones o configuraciones inalámbricas. Las técnicas se pueden aplicar a la codificación de vídeo como apoyo a cualquiera entre varias aplicaciones de multimedia, tales como difusiones de televisión por aire, transmisiones de televisión por cable, transmisiones de televisión por satélite, transmisiones de vídeo en tiempo real, por ejemplo, mediante Internet, codificación de vídeo digital para su almacenamiento en un medio de almacenamiento de datos, decodificación de vídeo digital almacenado en un medio de almacenamiento de datos, u otras aplicaciones. En algunos ejemplos, el sistema 10 puede estar configurado para prestar soporte a una transmisión de vídeo unidireccional o bidireccional, a fin de prestar soporte a aplicaciones tales como la transmisión de vídeo en tiempo real, la reproducción de vídeo, la difusión de vídeo y/o la videotelefonía.

**[0023]** En el ejemplo de la figura 1, el dispositivo de origen 12 incluye un origen de vídeo 18, un codificador de vídeo 20 y una interfaz de salida 22. En algunos casos, la interfaz de salida 22 puede incluir un modulador/demodulador (módem) y/o un transmisor. Por ejemplo, la interfaz de salida 22 puede incluir una unidad de paquetización del RTP, operable para encapsular datos en cargas útiles del RTP, de acuerdo a las técnicas descritas en el presente

documento. En el dispositivo de origen 12, el origen de vídeo 18 puede incluir un origen tal como un dispositivo de captura de vídeo, por ejemplo, una videocámara, un archivo de vídeo que contiene vídeo capturado previamente, una interfaz de alimentación de vídeo para recibir vídeo desde un proveedor de contenido de vídeo y/o un sistema de gráficos por ordenador para generar datos de gráficos por ordenador como el vídeo de origen, o una combinación de dichos orígenes. Como ejemplo, si el origen de vídeo 18 es una videocámara, el dispositivo de origen 12 y el dispositivo de destino 14 pueden formar los denominados teléfonos con cámara o videoteléfonos. Sin embargo, las técnicas descritas en esta divulgación pueden ser aplicables a la codificación de vídeo en general, y se pueden aplicar a aplicaciones inalámbricas y/o cableadas.

**[0024]** El dispositivo de origen 12 puede codificar el vídeo capturado, capturado previamente o generado por ordenador. Los datos de vídeo codificados se pueden transmitir directamente al dispositivo de destino 14 por medio de la interfaz de salida 22 del dispositivo de origen 12. Los datos de vídeo codificados, también o alternativamente, se pueden almacenar en el dispositivo de almacenamiento 27 para un posterior acceso por el dispositivo de destino 14 u otros dispositivos, para su decodificación y/o reproducción.

**[0025]** El dispositivo de destino 14 incluye una interfaz de entrada 28, un decodificador de vídeo 30 y un dispositivo de visualización 32. En algunos casos, la interfaz de entrada 28 puede incluir un receptor y/o un módem. Por ejemplo, la interfaz de entrada 28 puede incluir una unidad de desempaqueado del RTP, operable para desencapsular datos encapsulados en cargas útiles del RTP, de acuerdo a las técnicas descritas en el presente documento. La interfaz de entrada 28 del dispositivo de destino 14 recibe los datos de vídeo codificados por el enlace 16. Los datos de vídeo codificados, comunicados por el enlace 16, o proporcionados en el dispositivo de almacenamiento 27, pueden incluir una diversidad de elementos sintácticos generados por el codificador de vídeo 20 para su uso por un decodificador de vídeo, tal como el decodificador de vídeo 30, en la decodificación de los datos de vídeo. Dichos elementos sintácticos se pueden incluir con los datos de vídeo codificados transmitidos en un medio de comunicación, almacenados en un medio de almacenamiento o almacenados en un servidor de ficheros.

**[0026]** El dispositivo de visualización 32 puede estar integrado en, o ser externo a, el dispositivo de destino 14. En algunos ejemplos, el dispositivo de destino 14 puede incluir un dispositivo de visualización integrado y también estar configurado para interconectarse con un dispositivo de visualización externo. En otros ejemplos, el dispositivo de destino 14 puede ser un dispositivo de visualización. En general, el dispositivo de visualización 32 exhibe los datos de vídeo decodificados ante un usuario, y puede comprender cualquiera entre varios dispositivos de visualización, tales como una pantalla de cristal líquido (LCD), una pantalla de plasma, una pantalla de diodos orgánicos emisores de luz (OLED) u otro tipo de dispositivo de visualización.

**[0027]** Aunque no se muestra en la figura 1, en algunos aspectos, tanto el codificador de vídeo 20 como el decodificador de vídeo 30 pueden estar integrados con un codificador y decodificador de audio, y pueden incluir unidades MUX-DEMUX adecuadas, u otro tipo de hardware y software, para ocuparse de la codificación tanto de audio como de vídeo en un flujo de datos común o en flujos de datos distintos. Si procede, en algunos ejemplos, las unidades MUX-DEMUX pueden ajustarse al protocolo de multiplexador ITU H.223 o a otros protocolos, tales como el protocolo de datagramas de usuario (UDP).

**[0028]** Tanto el codificador de vídeo 20 como el decodificador de vídeo 30 pueden implementarse como cualquiera entre varios circuitos codificadores adecuados, tales como uno o más microprocesadores, procesadores de señales digitales (DSP), circuitos integrados específicos de la aplicación (ASIC), formaciones de compuertas programables in situ (FPGA), lógica discreta, software, hardware, firmware o cualquier combinación de los mismos. Cuando las técnicas se implementan parcialmente en software, un dispositivo puede almacenar instrucciones para el software en un medio adecuado no transitorio legible por ordenador, y ejecutar las instrucciones en hardware usando uno o más procesadores para realizar las técnicas de esta divulgación. Cada uno entre el codificador de vídeo 20 y el decodificador de vídeo 30 se puede incluir en uno o más codificadores o decodificadores, cualquiera de los cuales se puede integrar como parte de un codificador/decodificador (CÓDEC) combinado en un dispositivo respectivo.

**[0029]** Los esfuerzos de normalización de la HEVC se basaron en un modelo en evolución de un dispositivo de codificación de vídeo denominado modelo de prueba de la HEVC (HM). El HM supone varias capacidades adicionales de los dispositivos de codificación de vídeo respecto a los dispositivos existentes, de acuerdo a, por ejemplo, la norma ITU-T H.264/AVC. Por ejemplo, mientras que la norma H.264 proporciona nueve modalidades de codificación mediante intrapredicción, el HM puede proporcionar hasta treinta y tres modalidades de codificación mediante intrapredicción.

**[0030]** En general, el modelo de explotación del HM describe que una trama o imagen de vídeo puede dividirse en una secuencia de bloques arbolados o unidades máximas de codificación (LCU), que incluyen tanto muestras de luma como de croma. Un bloque arbolado tiene un fin similar al de un macrobloque de la norma H.264. Un fragmento incluye un determinado número de bloques arbolados consecutivos en orden de codificación. Una trama o imagen de vídeo puede dividirse en uno o más fragmentos. Cada bloque arbolado puede dividirse en unidades de codificación (CU) de acuerdo a un árbol cuádruple. Por ejemplo, un bloque arbolado, como un nodo raíz del árbol cuádruple, puede dividirse en cuatro nodos hijos, y cada nodo hijo puede, a su vez, ser un nodo padre y dividirse en otros cuatro nodos hijos. Un nodo hijo final, no dividido, como un nodo hoja del árbol cuádruple, comprende un nodo de codificación, es decir, un

bloque de vídeo codificado. Los datos sintácticos asociados a un flujo de bits codificado pueden definir un número máximo de veces que puede dividirse un bloque arbolado, y también pueden definir un tamaño mínimo de los nodos de codificación.

5 **[0031]** Una CU incluye un nodo de codificación y unidades de predicción (PU) y unidades de transformación (TU) asociadas al nodo de codificación. Un tamaño de la CU corresponde a un tamaño del nodo de codificación y debe ser de forma cuadrada. El tamaño de la CU puede variar desde 8x8 píxeles hasta el tamaño del bloque arbolado, con un máximo de 64x64 píxeles o más. Cada CU puede contener una o más PU y una o más TU. Los datos sintácticos asociados a una CU pueden describir, por ejemplo, la división de la CU en una o más PU. Las modalidades de división pueden diferir entre si la CU está codificada en modalidad de salto o directa, codificada en modalidad de intrapredicción o codificada en modalidad de interpredicción. Las PU pueden dividirse para tener forma no cuadrada. Los datos sintácticos asociados a una CU también pueden describir, por ejemplo, la división de la CU en una o más TU de acuerdo a un árbol cuádruple. Una TU puede tener forma cuadrada o no cuadrada.

15 **[0032]** La norma HEVC admite transformaciones de acuerdo a las TU, que pueden ser diferentes para diferentes CU. El tamaño de las TU se basa habitualmente en el tamaño de las PU dentro de una CU dada, definida para una LCU dividida, aunque puede que no sea siempre así. Las TU son habitualmente del mismo tamaño o de un tamaño más pequeño que las PU. En algunos ejemplos, las muestras residuales correspondientes a una CU pueden subdividirse en unidades más pequeñas utilizando una estructura de árbol cuádruple conocida como «árbol cuádruple residual» (RQT). Los nodos hojas del RQT pueden denominarse unidades de transformación (TU). Los valores de diferencias de píxeles, asociados a las TU, pueden transformarse para generar coeficientes de transformación, que pueden cuantizarse.

25 **[0033]** En general, una PU incluye datos relacionados con el proceso de predicción. Por ejemplo, cuando la PU se codifica en intramodalidad, la PU puede incluir datos que describen una modalidad de intrapredicción para la PU. En otro ejemplo, cuando la PU se codifica en intermodalidad, la PU puede incluir datos que definen un vector de movimiento para la PU. Los datos que definen el vector de movimiento para una PU pueden describir, por ejemplo, un componente horizontal del vector de movimiento, un componente vertical del vector de movimiento, una resolución para el vector de movimiento (por ejemplo, precisión de un cuarto de píxel o precisión de un octavo de píxel), una imagen de referencia a la que apunta el vector de movimiento y/o una lista de imágenes de referencia (por ejemplo, Lista 0, Lista 1 o Lista C) para el vector de movimiento.

35 **[0034]** En general, se usa una TU para los procesos de transformación y cuantización. Una CU determinada que tiene una o más PU también puede incluir una o más unidades de transformación (TU). Tras la predicción, el codificador de vídeo 20 puede calcular valores residuales correspondientes a la PU. Los valores residuales comprenden valores de diferencias de píxeles que pueden ser transformados en coeficientes de transformación, cuantizados y explorados usando las TU, para generar coeficientes de transformación en serie para la codificación por entropía. Esta divulgación usa habitualmente el término «bloque de vídeo» para referirse a un nodo de codificación de una CU. En algunos casos específicos, esta divulgación también puede usar el término «bloque de vídeo» para referirse a un bloque arbolado, es decir, una LCU o una CU, que incluye un nodo de codificación y unas PU y TU.

45 **[0035]** Una secuencia de vídeo incluye habitualmente una serie de tramas o imágenes de vídeo. Un grupo de imágenes (GOP) comprende, en general, una serie de una o más de las imágenes de vídeo. Un GOP puede incluir datos sintácticos, en una cabecera o en otro sitio, que describen una serie de imágenes incluidas en el GOP. Cada fragmento de una imagen puede incluir datos sintácticos de fragmento que describen una modalidad de codificación para el fragmento respectivo. El codificador de vídeo 20 actúa habitualmente sobre bloques de vídeo dentro de fragmentos de vídeo individuales con el fin de codificar los datos de vídeo. Un bloque de vídeo puede corresponder a un nodo de codificación dentro de una CU. Los bloques de vídeo pueden tener tamaños fijos o variables y pueden diferir en tamaño de acuerdo a una norma de codificación especificada.

50 **[0036]** En un ejemplo, el HM presta soporte a la predicción en diversos tamaños de PU. Si se supone que el tamaño de una CU particular es 2Nx2N, el HM presta soporte a la intrapredicción en tamaños de PU de 2Nx2N o NxN, y a la interpredicción en tamaños de PU simétricas de 2Nx2N, 2NxN, Nx2N o NxN. El HM también presta soporte a la división asimétrica para la interpredicción en tamaños de PU de 2NxnU, 2NxnD, nLx2N y nRx2N. En la división asimétrica, una dirección de una CU no está dividida, mientras que la otra dirección está dividida en un 25% y un 75%. La parte de la CU correspondiente a la división del 25% está indicada por una «n» seguida por una indicación de «arriba», «abajo», «izquierda» o «derecha». Así pues, por ejemplo, «2NxnU» se refiere a una CU de tamaño 2Nx2N que está dividida horizontalmente, con una PU de tamaño 2Nx0,5N encima y una PU de tamaño 2Nx1,5N debajo.

60 **[0037]** En la presente divulgación, «NxN» y «N por N» pueden usarse indistintamente para hacer referencia a las dimensiones de píxeles de un bloque de vídeo en términos de dimensiones verticales y horizontales, por ejemplo, 16x16 píxeles o 16 por 16 píxeles. En general, un bloque de tamaño 16x16 tendrá 16 píxeles en una dirección vertical (y = 16) y 16 píxeles en una dirección horizontal (x = 16). Asimismo, un bloque de tamaño NxN tiene, en general, N píxeles en una dirección vertical y N píxeles en una dirección horizontal, donde N representa un valor entero no negativo. Los píxeles de un bloque se pueden disponer en filas y columnas. Además, no es necesario que los bloques tengan necesariamente el mismo número de píxeles en la dirección horizontal que en la dirección vertical. Por ejemplo,

los bloques pueden comprender  $N \times M$  píxeles, donde  $M$  no es necesariamente igual a  $N$ .

**[0038]** Tras la codificación intrapredictiva o interpredictiva, usando las PU de una CU, el codificador de vídeo 20 puede calcular datos residuales para las TU de la CU. Las PU pueden comprender datos de píxeles en el dominio espacial (también denominado dominio del píxel) y las TU pueden comprender coeficientes en el dominio de la transformación tras la aplicación de una transformación, por ejemplo, una transformación discreta de coseno (DCT), una transformación de enteros, una transformación de ondículas o una transformación conceptualmente similar, a los datos de vídeo residuales. Los datos residuales pueden corresponder a diferencias de píxeles entre píxeles de la imagen no codificada y valores de predicción correspondientes a las PU. El codificador de vídeo 20 puede formar las TU incluyendo los datos residuales para la CU y, a continuación, transformar las TU para generar coeficientes de transformación para la CU.

**[0039]** Tras cualquier transformación para generar coeficientes de transformación, el codificador de vídeo 20 puede realizar la cuantización de los coeficientes de transformación. La cuantización se refiere, en general, a un proceso en el que coeficientes de transformación se cuantizan para reducir posiblemente la cantidad de datos usados para representar los coeficientes, proporcionando compresión adicional. El proceso de cuantización puede reducir la profundidad de bits asociada a algunos, o a la totalidad, de los coeficientes. Por ejemplo, un valor de  $n$  bits puede redondearse a la baja hasta un valor de  $m$  bits durante la cuantización, donde  $n$  es mayor que  $m$ .

**[0040]** En algunos ejemplos, el codificador de vídeo 20 puede utilizar un orden de exploración predefinido para explorar los coeficientes de transformación cuantizados, para producir un vector serializado que se pueda codificar por entropía. En otros ejemplos, el codificador de vídeo 20 puede realizar una exploración adaptativa. Después de explorar los coeficientes de transformación cuantizados para formar un vector unidimensional, el codificador de vídeo 20 puede realizar la codificación por entropía del vector unidimensional, por ejemplo, de acuerdo a la codificación de longitud variable adaptativa al contexto (CAVLC), la codificación aritmética binaria adaptativa al contexto (CABAC), la codificación aritmética binaria adaptativa al contexto basada en la sintaxis (SBAC), la codificación por entropía de división de intervalos de probabilidad (PIPE) u otra metodología de codificación por entropía. El codificador de vídeo 20 también puede realizar la codificación por entropía de los elementos sintácticos asociados a los datos de vídeo codificados, para su uso por el decodificador de vídeo 30 en la decodificación de los datos de vídeo.

**[0041]** Para realizar la codificación CABAC, el codificador de vídeo 20 puede asignar un contexto, dentro de un modelo de contexto, a un símbolo a transmitir. El contexto se puede referir, por ejemplo, a si los valores contiguos del símbolo son distintos de cero o no. Para realizar la CAVLC, el codificador de vídeo 20 puede seleccionar un código de longitud variable para un símbolo a transmitir. Las palabras de código en la VLC se pueden construir de modo que los códigos relativamente más cortos correspondan a símbolos más probables, mientras que los códigos más largos correspondan a símbolos menos probables. De esta forma, el uso de la VLC puede lograr un ahorro en bits con respecto, por ejemplo, al uso de palabras de código de igual longitud para cada símbolo a transmitir. La determinación de la probabilidad se puede basar en un contexto asignado al símbolo.

**[0042]** Los datos de vídeo codificados, tales como los datos de vídeo codificados de acuerdo a la norma HEVC u otras normas, pueden transmitirse entre dos dispositivos (por ejemplo, entre el dispositivo de origen 12 y el dispositivo de destino 14) usando varios procedimientos. Por ejemplo, los datos de vídeo pueden transmitirse a través de una o más redes utilizando varios protocolos de red. Algunos protocolos pueden especificar varios parámetros y/o reglas para la transmisión. Por ejemplo, algunos protocolos pueden procesar datos antes de la transmisión a través de una red y reprocesar los datos al recibirlos. En algunos ejemplos, el procesamiento de datos (por ejemplo, datos de vídeo codificados) puede incluir la separación de los datos en varios trozos (por ejemplo, empaquetando o entramando los datos). Un ejemplo de un protocolo para transmitir datos de vídeo para aplicaciones de transmisión continua en tiempo real es el protocolo de transporte en tiempo real (RTP).

**[0043]** El RTP es un protocolo de transporte que se especifica en el documento IETF RFC 3550, que, a partir del 30 de abril de 2014, está disponible en <http://www.ietf.org/rfc/rfc3550.txt>. En general, el RTP define un formato de paquete estandarizado para suministrar audio y/o vídeo por redes del IP. El RTP se puede utilizar para proporcionar diversos medios de transmisión continua, tales como servicios de telefonía, teleconferencias de vídeo, servicios de televisión u otros.

**[0044]** Para el transporte de datos de vídeo codificados de acuerdo a un códec de vídeo según el RTP, se necesita especificar un formato de carga útil del RTP para el códec de vídeo. Por ejemplo, el documento RFC 6184 (que a partir del 30 de abril de 2014 está disponible en <http://www.ietf.org/rfc/rfc6184.txt>) especifica el formato de carga útil del RTP para vídeo de norma H.264 y el documento RFC 6190 (que a partir del 30 de abril de 2014 está disponible en <http://www.ietf.org/rfc/rfc6190.txt>) especifica el formato de carga útil del RTP para vídeo de norma SVC. El documento RFC 4629 especifica el formato de carga útil del RTP para la norma UIT-T Rec. H.263.

**[0045]** Además, hay una nueva especificación de carga útil del RTP, a saber, el Formato de carga útil del RTP para la codificación de vídeo de alta eficacia (HEVC), que está siendo desarrollado por la Fuerza de tareas de ingeniería de Internet (IETF), Grupo de trabajo de carga útil de transporte de audio/vídeo. Un borrador reciente del formato de carga útil del RTP para el vídeo de norma HEVC está disponible, a partir del 30 de abril de 2014, en



<http://www.ietf.org/id/draft-ietf-pay-load-rtp-h265-02>.

**[0046]** El más reciente diseño del formato de carga útil del RTP de la HEVC, incluidas las mejoras descritas en la Solicitud Provisional Estadounidense N° 61/806.705, presentada el 29 de marzo de 2103, permite la transmisión de un flujo de bits de la HEVC en una sola sesión del RTP (por ejemplo, un solo flujo del RTP) o múltiples sesiones del RTP (por ejemplo, múltiples flujos del RTP). Un flujo del RTP puede ser una secuencia de paquetes del RTP transportados en una sola sesión del RTP. Una sesión del RTP puede corresponder a una dirección del IP y un par de puertos para recibir datos del RTP y del protocolo de control del RTP (RTCP). El RTCP generalmente puede proporcionar estadísticas fuera de banda e información de control para un flujo asociado del RTP.

**[0047]** Algunos conceptos y principios de funcionamiento del Formato de carga útil del RTP de la HEVC se heredan del documento RFC 6190 y siguen un diseño similar. Si solo se utiliza una sesión del RTP (por ejemplo, un flujo del RTP) para la transmisión del flujo de bits de la HEVC, la modalidad de transmisión se denomina transmisión de sesión única (SST) (o de flujo único); de lo contrario (por ejemplo, si se usa más de una sesión del RTP para la transmisión del flujo de bits de la HEVC), la modalidad de transmisión se denomina transmisión de múltiples sesiones (MST). La SST se utiliza habitualmente para escenarios de unidifusión punto a punto, mientras que la MST se utiliza para escenarios de multidifusión punto a multipunto, donde diferentes receptores requieren diferentes puntos de operación del mismo flujo de bits de la HEVC, para mejorar la eficacia de utilización del ancho de banda. La modalidad de transmisión, ya sea que se use la SST o la MST, se señala mediante el parámetro de tipo de medios, tx-mode, que se puede expresar como un parámetro del Protocolo de Descripción de Sesión (por ejemplo, durante la configuración de la sesión del RTP).

**[0048]** Otro parámetro que se puede especificar como un parámetro del SDP es el parámetro sprop-depack-buf-nalus. Sprop-depack-buf-nalus es un parámetro de sesión que especifica el número máximo de unidades de NAL que pueden preceder a una unidad de NAL en el almacén temporal de desempaqueado (por ejemplo, el almacén temporal del receptor del RTP) en el orden de recepción y seguir a la unidad de NAL en el orden de decodificación. El orden de decodificación, en general, puede indicar un orden en el cual las unidades de NAL han de ser decodificadas por un decodificador de vídeo. En consecuencia, un decodificador de vídeo y/o una unidad de desempaqueado pueden utilizar el orden de decodificación de las unidades de NAL recibidas para determinar el orden en el cual procesar las unidades de NAL. Por lo tanto, el valor de sprop-depack-buf-nalus puede indicar el número máximo de unidades de NAL que pueden transmitirse y/o recibirse fuera del orden de decodificación.

**[0049]** En un ejemplo, el valor del parámetro sprop-depack-buf-nalus es un entero en el intervalo de 0 a 32767, inclusive. Cuando no está presente, se deduce que el valor del parámetro sprop-depack-buf-nalus es igual a 0. Cuando la sesión del RTP depende de otras una o más sesiones del RTP (en este caso la modalidad de transmisión es igual a "MST"), el valor de parámetro sprop-depack-buf-nalus será mayor que 0. Un valor mayor que cero para el parámetro sprop-depack-buf-nalus indica que la sesión permite la paquetización intercalada. En otras palabras, cuando se usan múltiples flujos para transmitir unidades de datos, un almacén temporal receptor y/o un almacén temporal de desempaqueado pueden ser capaces de gestionar la paquetización intercalada (por ejemplo, gestionar unidades de datos transmitidas y/o recibidas fuera del orden de decodificación).

**[0050]** Los paquetes del RTP se transmiten en una sesión para transportar información (por ejemplo, datos de vídeo codificados) desde un dispositivo a otro. Un paquete del RTP incluye una cabecera del RTP y una carga útil del RTP. La cabecera del RTP incluye un campo Tipo de carga útil para especificar un identificador de carga útil. El identificador de carga útil indica el formato de la correspondiente carga útil del RTP. Según lo definido por la norma RTP, los identificadores de carga útil 96 a 127 están reservados para cargas útiles del RTP definidas dinámicamente durante una sesión. Es decir, un valor de identificador de carga útil de 96 a 127 puede correlacionar cargas útiles del RTP con un formato (o perfil) especificado durante el transcurso de la correspondiente sesión del RTP. En algunos ejemplos, un formato especificado para las cargas útiles del RTP de una sesión puede definirse usando un parámetro del SDP. Por ejemplo, un parámetro del SDP puede especificar que, para una sesión en particular, un valor de identificador de carga útil de 98 indica un perfil de la HEVC para las cargas útiles del RTP. Por lo tanto, los paquetes del RTP enviados durante la sesión pueden incluir cargas útiles del RTP que contienen datos de vídeo codificados utilizando la norma HEVC. Como tal, las cargas útiles del RTP pueden incluir unidades de NAL.

**[0051]** En un ejemplo de un perfil de la HEVC para cargas útiles del RTP, los dos primeros octetos de la carga útil del RTP pueden representar la cabecera de la carga útil del RTP. Para algunas cargas útiles del RTP que cumplen el formato de carga útil del RTP para la HEVC, la cabecera de carga útil del RTP consta de los mismos campos que la cabecera de la unidad de NAL para la HEVC.

**[0052]** La figura 2 es un diagrama conceptual que muestra la estructura de una cabecera de unidad de NAL de la HEVC. En general, la HEVC mantiene el concepto de unidad de NAL de la norma H.264, con unas pocas modificaciones. La semántica de los campos en la cabecera de la unidad de NAL es como se especifica en el documento HEVC WD10, y se describen brevemente a continuación por comodidad. Además del nombre y tamaño de cada campo, también se proporciona el correspondiente nombre del elemento sintáctico en el documento HEVC WD10. Con fines de descripción, los datos de carga útil de una unidad de NAL se refieren en este documento a la parte de la unidad de NAL que excluye la cabecera de la unidad de NAL. Es decir, una unidad de NAL puede constar

de una cabecera de unidad de NAL (por ejemplo, los octetos 1 y 2 de la unidad de NAL) y una carga útil de unidad de NAL (por ejemplo, los octetos 3 a N de la unidad de NAL).

5 **[0053]** El elemento sintáctico F, como se muestra en el ejemplo de la figura 1, es un bit único y se conoce como el prohibido\_cero\_bit. De acuerdo al documento HEVC WD10, F tiene un valor de cero. Es decir, el documento HEVC WD10 especifica que un valor de 1 para el elemento sintáctico F constituye una infracción de sintaxis. La inclusión de este bit en la cabecera de la unidad de NAL es para permitir el transporte de vídeo de norma HEVC por sistemas de transporte MPEG-2 (por ejemplo, para evitar las emulaciones de código de inicio).

10 **[0054]** Como se muestra en el ejemplo de la figura 2, una cabecera de unidad de NAL también incluye el elemento sintáctico Tipo. El elemento sintáctico Tipo tiene una longitud de 6 bits y se conoce como tipo\_unidad\_nal. Este campo especifica el tipo de unidad de NAL según lo definido en la Tabla 7-1 del documento HEVC WD10. Por una referencia de todos los tipos de unidades de NAL definidos actualmente y su semántica, remítase a la Sección 7.4.1 en el documento HEVC WD10.

15 **[0055]** El elemento sintáctico IdentificadorCapa, como se muestra en el ejemplo de la figura 2, también se incluye en la cabecera de la unidad de NAL. El elemento sintáctico IdentificadorCapa tiene una longitud de 6 bits y se conoce como nuh\_id\_capa. Actualmente, el documento HEVC WD10 especifica que el IdentificadorCapa debería ser igual a un valor de cero. En futuras extensiones de codificación de vídeo, ajustables a escala o tridimensionales, de la norma HEVC, el elemento sintáctico IdentificadorCapa se puede usar para identificar capas adicionales que pueden estar presentes en la secuencia de vídeo codificada, tales como una capa espacial ajustable a escala, una capa de calidad ajustable a escala, una vista de textura o una vista en profundidad.

20 **[0056]** Como se muestra en el ejemplo de la figura 2, la cabecera de la unidad de NAL también incluye el elemento sintáctico TID. El elemento sintáctico TID tiene una longitud de 3 bits y se conoce como nuh\_id\_temporal\_más1. El elemento sintáctico TID especifica el identificador temporal de la unidad de NAL más 1. El valor de IDTemporal es por tanto igual a TID menos 1. No se permite un valor de TID de 0 en el documento HEVC WD10, para garantizar que haya al menos un bit en la cabecera de la unidad de NAL igual a 1, para evitar emulaciones de código de inicio en la cabecera de la unidad de NAL.

25 **[0057]** En la especificación de carga útil de la HEVC para el RTP, se especifican cuatro tipos diferentes de estructuras de carga útil del RTP. Un receptor puede identificar el tipo de una carga útil del RTP mediante el campo Tipo en la cabecera de carga útil del RTP.

30 **[0058]** Las cuatro estructuras diferentes de carga útil del RTP para la HEVC son las siguientes:

35 

- Paquete de una sola unidad de NAL: Un paquete de una sola unidad de NAL contiene una sola unidad de NAL (por ejemplo, una cabecera de unidad de NAL y datos de carga útil de la unidad de NAL) en la carga útil del RTP. Anteriormente, la cabecera de unidad de NAL de la unidad de NAL también serviría como la cabecera de carga útil del RTP. Es decir, una carga útil del RTP que consiste en un paquete de una sola unidad de NAL no incluía una cabecera de carga útil del RTP y en cambio confiaba en que la cabecera de la unidad de NAL sirviera como cabecera de carga útil del RTP.

40 

- Paquete de agrupación (AP): Anteriormente, un AP contenía una o más unidades de NAL en la carga útil del RTP. Las unidades de NAL de un AP provienen de dentro de una unidad de acceso. Según lo definido por el documento HEVC WD10, una unidad de acceso es un conjunto de unidades de NAL que están asociadas entre sí de acuerdo a una regla de clasificación especificada, son consecutivas en el orden de decodificación y contienen exactamente una imagen codificada.

45 

- Unidad de fragmentación (FU): Una FU contiene un subconjunto de una sola unidad de NAL.
- Paquete del RTP que lleva Información de contenido de carga útil (PACI): Un paquete del RTP que lleva PACI contiene un cabecera de carga útil del RTP (que difiere de otras cabeceras de carga útil por su eficacia), una Estructura de extensión de cabecera de carga útil (PHES) y una carga útil de PACI.

50 **[0059]** Anteriormente, se especificaba la siguiente regla de paquetización: Cuando la modalidad de transmisión (por ejemplo, la modalidad tx) es igual a "MST" o el valor del parámetro sprop-depack-buf-nalus es mayor que 0, no se utilizan paquetes de una sola unidad de NAL. En otras palabras, cuando los datos del RTP se estaban recibiendo mediante una MST y/o cuando se permitía que un almacén temporal de desempaqueado recibiera paquetes del RTP fuera de orden, no estaban permitidas las unidades de NAL empaquetadas en cargas útiles del RTP sin un correspondiente número de orden de decodificación, y los anteriores formatos de paquetes del RTP para un paquete de una sola unidad de NAL no contenían información de número de orden de decodificación. Sin el número de orden de decodificación, las unidades de NAL no pueden ser repuestas en el orden correcto por el receptor del RTP. Si una sola unidad de NAL debía transmitirse en modalidad MST o cuando se especificaba un almacén temporal de desempaqueado (por ejemplo, mediante el parámetro sprop-depack-buf-nalus), se usaba un AP para encapsular la única unidad de NAL en un paquete del RTP. Sin embargo, encapsular una sola unidad de NAL en un AP da como resultado un aumento de la sobrecarga y una disminución del ancho de banda porque un AP incluye cuatro octetos

de información (a saber, un campo Tamaño de unidad de NAL y una cabecera de carga útil del RTP) que no son necesarios al transmitir una sola unidad de NAL.

**[0060]** La figura 3 es un diagrama conceptual que muestra la estructura de un formato de carga útil del protocolo de transporte en tiempo real (RTP) para un paquete de agrupación. Los paquetes de agrupación (AP) permiten la reducción de la sobrecarga de empaquetado para unidades de NAL pequeñas, tales como la mayoría de las unidades de NAL no de VCL (capa de codificación de vídeo), que tienen a menudo solo unos pocos octetos de tamaño.

**[0061]** Como se muestra en el ejemplo de la figura 3, un AP incluye un cabecera de carga útil del RTP, unidades de agrupación y relleno optativo del RTP. La cabecera de carga útil del RTP de un AP sigue el mismo formato que una cabecera de unidad de NAL, como se describe en la figura 2. Es decir, la cabecera de carga útil del RTP de un AP incluye el campo F, el campo Tipo, el campo IdentificadorCapa y el campo TID. En la cabecera de carga útil del RTP de un AP, el bit F es igual a 0 si el bit F de cada unidad de NAL agrupada en el AP es igual a cero. De lo contrario, el bit F es igual a 1. El valor del campo Tipo en la cabecera de carga útil del RTP de un AP es igual a 48. El valor del campo IdentificadorCapa en la cabecera de la carga útil del RTP de un AP es igual al valor más bajo de IdentificadorCapa entre todas las unidades de NAL agrupadas en el AP. El valor del campo TID en la cabecera de carga útil del RTP de un AP es igual al valor más bajo del campo TID entre todas las unidades de NAL agrupadas.

**[0062]** Un AP agrupa unidades de NAL dentro de una unidad de acceso. Es decir, un AP puede incluir una o más unidades de NAL de la misma unidad de acceso. Cada unidad de NAL a transportar en un AP está encapsulada en una unidad de agrupación. Las unidades de NAL agrupadas en un AP están en orden de decodificación de las unidades de NAL. Un AP puede llevar tantas unidades de agrupación como sea necesario.

**[0063]** La figura 4 es un diagrama conceptual que muestra la estructura de la primera unidad de agrupación en un paquete de agrupación. La primera unidad de agrupación en un AP incluye un campo optativo de Número de Orden de Decodificación Inferior (DONL) de 16 bits (en orden de octetos de red). La semántica del campo DONL puede ser la misma que la presentada en la solicitud estadounidense N° 14/228.164, presentada el 27 de marzo de 2014. Más específicamente, el valor del campo DONL, cuando está presente en una carga útil de paquete del RTP, es igual al valor de los 16 bits menos significativos del número de orden de decodificación de la correspondiente unidad de NAL.

**[0064]** La primera unidad de agrupación en el AP, como se muestra en la figura 4, también incluye un campo sin signo de 16 bits (el campo "Tamaño de NALU") que contiene información de tamaño (en orden de octetos de red) que indica el tamaño de la unidad de NAL en octetos. La información de tamaño dentro del campo Tamaño de NALU excluye los dos octetos de bits asociados al campo Tamaño de NALU, pero incluye los bits asociados a la cabecera de la unidad de NAL dentro de la propia unidad de NAL.

**[0065]** Como se muestra en la figura 4, el campo Tamaño de NALU es seguido por la propia unidad de NAL que, como se ha mencionado anteriormente, incluye la cabecera de la unidad de NAL y la carga útil de la unidad de NAL. Es decir, la unidad de agrupación consiste en el campo DONL que indica el número de orden del decodificador de la unidad de NAL contenida, un campo de tamaño (por ejemplo, el campo Tamaño de NALU), que indica el tamaño de la unidad de NAL contenida, y la unidad de NAL en sí.

**[0066]** Si la modalidad de transmisión es igual a "MST" y/o el valor del parámetro sprop-depack-buf-nalus es mayor que 0, el campo DONL está presente en la primera unidad de agrupación en un AP. Además, si la modalidad de transmisión es igual a "MST" y/o el valor del parámetro sprop-depack-buf-nalus es mayor que 0, cada unidad de agrupación posterior en el AP incluirá un campo de diferencia de número de orden de decodificación (DOND). Cuando está presente en una unidad de agrupación posterior de un AP, el valor del campo DOND indica la diferencia entre el valor del número de orden de decodificación de la unidad de NAL agrupada actual (por ejemplo, la unidad de NAL en la unidad de agrupación actual) y el valor del número de orden de decodificación de la unidad de NAL agrupada anterior en el mismo AP (por ejemplo, la unidad de NAL en la unidad de agrupación precedente).

**[0067]** Tal diseño de un formato de carga útil del RTP en general o, específicamente, del formato de carga útil del RTP de la HEVC, tiene el siguiente problema. Para encapsular una sola unidad de NAL en un paquete del RTP cuando se utiliza la transmisión de múltiples flujos (MST) y/o cuando se utiliza la paquetización intercalada (por ejemplo, tanto en la SST como en la MST) según lo indicado, por ejemplo, por el parámetro sprop-depack-buf-nalus, se ha de usar un AP (por ejemplo, no se pueden usar paquetes de una sola unidad de NAL). Es decir, como el receptor del RTP proporciona unidades de NAL a una unidad de decodificación de vídeo basándose, al menos en parte, en el número de orden de decodificación, las unidades de NAL enviadas fuera del orden de decodificación deberán enviarse con un correspondiente número de orden de decodificador. Sin embargo, el formato de paquete de una sola unidad de NAL para una carga útil del RTP no incluye una indicación del número de orden de decodificación. Por lo tanto, la única unidad de NAL debería enviarse en un paquete del RTP que tenga una carga útil del RTP que esté estructurada como un AP.

**[0068]** El uso de los AP para enviar una sola unidad de NAL en un paquete del RTP requiere la inclusión de los dos octetos asociados al campo de tamaño de unidad de NAL (por ejemplo, el campo Tamaño de NALU de la Figura 4), así como la repetición de la cabecera de dos octetos de la unidad de NAL. Cuando un AP contiene únicamente una

sola unidad de agrupación, el tamaño de la unidad de NAL agrupada no es necesario. Es decir, debido a que un módulo de desempaqueado no necesita diferenciar entre múltiples unidades de agrupación dentro del AP (por ejemplo, porque solo hay una unidad de agrupación), no es necesario especificar la longitud de la unidad de NAL agrupada. Además, la inclusión tanto de la cabecera de la carga útil del RTP (la cabecera de la carga útil mostrada en la figura 3) como de la cabecera de la unidad de NAL (el campo F, el campo Tipo, el campo IdentificadorCapa y el campo TID mostrados en la figura 4) es redundante. Es decir, un AP que incluye una sola unidad de NAL tendrá una cabecera de carga útil del RTP que es casi la misma que la cabecera de la unidad de NAL de la única unidad de NAL agrupada. La única diferencia será el valor del campo Tipo. En el cabecera de la carga útil del RTP, el valor del campo Tipo será 48 (por ejemplo, para indicar que la carga útil del RTP es un AP), mientras que en la cabecera de la unidad de NAL, el valor del campo Tipo puede ser diferente (por ejemplo, para indicar el Tipo de unidad de NAL). Por lo tanto, cuando se habilita la paquetización de intercalación y/o cuando se opera en una modalidad de transmisión de múltiples flujos (por ejemplo, cuando se está en modalidad MST en el RTP), se desperdician cuatro octetos por cada paquete que contenga una sola unidad de NAL. En otras palabras, por cada AP enviado que contiene solo una unidad de agrupación, cuatro octetos de datos son innecesarios.

**[0069]** Como tal, un problema con este diseño para el formato de carga útil del RTP es que, en la modalidad MST y/o cuando el valor del parámetro sprop-depack-buf-nalus es mayor que 0, y cuando se transmite una carga útil del RTP para una sola unidad de NAL, la carga útil para el paquete del RTP ha de encapsularse como un AP que contiene esa sola unidad de NAL. Esto da como resultado una carga útil del RTP que tiene un cabecera de carga útil del RTP con una longitud de 16 bits, un campo DONL con una longitud de 16 bits y un campo de Tamaño de NALU con una longitud de 16 bits, seguidos de la unidad de NAL (que a su vez incluye una cabecera de unidad de NAL distinta (es decir, distinta a la cabecera de carga útil del RTP) de 16 bits). En contraste, la carga útil de RTP asociada a paquetes de unidades individuales de NAL que se admiten en la modalidad SST (de acuerdo al formato anterior de carga útil del RTP para la HEVC), no incluye un campo DONL o un campo de Tamaño de NALU. Más bien, como se ha indicado anteriormente, la cabecera de carga útil del RTP de un paquete de una sola unidad de NAL son los primeros 2 octetos (por ejemplo, la cabecera de unidad de NAL) de la unidad de NAL contenida. En escenarios en los que la modalidad de MST se utiliza para la transmisión, y/o cuando el parámetro sprop-depack-buf-nalus tiene un valor mayor que 0, el envío de unidades de NAL individuales puede inhibir el ancho de banda y/o disminuir la eficacia de transmisión porque 4 octetos (esto es, la cabecera de la carga útil del RTP de 2 octetos del AP y el campo de Tamaño de NALU de 2 octetos de la única unidad de agrupación dentro del AP) se encapsulan y transmiten innecesariamente.

**[0070]** En vista de estos problemas, esta divulgación proporciona una estructura modificada de la carga útil del RTP del paquete de una sola unidad de NAL, de modo que el paquete de una sola unidad de NAL se pueda usar en la modalidad de MST y/o cuando el parámetro sprop-depack-buf-nalus tenga un valor mayor que 0. Es decir, la estructura modificada del paquete de una sola unidad de NAL puede permitir una transmisión más eficaz de unidades de NAL individuales cuando se habilita la paquetización intercalada y/o cuando se realiza una modalidad de transmisión de múltiples flujos (tal como la modalidad de MST para el RTP) manteniendo a la vez una transmisión eficaz de unidades de NAL individuales cuando se opera en la modalidad de SST con la paquetización intercalada inhabilitada. Más generalmente, se divulgan técnicas en las que un paquete de una sola unidad de NAL puede incluir información del recuento de orden de decodificación o del número de orden de decodificación (DON) (por ejemplo, el campo DONL de 2 octetos).

**[0071]** Como ejemplo, el campo DONL puede incluirse inmediatamente después de la cabecera de la unidad de NAL de 2 octetos e inmediatamente antes de la carga útil de la unidad de NAL. La información del DON está presente en cada paquete de una sola unidad de NAL si se realiza una transmisión de múltiples flujos (por ejemplo, si la modalidad de transmisión del RTP es igual a "MST") y/o si la intercalación está habilitada (por ejemplo, si el parámetro sprop-depack-buf-nalus es mayor que 0), y no está presente en otro caso. Con un cambio de este tipo en la estructura del paquetes de una sola unidad de NAL, los paquetes de una sola unidad de NAL se pueden usar tanto en una modalidad de unidifusión (con o sin intercalación) como en una modalidad de transmisión de múltiples flujos. De esta manera, una indicación de la información de DON para una unidad de NAL, cuando sea necesario, puede ser enviada junto con la unidad de NAL, disminuyendo a la vez la cantidad de información enviada. Es decir, incluir el campo DONL optativo en paquetes de una sola unidad de NAL puede aumentar la eficacia de la transmisión de datos de vídeo (por ejemplo, cuando se usa el RTP).

**[0072]** El formato de carga útil del RTP para un paquete de una sola unidad de NAL solo se puede usar durante la SST sin intercalación, de acuerdo al formato anterior de carga útil del RTP para la HEVC que se muestra en la figura 5. Como puede verse en la figura 5, no hay ninguna información de DON presente dentro de la carga útil del RTP.

**[0073]** La figura 6 es un diagrama conceptual que muestra un ejemplo de un formato de carga útil del RTP para un paquete de una sola unidad de NAL, de acuerdo a las técnicas de la divulgación. Como se muestra en la figura 6, la estructura de carga útil del RTP del paquete de una sola unidad de NAL incluye un campo DONL. El campo DONL es "optativo" en cuanto a que el campo DONL se señala para un paquete de una sola unidad de NAL cuando se realiza una transmisión de múltiples flujos (por ejemplo, cuando la modalidad de transmisión es igual a "MST") y/o cuando se habilita la intercalación (por ejemplo, el parámetro sprop-depack-buf-nalus tiene un valor mayor que 0). Es decir, para el RTP, si el valor del parámetro sprop-depack-buf-nalus es mayor que 0 y/o la modalidad de transmisión es igual a "MST", el campo DONL se incluye dentro de la carga útil del RTP. De lo contrario, el DONL no está presente. En otras

palabras, el campo DONL no está presente en el paquete modificado de una sola unidad de NAL cuando la modalidad de transmisión es igual a "SST" y el valor del parámetro sprop-depack-buf-nalus es igual a 0.

5 **[0074]** Como se muestra en el ejemplo de la figura 6, el campo DONL optativo puede encapsularse dentro de la propia unidad de NAL. Es decir, cuando está presente en el paquete modificado de una sola unidad de NAL, el campo DONL puede encapsularse inmediatamente después del encabezado de la unidad de NAL e inmediatamente antes de la carga útil de la unidad de NAL. De esta manera, la información en la cabecera de la unidad de NAL puede funcionar tanto como cabecera de carga útil del RTP como cabecera de la unidad de NAL.

10 **[0075]** La figura 7 es un diagrama conceptual que muestra otro ejemplo de un formato de carga útil del RTP para un paquete de una sola unidad de NAL, de acuerdo a las técnicas de la divulgación. El paquete de una sola unidad de NAL, mostrado en el ejemplo de la figura 7, también incluye el campo optativo DONL. Es decir, el campo DONL, como se muestra en la figura 7, puede estar presente cuando se realiza una transmisión de múltiples flujos (por ejemplo, la modalidad de transmisión es igual a "MST") y/o cuando se habilita la intercalación (por ejemplo, el valor del parámetro sprop-depack-buf-nalus no es igual a 0). El campo DONL, como se muestra en la figura 7, puede no estar presente en otro caso.

15 **[0076]** En el ejemplo de la figura 7, la unidad de NAL individual es parte de la carga útil del RTP a continuación del campo DONL. En este caso, los primeros 2 octetos de la unidad de NAL (por ejemplo, la cabecera de la unidad de NAL) se repiten antes del campo DONL (por ejemplo, como el encabezado de carga útil del RTP), ahorrando por ello 2 octetos en comparación con el uso de un AP para enviar una sola unidad de NAL. La ejemplar unidad única de NAL de la figura 7 también brinda la ventaja de no separar los primeros dos octetos de la unidad de NAL del resto de los datos de NAL con el campo DONL en medio. En otras palabras, el paquete modificado de una sola unidad de NAL puede incluir la decodificación de la información del número de orden en varias ubicaciones. El ejemplar paquete de una sola unidad de NAL de la figura 6 incluye el campo DONL entre la cabecera de la unidad de NAL y la carga útil de la unidad de NAL. El ejemplar paquete de una sola unidad de NAL de la figura 7 incluye el campo DONL antes de la cabecera de la unidad de NAL (por ejemplo, antes de la unidad de NAL) e incluye un duplicado de la cabecera de la unidad de NAL (por ejemplo, los dos primeros octetos de la unidad de NAL) antes del campo DONL. Duplicar la información de la cabecera de la unidad de NAL para crear una cabecera de carga útil del RTP, como se muestra en el ejemplo de la figura 7, puede proporcionar la necesaria cabecera de carga útil del RTP, evitando a la vez la separación de la cabecera de la unidad de NAL y la carga útil de la unidad de NAL.

20 **[0077]** Las técnicas para la codificación de una unidad de NAL única en un formato de carga útil del RTP de esta divulgación pueden realizarse mediante un codificador de vídeo, un decodificador de vídeo, un elemento de red consciente de los medios (MANE) y otro hardware de procesamiento de vídeo y/o red. Los siguientes diagramas describen estructuras ejemplares, que incluyen el codificador de vídeo 20, el decodificador de vídeo 30, el MANE 29, el dispositivo servidor 152, el dispositivo de encaminamiento 154A, el dispositivo de transcodificación 156, el dispositivo de encaminamiento 154B y el dispositivo cliente 158 que pueden implementar las técnicas de esta divulgación.

25 **[0078]** La figura 8 es un diagrama de bloques que ilustra un codificador ejemplar de vídeo 20 que puede implementar las técnicas descritas en esta divulgación. El codificador de vídeo 20 puede realizar la intracodificación y la intercodificación de bloques de vídeo dentro de fragmentos de vídeo. La intracodificación se basa en la predicción espacial para reducir o eliminar la redundancia espacial en el vídeo dentro de una trama o imagen de vídeo dada. La intercodificación se basa en la predicción temporal para reducir o eliminar la redundancia temporal en el vídeo dentro de tramas o imágenes adyacentes de una secuencia de vídeo. La intra-modalidad (modalidad I) puede referirse a cualquiera entre varias modalidades de compresión espacial. Las inter-modalidades, tales como la predicción unidireccional (modalidad P) o la predicción bidireccional (modalidad B), pueden referirse a cualquiera entre varias modalidades de compresión con base temporal.

30 **[0079]** En el ejemplo de la figura 8, el codificador de vídeo 20 incluye una memoria de datos de vídeo 34, una unidad de división 35, una unidad de procesamiento de predicción 41, una unidad de filtro 63, una memoria de imágenes 64, un sumador 50, una unidad de procesamiento de transformación 52, una unidad de cuantización 54 y una unidad de codificación por entropía 56. La unidad de procesamiento de predicción 41 incluye una unidad de estimación de movimiento 42, una unidad de compensación de movimiento 44 y una unidad de intrapredicción 46. Para la reconstrucción de bloques de vídeo, el codificador de vídeo 20 incluye también una unidad de cuantización inversa 58, una unidad de procesamiento de transformación inversa 60 y un sumador 62. La unidad de filtro 63 está destinada a representar uno o más filtros de bucle tales como un filtro de eliminación de bloques, un filtro de bucle adaptativo (ALF) y un filtro de desplazamiento adaptativo de muestras (SAO). Aunque la unidad de filtro 63 se muestra en la figura 8 como un filtro de bucle, en otras configuraciones, la unidad de filtro 63 puede implementarse como un filtro pos-bucle. La figura 8 también muestra el dispositivo de procesamiento posterior 57 que puede realizar un procesamiento adicional en los datos de vídeo codificados generados por el codificador de vídeo 20. Las técnicas de esta divulgación pueden implementarse en algunos casos mediante el codificador de vídeo 20. En otros casos, sin embargo, las técnicas de esta divulgación pueden implementarse mediante el dispositivo de procesamiento posterior 57.

**[0080]** La memoria de datos de vídeo 34 puede almacenar datos de vídeo a codificar por parte de los componentes del codificador de vídeo 20. Los datos de vídeo almacenados en la memoria de datos de vídeo 34 se pueden obtener, por ejemplo, a partir del origen de vídeo 18. La memoria de imágenes 64 puede ser una memoria de imágenes de referencia que almacena datos de vídeo de referencia para su uso en la codificación de datos de vídeo por el codificador de vídeo 20, por ejemplo, en modalidades de intracodificación o de intercodificación. La memoria de datos de vídeo 34 y la memoria de imágenes 64 pueden ser formadas por cualquiera entre varios dispositivos de memoria, tales como memoria dinámica de acceso aleatorio (DRAM), incluyendo DRAM síncrona (SDRAM), RAM magnetorresistiva (MRAM), RAM resistiva (RRAM) u otros tipos de dispositivos de memoria. La memoria de datos de vídeo 34 y la memoria de imágenes 64 pueden ser proporcionadas por el mismo dispositivo de memoria o por dispositivos de memoria independientes. En diversos ejemplos, la memoria de datos de vídeo 34 puede estar en un chip con otros componentes del codificador de vídeo 20, o fuera de chip con respecto a esos componentes.

**[0081]** Como se muestra en la figura 8, el codificador de vídeo 20 recibe datos de vídeo, y la unidad de división 35 divide los datos en bloques de vídeo. Esta división también puede incluir la división en fragmentos, mosaicos u otras unidades mayores, así como la división de bloques de vídeo, por ejemplo, de acuerdo a una estructura de árbol cuádruple de las LCU y las CU. El codificador de vídeo 20 ilustra, en general, los componentes que codifican bloques de vídeo dentro de un fragmento de vídeo a codificar. El fragmento se puede dividir en múltiples bloques de vídeo (y, posiblemente, en conjuntos de bloques de vídeo denominados mosaicos). La unidad de procesamiento de predicción 41 puede seleccionar una entre una pluralidad de posibles modalidades de codificación, tal como una entre una pluralidad de modalidades de intracodificación, o uno entre una pluralidad de modos de intercodificación, para el bloque de vídeo actual basándose en resultados de errores (por ejemplo, la velocidad de codificación y el nivel de distorsión). La unidad de procesamiento de predicción 41 puede proporcionar el bloque intracodificado o intercodificado resultante al sumador 50 para generar datos de bloques residuales y al sumador 62 para reconstruir el bloque codificado para su uso como imagen de referencia.

**[0082]** La unidad de intrapredicción 46, dentro de la unidad de procesamiento de predicción 41, puede realizar la codificación intrapredictiva del bloque de vídeo actual con respecto a uno o más bloques contiguos en la misma trama o fragmento que el bloque actual a codificar, para proporcionar compresión espacial. La unidad de estimación de movimiento 42 y la unidad de compensación de movimiento 44 de la unidad de procesamiento de predicción 41 realizan la codificación interpredictiva del bloque de vídeo actual con respecto a uno o más bloques predictivos de una o más imágenes de referencia para proporcionar una compresión temporal.

**[0083]** La unidad de estimación de movimiento 42 puede estar configurada para determinar la modalidad de interpredicción para un fragmento de vídeo, de acuerdo a un patrón predeterminado para una secuencia de vídeo. El patrón predeterminado puede designar fragmentos de vídeo en la secuencia como fragmentos P, fragmentos B o fragmentos GPB. La unidad de estimación de movimiento 42 y la unidad de compensación de movimiento 44 pueden estar sumamente integradas, pero se ilustran por separado con fines conceptuales. La estimación del movimiento, realizada por la unidad de estimación de movimiento 42, es el proceso de generación de vectores de movimiento, que estiman el movimiento para los bloques de vídeo. Un vector de movimiento, por ejemplo, puede indicar el desplazamiento de una PU de un bloque de vídeo dentro de una trama o imagen de vídeo actual en relación con un bloque predictivo dentro de una imagen de referencia.

**[0084]** Un bloque predictivo es un bloque que resulta corresponderse estrechamente con la PU del bloque de vídeo a codificar en términos de diferencia de píxel, que puede determinarse mediante una suma de diferencia absoluta (SAD), una suma de diferencia al cuadrado (SSD) u otras métricas de diferencia. En algunos ejemplos, el codificador de vídeo 20 puede calcular los valores para las posiciones fraccionarias de píxeles de imágenes de referencia almacenadas en la memoria de imágenes 64. Por ejemplo, el codificador de vídeo 20 puede interpolar valores de posiciones de un cuarto de píxel, posiciones de un octavo de píxel u otras posiciones fraccionarias de píxel de la imagen de referencia. Por lo tanto, la unidad de estimación de movimiento 42 puede realizar una búsqueda de movimiento relativa a las posiciones de píxel completo y las posiciones de píxel fraccionario, y emitir un vector de movimiento con una precisión de píxel fraccionario.

**[0085]** La unidad de estimación de movimiento 42 calcula un vector de movimiento para una PU de un bloque de vídeo en un fragmento intercodificado, comparando la posición de la PU con la posición de un bloque predictivo de una imagen de referencia. La imagen de referencia puede seleccionarse entre una primera lista de imágenes de referencia (Lista 0) o una segunda lista de imágenes de referencia (Lista 1), cada una de las cuales identifica una o más imágenes de referencia almacenadas en la memoria de imágenes 64. La unidad de estimación de movimiento 42 envía el vector de movimiento calculado, junto con otros elementos sintácticos, a la unidad de codificación por entropía 56 y a la unidad de compensación de movimiento 44.

**[0086]** La compensación de movimiento, realizada por la unidad de compensación de movimiento 44, puede implicar extraer o generar el bloque predictivo basándose en el vector de movimiento determinado mediante estimación de movimiento, realizando posiblemente interpolaciones hasta una precisión de subpíxel. Tras recibir el vector de movimiento para la PU del bloque de vídeo actual, la unidad de compensación de movimiento 44 puede localizar el bloque predictivo al que apunta el vector de movimiento en una de las listas de imágenes de referencia. El codificador de vídeo 20 forma un bloque de vídeo residual restando los valores de píxel del bloque predictivo a los valores de

píxel del bloque de vídeo actual que se está codificando, formando valores de diferencia de píxel. Los valores de diferencia de píxel forman datos residuales para el bloque, y pueden incluir componentes de diferencia tanto de luma como de croma. El sumador 50 representa el componente o los componentes que realizan esta operación de resta. La unidad de compensación de movimiento 44 también puede generar elementos sintácticos asociados a los bloques de vídeo y al fragmento de vídeo para su uso por el decodificador de vídeo 30 en la decodificación de los bloques de vídeo del fragmento de vídeo.

**[0087]** La unidad de intrapredicción 46 puede intrapredecir un bloque actual, como alternativa a la interpredicción llevada a cabo por la unidad de estimación de movimiento 42 y la unidad de compensación de movimiento 44, como se ha descrito anteriormente. En particular, la unidad de intrapredicción 46 puede determinar una modalidad de intrapredicción a usar para codificar un bloque actual. En algunos ejemplos, la unidad de intrapredicción 46 puede codificar un bloque actual usando diversas modalidades de intrapredicción, por ejemplo, durante pases de codificación independientes, y la unidad de intrapredicción 46 (o la unidad de selección de modalidad 40, en algunos ejemplos) puede seleccionar una modalidad de intrapredicción adecuada para usar, a partir de las modalidades probadas. Por ejemplo, la unidad de intrapredicción 46 puede calcular valores de velocidad-distorsión usando un análisis de velocidad-distorsión para las diversas modalidades de intrapredicción probadas, y seleccionar la modalidad de intrapredicción que tenga las mejores características de velocidad-distorsión entre las modalidades probadas. El análisis de velocidad-distorsión determina, en general, una magnitud de distorsión (o error) entre un bloque codificado y un bloque original no codificado, que fue codificado para generar el bloque codificado, así como una velocidad de bits (es decir, un número de bits) usada para generar el bloque codificado. La unidad de intrapredicción 46 puede calcular razones a partir de las distorsiones y velocidades para los diversos bloques codificados, para determinar qué modalidad de intrapredicción presenta el mejor valor de velocidad-distorsión para el bloque.

**[0088]** En cualquier caso, tras seleccionar una modalidad de intrapredicción para un bloque, la unidad de intrapredicción 46 puede proporcionar información, que indica la modalidad de intrapredicción seleccionada para el bloque, a la unidad de codificación por entropía 56. La unidad de codificación por entropía 56 puede codificar la información que indica la modalidad de intrapredicción seleccionada, de acuerdo a las técnicas de esta divulgación. El codificador de vídeo 20 puede incluir, en el flujo de bits transmitido, datos de configuración, que pueden incluir una pluralidad de tablas de índices de modalidad de intrapredicción y una pluralidad de tablas modificadas de índices de modalidad de intrapredicción (también denominadas tablas de correlación de palabras de código), definiciones de contextos de codificación para diversos bloques e indicaciones de una modalidad de intrapredicción más probable, una tabla de índices de modalidad de intrapredicción y una tabla modificada de índices de modalidad de intrapredicción, a usar para cada uno de los contextos.

**[0089]** Después de que la unidad de procesamiento de predicción 41 genera el bloque predictivo para el bloque de vídeo actual, ya sea mediante la interpredicción o la intrapredicción, el codificador de vídeo 20 forma un bloque de vídeo residual restando el bloque predictivo al bloque de vídeo actual. Los datos de vídeo residuales en el bloque residual pueden estar incluidos en una o más TU y aplicarse a la unidad de procesamiento de transformación 52. La unidad de procesamiento de transformación 52 transforma los datos de vídeo residuales en coeficientes de transformación residuales usando una transformación, tal como una transformación de coseno discreta (DCT) o una transformación conceptualmente similar. La unidad de procesamiento de transformación 52 puede convertir los datos de vídeo residuales, desde un dominio de píxel a un dominio de transformación, tal como un dominio de frecuencia.

**[0090]** La unidad de procesamiento de transformación 52 puede enviar los coeficientes de transformación resultantes a la unidad de cuantización 54. La unidad de cuantización 54 cuantiza los coeficientes de transformación para reducir más la velocidad de bits. El proceso de cuantización puede reducir la profundidad de bits asociada a algunos, o a la totalidad, de los coeficientes. El grado de cuantización se puede modificar ajustando un parámetro de cuantización. En algunos ejemplos, la unidad de cuantización 54 puede realizar, a continuación, una exploración de la matriz que incluye los coeficientes de transformación cuantizados. De forma alternativa, la unidad de codificación de entropía 56 puede realizar la exploración.

**[0091]** Tras la cuantización, la unidad de codificación por entropía 56 realiza la codificación por entropía de los coeficientes de transformación cuantizados. Por ejemplo, la unidad de codificación por entropía 56 puede realizar una codificación de longitud variable adaptativa al contexto (CAVLC), una codificación aritmética binaria adaptativa al contexto (CABAC), una codificación aritmética binaria adaptativa al contexto basada en la sintaxis (SBAC), una codificación de entropía por división de intervalos de probabilidad (PIPE) u otros procedimientos o técnicas de codificación por entropía. Tras la codificación por entropía por la unidad de codificación por entropía 56, el flujo de bits codificado se puede transmitir al decodificador de vídeo 30, o archivarse para su posterior transmisión o recuperación por el decodificador de vídeo 30. La unidad de codificación por entropía 56 también puede realizar la codificación por entropía de los vectores de movimiento y los otros elementos sintácticos para el fragmento de vídeo actual que se está codificando.

**[0092]** La unidad de cuantificación inversa 58 y la unidad de procesamiento de transformación inversa 60 aplican una cuantificación inversa y una transformación inversa, respectivamente, para reconstruir el bloque residual en el dominio del píxel, para su posterior uso como bloque de referencia de una imagen de referencia. La unidad de compensación de movimiento 44 puede calcular un bloque de referencia añadiendo el bloque residual a un bloque

predictivo de una de las imágenes de referencia dentro de una de las listas de imágenes de referencia. La unidad de compensación de movimiento 44 también puede aplicar uno o más filtros de interpolación al bloque residual reconstruido para calcular valores fraccionarios de píxel, para su uso en la estimación de movimiento. El sumador 62 añade el bloque residual reconstruido al bloque predictivo con compensación de movimiento generado por la unidad de compensación de movimiento 44 para generar un bloque de referencia para su almacenamiento en la memoria de imágenes 64. La unidad de estimación de movimiento 42 y la unidad de compensación de movimiento 44 pueden usar el bloque de referencia como bloque de referencia para realizar la interpredicción de un bloque en una trama o imagen de vídeo posterior.

**[0093]** De acuerdo a las técnicas descritas en el presente documento, el codificador de vídeo 20 y/o el dispositivo de procesamiento posterior 57 pueden encapsular datos de vídeo codificados para su transmisión a uno o más dispositivos (por ejemplo, usando el RTP). Por ejemplo, el dispositivo de procesamiento posterior 57 puede recibir datos de vídeo de norma HEVC codificados (por ejemplo, unidades de NAL) y generar paquetes (por ejemplo, paquetes del RTP) que tienen cargas útiles que cumplen un formato de carga útil particular para la HEVC (por ejemplo, el formato de carga útil del RTP para la HEVC), encapsulando los datos de vídeo en un paquete de una sola unidad de capa de abstracción de red (NAL) (por ejemplo, para una sesión del RTP). El dispositivo de procesamiento posterior 57 también puede encapsular la información del número de orden de decodificación (por ejemplo, el DONL) en el paquete de una sola unidad de NAL, en función de al menos uno entre: la sesión es una transmisión de múltiples flujos (por ejemplo, una sesión del RTP en una modalidad de transmisión de múltiples flujos (MST)), o un número máximo, de unidades de NAL que pueden preceder a la unidad de NAL en un almacén temporal de desempaqueado en el orden de recepción y seguir a la unidad de NAL en el orden de decodificación, es mayor que 0.

**[0094]** La figura 9 es un diagrama de bloques que ilustra un decodificador ejemplar de vídeo 30 que puede implementar las técnicas descritas en esta divulgación. En el ejemplo de la figura 9, el decodificador de vídeo 30 incluye una memoria de datos de vídeo 83, una unidad de decodificación por entropía 80, una unidad de procesamiento de predicción 81, una unidad de cuantización inversa 86, una unidad de transformación inversa 88, un sumador 90, una unidad de filtro 91 y una memoria de imágenes 92. La unidad de procesamiento de predicción 81 incluye una unidad de compensación de movimiento 82 y una unidad de procesamiento de intrapredicción 84. En algunos ejemplos, el decodificador de vídeo 30 puede realizar un pase de decodificación, generalmente recíproco al pase de codificación descrito con respecto al codificador de vídeo 20 de la figura 8.

**[0095]** La memoria de datos de vídeo 83 puede almacenar datos de vídeo, tales como un flujo de bits de vídeo codificado, a decodificar por parte de los componentes del decodificador de vídeo 30. Los datos de vídeo almacenados en la memoria de datos de vídeo 83 pueden obtenerse, por ejemplo, a partir de un medio legible por ordenador, por ejemplo, desde un origen de vídeo local, tal como una cámara, mediante comunicación de datos de vídeo por red cableada o inalámbrica, o accediendo a medios de almacenamiento físico de datos. La memoria de datos de vídeo 83 puede formar un almacén temporal de imágenes codificadas (CPB) que almacene datos de vídeo codificados a partir de un flujo de bits de vídeo codificado. La memoria de imágenes 92, en algunos ejemplos, puede ser una memoria de imágenes de referencia que almacena datos de vídeo de referencia para su uso en la decodificación de datos de vídeo por el decodificador de vídeo 30, por ejemplo, en modalidades de intracodificación o intercodificación. La memoria de datos de vídeo 83 y la memoria de imágenes de referencia 92 pueden ser formadas por cualquiera entre varios dispositivos de memoria, tales como memoria dinámica de acceso aleatorio (DRAM), incluyendo DRAM síncrona (SDRAM), RAM magnetorresistiva (MRAM), RAM resistiva (RRAM) u otros tipos de dispositivos de memoria. La memoria de datos de vídeo 83 y la memoria de imágenes 92 pueden ser proporcionadas por el mismo dispositivo de memoria o por dispositivos de memoria independientes. En diversos ejemplos, la memoria de datos de vídeo 83 puede estar en un chip con otros componentes del decodificador de vídeo 30, o fuera de chip con respecto a esos componentes.

**[0096]** Durante el proceso de decodificación, el decodificador de vídeo 30 recibe un flujo de bits de vídeo codificado, que representa bloques de vídeo de un fragmento de vídeo codificado y elementos sintácticos asociados, desde el codificador de vídeo 20. El decodificador de vídeo 30 puede recibir el flujo de bits de vídeo codificado desde una entidad de red 79. La entidad de red 79 puede ser, por ejemplo, un servidor, un MANE, un editor/empalmador de vídeo, un receptor del RTP u otro dispositivo de ese tipo, configurado para implementar una o más de las técnicas descritas anteriormente. La entidad de red 79 puede incluir o no el codificador de vídeo 20. Como se ha descrito anteriormente, algunas de las técnicas descritas en esta divulgación pueden ser implementadas por la entidad de red 79 antes de que la entidad de red 79 transmita el flujo de bits de vídeo codificado al decodificador de vídeo 30. En algunos sistemas de decodificación de vídeo, la entidad de red 79 y el decodificador de vídeo 30 pueden formar parte de dispositivos independientes, mientras que, en otros casos, el mismo dispositivo que comprende el decodificador de vídeo 30 puede desempeñar la funcionalidad descrita con respecto a la entidad de red 79.

**[0097]** Durante el proceso de decodificación, el decodificador de vídeo 30 recibe un flujo de bits de vídeo codificado, que representa bloques de vídeo de un fragmento de vídeo codificado y elementos sintácticos asociados, desde el codificador de vídeo 20. Los bloques de vídeo pueden, por ejemplo, encaminarse desde el codificador de vídeo 20 al decodificador de vídeo 30 mediante uno o más MANE, tales como el MANE 29 en la figura 1 o la entidad de red 79 en la figura 9. La unidad de decodificación por entropía 80 del decodificador de vídeo 30 decodifica por entropía el flujo de bits para generar coeficientes cuantizados, vectores de movimiento y otros elementos sintácticos. La unidad



de decodificación por entropía 80 remite los vectores de movimiento y otros elementos sintácticos a la unidad de procesamiento de predicción 81. El decodificador de vídeo 30 puede recibir los elementos sintácticos en el nivel de fragmento de vídeo y/o el nivel de bloque de vídeo.

5 **[0098]** Cuando el fragmento de vídeo se codifica como un fragmento intracodificado (I), la unidad de procesamiento de intrapredicción 84 de la unidad de procesamiento de predicción 81 puede generar datos de predicción para un  
 10 bloque de vídeo del fragmento de vídeo actual basándose en una modalidad de intrapredicción señalizada, y datos de bloques previamente decodificados de la trama o imagen actual. Cuando la trama de vídeo se codifica como un  
 15 fragmento intercodificado (es decir, B, P o GPB), la unidad de compensación de movimiento 82 de la unidad de procesamiento de predicción 81 genera bloques predictivos para un bloque de vídeo del fragmento de vídeo actual, basándose en los vectores de movimiento y en otros elementos sintácticos recibidos desde la unidad de decodificación por entropía 80. Los bloques predictivos se pueden generar a partir de una de las imágenes de referencia dentro de una de las listas de imágenes de referencia. El decodificador de vídeo 30 puede construir las listas de tramas de referencia, la Lista 0 y la Lista 1, usando técnicas de construcción predeterminadas, basándose en las imágenes de referencia almacenadas en la memoria de imágenes 92.

20 **[0099]** La unidad de compensación de movimiento 82 determina la información de predicción para un bloque de vídeo del fragmento de vídeo actual, analizando sintácticamente los vectores de movimiento y otros elementos sintácticos, y usa la información de predicción para generar los bloques predictivos para el bloque de vídeo actual que se está decodificando. Por ejemplo, la unidad de compensación de movimiento 82 usa algunos de los elementos sintácticos recibidos para determinar una modalidad de predicción (por ejemplo, intrapredicción o interpredicción) usada para codificar los bloques de vídeo del fragmento de vídeo, un tipo de fragmento de interpredicción (por ejemplo, fragmento B, fragmento P o fragmento GPB), información de construcción para una o más de las listas de imágenes de referencia para el fragmento, vectores de movimiento para cada bloque de vídeo intercodificado del fragmento, el estado de interpredicción para cada bloque de vídeo intercodificado del fragmento y otra información para decodificar los bloques de vídeo en el fragmento de vídeo actual.

30 **[0100]** La unidad de compensación de movimiento 82 también puede realizar la interpolación basándose en filtros de interpolación. La unidad de compensación de movimiento 82 puede usar filtros de interpolación como los usados por el codificador de vídeo 20 durante la codificación de los bloques de vídeo, para calcular valores interpolados para píxeles fraccionarios de los bloques de referencia. En este caso, la unidad de compensación de movimiento 82 puede determinar los filtros de interpolación usados por el codificador de vídeo 20 a partir de los elementos sintácticos recibidos y usar los filtros de interpolación para generar bloques predictivos.

35 **[0101]** La unidad de cuantización inversa 86 realiza la cuantización inversa, es decir, la decuantización, de los coeficientes de transformación cuantizados proporcionados en el flujo de bits y decodificados por la unidad de decodificación por entropía 80. El proceso de cuantización inversa puede incluir el uso de un parámetro de cuantización calculado por el codificador de vídeo 20 para cada bloque de vídeo en el fragmento de vídeo, con el fin de determinar un grado de cuantización y, asimismo, un grado de cuantización inversa que se debería aplicar. La  
 40 unidad de transformación inversa 88 aplica una transformación inversa, por ejemplo, una DCT inversa, una transformación entera inversa, o un proceso de transformación inversa conceptualmente similar, a los coeficientes de transformación, con el fin de generar bloques residuales en el dominio del píxel.

45 **[0102]** Después de que la unidad de compensación de movimiento 82 genera el bloque predictivo para el bloque de vídeo actual, basándose en los vectores de movimiento y otros elementos sintácticos, el decodificador de vídeo 30 forma un bloque de vídeo decodificado sumando los bloques residuales procedentes de la unidad de transformación inversa 88 a los correspondientes bloques predictivos generados por la unidad de compensación de movimiento 82. El sumador 90 representa el componente o los componentes que realizan esta operación de suma. Si se desea, también pueden usarse filtros de bucle (ya sea en el bucle de codificación o después del bucle de codificación) para allanar las transiciones de píxeles o mejorar de otro modo la calidad del vídeo. La unidad de filtro 91 está destinada a representar uno o más filtros de bucle tales como un filtro de eliminación de bloques, un filtro de bucle adaptativo (ALF) y un filtro de desplazamiento adaptativo de muestras (SAO). Aunque la unidad de filtro 91 se muestra en la figura 9 como un filtro en bucle, en otras configuraciones, la unidad de filtro 91 puede implementarse como un filtro pos-bucle. Los bloques de vídeo decodificados en una trama o imagen dada son a continuación almacenados en la  
 50 memoria de imágenes 92, que almacena imágenes de referencia usadas para la posterior compensación de movimiento. La memoria de imágenes 92 almacena también vídeo decodificado para su presentación posterior en un dispositivo de visualización, tal como el dispositivo de visualización 32 de la figura 1.

60 **[0103]** De acuerdo a las técnicas descritas en el presente documento, la entidad de red 79 y/o el decodificador de vídeo 30 pueden desencapsular los datos de vídeo codificados que fueron encapsulados para su transmisión a otros uno o más dispositivos (por ejemplo, utilizando el RTP). Por ejemplo, la entidad de red 79 puede recibir uno o más paquetes (por ejemplo, paquetes del RTP) que incluyen datos de vídeo de norma HEVC codificados (por ejemplo, unidades de NAL). Los paquetes pueden tener cargas útiles que cumplen un formato de carga útil particular para la HEVC (por ejemplo, el formato de carga útil del RTP para la HEVC). Para procesar los datos de vídeo, la entidad de red 79 puede desencapsular los datos de vídeo encapsulados en un paquete de una sola unidad de capa de abstracción de red (NAL). La entidad de red 79 y/o el decodificador de vídeo 30 también pueden desencapsular la  
 65

información del número de orden de decodificación (por ejemplo, un DONL) encapsulada en el paquete de una sola unidad de NAL, basándose en al menos uno entre: la sesión es una transmisión de múltiples flujos (por ejemplo, una sesión del RTP en una modalidad de transmisión de múltiples flujos (MST)), o un número máximo, de unidades de NAL que pueden preceder a la unidad de NAL en un almacén temporal de desempaquetado en el orden de recepción y seguir a la unidad de NAL en el orden de decodificación, es mayor que 0. Después de obtener los datos de vídeo (por ejemplo, la unidad de NAL) y la información de DON, el decodificador de vídeo 30 puede procesar los datos de vídeo codificados.

**[0104]** La figura 10 es un diagrama de bloques que ilustra un conjunto ejemplar de dispositivos que forman parte de la red 150. En este ejemplo, la red 150 incluye los dispositivos de encaminamiento 154A, 154B (dispositivos de encaminamiento 154) y el dispositivo de transcodificación 156. Los dispositivos de encaminamiento 154 y el dispositivo de transcodificación 156 están concebidos para representar un pequeño número de dispositivos que pueden formar parte de la red 150. Otros dispositivos de red, tales como conmutadores, concentradores, pasarelas, cortafuegos, puentes y otros dispositivos de ese tipo también pueden estar incluidos dentro de la red 150. Además, pueden proporcionarse dispositivos de red adicionales a lo largo de un trayecto de red entre un dispositivo servidor 152 y un dispositivo cliente 158. El dispositivo servidor 152 puede corresponder a un dispositivo de origen 12 (figura 1), mientras que el dispositivo cliente 158 puede corresponder a un dispositivo de destino 14 (figura 1), en algunos ejemplos. Los dispositivos de encaminamiento 154, por ejemplo, pueden ser MANE configurados para encaminar datos de medios.

**[0105]** En general, los dispositivos de encaminamiento 154 implementan uno o más protocolos de encaminamiento para intercambiar datos de red a través de la red 150. En general, los dispositivos de encaminamiento 154 ejecutan protocolos de encaminamiento para descubrir rutas a través de la red 150. Al ejecutar dichos protocolos de encaminamiento, un dispositivo de encaminamiento 154B puede descubrir una ruta de red desde sí mismo hasta el dispositivo servidor 152, mediante el dispositivo de encaminamiento 154A. Los diversos dispositivos de la figura 10 representan ejemplos de dispositivos que pueden implementar las técnicas de esta divulgación y pueden configurarse para procesar datos del RTP de acuerdo a las técnicas de esta divulgación.

**[0106]** Por ejemplo, uno o más entre el dispositivo servidor 152, los dispositivos de encaminamiento 154, el dispositivo de transcodificación 156 o el dispositivo cliente 158 pueden procesar datos de vídeo en una carga útil de unidad de datos (por ejemplo, en una carga útil del protocolo de transporte en tiempo real (RTP)) encapsulando los datos de vídeo en un paquete de una sola unidad de capa de abstracción de red (NAL) (por ejemplo, para una sesión del RTP) y encapsulando la información del número de orden de decodificación en el paquete de una sola unidad de NAL, en función de al menos uno entre: la sesión es una transmisión de múltiples flujos (por ejemplo, la sesión del RTP está en modalidad de transmisión de múltiples flujos (MST)), o un número máximo, de unidades de NAL que pueden preceder a la unidad de NAL en un almacén temporal de desempaquetado en el orden de recepción y seguir a la unidad de NAL en el orden de decodificación, es mayor que 0.

**[0107]** El paquete de una sola unidad de NAL, como parte de la sesión del RTP, puede transmitirse a otros uno o más entre el dispositivo servidor 152, los dispositivos de encaminamiento 154, el dispositivo de transcodificación 156 o el dispositivo cliente 158. Al recibir el paquete del RTP que incluye una carga útil del RTP formateada como un paquete de una sola unidad de NAL, el dispositivo receptor puede procesar los datos de vídeo desencapsulando los datos de vídeo que están encapsulados en el paquete de una sola unidad de NAL y desencapsulando la información del número de orden de decodificación que está encapsulada en el paquete de una sola unidad de NAL, en función de al menos uno entre: la sesión del RTP está en una modalidad de transmisión de múltiples flujos (MST), o un número máximo, de unidades de NAL que pueden preceder a la unidad de NAL en un almacén temporal de desempaquetado en orden de recepción y seguir a la unidad de NAL en orden de decodificación, es mayor que 0.

**[0108]** La figura 11 es un diagrama de flujo que ilustra operaciones ejemplares para encapsular datos de vídeo en un formato de carga útil del RTP, de acuerdo a las técnicas de la divulgación. Con fines de ilustración solamente, las operaciones ejemplares de la figura 11 se describen a continuación dentro del contexto de la figura 1.

**[0109]** En el ejemplo de la figura 11, una unidad de encapsulación del RTP (por ejemplo, la interfaz de salida 22) puede recibir datos de vídeo (180). Por ejemplo, los datos de vídeo pueden ser codificados (por ejemplo, por el codificador de vídeo 20) en una sola unidad de NAL, de acuerdo a la norma HEVC u otro esquema de codificación de vídeo. En algunos ejemplos, la unidad de NAL puede incluir datos de carga útil de unidad de NAL y una cabecera de unidad de NAL. Para generar una carga útil del RTP, la interfaz de salida 22 puede encapsular los datos de vídeo en un paquete de una sola unidad de NAL (182).

**[0110]** La interfaz de salida 22 puede determinar si la transmisión del RTP está en modalidad MST (184). Si la transmisión está en modalidad MST (rama "SÍ" de 184), la interfaz de salida 22 puede determinar si un número máximo, de unidades de NAL que pueden preceder a la unidad de NAL en un almacén temporal de desempaquetado en el orden de recepción y seguir a la unidad de NAL en orden de decodificación, es igual a cero (186). Por ejemplo, la interfaz de salida 22 puede determinar si el valor del parámetro sprop-depack-buf-nalus de la transmisión del RTP es igual a cero. Si el valor es igual a cero (rama "SÍ" de 186), la interfaz de salida 22 puede abstenerse de encapsular la información del número de orden de decodificación en el paquete de una sola unidad de NAL (188).

**[0111]** En el ejemplo de la figura 11, si la transmisión del RTP está en modalidad MST (rama "NO" de 184) y/o si el número máximo, de unidades de NAL que pueden preceder a la unidad de NAL en el almacén temporal de desempaqueado en el orden de recepción y seguir a la unidad de NAL en orden de decodificación, es mayor que cero (rama "SÍ" de 186), la interfaz de salida 22 puede encapsular la información del número de orden de decodificación en el paquete de una sola unidad de NAL (190). En algunos ejemplos, para encapsular la información del número de orden de decodificación en el paquete de una sola unidad de NAL, la interfaz de salida 22 puede encapsular la información del número de orden de decodificación en el paquete de una sola unidad de NAL, entre la cabecera de la unidad de NAL y los datos de la carga útil de la unidad de NAL. En algunos ejemplos, la interfaz de salida 22 puede encapsular la información del número de orden de decodificación en el paquete de una sola unidad de NAL antes de la unidad de NAL y encapsular una cabecera de carga útil del RTP antes de la información del número de orden de decodificación. La cabecera encapsulada de carga útil del RTP puede comprender información incluida en la cabecera de la unidad de NAL.

**[0112]** La figura 12 es un diagrama de flujo que ilustra operaciones ejemplares para desencapsular datos de vídeo encapsulados en un formato de carga útil del RTP, de acuerdo a las técnicas de la divulgación. Con fines de ilustración solamente, las operaciones ejemplares de la figura 12 se describen a continuación dentro del contexto de la figura 1.

**[0113]** En el ejemplo de la figura 12, una unidad de desencapsulación del RTP (por ejemplo, la interfaz de entrada 28) puede recibir un paquete del RTP (200). Por ejemplo, el paquete del RTP puede incluir una carga útil del RTP formateada como un paquete de una sola unidad de NAL. Es decir, la carga útil de RTP puede incluir una sola unidad de NAL en la carga útil de RTP. En algunos ejemplos, la unidad de NAL puede incluir datos de carga útil de unidad de NAL y una cabecera de unidad de NAL. Para obtener los datos de vídeo encapsulados en el mismo, la interfaz de entrada 28 puede desencapsular los datos de vídeo que están encapsulados en el paquete de una sola unidad de la capa de abstracción de red (NAL) (202).

**[0114]** La interfaz de entrada 28 puede determinar si la transmisión del RTP está en modalidad de MST (204). Si la transmisión está en modalidad de MST (rama "SÍ" de 204), la interfaz de entrada 28 puede determinar si un número máximo, de unidades de NAL que pueden preceder a la unidad de NAL en un almacén temporal de desempaqueado en el orden de recepción y seguir a la unidad de NAL en el orden de decodificación, es igual a cero (206). Por ejemplo, la interfaz de entrada 28 puede determinar si el valor del parámetro sprop-depack-buf-nalus de la transmisión del RTP es igual a cero. Si el valor es igual a cero (rama "SÍ" de 206), la interfaz de entrada 28 puede abstenerse de desencapsular la información del número de orden de decodificación del paquete de una sola unidad de NAL (208).

**[0115]** En el ejemplo de la figura 12, si la transmisión del RTP está en modalidad de MST (rama "NO" de 204) y/o si el número máximo, de unidades de NAL que pueden preceder a la unidad de NAL en el almacén temporal de desempaqueado en el orden de recepción y seguir a la unidad de NAL en el orden de decodificación, es mayor que cero (rama "SÍ" de 206), la interfaz de entrada 28 puede desencapsular la información del número de orden de decodificación, encapsulada en el paquete de una sola unidad de NAL (210). En algunos ejemplos, la información del número de orden de decodificación puede encapsularse en el paquete de una sola unidad de NAL, entre la cabecera de la unidad de NAL y los datos de la carga útil de la unidad de NAL. Para desencapsular la información del número de orden de decodificación en el paquete de una sola unidad de NAL, la interfaz de entrada 28 puede desencapsular la información del número de orden de decodificación, encapsulada en el paquete de una sola unidad de NAL, entre la cabecera de la unidad de NAL y los datos de carga útil de la unidad de NAL. En algunos ejemplos, la información del número de orden de decodificación puede encapsularse en el paquete de una sola unidad de NAL antes de que la unidad de NAL y una cabecera de carga útil del RTP puedan ser encapsuladas en el paquete de una sola unidad de NAL, antes de la información del número de orden de decodificación. La cabecera encapsulada de carga útil del RTP puede comprender información incluida en la cabecera de la unidad de NAL. Para desencapsular la información del número de orden de decodificación, encapsulada en el paquete de una sola unidad de NAL antes de la unidad de NAL, la interfaz de entrada 28 puede desencapsular la información del número de orden de decodificación, encapsulada en el paquete de una sola unidad de NAL antes de la unidad de NAL y desencapsular la cabecera de carga útil del RTP, encapsulada en el paquete de una sola unidad de NAL antes de la información del número de orden de decodificación.

**[0116]** En uno o más ejemplos, las funciones descritas pueden implementarse en hardware, software, firmware o cualquier combinación de los mismos. Si se implementan en software, las funciones se pueden almacenar en, o transmitir por, un medio legible por ordenador, como una o más instrucciones o código, y ejecutarse mediante una unidad de procesamiento basada en hardware. Los medios legibles por ordenador pueden incluir medios de almacenamiento legibles por ordenador, que correspondan a un medio tangible tal como medios de almacenamiento de datos, o medios de comunicación que incluyan cualquier medio que facilite la transferencia de un programa informático desde un lugar a otro, por ejemplo, de acuerdo a un protocolo de comunicación. De esta manera, los medios legibles por ordenador pueden corresponder en general a (1) medios de almacenamiento tangibles legibles por ordenador que sean no transitorios o (2) un medio de comunicación tal como una señal o una onda portadora. Los medios de almacenamiento de datos pueden ser medios disponibles cualesquiera a los que se pueda acceder desde uno o más ordenadores o uno o más procesadores para recuperar instrucciones, código y/o estructuras de datos para la implementación de las técnicas descritas en esta divulgación. Un producto de programa informático puede incluir un medio legible por ordenador.

5 [0117] A modo de ejemplo, y no de limitación, tales medios de almacenamiento legibles por ordenador pueden comprender RAM, ROM, EEPROM, CD-ROM u otro almacenamiento de disco óptico, almacenamiento de disco magnético u otros dispositivos de almacenamiento magnético, memoria flash o cualquier otro medio que pueda usarse para almacenar código de programa deseado en forma de instrucciones o estructuras de datos y al que pueda accederse mediante un ordenador. Además, cualquier conexión recibe adecuadamente la denominación de medio legible por ordenador. Por ejemplo, si las instrucciones se transmiten desde una sede de la Red, un servidor u otro origen remoto usando un cable coaxial, un cable de fibra óptica, un par trenzado, una línea de abonado digital (DSL) o tecnologías inalámbricas tales como infrarrojos, radio y microondas, entonces el cable coaxial, el cable de fibra óptica, el par trenzado, la DSL o las tecnologías inalámbricas tales como infrarrojos, radio y microondas se incluyen en la definición de medio. Sin embargo, debería entenderse que los medios de almacenamiento legibles por ordenador y los medios de almacenamiento de datos no incluyen conexiones, ondas portadoras, señales u otros medios transitorios, sino que, en cambio, se orientan a medios de almacenamiento tangibles no transitorios. El término disco, como se usa en el presente documento, incluye el disco compacto (CD), el disco láser, el disco óptico, el disco versátil digital (DVD), el disco flexible y el disco Blu-ray, donde algunos discos normalmente emiten datos magnéticamente, mientras que otros discos emiten datos ópticamente con láseres. Las combinaciones de lo anterior también deberían incluirse dentro del alcance de los medios legibles por ordenador.

20 [0118] Las instrucciones pueden ser ejecutadas por uno o más procesadores, tales como uno o más procesadores de señales digitales (DSP), microprocesadores de propósito general, circuitos integrados específicos de la aplicación (ASIC), formaciones de compuertas programables in situ (FPGA) u otros circuitos lógicos, integrados o discretos, equivalentes. En consecuencia, el término "procesador", como se usa en el presente documento, se puede referir a cualquiera de las estructuras anteriores o a cualquier otra estructura adecuada para la implementación de las técnicas descritas en el presente documento. Además, en algunos aspectos, la funcionalidad descrita en el presente documento se puede proporcionar dentro de módulos de hardware y/o software dedicados, configurados para la codificación y la decodificación, o incorporados en un códec combinado. Además, las técnicas se podrían implementar totalmente en uno o más circuitos o elementos lógicos.

30 [0119] Las técnicas de la presente divulgación se pueden implementar en una amplia variedad de dispositivos o aparatos, incluidos un equipo manual inalámbrico, un circuito integrado (IC) o un conjunto de IC (por ejemplo, un conjunto de chips). Diversos componentes, módulos o unidades se describen en esta divulgación para enfatizar aspectos funcionales de dispositivos configurados para realizar las técnicas divulgadas, pero no requieren necesariamente su realización mediante diferentes unidades de hardware. En cambio, como se ha descrito anteriormente, diversas unidades se pueden combinar en una unidad de hardware de códec, o ser proporcionadas por un grupo de unidades de hardware interoperativas, incluyendo uno o más procesadores, como se ha descrito anteriormente, conjuntamente con software y/o firmware adecuados.

35 [0120] Se han descrito diversos ejemplos. Estos y otros ejemplos están dentro del alcance de las siguientes reivindicaciones.

**REIVINDICACIONES**

1. Un procedimiento para procesar datos de vídeo en una carga útil del protocolo de transporte en tiempo real (RTP), comprendiendo el procedimiento:
- 5 encapsular (182) datos de vídeo en un paquete del RTP para una sesión del RTP, en donde el paquete del RTP incluye una cabecera del RTP y una carga útil del RTP, y en donde la cabecera del RTP incluye un campo de tipo que indica que el paquete del RTP contiene solo una unidad de NAL en la carga útil del RTP; y
- 10 encapsular (190) la información del número de orden de decodificación en el paquete del RTP si una modalidad de transmisión del RTP para la sesión del RTP representa una transmisión de múltiples flujos (MST), en la que se utilizan múltiples sesiones del RTP para la transmisión de los datos de vídeo, y/o si un número máximo, de unidades de NAL que pueden preceder a la unidad de NAL en un almacén temporal de desempaqueado en el orden de recepción y seguir a la unidad de NAL en el orden de decodificación, es mayor que 0,
- 15 en donde la información del número de orden de decodificación representa un número de orden de decodificación para la unidad de NAL contenida en el paquete del RTP.
- 20 2. El procedimiento de la reivindicación 1, en el que la única unidad de NAL comprende un cabecera de unidad de NAL y datos de carga útil de la unidad de NAL, y en el que encapsular la información del número de orden de decodificación en el paquete del RTP comprende encapsular la información del número de orden de decodificación en el paquete del RTP entre la cabecera de la unidad de NAL y los datos de la carga útil de la
- 25 unidad de NAL.
3. El procedimiento de la reivindicación 1, en el que la única unidad de NAL comprende una cabecera de unidad de NAL y datos de carga útil de la unidad de NAL, y en el que encapsular la información del número de orden de decodificación en el paquete del RTP antes de la única unidad de NAL, comprendiendo además el procedimiento:
- 30 encapsular, en el paquete del RTP, una cabecera de carga útil del RTP antes de la información del número de orden de decodificación, en donde la cabecera de carga útil del RTP comprende información incluida en la cabecera de la unidad de NAL.
- 35 4. El procedimiento de la reivindicación 1, que comprende además:
- 40 abstenerse de encapsular la información del número de orden de decodificación en el paquete de una sola unidad de NAL, basándose en que la sesión del RTP está en una modalidad de transmisión de flujo único (SST) y el número máximo, de unidades de NAL que pueden preceder a la unidad de NAL en el almacén temporal de desempaqueado en orden de recepción y seguir a la unidad de NAL en orden de decodificación, es igual a 0.
- 45 5. El procedimiento de la reivindicación 1, en el que el número máximo de unidades de NAL que pueden preceder a la unidad de NAL en el almacén temporal de desempaqueado en orden de recepción y seguir a la unidad de NAL en orden de decodificación se representa mediante un valor de un elemento sintáctico especificado durante la configuración de la sesión del RTP.
- 50 6. El procedimiento de la reivindicación 5, en el que el elemento sintáctico cuyo valor representa el número máximo de unidades de NAL que pueden preceder a la unidad de NAL en el almacén temporal de desempaqueado en orden de recepción y seguir a la unidad de NAL en orden de decodificación es el parámetro sprop-depack-buf-nalus.
- 55 7. Un procedimiento para procesar datos de vídeo en una carga útil del protocolo de transporte en tiempo real (RTP), comprendiendo el procedimiento:
- 60 desencapsular (202) los datos de vídeo que están encapsulados en un paquete del RTP para una sesión del RTP, en donde el paquete del RTP incluye una cabecera del RTP y una carga útil del RTP, y en donde la cabecera del RTP incluye un campo de tipo que indica que el paquete del RTP contiene solo una unidad de NAL en la carga útil del RTP; y
- 65 desencapsular (210) la información del número de orden de decodificación que está encapsulada en el paquete del RTP si una modalidad de transmisión del RTP para la sesión del RTP representa la modalidad de transmisión de múltiples flujos (MST), en la que se usan varias sesiones del RTP para la transmisión de los datos de vídeo, y/o si un número máximo, de unidades de NAL que pueden preceder a la unidad de

NAL en un almacén temporal de desempaquetado en orden de recepción y seguir a la unidad de NAL en orden de decodificación, es mayor que 0,

5 en donde la información del número de orden de decodificación representa un número de orden de decodificación para la unidad de NAL contenida en el paquete del RTP.

8. El procedimiento de la reivindicación 7, en el que la única unidad de NAL comprende una cabecera de unidad de NAL y datos de carga útil de la unidad de NAL, y en el que la desencapsulación de la información del número de orden de decodificación, que está encapsulada en el paquete del RTP, comprende la desencapsulación de la información del número de orden de decodificación, que está encapsulada entre la cabecera de la unidad de NAL y los datos de carga útil de la unidad de NAL.

9. El procedimiento de la reivindicación 7, en el que la única unidad de NAL comprende una cabecera de unidad de NAL y datos de carga útil de la unidad de NAL, y en el que la desencapsulación de la información del número de orden de decodificación que está encapsulada en el paquete de la única unidad de NAL comprende la desencapsulación de la información del número de orden de decodificación, que está encapsulada antes de la única unidad de NAL, comprendiendo además el procedimiento:

20 desencapsular, a partir del paquete del RTP, una cabecera de carga útil del RTP que está encapsulada en el paquete del RTP antes de la información del número de orden de decodificación, en donde la cabecera de carga útil del RTP comprende información incluida en la cabecera de la unidad de NAL.

10. El procedimiento de la reivindicación 7, que comprende además:  
25 abstenerse de desencapsular la información del número de orden de decodificación del paquete de una sola unidad de NAL, basándose en que la sesión del RTP está en una modalidad de transmisión de flujo único (SST) y el número máximo, de unidades de NAL que pueden preceder a la unidad de NAL en el almacén temporal de desempaquetado en orden de recepción y seguir a la unidad de NAL en orden de decodificación, es igual a 0.

11. El procedimiento de la reivindicación 7, en el que el número máximo de unidades de NAL que pueden preceder a la unidad de NAL en el almacén temporal de desempaquetado en orden de recepción y seguir a la unidad de NAL en orden de decodificación se representa mediante un valor de un elemento sintáctico especificado durante la configuración de la sesión del RTP.

12. El procedimiento de la reivindicación 11, en el que el elemento sintáctico cuyo valor representa el número máximo de unidades de NAL que pueden preceder a la unidad de NAL en el almacén temporal de desempaquetado en orden de recepción y seguir a la unidad de NAL en orden de decodificación es el parámetro sprop-depack-buf-nalus.

13. Un aparato configurado para procesar datos de vídeo en una carga útil del protocolo de transporte en tiempo real (RTP), comprendiendo el aparato:

45 medios para encapsular datos de vídeo en un paquete del RTP para una sesión del RTP, en donde el paquete del RTP incluye una cabecera del RTP y una carga útil del RTP, y en donde la cabecera del RTP incluye un campo de tipo que indica que el paquete del RTP contiene solo una unidad de NAL en la carga útil del RTP; y

50 medios para encapsular la información del número de orden de decodificación en el paquete del RTP si una modalidad de transmisión del RTP para la sesión RTP representa transmisión de múltiples flujos (MST) en la que se utilizan múltiples sesiones del RTP para la transmisión de los datos de vídeo, y/o si un número máximo, de las unidades de NAL que pueden preceder a la unidad de NAL en un almacén temporal de desempaquetado en orden de recepción y seguir a la unidad de NAL en orden de decodificación, es mayor que 0,

55 en donde la información del número de orden de decodificación representa un número de orden de decodificación para la unidad de NAL contenida en el paquete de la unidad del RTP.

14. Un aparato configurado para procesar datos de vídeo en una carga útil del protocolo de transporte en tiempo real (RTP), comprendiendo el aparato:

60 medios para desencapsular datos de vídeo en una sesión del RTP, en donde el paquete del RTP incluye una cabecera del RTP y una carga útil del RTP, y en donde la cabecera del RTP incluye un campo de tipo que indica que el paquete del RTP contiene solo una unidad de NAL en la carga útil del RTP; y

65 medios para desencapsular la información del número de orden de decodificación en el paquete del RTP

5 si una modalidad de transmisión del RTP para la sesión del RTP representa una transmisión de múltiples flujos (MST) en la que se utilizan múltiples sesiones del RTP para la transmisión de los datos de vídeo, y/o si un número máximo, de las unidades de NAL que pueden preceder a la unidad de NAL en un almacén temporal de desempaqueado en orden de recepción y seguir a la unidad de NAL en orden de decodificación, es mayor que 0,

en donde la información del número de orden de decodificación representa un número de orden de decodificación para la unidad de NAL contenida en el paquete del RTP.

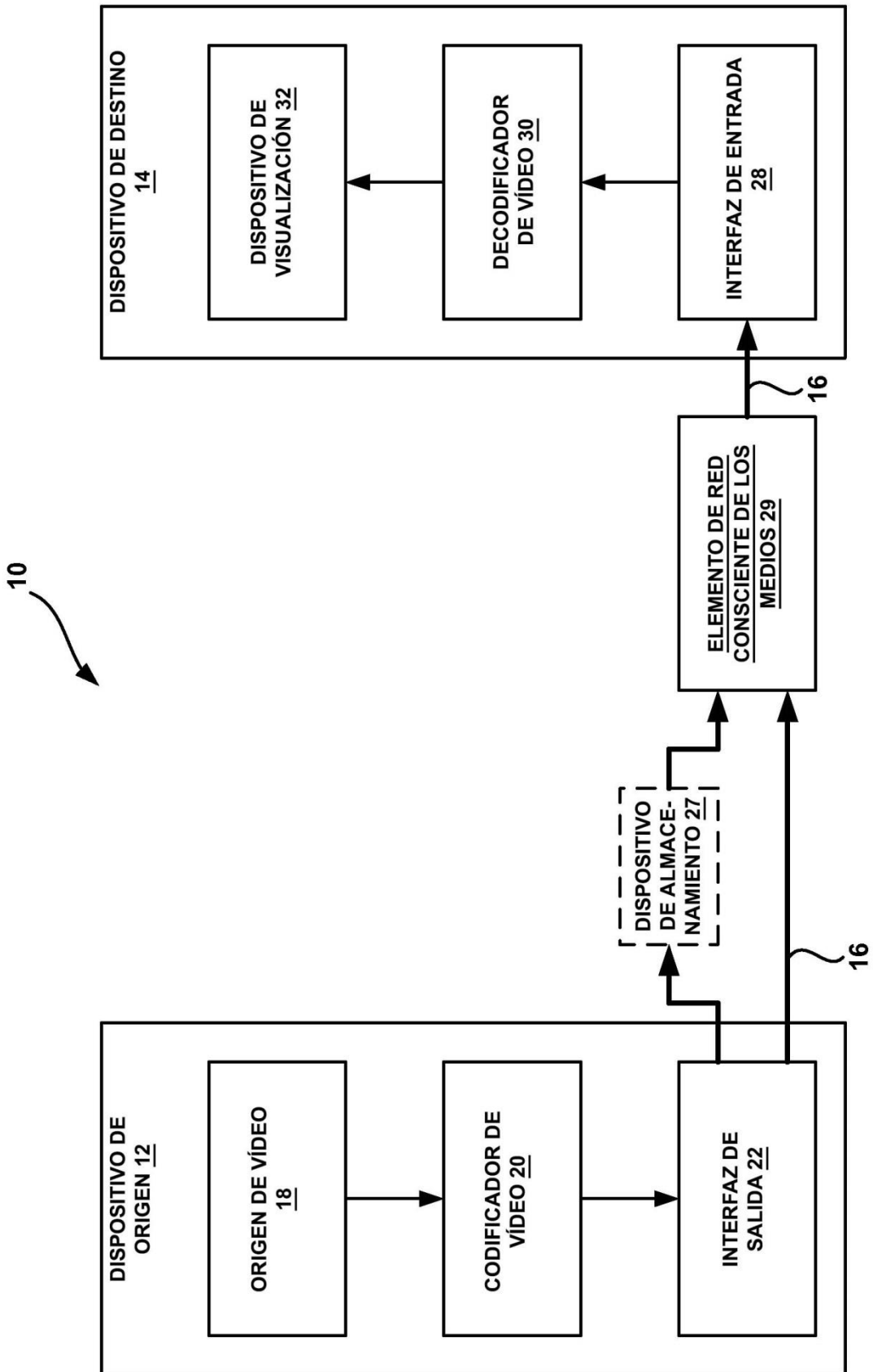
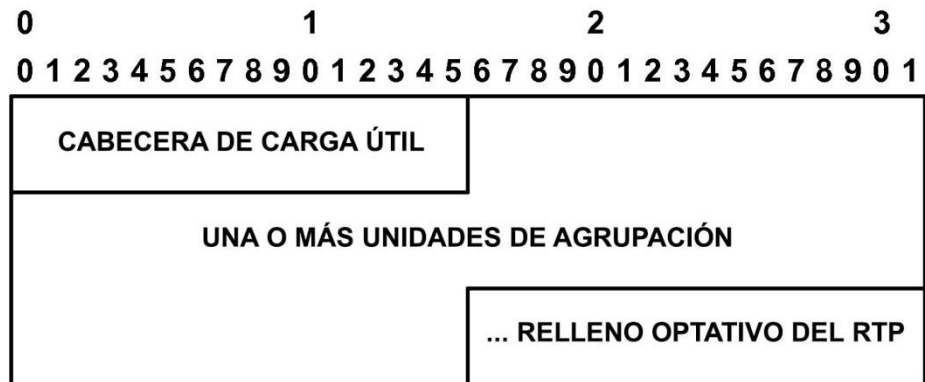


FIG. 1



|          |             |          |          |          |          |          |                          |          |          |          |          |          |            |          |          |
|----------|-------------|----------|----------|----------|----------|----------|--------------------------|----------|----------|----------|----------|----------|------------|----------|----------|
| <b>0</b> | <b>1</b>    | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b> | <b>6</b> | <b>7</b>                 | <b>0</b> | <b>1</b> | <b>2</b> | <b>3</b> | <b>4</b> | <b>5</b>   | <b>6</b> | <b>7</b> |
| <b>F</b> | <b>Tipo</b> |          |          |          |          |          | <b>IdentificadorCapa</b> |          |          |          |          |          | <b>TID</b> |          |          |

**FIG. 2**

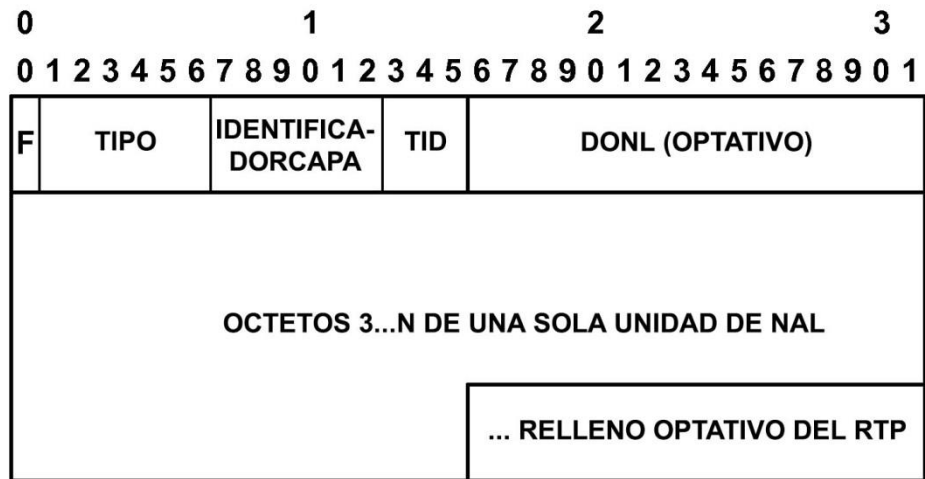


**FIG. 3**

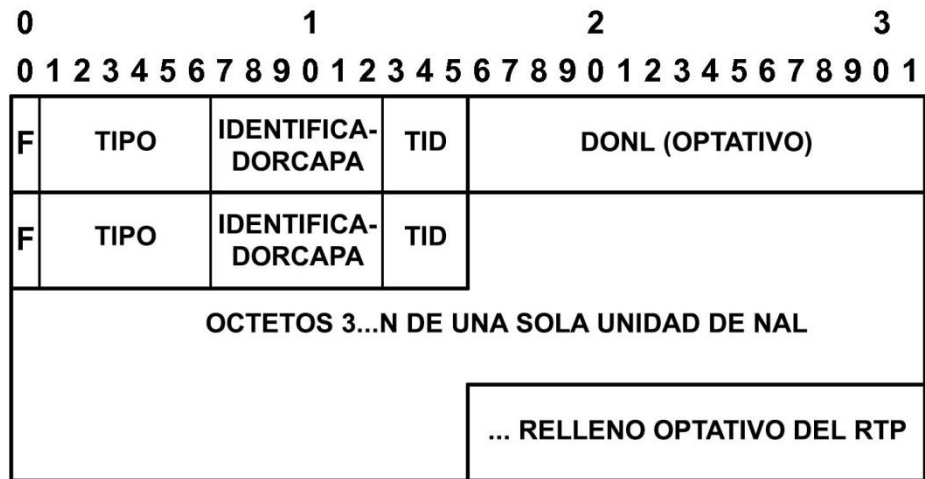


**FIG. 4**





**FIG. 6**



**FIG. 7**

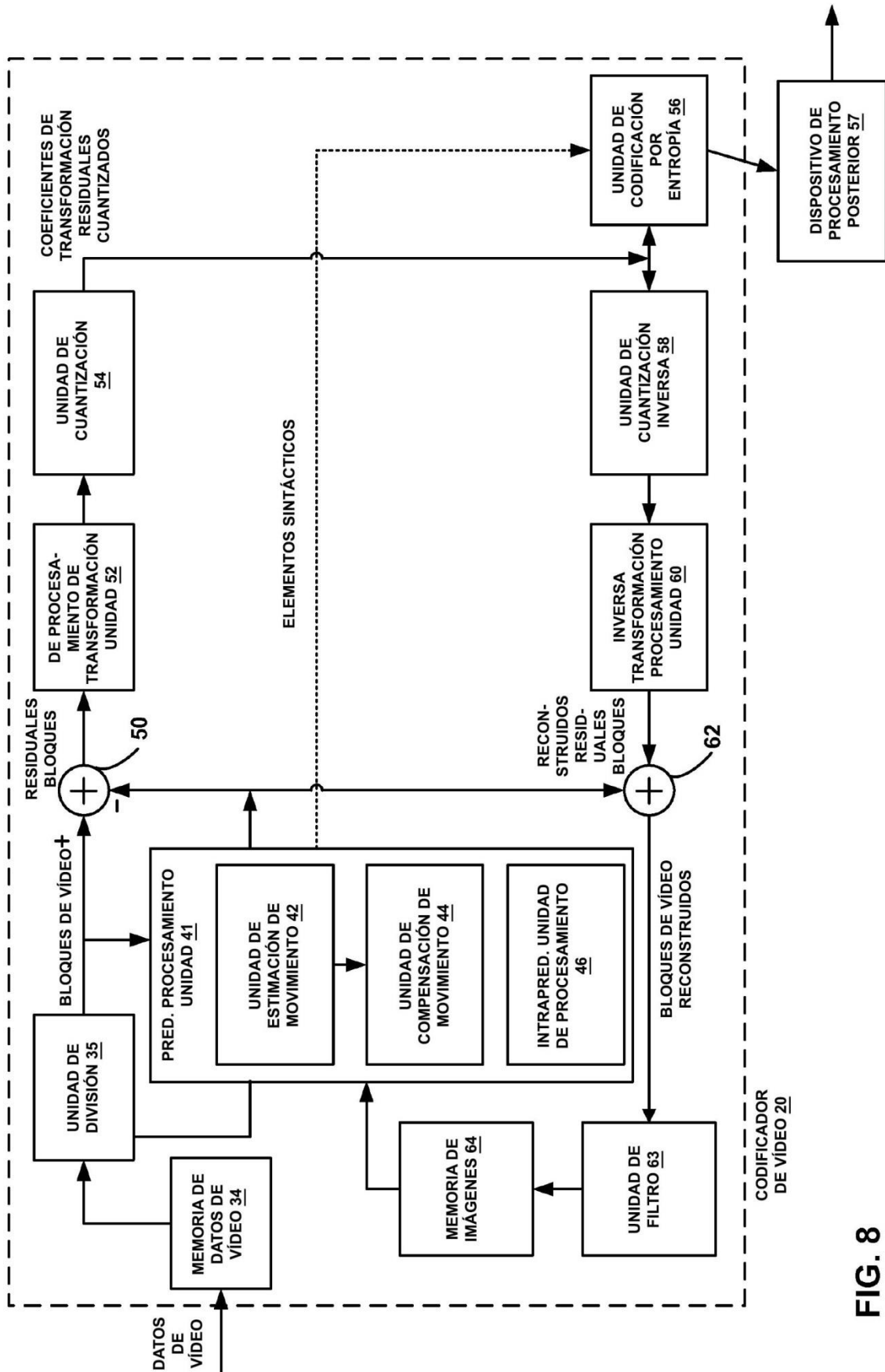


FIG. 8

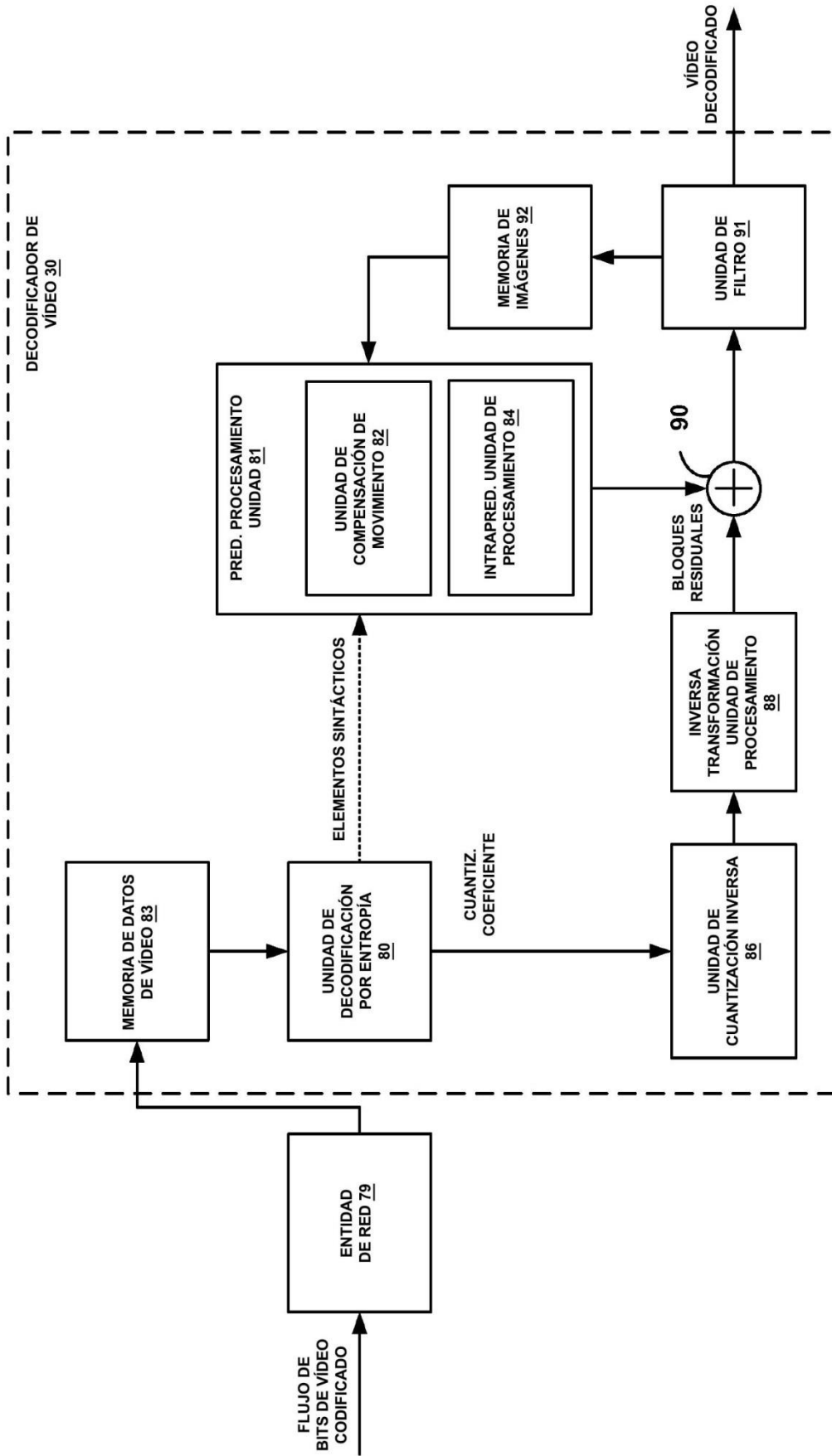


FIG. 9



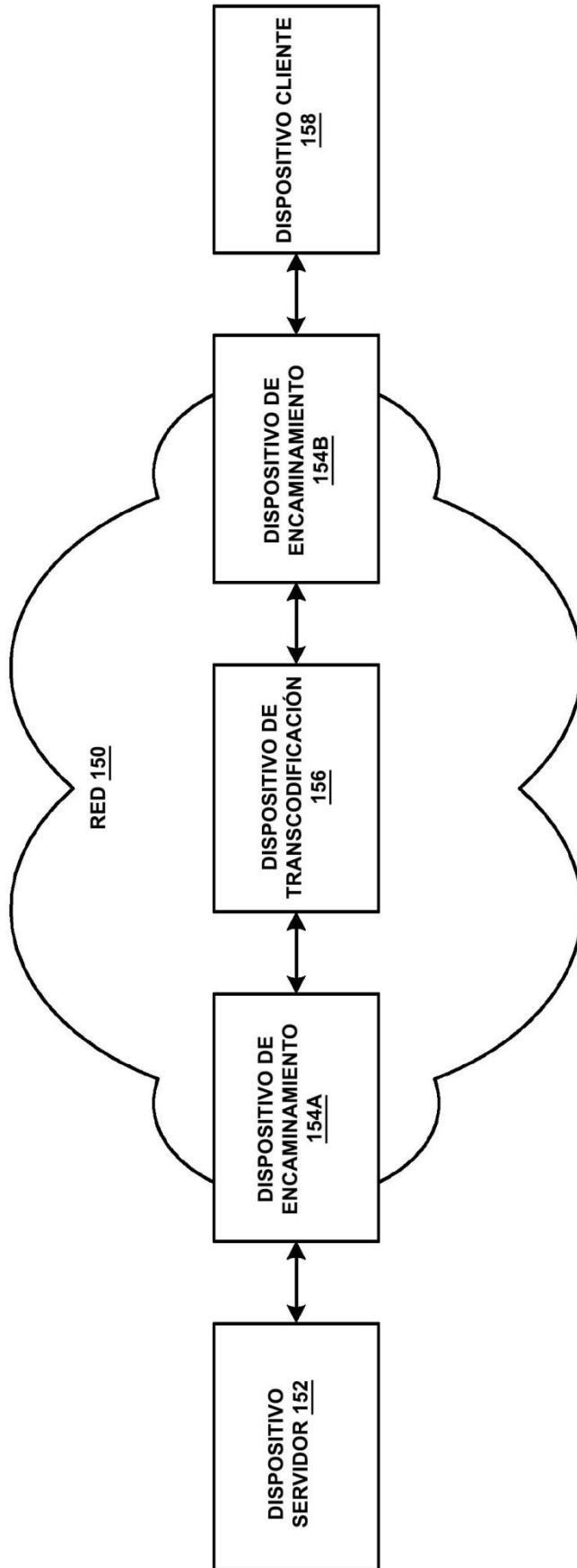


FIG. 10

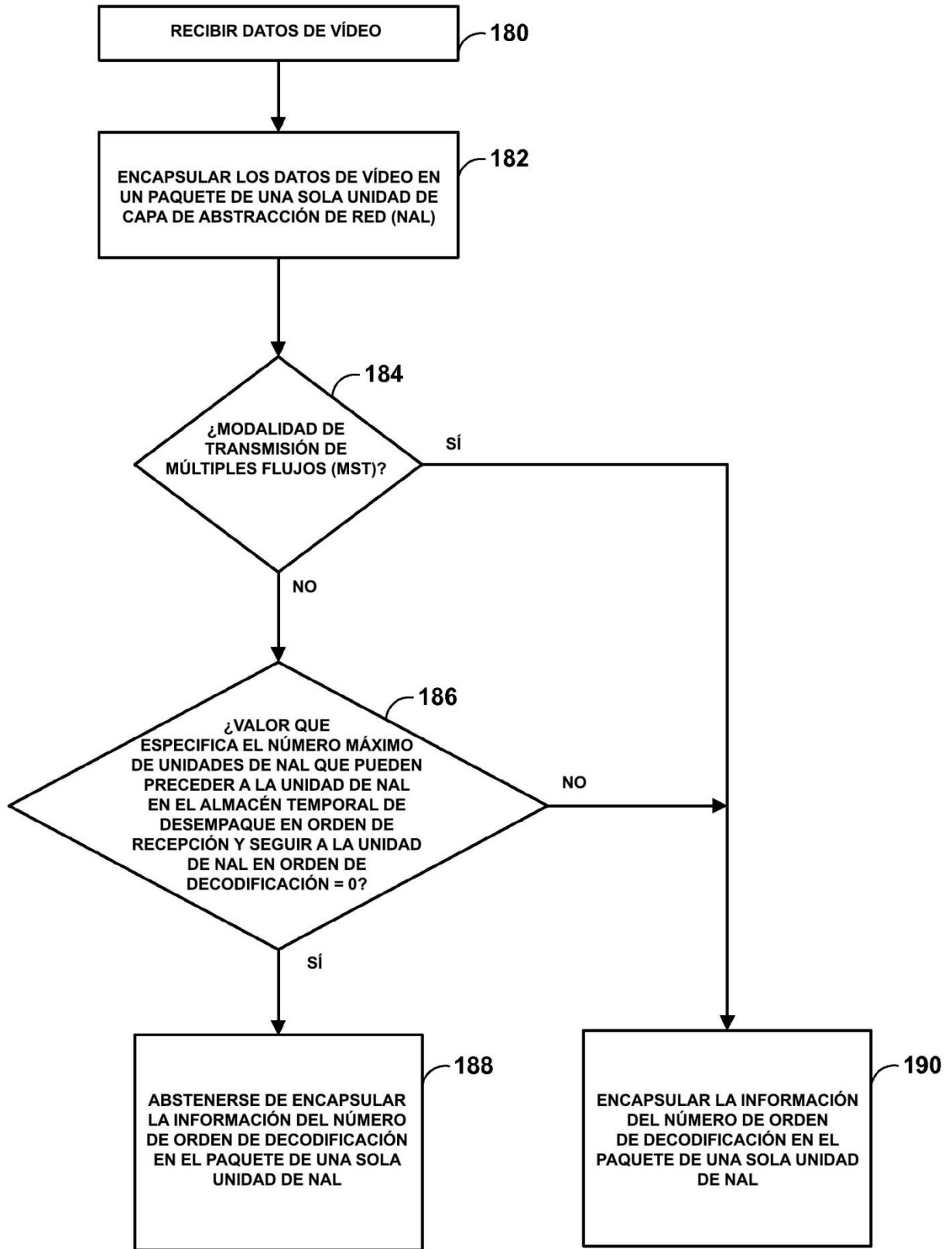


FIG. 11

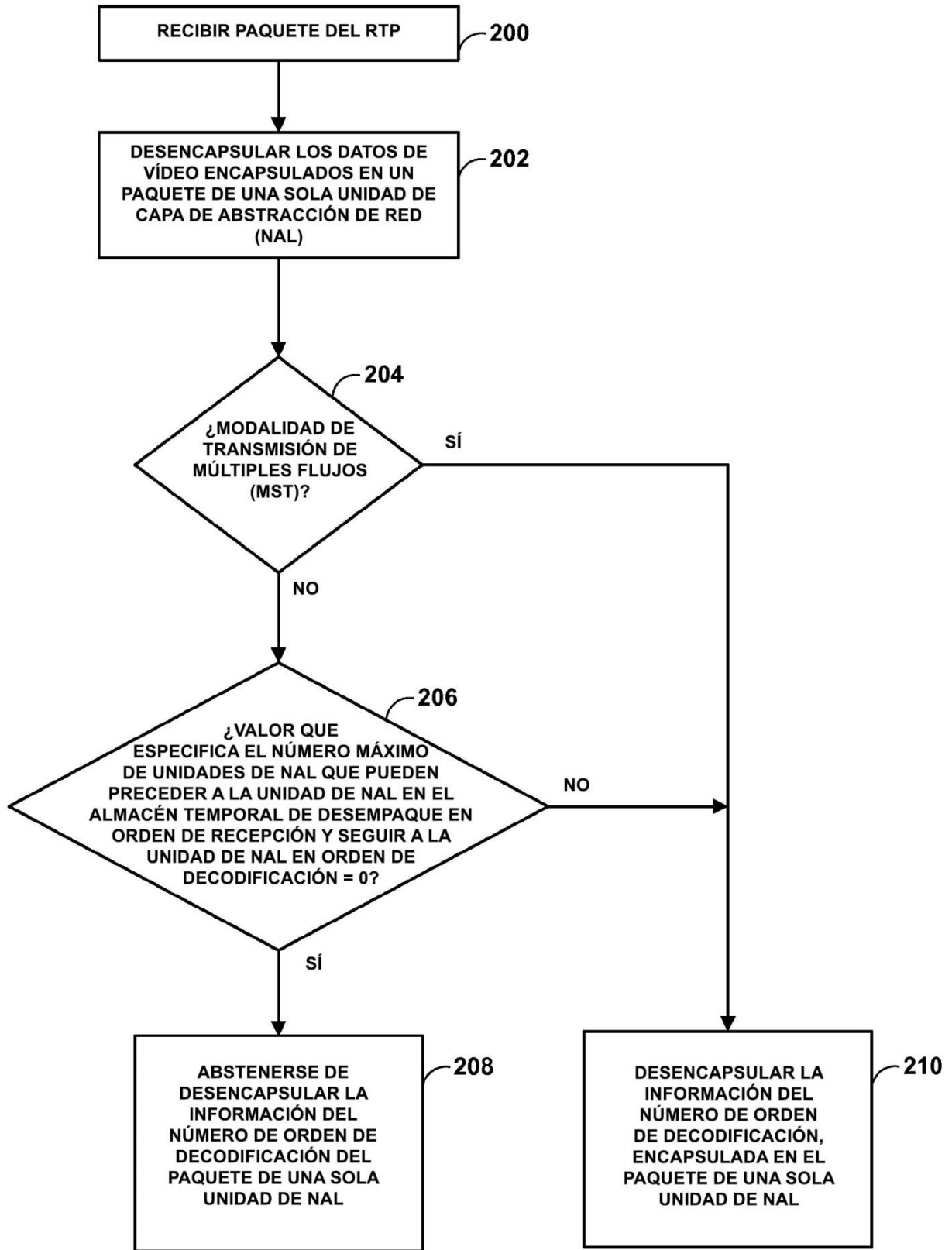


FIG. 12