(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 642 829**

(51) Int. CI.:

G06F 13/40 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **28.06.2014 PCT/CN2014/081070**

(87) Fecha y número de publicación internacional: **31.12.2014 WO14206356**

(96) Fecha de presentación y número de la solicitud europea: **28.06.2014 E 14816848 (7)**

(97) Fecha y número de publicación de la concesión europea: **23.08.2017 EP 2997483**

(54) Título: **Sistema y procedimiento para estructuras extendidas de interconexión de componentes periféricos exprés**

(30) Prioridad:

28.06.2013 US 201313931640

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
20.11.2017

(73) Titular/es:

HUAWEI TECHNOLOGIES CO., LTD. (100.0%)
Huawei Administration Building, Bantian Longgang District , Shenzhen, Guangdong 518129, CN

(72) Inventor/es:

SHAO, WESLEY

(74) Agente/Representante:

LEHMANN NOVO, María Isabel

DESCRIPCIÓN

Sistema y procedimiento para estructuras extendidas de interconexión de componentes periféricos exprés

SECTOR TÉCNICO

La presente invención se refiere, en general, a la industria informática; y más específicamente a sistemas, procedimientos, productos de programa informático y aparatos para estructuras extendidas de interconexión de componentes periféricos exprés (PCIe, peripheral component interconnect express).

ANTECEDENTES

La interconexión de componentes periféricos exprés (PCIe) es un estándar de bus informático de expansión en serie de alta velocidad ampliamente utilizado para acoplar diversos dispositivos de hardware (por ejemplo, dispositivos de almacenamiento, tarjetas de red, tarjetas de sonido y similares) a una unidad central de proceso (CPU, central processing unit) anfitrión. Dado que las configuraciones de memoria de la CPU anfitrión pueden ser específicas por fabricante, PCIe proporciona un estándar de entrada/salida (I/O, input/output) para conectar diversos dispositivos a la CPU. PCIe permite diversas mejoras sobre los estándares de bus anteriores (por ejemplo, PCI y PCI-eXtended). Por ejemplo, PCIe permite generalmente una mayor capacidad máxima del bus del sistema, una menor cantidad de patillas I/O, una menor dimensión del bus, funcionalidad nativa de conexión en caliente y otras ventajas.

Un problema con el estándar de bus PCIe es que cada estructura PCIe está limitada a una cantidad finita de recursos. Por ejemplo, cada espacio de memoria de direcciones de 32 bits de la estructura PCIe no puede superar los 4 GB de tamaño, y cada estructura puede tener solamente un máximo de 256 números de bus. Dado que PCIe funciona en conexiones en serie punto a punto, estas limitaciones constriñen directamente el número máximo de nodos (es decir, dispositivos) que se pueden acoplar a una estructura PCIe. Es decir, los números de bus para diversos dispositivos no pueden solapar, y cada dispositivo acoplado requiere un conjunto de números de bus únicos para funcionar. Varios números de bus en una estructura PCIe pueden estar reservados para usos particulares (por ejemplo, números de bus internos de los conmutadores PCIe, funcionalidad de conexión en caliente o similares), limitando adicionalmente el número de números de bus disponibles.

Además, la aparición de un fallo en cualquier componente acoplado a una estructura PCIe puede impactar sobre cualesquiera otros componentes acoplados más arriba o más abajo del componente averiado. A medida que aumenta el número de componentes y de controladores de software acoplados a la estructura PCIe, el tratamiento de los fallos se complica y la propagación de cualesquiera fallos puede conducir a un colapso de todo el sistema.

El documento US 2011/0131362 A1 describe un punto extremo integrado que tiene un puerto virtual acoplado entre una estructura de arriba y una estructura de dispositivo integrado que incluye una lógica multifunción para manejar diversas funciones para uno o varios bloques de propia intelectual (IP, intellectual property) acoplados a la estructura de dispositivo integrado. La estructura de dispositivo integrado tiene un canal principal para comunicar datos e información de comandos entre el bloque IP y la estructura de arriba, y un canal de banda lateral para comunicar información de banda lateral entre el bloque IP y la lógica multifunción.

COMPENDIO DE LA INVENCIÓN

Estos y otros problemas se resuelven o evitan en general, y se consiguen en general ventajas técnicas, mediante realizaciones preferidas de la presente invención, que proporcionan una estructura extendida de interconexión de componentes periféricos exprés.

De acuerdo con una realización de ejemplo, una topología de interconexión de componentes periféricos exprés incluye una estructura PCIe anfitrión que comprende un complejo raíz anfitrión. La estructura PCIe anfitrión incluye un primer conjunto de números de bus y un primer espacio de entrada/salida con mapeo a memoria (MMIO, memory mapped input/output) en una unidad central de proceso (CPU) anfitrión. Además, está dispuesta una estructura PCIe extendida, que incluye un punto extremo de complejo raíz (RCEP, root complex endpoint) como parte de un punto extremo de la estructura PCIe anfitrión, donde el RCEP es un puente entre la estructura PCIe extendida y la estructura PCIe anfitrión. La estructura PCIe extendida incluye asimismo un segundo conjunto de números de bus y un segundo espacio MMIO, respectivamente separado del primer conjunto de números de bus y del primer espacio MMIO. El segundo espacio MMIO está mapeado al primer espacio MMIO y el RCEP está configurado para aislar fallos originados en la estructura PCIe extendida.

De acuerdo con otra realización de ejemplo, proporcionada como antecedente para comprender la invención, una topología de interconexión de componentes periféricos exprés (PCIe) incluye una estructura PCIe extendida. La estructura PCIe extendida incluye un punto extremo de complejo raíz (RCEP). El RCEP está configurado para ser parte de un punto extremo de una estructura PCIe de primer nivel. Además, la estructura PCIe extendida comprende un espacio de entrada/salida con mapeo a memoria (MMIO) y un conjunto de números de bus.

De acuerdo con otra realización más de ejemplo, un procedimiento para conectar periféricos incluye proporcionar un punto extremo de complejo raíz (RCEP) que aloja una estructura extendida de interconexión de componentes periféricos exprés (PCIe) como parte de un punto extremo de una estructura PCIe anfitrión, donde el RCEP es un

puente entre la estructura PCIe extendida y la estructura PCIe anfitrión. La estructura PCIe extendida tiene un primer espacio MMIO que está separado de un segundo espacio MMIO de la estructura PCIe anfitrión. El procedimiento comprende además mapear el primer espacio MMIO al segundo espacio MMIO. El procedimiento comprende además interceptar, mediante el RCEP, la aparición de fallos más abajo en la estructura PCIe extendida.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Para una comprensión más completa de la presente invención, y de las ventajas de la misma, se hace referencia a continuación a las siguientes descripciones, tomadas junto con los dibujos adjuntos, en los cuales:

la figura 1 es un diagrama de bloques de una estructura de interconexión de componentes periféricos exprés (PCIe) de acuerdo con varias realizaciones de ejemplo;

las figuras 2A y 2B son diagramas de bloques de mapeo y direccionamiento del espacio de configuración PCIe para una estructura extendida, de acuerdo con diversas realizaciones de ejemplo;

las figuras 3A y 3B son diagramas de bloques de mapeo y direccionamiento del espacio de memoria de 32 bits para una estructura extendida, de acuerdo con diversas realizaciones de ejemplo.

las figuras 4A y 4B son diagramas de bloques de mapeo y direccionamiento del espacio de memoria de 64 bits para una estructura extendida, de acuerdo con diversas realizaciones de ejemplo; y

la figura 5 es un diagrama de sistema, de un sistema informático que tiene estructuras PCIe, de acuerdo con diversas realizaciones de ejemplo.

DESCRIPCIÓN DETALLADA DE REALIZACIONES ILUSTRATIVAS

A continuación se discuten en mayor detalle realizaciones de ejemplo que abarcan diversos aspectos de la innovación conseguida. Sin embargo, se deberá apreciar que la presente invención da a conocer muchos conceptos aplicables únicos y nuevos, que se pueden realizar en una amplia variedad de contextos específicos. Por consiguiente, las realizaciones específicas descritas en la presente memoria son tan sólo ilustrativas de maneras específicas de realizar, utilizar e implementar diversos aspectos de la presente invención, y no limitan necesariamente el alcance de la misma, salvo que se reivindique lo contrario.

Las siguientes diversas realizaciones a modo de ejemplo se describen en un contexto específico, a saber, una estructura estándar de bus de interconexión de componentes periféricos exprés (PCIe). Sin embargo, tal como se apreciará, dichas realizaciones de ejemplo se pueden extender asimismo a otras estructuras (por ejemplo, topologías de árbol invertido con restricciones de recursos).

Tal como se describe en la presente memoria, un complejo raíz es una estructura de hardware que sirve como puente entre una estructura PCIe y una unidad central de proceso (CPU) anfitrión. El complejo raíz puede estar integrado como parte de la CPU. Por ejemplo, la figura 5 muestra un complejo raíz 502 como parte de la CPU 500. El complejo raíz gestiona y transmite diversas solicitudes entre la CPU y los dispositivos (por ejemplo, los dispositivos 506 y 508) conectados a la misma estructura PCIe (por ejemplo, la estructura PCIe 504). El complejo raíz mapea asimismo varios dispositivos conectados (incluyendo posiciones de almacenamiento de los dispositivos, tales como registros y posiciones de memoria) sobre el espacio de memoria de la PCIe. Estos tipos de mapeo y espacio de memoria se conocen como espacio de entrada/salida con mapeo a memoria (MMIO).

Asimismo, tal como se utiliza en la presente memoria, el espacio MMIO puede incluir una parte de memoria direccionable que utiliza direcciones de 32 bits, que está limitada generalmente a los primeros 4 GB del espacio MMIO. El espacio MMIO puede incluir además una parte de memoria direccionable que utiliza direcciones de 64 bits, que se pueden mapear al espacio MMIO sobre los primeros 4 GB. Varias realizaciones de ejemplo descritas en la presente memoria incluyen una o varias estructuras adicionales de hardware de complejo raíz, como parte de los puntos extremos de la estructura PCIe anfitrión. Añadiendo funcionalidades de complejo raíz a puntos extremos (denominados punto extremo de complejo raíz (RCEP)) se pueden conectar estructuras PCIe adicionales para formar las estructuras PCIe extendidas que no se limitan a los recursos finitos de la PCIe anfitrión.

Por ejemplo, la figura 5 muestra un RCEP 508 y una estructura PCIe extendida 510 que conectan puntos extremos adicionales (por ejemplo, los dispositivos 512 y 514) a la CPU anfitrión 500. De acuerdo con varias realizaciones a modo de ejemplo, estos RCEP pueden ser utilizados para aislar fallos con el fin de que no contaminen dispositivos o estructuras limítrofes. Además, otras realizaciones de ejemplo dan a conocer que los RCEP y las estructuras PCIe extendidas pueden ser compatibles con controladores de dispositivos periféricos existentes; eliminando de ese modo la necesidad de actualizaciones de software para controladores de dispositivos existentes.

La figura 1 muestra un diagrama de bloques de una estructura PCIe de primer nivel 100 conectada a una estructura PCIe extendida 118, de acuerdo con varias realizaciones a modo de ejemplo. La estructura PCIe 100 puede incluir un complejo raíz anfitrión 102 que gestiona y transmite diversas solicitudes entre la CPU anfitrión y diversos dispositivos conectados a la estructura PCIe 100. Generalmente, el complejo raíz 102 posee un conjunto de 256 números de bus únicos y espacio MMIO, y mapea varios dispositivos conectados (incluyendo sus registros

asociados y/o hardware de memoria) a su espacio MMIO. Además, el complejo raíz 102 puede incluir una parte raíz 103, que es una estructura de puente de hardware que permite conexiones PCI a PCI entre la CPU anfitrión y los dispositivos conectados a la estructura PCIe 100 (por ejemplo, los puntos extremos 106 y 108). Aunque la figura 1 muestra el complejo raíz 102 teniendo solamente un puerto raíz 103, un típico complejo raíz 102 puede tener múltiples puertos raíz 103 (por ejemplo, entre cuatro y seis puertos raíz).

Tal como se muestra en la figura 1, el complejo raíz 102 está conectado eléctricamente a los puntos extremos 106 y 108 a través del puerto raíz 103 y del conmutador 104. Habitualmente, los puntos extremos 106 y 108 son estructuras que terminan la estructura PCIe 100. Dichos puntos extremos 106 y 108 pueden incluir dispositivos de almacenamiento, dispositivos de red, tarjetas de sonido, tarjetas de video y similares. Se muestra también en la figura 1 un conmutador 104, que puede incluir por lo menos un puerto de arriba (puerto 104A) y numerosos puertos de abajo (por ejemplo, puertos 104B) que interconectan diversos puntos extremos (por ejemplo, 106 y 108) al complejo raíz 102. Tal como se utiliza en general en la presente memoria, un puerto de arriba se refiere a un puerto que está dirigido hacia el complejo raíz anfitrión (por ejemplo, el complejo raíz 102) mientras que un puerto de abajo se refiere a un puerto que está dirigido alejándose del mismo.

El conmutador 104 puede incluir buses internos que permiten conectar múltiples dispositivos a un solo puerto raíz 103 manteniendo al mismo tiempo una conexión en serie punto a punto utilizada por el estándar PCIe. Aunque la figura 1 muestra solamente un conmutador 104 que conecta dos puntos extremos 106 y 108 al puerto raíz 103, varias realizaciones a modo de ejemplo contemplan múltiples puertos raíz conectados a cualquier número de conmutadores, donde cada conmutador se puede conectar a cualquier número variable de puntos extremos. Alternativamente, o conjuntamente, un puerto raíz 103 puede estar conectado directamente a un punto extremo. Por consiguiente, la configuración mostrada en la figura 1 y otras figuras de la presente memoria, tiene solamente propósitos ilustrativos y no pretende limitar o reducir de otro modo el alcance de la presente invención, salvo que se reivindique explícitamente lo contrario.

Tal como se muestra en la figura 1, el punto extremo 106 puede ser un punto extremo de complejo raíz (RCEP) de acuerdo con realizaciones a modo de ejemplo, que proporciona un puente de hardware entre la estructura PCIe extendida 118 y una estructura PCIe de primer nivel 100. Es decir, el RCEP 106 puede alojar una estructura PCIe extendida 118. De acuerdo con dichas realizaciones, el RCEP 106 puede ser similar, en construcción y lógica, a un complejo raíz anfitrión (por ejemplo, el complejo raíz 102). De este modo, el RCEP 106 puede incluir su propio conjunto de números de bus, espacio de configuración PCIe, y espacio MMIO -diferentes de los del complejo raíz 102. Por lo tanto, se pueden conectar dispositivos adicionales al RCEP 106 incluso cuando el complejo raíz 102 no tiene la cantidad suficiente de recursos disponibles (por ejemplo, números de bus).

Por ejemplo, los puntos extremos 114 y 116 se pueden conectar eléctricamente al RCEP 106 a través del conmutador 112 y del puerto raíz 110. Tal como se ha indicado anteriormente, los puntos extremos 114 y 116 pueden ser casi cualquier tipo de dispositivos periféricos, incluyendo dispositivos de almacenamiento, dispositivos de red, tarjetas de sonido, tarjetas de video y similares. De manera similar a las estructuras PCIe anteriores, los puntos extremos 114 y 116 pueden simplemente terminar la estructura PCIe extendida 118. Alternativamente, y de acuerdo con realizaciones a modo de ejemplo, los puntos extremos 114 y/o 116 pueden incluir otro RCEP que tenga su propio conjunto de números de bus y espacio MMIO; por lo tanto, formando otra estructura PCIe extendida. Por consiguiente, los RCEP añaden esencialmente funcionalidad de puerta de enlace a un punto extremo PCIe; y por lo tanto, teóricamente, permiten acoplar un número virtualmente ilimitado de nodos al complejo raíz anfitrión.

Aunque la figura 1 muestra la estructura PCIe extendida 118 teniendo solamente un punto raíz 110, un conmutador 112 y dos puntos extremos 114 y 116, diversas realizaciones pueden incluir una estructura PCIe extendida con múltiples puertos raíz, múltiples conmutadores por estructura extendida y múltiples puntos extremos por conmutador.

De acuerdo con realizaciones a modo de ejemplo, el MMIO del RCEP 106 puede incluir una parte direccionable que utiliza direcciones de 32 bits (denominada espacio de memoria de 32 bits) y una parte que utiliza direcciones de 64 bits (denominada espacio de memoria de 64 bits). De acuerdo con dichas realizaciones, el espacio de configuración PCIe, el espacio de memoria de 32 bits y el espacio de memoria de 64 bits del RCEP 106 se pueden mapear al espacio MMIO de 64 bits de un estructura PCIe anfitrión 100 (es decir, la parte del espacio MMIO de la estructura PCIe 100 que es direccionable utilizando direcciones de 64 bits). Por lo tanto, de acuerdo con dichas realizaciones, se puede acceder al RCEP 106 desde el espacio MMIO de una estructura PCIe anfitrión 100. El mapeo y la enumeración de la estructura extendida del RCEP 106 se pueden realizar utilizando controladores de punto extremo asociados con el RCEP 106, tal como se explica en mayor detalle a continuación.

De acuerdo con otras realizaciones a modo de ejemplo, el RCEP 106 puede incluir asimismo un mecanismo de tratamiento de fallos que resuelve cualesquiera fallos que se produzcan en sus dispositivos situados abajo (por ejemplo, los puntos extremos 114 y 116). Por lo tanto, en dichas realizaciones, los fallos pueden ser contenidos por el RCEP 106 y no se propagan hacia arriba a la estructura PCIe 100, y el RCEP 106 puede actuar como una barrera de fallos. Además, el RCEP 106 puede generar una interrupción de error para notificar cualesquiera fallos al complejo raíz anfitrión 102. El dichas realizaciones, estas interrupciones de error pueden ser utilizadas como un mecanismo de notificación, y cualesquiera fallos PCIe que se produzcan en los dispositivos de abajo del RCEP 106 puede ser tratados por el RCEP 106 y no pasados hacia arriba al complejo raíz anfitrión 102. Los detalles

específicos relativos a la notificación de interrupciones y el tratamiento de fallos pueden ser específicos de cada implementación y variar entre plataformas informáticas/complejos raíz. Por ejemplo, los actuales estándares PCIe dejan los detalles de implementación relativos a cómo el complejo raíz trata los fallos, abiertos a diferentes implementaciones que varían por fabricante, de plataformas informáticas/complejos raíz. Por lo tanto, los detalles de implementación de los mecanismos de tratamiento de fallos y notificación de errores del RCEP 106 pueden, análogamente, estar abiertos a diferentes implementaciones en función de las configuraciones aplicables de plataforma informática/complejo raíz.

El mapeo y al acceso al espacio de configuración PCIe para la estructura extendida 118 se pueden realizar utilizando cualquier configuración adecuada. Por ejemplo, las figuras 2A y 2B muestran un ejemplo de mapeo y acceso de espacio de configuración PCIe de acuerdo con realizaciones a modo de ejemplo, que soporta la utilización de MMIO para acceder al espacio de configuración PCI. Generalmente, el espacio de direcciones físicas 200 es el espacio de direcciones físicas de la CPU anfitrión, que puede tener, por ejemplo, 2^n-1 octetos de tamaño (donde "n" es el número de bits de dirección física implementados por la CPU anfitrión). Tal como se muestra, el espacio de direcciones físicas 200 puede incluir una parte del espacio 202 direccionable mediante direcciones físicas de 32 bits, con una parte restante 204 direccionable mediante direcciones físicas de 64 bits. En dicha realización, la parte 202 puede ocupar los primeros 4 GB de espacio de direcciones 200, mientras que la parte 204 puede ocupar el resto del espacio de direcciones 200. Por otra parte, la parte 204 puede incluir además una parte asignada a un espacio MMIO de 64 bits 206 de la estructura PCIe anfitrión, de primer nivel (por ejemplo, la estructura PCIe 100). Por supuesto, se contemplan en la presente memoria otras asignaciones de memoria, configuraciones y mapeos del espacio MIMO a la estructura PCIe; y por lo tanto, cualquier implementación específica descrita en la presente memoria se utiliza solamente con fines ilustrativos -salvo que se reivindique explícitamente lo contrario.

De acuerdo con realizaciones a modo de ejemplo, la totalidad (o cualquier parte) de las funciones de un dispositivo conectado a la estructura PCIe 100, o la estructura extendida 118, se pueden mapear a sus propios 256 MB dedicados del espacio de configuración de la respectiva estructura. Dicho espacio puede ser direccionable, por ejemplo, conociendo los números de función de bus PCI de 8 bits, dispositivo de 5 bits y función de 3 bits para una función de dispositivo particular. Este tipo de direccionamiento de funciones se puede denominar en la presente memoria direccionamiento de bus/dispositivo/función (BDF), que permite un total de 256 números de bus, 32 dispositivos y 8 funciones para cada estructura PCIe. Generalmente, en dichas realizaciones, el número de dispositivo se puede ajustar a 0 de acuerdo con las estructuras estándar de bus de PCIe debido a la estructura de la conexión punto a punto en serie del PCIe. Además, de acuerdo con dichas realizaciones, cada función de dispositivo se puede habilitar para 4 KB de registros de configuraciones.

Tal como se muestra, el espacio de configuración PCIe para la estructura PCIe de primer nivel (por ejemplo, estructura 100) puede ocupar 256 MB del espacio de direcciones 208 en parte del espacio 202. En dichas realizaciones, el espacio de configuración PCIe para la PCIe extendida (por ejemplo, la estructura extendida 118) se puede mapear a 256 MB del espacio de direcciones 210 en el espacio MMIO de 64 bits 206. Además, cualesquiera transacciones MMIO en el espacio de direcciones 208 ó 210 pueden ser tratadas como transacciones de acceso de configuración PCIe para la estructura anfitrión PCIe 100 o bien la estructura extendida 118, respectivamente, mediante su correspondiente complejo raíz.

En dichas realizaciones, las direcciones del espacio de configuración PCIe extendido 210 se pueden iniciar en el valor base 212. Por lo tanto, los registros del espacio de configuración de una función de dispositivo PCIe situada en un número de bus B, número de dispositivo D y número de función F se pueden iniciar, por ejemplo, en Base + (B + D + F)*4K. Alternativamente, se contemplan asimismo otras configuraciones adecuadas para direccionar registros de espacio de dispositivos, y la descripción de direccionamiento BDF en este caso se utiliza solamente con fines ilustrativos.

La figura 2B muestra un formato de ejemplo de dirección física de dispositivo PCIe 220 de acuerdo con realizaciones de ejemplo. El formato 220 puede incluir un campo de base 222 (situado, por ejemplo, entre los bits n-1 y 28), un campo de bus 224 (situado, por ejemplo, entre los bits 27 y 20), un campo de dispositivo 226 (situado, por ejemplo, entre los bits 19 y 15) y un campo de función 228 (situado, por ejemplo, entre los bits 14 y 12). Además, en dichas realizaciones, los bits 11 a 0 pueden contener un desplazamiento del registro de espacio de configuración (por ejemplo, en el interior del dispositivo) al que se está accediendo, que se puede incluir, por ejemplo, con fines de alineamiento de memoria.

El mapeo y acceso al espacio de memoria de 32 bits para la estructura PCIe extendida 118 se puede realizar utilizando cualquier configuración adecuada. Por ejemplo, las figuras 3A y 3B muestran un ejemplo de mapeo y acceso al espacio de memoria de 32 bits PCIe, de acuerdo con realizaciones a modo de ejemplo. Generalmente, el acceso al espacio de memoria de 32 bits se puede utilizar para bancos de registros de dispositivo a los que se puede acceder con direcciones de 32 bits. En dichas realizaciones, las transacciones correspondientes se pueden marcar como acceso a espacio de memoria PCIe de 32 bits, por ejemplo, en una cabecera de paquete PCIe. Por supuesto, se contemplan en la presente memoria otras asignaciones de memoria, configuraciones y mapeo del espacio MIMO a la estructura PCIe; y por lo tanto, cualquier implementación específica descrita en la presente memoria se utiliza solamente con fines ilustrativos -salvo que se reivindique explícitamente lo contrario.

Bajo los estándares PCIe actuales, el tamaño máximo para un espacio de direcciones de 32 bits es de 4 GB. Además, de acuerdo con los actuales estándares PCIe, en ciertas plataformas informáticas (por ejemplo, plataformas x86), la memoria de 32 bits para la estructura PCIe de primer nivel 100 se puede compartir con su espacio de configuración PCIe y, por lo tanto, puede tener solamente 256 MB de tamaño.

De acuerdo con realizaciones a modo de ejemplo, la estructura PCIe extendida 118 puede tener su propio espacio de memoria de 32 bits diferente del espacio de direcciones físicas 200 de la CPU anfitrión. Además, tal como se muestra en la figura 3A, el RCEP 106 puede mapear el espacio de direcciones de 32 bits de la estructura extendida 118 a una ventana de direcciones de 64 bits superior en el espacio MMIO de 64 bits 206 del anfitrión, tal como se muestra mediante el espacio de memoria 302. El dichas realizaciones, el espacio de memoria 302 puede ser independiente del espacio de memoria de 32 bits 300 para la estructura de primer nivel (por ejemplo, la estructura PCIe 100). Además, el espacio de memoria 300 puede estar situado en el espacio direccionable de 32 bits 202 de la CPU anfitrión, que puede estar dentro del mismo espacio de 4G inferior que el espacio de configuración PCIe 208. Tal como se muestra, el espacio de memoria 302 se inicia en una dirección base 304 y puede tener hasta 4 GB de tamaño. Por lo tanto, de acuerdo con dichas realizaciones, un banco de registros de dispositivo (por ejemplo, el banco de registros 308) se puede direccionar conociendo un desplazamiento apropiado (por ejemplo, desplazamiento 306) del banco de registros. Es decir, la dirección del banco de registros 308 puede ser base 304 + desplazamiento 306. Alternativamente, se contemplan asimismo otros esquemas adecuados para direccionar el banco de registros 308.

La figura 3B muestra formatos 310 y 312 de dirección física de bancos de registros para el espacio de direcciones de anfitrión y el espacio de direcciones de estructura extendida, respectivamente, de acuerdo con realizaciones de ejemplo. El formato 310 puede incluir un campo de base 314 (por ejemplo, ocupando los bits n-1 a 32) y un campo de desplazamiento 316 (por ejemplo, ocupando los bits 31 a 0). El dichas realizaciones, el formato 310 puede ser utilizado para direccionar un banco de registros de dispositivo mapeado al espacio MMIO 206. El formato 312 puede contener solamente un campo de desplazamiento 318 (por ejemplo, ocupando 32 bits) para direccionar el banco de registros de dispositivo en la estructura extendida (por ejemplo, estructura 118). Además, antes de poner un paquete PCIe en la estructura extendida 118, el RCEP 106 puede vaciar el formato 310 de 32 bits superior (que puede tener 64 bits de longitud) y marcar el paquete PCIe vaciado como un paquete de direcciones de 32 bits.

El mapeo y acceso al espacio de memoria de 64 bits para la estructura PCIe extendida 118 se puede realizar utilizando cualquier configuración adecuada. Por ejemplo, las figuras 4A y 4B muestran un ejemplo de mapeo y acceso al espacio de memoria de 64 bits, de acuerdo con realizaciones a modo de ejemplo. Generalmente, el espacio de memoria de 64 bits se puede utilizar para bancos de registros de dispositivo a los que se puede acceder con direcciones de 64 bits o direcciones de 32 bits. Se debe observar que si se asigna un intervalo de 32 bits a una dirección de 64 bits, el hardware PCIe puede utilizar un esquema de configuración de espacio de memoria PCI de 32 bits para acceder (por ejemplo, la configuración descrita en las figuras 3A y 3B). En dichas realizaciones, las transacciones correspondientes se pueden marcar como acceso a espacio de memoria PCIe de 64 bits en una cabecera de paquete PCIe. Además, por razones históricas, el espacio de memoria de 64 bits puede estar situado generalmente sobre los 4 GB inferiores del espacio de direcciones físicas. Por supuesto, se contemplan en la presente memoria otras asignaciones de memoria, configuraciones y mapeo del espacio MIMO a la estructura PCIe; y por lo tanto, cualquier implementación específica descrita en la presente memoria se utiliza solamente con fines ilustrativos -salvo que se reivindique explícitamente lo contrario.

En dichas realizaciones, la estructura PCIe extendida 118 puede tener su propio espacio de memoria de 64 bits diferente del espacio de direcciones físicas 200 de la CPU anfitrión. Tal como se muestra en la figura 4A, el RCEP 106 puede mapear el espacio de direcciones de 64 bits de la estructura extendida 118 al espacio MMIO de 64 bits 206 de la CPU anfitrión, tal como se muestra mediante el espacio de memoria 402. Además, el espacio de memoria 402 puede ser independiente del espacio de memoria PCIe de 64 bits 400 de la estructura de primer nivel 100, que puede estar asimismo situado en el espacio MMIO de 64 bits 206. En estas realizaciones, el espacio de memoria 402 se puede iniciar en una dirección base 404 y puede tener, por ejemplo, un tamaño de m veces 4 GB (de m es una constante, que significa que este espacio 402 puede tener un tamaño de una o varias veces 4 GB). Además, un banco de registros de dispositivo (por ejemplo, el banco de registros 408) puede ser direccionado conociendo un desplazamiento apropiado (por ejemplo, el desplazamiento 406) del banco de registros. De este modo, de acuerdo con dichas realizaciones, una dirección del banco de registros 408 puede ser base 404 + desplazamiento 406. Alternativamente, se contemplan asimismo otros esquemas adecuados para direccionar el banco de registros 408.

La figura 4B muestra formatos 410 y 416/418 de direcciones físicas del banco de registros para el espacio de direcciones de la CPU anfitrión y el espacio de direcciones de la estructura extendida, respectivamente, de acuerdo con realizaciones de ejemplo. El formato 410 puede incluir un campo de base 412 (por ejemplo, ocupando los bits n-1 a p) y un campo de desplazamiento 414 (por ejemplo, ocupando los bits p-1 a 0). El valor de la constante p puede ser de 32 + ln(m), que se puede redondear al siguiente valor entero superior. El formato 410 se puede utilizar para direccionar un banco de registros de dispositivo mapeado al espacio MMIO de 64 bits 206.

En ciertas realizaciones a modo de ejemplo, direccionar el banco de registros de dispositivo en la estructura extendida 118 se puede realizar utilizando el formato 416. Por ejemplo, el formato 416 se puede utilizar si la dirección física de base está alineada en tamaño con la configuración del espacio de memoria de la estructura

extendida 118. Utilizando el formato 416, el RCEP 106 puede vaciar los bits superiores (por ejemplo, bits 63 a p) del formato 410 para formar una dirección de 64 bits para la estructura extendida 118. En otras realizaciones a modo de ejemplo, el formato 418 puede ser utilizado si la dirección física de base no está alineada en tamaño con el espacio de memoria de la estructura extendida 118. En dichas realizaciones, para compensar el desalineamiento de las direcciones físicas, se puede añadir un desplazamiento 420 a la dirección de base 422 del sistema de 64 bits. Además, si la dirección de base está por lo menos alineada en 4 GB (por ejemplo, los 32 bits inferiores son 0), el ajuste de tamaño puede llevarse a cabo solamente para los 32 bits superiores de la dirección de base.

En una realización a modo de ejemplo, el espacio de configuración PCIe 210, el espacio de memoria de 32 bits 302 y el espacio de memoria de 64 bits 402 de la estructura extendida 118 pueden solapar en la CPU anfitrión. En dichas realizaciones, el RCEP 106 puede solicitar a la CPU anfitrión una ventana de mapeo común lo suficientemente grande para acomodar todos los intervalos de direcciones deseados (por ejemplo, espacios 210, 302 y 402), y el RCEP 106 puede a continuación dividir la ventana de mapeo común como se requiera, en varios intervalos de direcciones deseados.

De acuerdo con realizaciones a modo de ejemplo, la estructura extendida 118 puede soportar interrupciones de dispositivo, que se pueden tratar utilizando cualquier procedimiento adecuado. Por ejemplo, la estructura extendida 118 puede utilizar una configuración de interrupción señalizada por mensaje (MSI, message signaled interrupt). En dichas realizaciones, las MSI que se originan en dispositivos conectados a la estructura extendida 118 (por ejemplo, los puntos extremos 114 y 116) se pueden suministrar a puertos raíz aplicables (por ejemplo, el puerto raíz 110), de acuerdo con el estándar de bus PCIe. Además, el puerto raíz 110 del RCEP 106 puede tener una ventana de direcciones preasignada para MSI. En dichas realizaciones, cuando una dirección de escritura de memoria coincide con la ventana de direcciones MSI preasignada, la transacción se puede reconocer como una interrupción. Además, el puerto raíz 110 puede recoger todas las MSI que se originan en sus estructuras situadas más abajo, y depositarlas en una cola (donde la cola puede estar situada en la memoria de la CPU anfitrión en el espacio de direcciones físicas 200). El puerto raíz 110 puede señalizar a continuación una interrupción independiente, que puede ser asimismo una MSI, a su puerto raíz de arriba (por ejemplo, el puerto raíz anfitrión 103). El puerto raíz anfitrión 102 puede activar a continuación un gestor de software apropiado, en función de la interrupción recibida. Un gestor de interrupción del puerto raíz 110 del RCEP 106 puede examinar a continuación la cola MSI en la memoria principal, determinar el dispositivo de origen (por ejemplo, el punto extremo 114 ó 116) y despachar el gestor de interrupciones apropiado del controlador del dispositivo. Por supuesto, en la presente memoria se contemplan otros esquemas para manejar interrupciones de dispositivo; y por lo tanto, cualquier implementación específica descrita en la presente memoria se utiliza solamente con fines ilustrativos -salvo que se reivindique explícitamente lo contrario.

De acuerdo con otras realizaciones a modo de ejemplo, la estructura extendida 118 puede soportar asimismo acceso directo a memoria (DMA, direct memory access), que se puede tratar utilizando cualquier procedimiento adecuado. En dichas realizaciones, las transacciones DMA pueden incluir solicitudes de lectura, finalizaciones de lectura y solicitudes de escritura. Los paquetes PCIe pueden llevar una dirección física de sistema o una dirección virtual de IO (entrada/salida) traducida por una IOMMU (unidad de gestión memoria de entrada/salida (input/output memory management unit)). Además, los ID de solicitante PCIe pueden ser por estructura. De este modo, el ID de solicitante se puede sustituir por el ID del RCEP 106 cuando una solicitud atraviesa una delimitación de estructura PCIe y sube a una estructura PCIe anfitrión 100. Es decir, en la estructura extendida 118, el ID de solicitante puede ser el ID del dispositivo de punto extremo (por ejemplo, el punto extremo 114 ó 116). Cuando la solicitud es transmitida hacia arriba al complejo raíz 102, el ID de solicitante se puede sustituir por el ID del RCEP 106.

En dichas realizaciones, las escrituras DMA se refieren al desplazamiento de datos de un dispositivo (por ejemplo, el punto extremo 114 ó 116) a la memoria de la CPU anfitrión. El RCEP 106 puede sustituir el ID de dispositivo con el ID del RCEP 106 cuando la solicitud se pasa hacia arriba mediante el RCEP 106 al complejo raíz 102 y a la CPU anfitrión. Además, las lecturas DMA se refieren al desplazamiento de datos desde la memoria de la CPU anfitrión al dispositivo. En dichas realizaciones, el RCEP 106 puede utilizar un marcador de hardware para rastrear todas las solicitudes de lectura asignando etiquetas de transacción (por ejemplo, como parte de paquetes de solicitud) para las transacciones a la estructura 100. Estas etiquetas de transacción se pueden vincular a entradas de marcador del RCEP 106 y pueden ser utilizadas para registrar los ID de solicitante de paquetes de solicitud de lectura originados en la estructura extendida 118. La finalización de los datos recibidos por el RCEP 106 desde el complejo raíz 102 puede llevar la misma etiqueta de transacción que la correspondiente solicitud de lectura, según dichas realizaciones. Por lo tanto, en dichas realizaciones, se pueden utilizar etiquetas de transacción para cotejarlas con entradas de marcador para determinar el ID de dispositivo apropiado utilizado en la estructura extendida 118. Por supuesto, se contemplan en la presente memoria otros esquemas para manejar solicitudes DMA; y por lo tanto, cualquier implementación específica descrita en la presente memoria se utiliza solamente con fines ilustrativos -salvo que se reivindique explícitamente lo contrario.

Por lo tanto, utilizando los diversos mecanismos de acceso a configuración PCIe, acceso de memoria, DMA e interrupciones, descritos en las realizaciones a modo de ejemplo de los párrafos anteriores, se pueden utilizar dispositivos RCEP para alojar estructuras PCIe extendidas y dispositivos adicionales conectados (por ejemplo, dispositivos RCEP adicionales y/o dispositivos periféricos) en un complejo raíz anfitrión. Los dispositivos RCEP pueden ser similares a una típica lógica de complejo raíz PCIe. Cada estructura PCIe extendida puede tener su propio espacio MMIO y un conjunto de números de bus. Por lo tanto, el número total de dispositivos que se pueden

acoplar a todos los conectados a una CPU anfitrión puede no limitarse a la cantidad de números de bus disponibles de la estructura PCIe anfitrión. En dichas realizaciones, el espacio MMIO de cada estructura extendida se puede mapear al espacio MMIO de 64 bits de su estructura padre (por ejemplo, la estructura padre para la estructura extendida 118 es la estructura de primer nivel 100) para facilitar el acceso. Además, la enumeración de estructura de las estructuras extendidas se puede conseguir por medio del controlador de dispositivo de punto extremo RCEP. En dichas realizaciones, cada RCEP puede tratar los fallos que se originan en la estructura de complejo raíz extendida aplicable. Por lo tanto, se puede conseguir el aislamiento de los fallos, de tal modo que los fallos de estructuras más abajo pueden ser interceptados en el límite de la estructura de una estructura PCIe extendida y no propagarse hacia arriba.

De acuerdo con realizaciones a modo de ejemplo, casi cualquier dispositivo periférico (por ejemplo tarjetas de sonido, tarjetas de video, controladores de red, tarjetas de memoria y similares) se puede conectar a la estructura extendida y no tiene que cambiar su software controlador. En dichas realizaciones, la estructura PCIe extendida interactúa con los dispositivos periféricos del mismo modo que una estructura PCIe de primer nivel. Además, el software de la CPU anfitrión (por ejemplo, un sistema operativo) y los controladores de RCEP se pueden modificar y/o crear para establecer mapeo de registros, mapeo de direcciones DMA, implementar gestores de interrupciones a través del RCEP, y otras funciones similares. Por lo tanto, no es necesario que los dispositivos periféricos sean notificados de que están conectados a una estructura PCIe extendida en lugar de a una estructura PCIe de primer nivel. De este modo, la estructura PCIe extendida y el RCEP pueden ser compatibles con controladores de dispositivo periférico existentes.

Aunque esta invención se ha descrito haciendo referencia realizaciones ilustrativas a modo de ejemplo, esta descripción no se deberá interpretar en sentido limitativo. Diversas modificaciones y combinaciones de las realizaciones ilustrativas a modo de ejemplo, así como otras realizaciones de la invención, serán evidentes para los expertos en la materia tras hacer referencia a la descripción. Por lo tanto, se prevé que las reivindicaciones adjuntas abarquen cualesquiera de dichas modificaciones o realizaciones.

REIVINDICACIONES

1. Una topología extendida de interconexión de componentes periféricos exprés (PCIe), que comprende:

una estructura PCIe anfitrión (100, 504) que comprende un complejo raíz anfitrión (102, 502), teniendo la estructura PCIe anfitrión (100, 504) un primer conjunto de números de bus y un primer espacio de entrada/salida con mapeo de memoria (MMIO) en una unidad central de proceso (CPU) anfitrión; y

una estructura PCIe extendida (118, 510) que comprende un punto extremo de complejo raíz (RCEP) (106, 508) como parte de un punto extremo de la estructura PCIe anfitrión (100, 504), donde el RCEP (106, 508) es un puente entre la estructura PCIe extendida (118, 510) y la estructura PCIe anfitrión (100, 504), teniendo la estructura PCIe extendida (118, 510) un segundo conjunto de números de bus y un segundo espacio MMIO diferente del primer conjunto de números de bus y del primer espacio MMIO, respectivamente;

donde segundo espacio MMIO está mapeado al primer espacio MMIO; y

el RCEP (106, 508) está configurado para aislar los fallos originados en la estructura PCIe extendida (118, 510).

2. La topología PCIe según la reivindicación 1, en la que un espacio de memoria de 32 bits de la estructura PCIe extendida (118, 510) está mapeado a un espacio MMIO de 64 bits del primer espacio MMIO,

en el que un espacio de memoria de 64 bits de la estructura PCIe extendida (118, 510) está mapeado a un espacio MMIO de 64 bits del primer espacio MMIO; o

en el que un espacio de configuración PCIe de la estructura PCIe extendida (118, 510) está mapeado a un espacio MMIO de 64 bits del primer espacio MMIO.

3. La topología de PCIe según la reivindicación 2, en la que cada función de dispositivo de la estructura PCIe extendida (118, 510) está mapeada a cuatro kilobytes respectivos del espacio MMIO de 64 bits.

4. La topología de PCIe según la reivindicación 2, en la que registros de configuración para cada función de dispositivo asociada con la estructura PCIe extendida (118, 510) están configurados para ser direccionados utilizando direccionamiento de base/dispositivo/función.

5. La topología de PCIe según la reivindicación 1, en la que el segundo conjunto de números de bus incluye hasta (256) números de bus únicos para la estructura PCIe extendida (118, 510).

6. La topología de PCIe según la reivindicación 1, en la que la estructura PCIe extendida (118, 510) interactúa con un dispositivo periférico del mismo modo que la estructura PCIe anfitrión (100, 504).

7. Un procedimiento para conectar un dispositivo periférico, que comprende:

alojar, mediante un punto extremo de complejo raíz (RCEP) (106, 508) como parte de un punto extremo de una estructura PCIe anfitrión (100, 504), una estructura extendida de interconexión de componentes periféricos exprés (PCIe) (118, 510), donde el RCEP (106, 508) es un puente entre la estructura PCIe extendida (118, 510) y la estructura PCIe anfitrión (100, 504), donde la estructura PCIe extendida (118, 510) tiene un primer espacio MMIO que es diferente de un segundo espacio MMIO de la estructura PCIe anfitrión (100, 504); y

mapear el primer espacio MMIO al segundo espacio MMIO;

interceptar, mediante el RCEP (106, 508), fallos que se producen más abajo en la estructura PCIe extendida (118, 510).

8. El procedimiento según la reivindicación 7, en el que la estructura PCIe extendida (118, 510) comprende además un primer conjunto de hasta 256 números de bus únicos, que es diferente de un segundo conjunto de números de bus de la estructura PCIe anfitrión.

9. El procedimiento según la reivindicación 7, que comprende además mapear un espacio de configuración PCIe, un espacio de memoria de 32 bits y un espacio de memoria de 64 bits de la estructura PCIe extendida (118, 510) al segundo espacio MMIO.

10. El procedimiento según la reivindicación 7, que comprende además tratar, mediante el dispositivo RCEP (106, 508), interrupciones de dispositivo originadas en la estructura PCIe extendida (118, 510) utilizando interrupciones señalizadas por mensaje (MSIs).

11. El procedimiento según la reivindicación 7, que comprende además manejar una solicitud de acceso directo a memoria (DMA) procedente de un dispositivo conectado a la estructura PCIe extendida (118, 510) mediante sustituir un ID de dispositivo de origen en una solicitud DMA con un ID de RCEP cuando la solicitud DMA se transmite hacia arriba a la estructura PCIe anfitrión (100, 504).

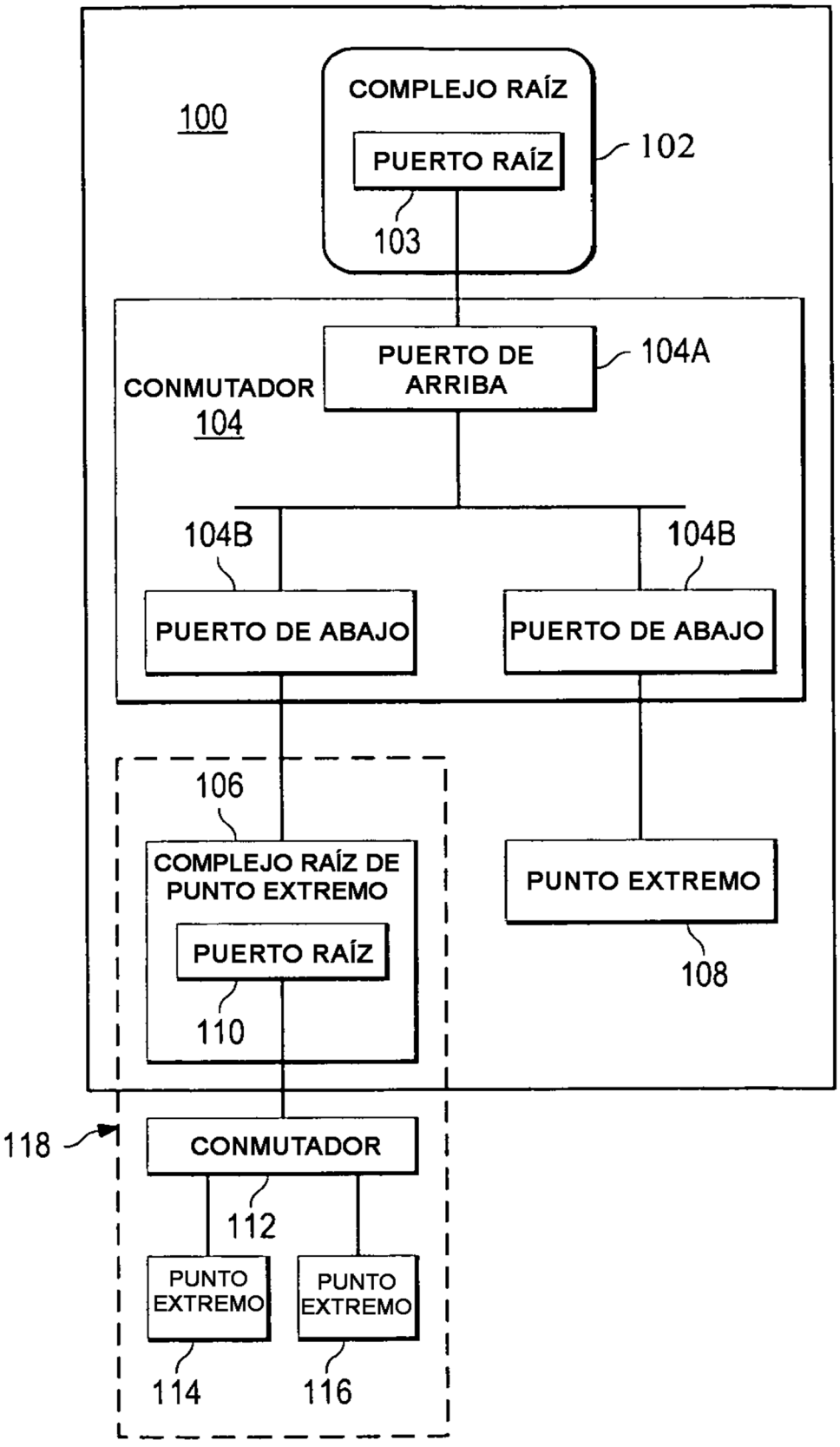
100
COMPLEJO RAÍZ
PUERTO RAÍZ
102
103
CONMUTADOR
104
PUERTO DE ARRIBA
104A
104B
104B
PUERTO DE ABAJO
PUERTO DE ABAJO
106
COMPLEJO RAÍZ DE PUNTO EXTREMO
PUERTO RAÍZ
110
PUNTO EXTREMO
108
118
CONMUTADOR
112
PUNTO EXTREMO
PUNTO EXTREMO
114
116

FIG. 1

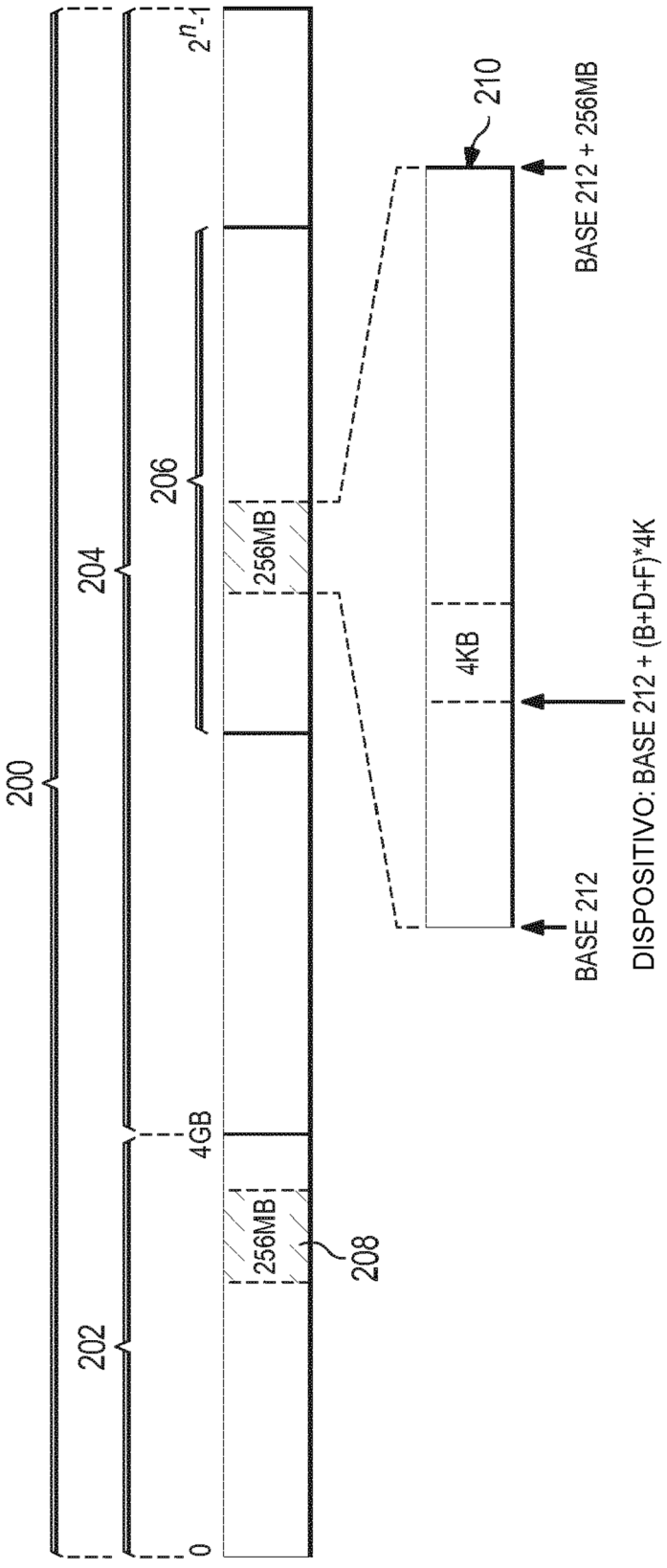
200
202
204
206
208
210
0
4GB
2^{n-1}
256MB
256MB
4KB
BASE 212
BASE 212 + 256MB
DISPOSITIVO: BASE 212 + (B+D+F)*4K

FIG. 2A

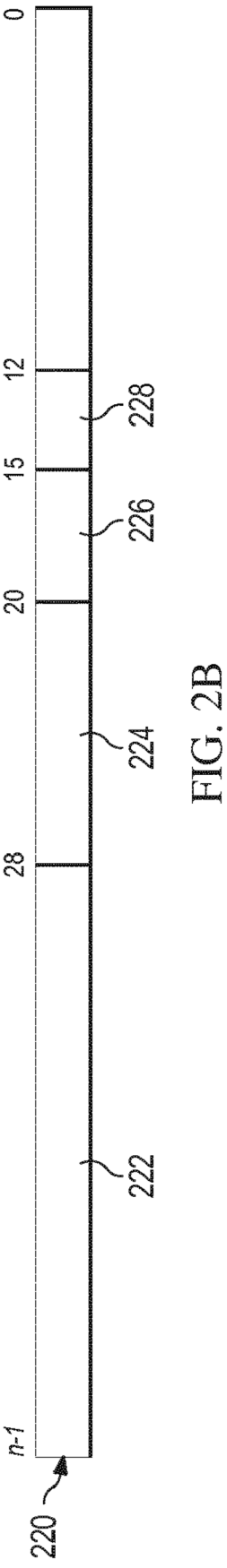
0
12
15
20
28
n-1
220
222
224
226
228

FIG. 2B

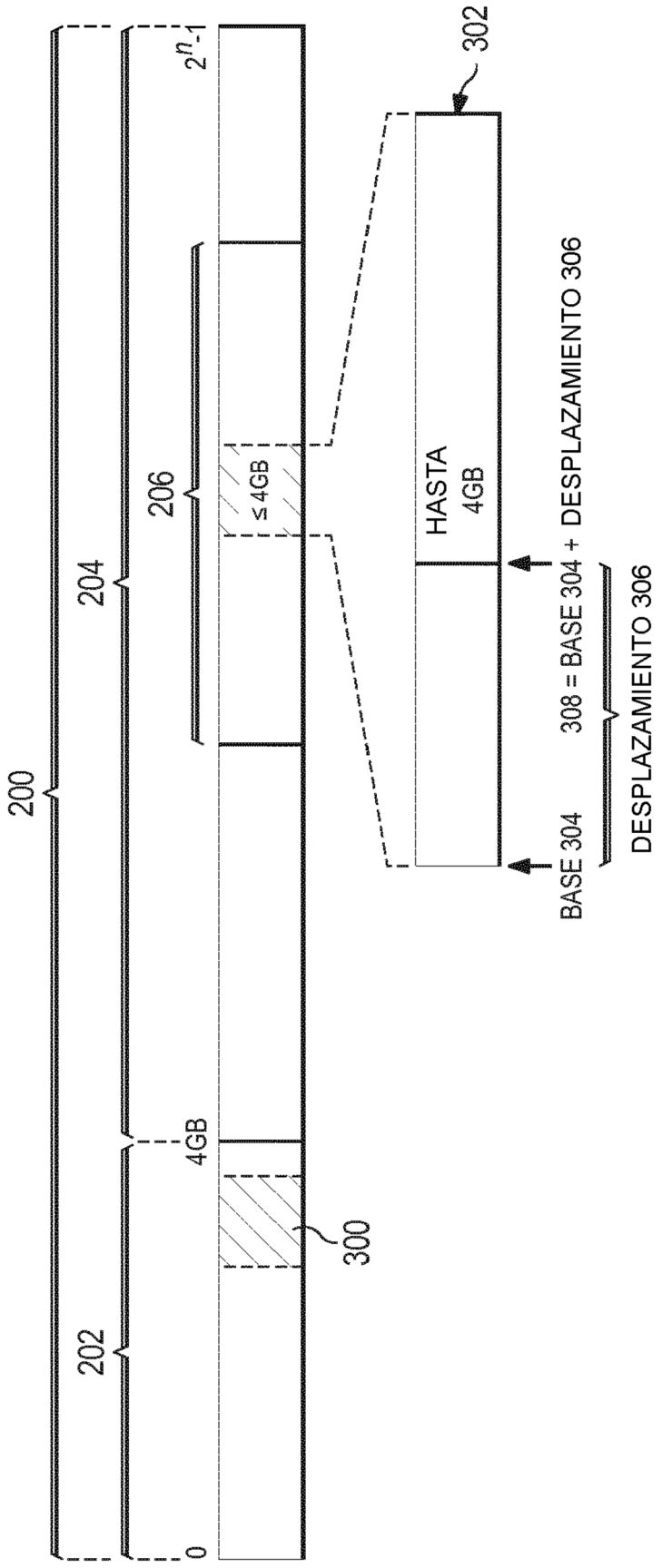
200
202
204
206
0
4GB
2^{n}-1
≤ 4GB
300
302
HASTA
4GB
BASE 304
308 = BASE 304 + DESPLAZAMIENTO 306
DESPLAZAMIENTO 306

FIG. 3A

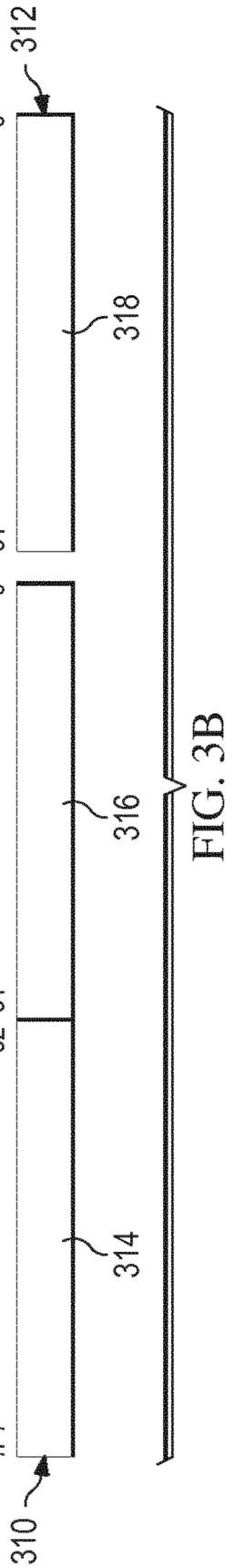
0
312
31
318
0
316
32 31
314
n-1
310

FIG. 3B

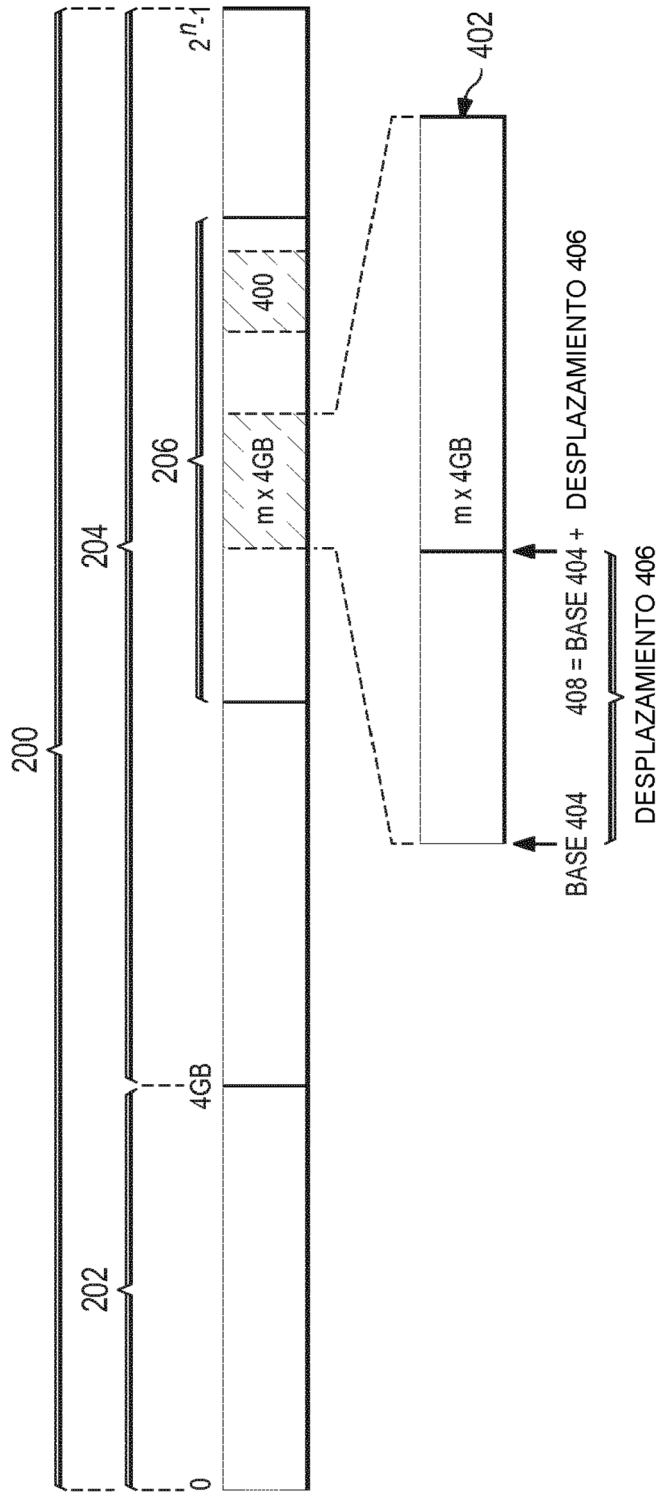
200
202
204
206
0
4GB
2^{n-1}
400
m x 4GB
402
m x 4GB
BASE 404
408 = BASE 404 + DESPLAZAMIENTO 406
DESPLAZAMIENTO 406

FIG. 4A

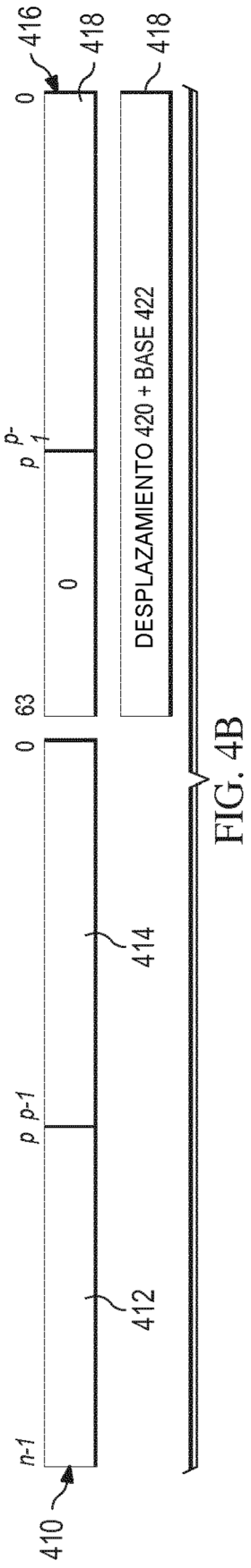
410
n-1
412
p p-1
414
0 63
0
416
418
p p-1
0
DESPLAZAMIENTO 420 + BASE 422
418

FIG. 4B

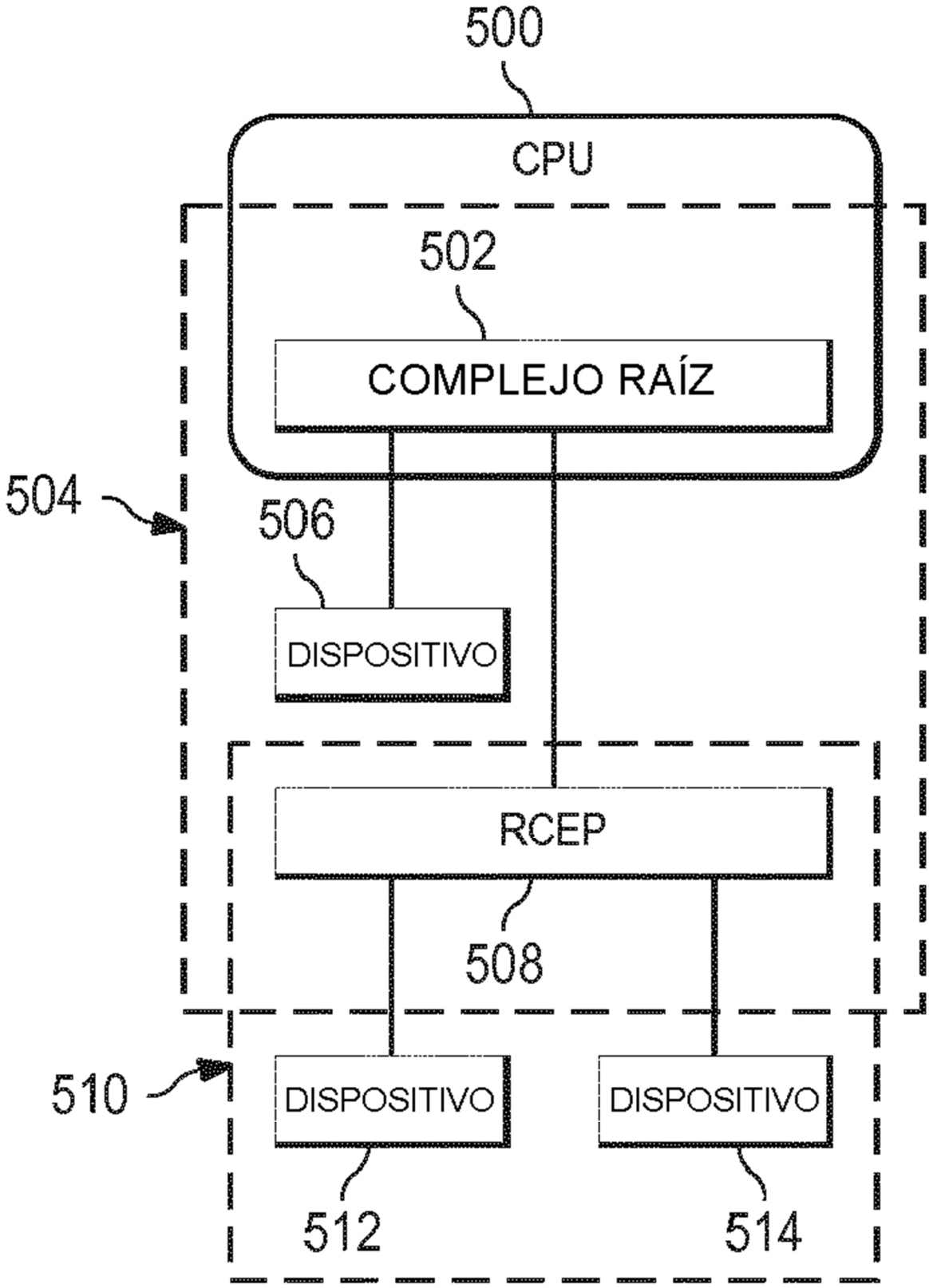
500
CPU
502
COMPLEJO RAÍZ
504
506
DISPOSITIVO
RCEP
508
510
DISPOSITIVO
DISPOSITIVO
512
514

FIG. 5