(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 370 749**

(51) Int. CI.:
G06F 12/08 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **08756742 .6**
(96) Fecha de presentación: **05.06.2008**
(97) Número de publicación de la solicitud: **2156302**
(97) Fecha de publicación de la solicitud: **24.02.2010**

(54) Título: **REDUCCIÓN DE LATENCIA PARA MEMORIA TEMPORAL BASADA EN BUS COHERENTE DE MEMORIA TEMPORAL.**

(30) Prioridad:
05.06.2007 US 758219

(45) Fecha de publicación de la mención BOPI:
22.12.2011

(45) Fecha de la publicación del folleto de la patente:
22.12.2011

(73) Titular/es:
APPLE INC.
1 INFINITE LOOP
CUPERTINO CA 95014-2084, US

(72) Inventor/es:
LILLY, Brian P.;
SUBRAMANIAN, Sridhar P. y
GUNNA, Ramesh

(74) Agente: **Fàbrega Sabaté, Xavier**

DESCRIPCIÓN

Reducción de latencia para memoria temporal basada en bus coherente de memoria temporal

Campo de la invención

Esta invención se refiere a procesadores y sistemas coherentes que incluyen procesadores.

Descripción de la Técnica Relacionada

Los sistemas informáticos por lo general han implementado uno o más niveles de memoria temporal para reducir la latencia de memoria. Las memorias temporales son memorias más pequeñas y de mayor velocidad que la memoria en el sistema de memoria principal. Por lo general, las memorias temporales almacenan datos utilizados recientemente. Por ejemplo, las memorias temporales a menudo se implementan para el acceso del procesador y almacenan datos leídos/escritos recientemente por los procesadores en los sistemas informáticos. Las memorias temporales a veces también se implementan para otros dispositivos de alta velocidad en el sistema informático. Además de almacenar datos utilizados recientemente, las memorias temporales pueden utilizarse para almacenar datos previamente capturados que se espera que sean utilizados por el procesador (u otro dispositivo).

Las memorias temporales almacenan copias de datos que también se almacenan en la memoria principal. En sistemas multiprocesador e incluso en sistemas de un único procesador en los que otros dispositivos acceden a la memoria principal pero no acceden a una determinada memoria temporal, se plantea la cuestión de la coherencia de memoria temporal. Es decir, un productor de datos determinado puede escribir una copia de los datos en la memoria temporal, y se retrasa la actualización a la copia de la memoria principal. En memorias temporales de escritura a través de memoria temporal, una operación de escritura es enviada a memoria en respuesta a la escritura a la línea de memoria temporal, pero la escritura se retrasa en el tiempo. En una memoria temporal de reescritura, las escrituras se efectúan en la memoria temporal y no se reflejan en la memoria hasta que el bloque de memoria temporal actualizada se sustituya en la memoria temporal (y se vuelve a escribir a la memoria principal en respuesta a la sustitución).

Debido a que las actualizaciones en la memoria principal no se han efectuado en el momento en que las actualizaciones se efectúan en la memoria temporal, un consumidor de datos determinado puede leer la copia de datos en la memoria principal y obtener datos "antiguos" (datos que aún no ha sido actualizados). Una copia almacenada temporalmente en una memoria temporal distinta a la que se acopla un productor de datos también puede tener datos antiguos. Además, si múltiples productores de datos están escribiendo en las mismas ubicaciones de memoria, diferentes consumidores de datos podían observar las escrituras en diferentes órdenes.

La coherencia de memoria temporal soluciona estos problemas al asegurar que se pueden mantener varias copias de los mismos datos (desde la misma ubicación de memoria) evitando "datos antiguos" y mediante el establecimiento de un orden "global" de lecturas/escrituras en las ubicaciones de memoria por diferentes productores/consumidores. Si a una escritura le sigue una lectura en el orden global, la lectura de datos refleja la escritura.

Los esquemas de coherencia de memoria temporal crean una sobrecarga en las operaciones de lectura/escritura de memoria. Por lo general, las memorias temporales efectuarán un seguimiento de un estado de sus copias de acuerdo con el esquema de coherencia. Por ejemplo, el conocido esquema Modificado, Exclusivo, Compartido, No válido (MESI) incluye un estado modificado (la copia se modifica con respecto a la memoria principal y otras copias); un estado exclusivo (la copia es la única copia además de la memoria principal); un estado compartido (puede haber una o más copias además de la copia de la memoria principal); y el estado no válido (la copia no es válida). El esquema MOESI agrega un estado En Propiedad en el que la memoria temporal es responsable de proporcionar los datos para una solicitud (ya sea volviendo a escribir en la memoria principal antes de proporcionar los datos al solicitante, o proporcionando directamente los datos al solicitante), pero puede haber otras copias en otras memorias temporales. Así, la sobrecarga del esquema de coherencia de memoria temporal incluye las comunicaciones entre las memorias temporales para mantener o actualizar el estado de coherencia. Estas comunicaciones pueden aumentar la latencia de las operaciones de lectura/escritura de memoria.

La sobrecarga es dependiente de la estructura del sistema informático. Más concretamente, la sobrecarga depende de la forma de interconexión entre las distintas memorias temporales y los productores/consumidores de datos. En un sistema de bus compartido, con frecuencia se suele implementar el examen para mantener la coherencia. Una solicitud de memoria dada transmitida en el bus es capturada por otras memorias temporales, que comprueban si una copia de los datos solicitados se encuentra almacenada en la memoria temporal. Las memorias temporales pueden actualizar el estado de sus copias (y proporcionar los datos, si la memoria temporal tiene la copia más actualizada). En general, en un EP1280062 divulga un sistema que comprende una pluralidad de agentes configurados para almacenar temporalmente datos, en el que la pluralidad de agentes están acoplados a una interconexión y una memoria temporal acoplada a la interconexión, en el que se configura un primer agente de la

pluralidad de agentes para iniciar una transacción en la interconexión mediante la transmisión de una solicitud de memoria, y en el que otros agentes de la pluralidad de agentes están configurados para examinar la solicitud de memoria de la interconexión y proporcionar una respuesta en una fase de respuesta de la transacción en la interconexión, y en el que la memoria temporal está configurada para detectar un acierto para la solicitud de memoria y proporcionar datos para la transacción al primer agente, en caso de que no se afirme una señal exclusiva. sistema de examen, los agentes de examen proporcionan una respuesta en la fase de respuesta de la transacción. Una fuente para la memoria temporal de datos puede determinarse de la respuesta (p. ej., el sistema de memoria principal o una memoria temporal con una copia más coherente). Debido a que la respuesta de examen se utiliza para determinar la fuente de los datos para una transacción de memoria, la transferencia de datos se retrasa a la respuesta de examen, y así se puede aumentar la latencia de memoria en casos en los que los datos podrían de otra manera ser proporcionados antes de la respuesta de examen (p. ej., debido a un acierto de memoria temporal).

Resumen de la invención

En una forma de realización, un sistema comprende una pluralidad de agentes acoplados a una interconexión y una memoria temporal acoplada a la interconexión. La pluralidad de agentes están configurados para almacenar datos temporalmente. Un primer agente de la pluralidad de agentes está configurado para iniciar una transacción en la interconexión mediante la transmisión de una solicitud de memoria, y otros agentes de la pluralidad de agentes están configurados para examinar la solicitud de memoria de la interconexión. Los otros agentes proporcionan una respuesta en una fase de respuesta de la transacción en la interconexión. La memoria temporal está configurada para detectar un acierto para la solicitud de memoria y proporcionar datos para la transacción al primer agente antes de la fase de respuesta e independientemente de la respuesta.

En un sistema que comprende una pluralidad de agentes acoplados a una interconexión y una memoria temporal acoplada a la interconexión, en el que la pluralidad de agentes también están configurados para almacenar datos temporalmente, se contempla un procedimiento. El procedimiento comprende iniciar una transacción mediante la transmisión de una solicitud de memoria en la interconexión de un primer agente de la pluralidad de agentes; examinar la solicitud de memoria de la interconexión por otros agentes de la pluralidad de agentes; proporcionar una respuesta en una fase de respuesta de la transacción en la interconexión por los otros agentes; detectar un acierto para la solicitud de memoria en la memoria temporal; y proporcionar datos para la transacción al primer agente antes de la fase de respuesta e independientemente de la respuesta, proporcionándose los datos por la memoria temporal en respuesta a la detección del acierto.

En otra forma de realización, un sistema comprende una pluralidad de agentes configurados para almacenar datos temporalmente, en el que la pluralidad de agentes están acoplados a una interconexión; y una memoria temporal acoplada a la interconexión. La memoria temporal y la pluralidad de agentes están configurados para mantener estados de coherencia tal que, si la memoria temporal detecta un acierto para una solicitud de memoria transmitida en la interconexión, la memoria temporal es capaz de proporcionar los datos independientemente del estado de los datos en la pluralidad de agentes. La memoria temporal está configurada para proporcionar datos antes de la fase de respuesta correspondiente a la solicitud de memoria si se detecta el acierto.

Breve descripción de los dibujos

La siguiente descripción detallada hace referencia a los dibujos adjuntos, que se describen brevemente a continuación.

Fig. 1 es un diagrama de bloques de una forma de realización de un sistema que incluye uno o más agentes y una memoria temporal de nivel dos (L2).

Fig. 2 es un diagrama de temporización que ilustra la operación de una forma de realización del sistema.

Fig. 3 es un diagrama de temporización que ilustra la operación de otra forma de realización del sistema.

Fig. 4 es un diagrama de flujo que ilustra la operación de una forma de realización de la memoria temporal L2 en respuesta a un examen.

Fig. 5 es un diagrama de flujo que ilustra la operación de una forma de realización de un agente en respuesta a un examen.

Fig. 6 es un diagrama de flujo que ilustra la operación de una forma de realización de un en respuesta al desalojo de un bloque de memoria temporal.

Mientras que la invención es susceptible a diversas modificaciones y formas alternativas, formas de realización específicas de la misma se muestran a modo de ejemplo en los dibujos y se describirán en detalle en el presente documento.

Descripción detallada

En cuanto a Fig. 1, se muestra un diagrama de bloques de una forma de realización de un sistema 10. En la forma de realización ilustrada, el sistema 10 incluye una pluralidad de agentes como los agentes 12A-12D. El sistema también incluye un conmutador de dirección 14, una interconexión de dirección 16, una interconexión de respuesta 18, una memoria temporal 22 de nivel 2 (L2) y un árbitro de datos 24. Los agentes 12A-12B y la memoria temporal L2 22 están acoplados al conmutador de dirección 14 (donde el agente 12B está acoplado a través de un biestable 20A en la forma de realización ilustrada). El conmutador de dirección 14 está adicionalmente acoplado a la interconexión de dirección 16, que está acoplada a los agentes 12A-12D (a través de los biestables 20B-201 en la forma de realización ilustrada). Visto de otra manera, los biestables 20B-201 pueden ser parte de la interconexión de dirección 16. La memoria temporal L2 22 también está acoplada a la interconexión de dirección 16, pero no a través de biestables en la forma de realización ilustrada. Los agentes 12A-12D también están acoplados a la interconexión de respuesta 18 (a través de los biestables 20J-20N y 20P-20R, en la forma de realización ilustrada). Visto de otra manera, los biestables 20J-20N y 20P-20R pueden ser parte de la interconexión de respuesta 18. La memoria temporal L2 22 también está acoplada a la interconexión de respuesta 18 (nuevamente, sin biestables en la forma de realización ilustrada). Los agentes 12A-12D y la memoria temporal L2 22 están acoplados al árbitro de datos 24. En una forma de realización, el sistema 10 puede estar integrado en un único chip de circuito integrado. En otras formas de realización, diversos componentes del sistema 10 pueden implementarse en diferentes circuitos integrados. Puede utilizarse cualquier nivel de integración en diversas formas de realización.

Los agentes 12A-12B pueden configurarse para transmitir solicitudes a ser transmitidas en la interconexión de dirección 16 al conmutador de dirección 14. Cada solicitud puede incluir la dirección de la transacción y el comando (que identifica la transacción a llevar a cabo). Pueden soportarse diversos comandos, como comandos de lectura y de escritura coherentes, comandos de lectura y de escritura no coherentes, comandos de propiedad coherente, comandos de sincronización, comandos de administración de memoria temporal, etc. Las solicitudes también pueden incluir otra información en diversas formas de realización. Por ejemplo, en una forma de realización, las solicitudes pueden incluir un nivel de prioridad de la solicitud (para su arbitraje) y también una indicación de si los datos para esta solicitud deben ser copiados a la memoria temporal de nivel 2.

Los agentes 12A-12B puede ser denominados agentes fuente, ya que pueden iniciar las transacciones en el sistema 10 transmitiendo una solicitud para la interconexión de dirección 16. Agentes fuente de ejemplo pueden incluir procesadores, memorias temporal de reescritura externas como la memoria temporal L2 22 (qué proceden transacciones de escritura para escribir bloques de memoria temporal desalojados que se han modificado en memoria), y puentes de entrada/salida (E/S) (que proceden transacciones de parte de dispositivos periféricos a los que están acoplados). Como lo ilustran las elipses en Fig. 1, diversas formas de realización pueden incluir más de dos agentes fuente (o agentes fuente/destino, que se describen a continuación). Otros agentes no pueden proceder transacciones, pero pueden ser el destino de una transacción (es decir, el agente que recibe la transacción y es responsable de los datos de la transacción). Dichos agentes se denominan agentes destino. Para las transacciones de lectura, el agente destino puede suministrar los datos a menos que otro agente tenga una copia más reciente de los datos en memoria temporal. Para las transacciones de escritura, el agente destino puede recibir los datos de escritura suministrados por el agente fuente. Agentes destino pueden incluir, por ejemplo, controladores de memoria y puentes de E/S. Algunos agentes pueden ser al mismo tiempo un agente fuente para algunas transacciones y un agente destino para otras transacciones. Agentes fuente/destino de ejemplo pueden incluir el puente de E/S o la memoria temporal externa anteriormente mencionados. Generalmente, un agente puede comprender cualquier circuito que esté configurado para comunicar a través de transacciones en la interconexión de dirección 16 y la interconexión de respuesta 18 (y la interconexión de datos, no mostrada en Fig. 1). Agentes fuente a menudo pueden incluir memorias temporales internas (p. ej., memorias temporales de nivel uno (L1)). Es decir, por lo menos algunos de los agentes 12A-12D en el sistema está configurados para almacenar temporalmente datos que han leído de la memoria.

En una forma de realización, cada agente fuente 12A-12B (o agente fuente/destino, aunque para mayor brevedad en esta descripción se utilizará agente fuente) y la memoria temporal L2 22 pueden utilizar una señal de solicitud para indicar que el agente fuente 12A-12B/la memoria temporal L2 22 está transmitiendo una solicitud. El conmutador de dirección 14 también puede afirmar una señal de concesión a un agente fuente 12A-12B dado/una memoria temporal L2 22 dada para indicar que una solicitud transmitida por ese agente fuente 12A-12B/esa memoria temporal L2 22 ha sido concedida en la interconexión de dirección 16. El conmutador de dirección 14 puede incluir una pluralidad de ubicaciones de almacenamiento configuradas para almacenar solicitudes transmitidas por los agentes fuente hasta que las solicitudes son concedidas en la interconexión de dirección 16. En una forma de realización, las ubicaciones de almacenamiento pueden comprender una pluralidad de colas. Cada cola puede corresponder a un agente fuente particular y puede estar dedicada a almacenar solicitudes transmitidas por ese agente fuente. Es decir, puede haber una correspondencia uno a uno entre las colas y los agentes fuente. La cola de un agente fuente dado puede almacenar una pluralidad de solicitudes transmitidas al conmutador de dirección 14 por el agente fuente dado. Cada agente fuente puede tener en cuenta la cantidad de entradas de cola en la cola correspondiente a ese agente fuente y no puede transmitir más solicitudes que las entradas de cola.

El conmutador de dirección 14 también puede configurarse para arbitrar entre las solicitudes en las colas para seleccionar una solicitud a transmitir en la interconexión de dirección 16. Puede emplearse cualquier esquema de arbitraje. Por ejemplo, en algunas formas de realización, cada solicitud puede tener un nivel de prioridad asignado a él. El sistema de arbitraje puede ser un esquema de prioridad estricta (selección de la solicitud de prioridad más alta) con mecanismos de prevención de privación para evitar la privación de solicitudes de prioridad baja. El conmutador de dirección 14 puede conducir la solicitud seleccionada en la interconexión de dirección 16.

La interconexión de dirección 16 puede comprender cualquier medio de comunicación, en diversas formas de realización. Por ejemplo, la interconexión de dirección 16 puede comprender una interfaz de paquetes, en la que una solicitud se transmite como un paquete durante uno o más ciclos de reloj en la interconexión de dirección 16. En particular, en una forma de realización, el paquete de dirección puede ser transmitido en un ciclo de reloj en la interconexión de dirección 16. Estas formas de realización pueden aislar el conmutador de dirección 14, de alguna manera, del protocolo de la fase de dirección de una transacción. Otras formas de realización pueden implementar la dirección de interconexión 16 como un bus, con una dirección transferida junto con diversas señales de control para indicar el comando y otra información de control transferida durante la fase de dirección. Generalmente, la interconexión de dirección puede considerarse lógicamente como una sola unidad, a ser concedida a un agente fuente para una transferencia de transacción dada como un todo.

Las solicitudes se emiten a los agentes 12A-12D y la memoria temporal L2 22 en la interconexión de dirección 16. En algunas formas de realización, el tiempo de vuelo en la interconexión de dirección 16 al agente más lejano 12A-12D (en términos de distancia física) podrá superar un ciclo de reloj del reloj asociado con la interconexión de dirección 16. Los biestables 20B-201 pueden utilizarse para capturar la solicitud y continuar su propagación a los agentes 12A-12D. Así, el número de biestables 20B-201 incluido entre el conmutador de dirección 14 y un agente determinado 12A-12D puede basarse en el tiempo de vuelo al agente más lejano (en cantidad de ciclos de reloj de la señal de reloj utilizada para la interconexión de dirección 16). En la forma de realización ilustrada, el tiempo de vuelo es superior a dos ciclos de reloj y así se utilizan dos biestables. Otras formas de realización pueden incluir cero biestables (si el tiempo de vuelo es inferior a un ciclo de reloj), un biestable (si el tiempo de vuelo es superior a un ciclo de reloj pero es inferior a dos ciclos de reloj) o más de dos biestables (dependiendo del tiempo de vuelo). Para garantizar que una solicitud dada es lógicamente recibida por cada agente 12A-12D en el mismo ciclo de reloj, se puede proporcionar una cantidad igual de biestables 20B-201 entre el conmutador de dirección 14 y cada agente 12A-12D a pesar de que algunos agentes pueden estar físicamente más cerca al conmutador de dirección 14 y la solicitud puede ser físicamente capaz de llegar al agente más cercano en un menor tiempo de vuelo. Así, la interconexión de dirección 16 se conoce como una interconexión implementada por etapas, ya que los comandos de dirección transmitidos son implementados por etapas a través de biestables a los destinos. Los biestables 20B-201 a los agentes más lejanos se pueden distribuir físicamente a lo largo de la distancia entre el conmutador de dirección 14 y los agentes más lejanos. Fig. 1 no intenta ilustrar la distribución física de los flops 20B-201, para mayor simplicidad en el dibujo.

Puesto que cada agente 12A-12D lógicamente recibe las solicitudes transmitidas en la interconexión de dirección 16 en el mismo ciclo de reloj, la interconexión de dirección 16 puede, en algunas formas de realización, ser el punto de coherencia en el espacio para las transacciones coherentes. Es decir, el orden de las solicitudes transmitidas con éxito en la interconexión de dirección 16 puede definir el orden de las transacciones con fines de coherencia.

Los agentes 12A-12D y la memoria temporal L2 22 también pueden estar acoplados a la interconexión de respuesta 18 para comunicar la fase de respuesta de las transacciones iniciadas a través de solicitudes en la interconexión de dirección 16. La fase de respuesta puede incluir, por ejemplo, las respuestas de los agentes de almacenamiento temporal para transacciones coherentes. Las respuestas pueden proporcionar información que indica qué estado de coherencia se debería establecer en el receptor de datos correspondientes a una transacción. En algunas formas de realización, puede emplearse un protocolo "de reintento" en el que un agente de respuesta puede indicar que una solicitud debe ser reintentada más tarde (y cancela la transacción actual). Otras formas de realización no pueden emplear reintentos. En la forma de realización ilustrada, también puede implementarse por etapas la fase de respuesta (a través de los biestables 20J-20N y 20P-20R). La implementación por etapas puede tener una longitud de dos biestables (total, de transmisor a receptor) en la presente forma de realización, para un retraso de dos relojes desde conducir una respuesta hasta recibir una respuesta. El retraso se ilustra en Fig. 2 como dos biestables desde cada agente hasta la interconexión de respuesta 18, pero no pretende implicar un retraso total de 4 relojes para la forma de realización ilustrada.

Una vez que se ha producido la fase de respuesta para una determinada transacción, se conoce la fuente de los datos para la transacción (p. ej., puede ser un controlador de memoria, o puede ser un agente de almacenamiento temporal que tiene una copia más coherente de los datos). En una forma de realización, la fase de respuesta puede incluir una señal de acierto (que indica que existe una copia en por lo menos una memoria temporal) y una señal de acierto modificada (HitM) que indica que existe una copia modificada en una memoria temporal. La copia modificada puede encontrarse en el estado M, o el estado O en el protocolo MOESI.

Mientras que el haber recibido lógicamente cada agente 12A-12D las solicitudes transmitidas en la interconexión de dirección 16 en el mismo ciclo de reloj puede simplificar la administración de coherencia en el sistema 10, la latencia de memoria también puede ser mayor de lo que sería si no se llevara a cabo la implementación por etapas. Concretamente, la fase de respuesta de la transacción iniciada por la solicitud puede retrasarse tanto por la implementación por etapas de la transmisión de la solicitud como la implementación por etapas de la respuesta. Así, no se conoce la fuente de los datos (p. ej., el controlador de memoria o la memoria temporal que tiene una copia más coherente que la memoria) para un período de tiempo superior que si no se lleva a cabo la implementación por etapas (p. ej., 4 relojes más tarde, en esta forma de realización). Para las solicitudes de lectura (que solicitan una transferencia de datos al solicitante), la latencia puede reducir el rendimiento porque el solicitante puede necesitar los datos para continuar procesando. Por ejemplo, un procesador puede haber ejecutado una instrucción de carga que causó la solicitud de lectura, y cualquier instrucción que dependa del resultado de la carga puede retrasarse hasta que sean proporcionados los datos.

La memoria temporal L2 22 también puede ser un agente en la interconexión de dirección 16, pero no puede implementarse por etapas como otros agentes. En otras formas de realización, la memoria temporal L2 22 puede implementarse por etapas, pero puede implementarse por etapas a través de una cantidad de biestables inferior al de otros agentes. En consecuencia, la memoria temporal L2 22 puede recibir cada solicitud en la interconexión de dirección 16 antes que otros agentes (p. ej., un ciclo de reloj de bus más antes). La memoria temporal L2 22 puede examinar la transmisión y puede, para operaciones de lectura, suministrar los datos si la memoria temporal L2 22 detecta un acierto para los datos. Particularmente, la memoria temporal L2 22 puede suministrar los datos antes de que se produzca la fase de respuesta de la transacción (aunque los datos no se pueden transferir completamente antes de la fase de respuesta), o por lo menos antes de que la respuesta implementada por etapas sea proporcionada a todos los agentes. Así, se podrá reducir la latencia de memoria en casos en los que se detecte un acierto de memoria temporal L2.

En una forma de realización, un conjunto de normas de almacenamiento temporal de datos para la memoria temporal L2 y las memorias temporales L1 en los agentes 12A-12D pueden permitir a la memoria temporal L2 22 proporcionar datos en respuesta a un acierto, sin conocer la respuesta de otros agentes. Las normas pueden ser diseñadas para garantizar que, si la memoria temporal L2 almacena una copia de un bloque, esa copia sea la copia más coherente en el sistema 10 (es decir, sea la más actualizada). Así, la fase de respuesta será consistente con el suministro de datos de la memoria temporal L2 22. Visto de otra manera, si se detecta un acierto memoria temporal L2, memorias temporales L1 pueden tener los datos en sólo el estado compartido o no válido, para una forma de realización que implemente los esquemas de coherencia MESI o MOESI.

En una forma de realización, la memoria temporal L2 22 puede actuar como una memoria temporal víctima para las solicitudes de datos, y puede incluir solicitudes de captura de instrucciones de procesadores. Si una solicitud acertara en la memoria temporal L2, y el estado resultante en el solicitante causara que el solicitante proporcionara los datos en respuesta a un examen posterior, la memoria temporal L2 podría invalidar la entrada de memoria temporal L2 que almacena el bloque. Visto de otra manera, si el estado resultante en el solicitante es un estado potencialmente modificable (p. ej., exclusivo, modificado, o en propiedad), la memoria temporal L2 puede invalidar su entrada de memoria temporal. Visto todavía de otra manera, si el estado resultante en el solicitante indica que el solicitante tiene la copia más coherente, la memoria temporal L2 puede invalidar su entrada de memoria temporal.

En una forma de realización, las memorias temporales L1 pueden configurarse para reescribir bloques exclusivos que son desalojados, similar a la reescritura llevada a cabo para los bloques modificados o en propiedad. La memoria temporal L2 22 puede asignar una entrada para los bloques exclusivos, y así puede capturar una copia que podría haber sido invalidada con la memoria temporal L1 tomó la copia exclusiva. Si la indicación de copia L2 no se establece para una solicitud, como se ha mencionado anteriormente, un llenado de memoria temporal de datos L1 desde memoria generalmente no puede causar un llenado en la memoria temporal L2 22. Los llenados de memoria temporal de instrucciones pueden almacenarse temporalmente en la memoria temporal L2 22, ya que la memoria temporal de instrucciones por lo general no se modifica.

El árbitro de datos 24 puede configurarse para arbitrar entre solicitudes de uso de una interconexión de datos (no mostrado en Fig. 1). La memoria temporal L2 22 puede afirmar una solicitud al árbitro de datos 24 para suministrar datos para una solicitud de memoria, incluyendo el suministro de datos a una solicitud de memoria en respuesta a la detección de un acierto antes de la fase de respuesta. En algunas formas de realización, la memoria temporal L2 22 puede configurarse para afirmar una solicitud al árbitro de datos 24 de forma especulativa, mientras esté ocurriendo la búsqueda de etiquetas para una solicitud. Si se concede la interconexión de datos a la memoria temporal L2 22 y la especulación es correcta, los datos pueden suministrarse con una latencia incluso menor.

La memoria temporal L2 22 puede proporcionar una señal de acierto L2 en la interconexión de respuesta 18, señalizando que la memoria temporal L2 22 proporcionará los datos. La memoria temporal L2 22 también puede proporcionar una señal de acierto L2 temprano antes de la fase de respuesta, indicando al solicitante que los datos pueden proporcionarse temprano para la solicitud (si a la memoria temporal L2 22 se le concede el uso de la interconexión de datos). En una forma de realización, la interconexión de datos puede ser una barra cruzada

conmutada jerárquica, aunque puede utilizarse cualquier interconexión en diversas formas de realización (p. ej., bus compartido, punto a punto, etc.). Las señales acierto L2 temprano y acierto L2 pueden ser respuestas codificadas, en otras formas de realización.

En algunas formas de realización, un tiempo de vuelo de una solicitud desde un agente fuente 12A-12B hasta el conmutador de dirección 14 también puede superar un ciclo de reloj. En algunas formas de realización, el conmutador de dirección 14 puede colocarse físicamente más cerca de los agentes fuente que se espera que tengan el mayor ancho de banda para solicitudes (p. ej., agentes de procesador pueden por lo general tener mayor ancho de banda para solicitudes que agentes de memoria temporal son agentes de E/S). En la forma de realización de Fig. 1, el tiempo de vuelo de las solicitudes desde el agente fuente 12B podrá ser superior a un ciclo de reloj, y así el biestable 20A puede utilizarse para capturar la solicitud y continuar su propagación hasta el conmutador de dirección 14. Del mismo modo, la señal de concesión devuelta por el conmutador de dirección 14 puede ser capturada por el biestable 20A y propagada en el siguiente ciclo de reloj.

Ya que, en la presente forma de realización, la interconexión de dirección 16 es el punto de coherencia para las transacciones coherentes (y también puede definir el orden para las solicitudes como un todo), no hay ningún orden entre las solicitudes transmitidas al conmutador de dirección 14 desde diferentes agentes. En consecuencia, si un biestable como el biestable 20A se utiliza para un tiempo de vuelo desde un agente fuente, no se necesita insertar biestables para otros agentes cuyo tiempo de vuelo para las solicitudes sea inferior a un ciclo de reloj.

Como se ha mencionado anteriormente, los agentes fuente pueden recibir solicitudes en la interconexión de dirección 16, en algunas formas de realización, para determinar qué solicitud de entre múltiples solicitudes pendientes en el conmutador de dirección 14 de un agente determinado se concedió realmente en la interconexión de dirección 16. Además, en algunas formas de realización, los agentes fuente que también pueden almacenar datos temporalmente (y así puedan participar en transacciones coherentes) también pueden examinar otras solicitudes de agentes fuente en la interconexión de dirección 16 para fines de coherencia. Los agentes destino, como los agentes 12C-12D, son acoplados a la interconexión de dirección 16 para recibir las solicitudes para las que son el destino.

En una forma de realización, el conmutador de dirección 14 también se puede configurar para administrar el control de flujo a diversos agentes destino 12 C-12D. Por ejemplo, el conmutador de dirección 14 puede configurarse para determinar qué agente destino es abordado por cada solicitud (p. ej., a través de la descodificación de grano grueso de la dirección de la solicitud y mapeando la dirección a un agente destino basado en la descodificación). El conmutador de dirección 14 puede tener en cuenta la cantidad de solicitudes que pueden ponerse en cola en un agente destino (después de la recepción de las solicitudes desde la interconexión de dirección 16) y puede asegurar que las colas de entradas del agente destino no sean desbordabas por las solicitudes. Si una determinada solicitud está dirigida a un determinado agente destino cuya cola de entrada está llena, el conmutador de dirección 14 puede garantizar que dicha solicitud no sea seleccionada como la ganadora del arbitraje hasta que haya disponible una cola de entradas en dicho agente destino. El conmutador de dirección 14 puede que no bloquee otras solicitudes en la misma situación. Es decir, el conmutador de dirección 14 aún puede seleccionar otra solicitud dirigida a otro agente destino si una solicitud anterior o una solicitud de prioridad más alta no reúne los requisitos para ganar el arbitraje debido a que el agente destino no puede recibir la solicitud. En algunas formas de realización, el conmutador de dirección 14 también puede intentar implementar justicia u optimizar el acceso a un agente destino entre los agentes fuente.

Hay que reseñar que, mientras que los biestables 20A-20N y 20P-20R se ilustran en la forma de realización de Fig. 1, generalmente puede utilizarse cualquier dispositivo de almacenamiento con reloj como los dispositivos 20A-20N y 20P-20R. Por ejemplo, pueden utilizarse registros, circuitos biestables, etc. Un dispositivo de almacenamiento con reloj puede comprender cualquier dispositivo de almacenamiento que esté configurado para capturar un valor de almacenamiento en respuesta a una señal de reloj. En la presente forma de realización, la entrada de señal de reloj para los biestables 20A-20N y 20P-20R puede ser el reloj utilizado para la interconexión de dirección 16. Algunos agentes pueden operar internamente con múltiplos del reloj. Otros flops descritos en la presente memoria también podrán implementarse con cualquier dispositivo de almacenamiento con reloj. Generalmente, cada biestable 20A-20N y 20P-20R puede tener un ancho de bits igual a la anchura de su entrada. Por ejemplo, el biestable 20A puede tener el ancho de la interfaz de solicitud/concesión al conmutador de dirección 14 y el ancho de los biestables 20B-20N y 20P-20R puede ser el ancho de la interconexión de dirección 16.

Mientras que un esquema de arbitraje de prioridad estricta se utiliza como un ejemplo en lo anteriormente indicado, otras formas de realización pueden implementar otros esquemas de arbitraje. Por ejemplo, otros esquemas de arbitraje pueden incluir round-robin, round-robin ponderado por prioridades, combinaciones de round-robin y esquemas de prioridad, etc.

En cuanto a Fig. 2, se muestra un diagrama de temporización que ilustra la operación de una forma de realización del sistema 10 para una transacción. Los ciclos de reloj de bus (BClk) están delimitados por líneas discontinuas verticales y están etiquetados en la parte superior (BClk1 a BClk8). En Fig. 2, una "D" o "R" entre paréntesis puede indicar cuándo un determinado valor o señal es conducido (D) o recibido (R).

El conmutador de dirección 14 puede conducir una solicitud dirigida a la dirección A1 en la interconexión de dirección 16 en BClk1 (número de referencia 30). La interconexión de dirección 16 puede ser procesada plenamente por entubamiento en esta forma de realización, y así otra solicitud (dirigida a la dirección A2) puede opcionalmente ser conducida en BClk2 (número de referencia 32). La solicitud para A1 se recibe en un agente en BClk3 (número de referencia 34). Sin embargo, la memoria temporal L2 22 recibe la solicitud para A1 en BClk1 (número de referencia 36).

La memoria temporal L2 22 puede llevar a cabo una lectura de etiquetas para la solicitud en BClk2 (número de referencia 38) y puede llevar a cabo una correspondientes lectura de datos en BClk3 (número de referencia 40). En una forma de realización, la memoria temporal L2 22 puede incluir una protección por código de corrección de errores (ECC) para las memorias de etiquetas y datos, y así puede llevar a cabo comprobaciones ECC en los ciclos de reloj después de la lectura de etiquetas y la lectura de datos, respectivamente. La memoria temporal L2 22 puede detectar un acierto de memoria temporal para la solicitud, y puede conducir una respuesta de acierto L2 temprano en BClk3 (número de referencia 42). El agente solicitante puede recibir la señal de acierto L2 temprano en BClk4, y por lo tanto puede ser informado que potencialmente los datos se proporcionarán temprano (antes de la fase de respuesta) para la actual transacción (número de referencia 44). En otras formas de realización, la respuesta de acierto L2 temprano puede ser conducida y recibida en el mismo ciclo de reloj.

La memoria temporal L2 22 también puede conducir una solicitud para el árbitro de datos 24 en BClk3 (número de referencia 46) y puede recibir una concesión desde el árbitro de datos 24 en BClk4 (número de referencia 48). La memoria temporal L2 22 suministra datos en BClk6 a BClk8 (y tal vez ciclos de reloj adicionales, no mostrados--números de referencias 50, 52 y 54). Por ejemplo, en una forma de realización, un bloque de memoria temporal de datos podrá transferirse en 4 ciclos de reloj de bus (4 pulsaciones). Otras formas de realización pueden utilizar más o menos pulsaciones para transferir un bloque de memoria temporal.

En la forma de realización ilustrada, un agente receptor 12A-12D puede determinar su respuesta a la solicitud (recibida en BClk 3, número de referencia 34) y puede conducir la respuesta en BClk5 (número de referencia 56). La respuesta puede ser recibida en el agente fuente de la transacción en BClk7 (número de referencia 58), que es después de los datos de memoria temporal L2 que se están suministrando en este caso. Si se detecta un fallo L2, el agente que debe proporcionar los datos puede afirmar su solicitud de datos en BClk8 (número de referencia 60), o tal vez incluso más tarde, dependiendo de la latencia para el agente suministrador para obtener los datos (p. ej., de memoria, si el agente suministrador es el controlador de memoria, o de una memoria temporal, si el agente suministrador es un agente de examen). Así, podrá reducirse la latencia en un mínimo de 5 ciclos de reloj en esta forma de realización (la diferencia entre la solicitud de arbitraje de datos L2 en BClk3 y la solicitud de arbitraje de datos de agente en BClk8).

Hay que reseñar que, si bien en Fig. 2 se muestran temporizaciones específicas, las temporizaciones pueden variar de una forma de realización a otra forma de realización, como se desee.

Fig. 3 es un diagrama de temporización que ilustra la operación de una segunda forma de realización del sistema 10 para una transacción. Las temporizaciones en Fig. 3 para la solicitud, la memoria temporal L2 que recibe la solicitud y lleva a cabo la lectura de etiquetas y la lectura de datos, el agente que recibe la solicitud y conduce la respuesta, la respuesta que se recibe, y la conducción y la recepción del acierto L2 temprano pueden ser los mismos que en Fig. 2 (números de referencia 30, 32, 34, 36, 38, 40, 42, 44, 56, 58 y 60). Sin embargo, en esta forma de realización, la memoria temporal L2 22 puede configurarse para solicitar de forma especulativa la interconexión de datos del árbitro de datos, en respuesta a la recepción de la solicitud (número de referencia 70). Si la solicitud es concedida (número de referencia 72), entonces también los datos pueden ser suministrados antes (números de referencia 74, 76, 78 y 80). Si la solicitud es concedida por la memoria temporal L2 22 detecta un fallo, puede suministrarse una cancelación con los primeros datos transferidos para indicar al agente fuente que los datos no son válidos (número de referencia 82). O bien, simplemente podrá cancelarse la transferencia, y no podrán conducirse datos.

En cuanto a Fig. 4, se muestra un diagrama de flujo que ilustra la operación de una forma de realización de la memoria temporal L2 22 en respuesta al examen de una solicitud de memoria de la interconexión de dirección 16. Mientras que los bloques se muestran en un orden determinado para proporcionar la comprensión, pueden utilizarse otros órdenes. Los bloques pueden llevarse a cabo en paralelo en lógica combinatoria en la memoria temporal L2 22. Bloques, combinaciones de bloques y/o el diagrama de flujo como un todo, pueden procesarse por entubamiento en múltiples ciclos de reloj.

La memoria temporal L2 22 puede leer la memoria de etiquetas y determinar si el examen es un acierto (bloque de decisión 90). Si el examen es un fallo (bloque de decisión 90, rama “no”), y la solicitud de examen no es una reescritura (bloque de decisión 92, rama “no”), la memoria temporal L2 22 no adoptará ninguna acción adicional para el examen. Si el examen es un fallo (bloque de decisión 90, rama “no”) pero el examen es una reescritura (bloque de decisión 92, rama "sí"), la memoria temporal L2 22 puede asignar una entrada y aceptar los datos de reescritura en

la entrada asignada (bloque 94). El estado para la entrada puede ser exclusivo, modificado, o en propiedad (el mismo estado que tenía el agente de desalojo para los datos).

Si el examen es un acierto (bloque de decisión 90, rama "sí"), y el examen es una solicitud de lectura (bloque de decisión 96, rama “sí”), la memoria temporal L2 22 puede responder con datos. Así, la memoria temporal L2 22 puede afirmar la respuesta de acierto L2 temprano (bloque 98). Además, la memoria temporal L2 22 puede afirmar la respuesta de acierto L2 durante la fase de respuesta y también puede afirmar HitM si la memoria temporal L2 22 está almacenando una copia modificada. Si el solicitante toma la línea en un estado Exclusivo (E) En Propiedad (O), o Modificado (M) (bloque de decisión 100, rama “sí”), la memoria temporal L2 22 puede invalidar la entrada L2 de acierto (bloque 102). El solicitante podrá exigir el bloque en un estado E, O, o M transmitiendo un tipo de solicitud especial (p, ej., una solicitud de lectura para tener en propiedad o de lectura para modificar). Además, el solicitante podrá adquirir el bloque en un estado E, O o M si las respuestas en la fase de respuesta indican que el bloque puede ser tomado en ese estado, incluso si la solicitud no lo exigiera (p. ej., una solicitud de lectura compartida o lectura de bloque que no se almacena temporalmente en memorias temporales de otros agentes). Si el solicitante no adquiere el bloque en un estado E, O, o M, entonces la memoria temporal L2 22 puede actualizar su estado a compartido con la excepción de una solicitud de lectura para compartir si la memoria temporal L2 22 tiene el bloque en un estado en propiedad o modificado (en cuyo caso el estado se actualiza a en propiedad).

Porque la memoria temporal L2 22 detecta un acierto, la memoria temporal L2 22 puede arbitrar para la interconexión de datos (bloque 104). En algunas formas de realización, la memoria temporal L2 22 puede arbitrar de forma especulativa para el bus de datos en respuesta al examen de una solicitud de transacción coherente, almacenable temporalmente. Otras formas de realización sólo podrán arbitrar en respuesta a una detección del acierto. Si el árbitro de datos 24 concede la interconexión de datos a la memoria temporal L2 (bloque de decisión 106, rama "sí"), la memoria temporal L2 puede conducir los datos al solicitante (bloque 108). De otra manera (bloque de decisión 106, rama “no”), la memoria temporal L2 puede continuar el arbitraje de la interconexión de datos.

Si el examen es un acierto (bloque de decisión 90, rama "sí") y el examen no es una solicitud de lectura (bloque de decisión 96, rama “no”), pero la solicitud es una solicitud de reescritura (bloque de decisión 110, rama "sí"), la memoria temporal L2 22 puede aceptar los datos de reescritura en la entrada de acierto (bloque 112) y puede actualizar el estado de la entrada de acierto a modificado, exclusivo, o en propiedad (el mismo estado que la memoria temporal que transmitió la reescritura). Para solicitudes de no lectura, no reescritura (bloque de decisión 110, rama “no”), la memoria temporal L2 22 puede invalidar la entrada de memoria temporal L2 de acierto (bloque 114). Las solicitudes de no lectura, no reescritura pueden incluir solicitudes para cambiar un bloque a sucio (estado modificado), como un acierto de almacenamiento a un bloque no exclusivo, no modificado; solicitudes para tener en propiedad un bloque; solicitudes de invalidación; solicitudes de escritura no- almacenables temporalmente, etc.

En cuanto a Fig. 5, se muestra un diagrama de flujo que ilustra la operación de una forma de realización de un agente (p. ej., uno de los agentes 12A-12D) que almacena datos temporalmente en respuesta al examen de una solicitud de memoria de la interconexión de dirección 16. Mientras que los bloques se muestran en un orden determinado para proporcionar la comprensión, pueden utilizarse otros órdenes. Los bloques pueden llevarse a cabo en paralelo en lógica combinatoria en el agente. Los bloques, las combinaciones de bloques y/o el diagrama de flujo como un todo pueden procesarse por entubamiento en múltiples ciclos de reloj.

Si la solicitud no es una lectura (bloque de decisión 120, rama “no”), el examen puede procesarse con normalidad (bloque 122). El procesamiento puede incluir invalidar el bloque, reescribir el bloque en memoria si ha sido modificado, etc. Si la solicitud es una lectura (bloque de decisión 120, rama “sí”) y el examen acierta en el agente en estado E, O o M (bloque de decisión 124, rama "sí"), entonces el agente puede proporcionar los datos para la transacción (bloque 126). En cualquier caso, el agente puede actualizar el estado del bloque de memoria temporal en la memoria temporal (si existe) y proporcionar la respuesta en la fase de respuesta (bloque 128).

En cuanto a Fig. 6, se muestra un diagrama de flujo que ilustra la operación de una forma de realización de un agente (p. ej., uno de los agentes 12A-12D) que almacena datos temporalmente en respuesta al desalojo de un bloque de su memoria temporal. Mientras que los bloques se muestran en un orden determinado para proporcionar la comprensión, pueden utilizarse otros órdenes. Los bloques pueden llevarse a cabo en paralelo en lógica combinatoria en el agente. Los bloques, las combinaciones de bloques y/o el diagrama de flujo como un todo pueden procesarse por entubamiento en múltiples ciclos de reloj.

Si el estado del bloque es E, O o M (bloque de decisión 130, rama “sí”), el agente puede generar una solicitud de reescritura para reescribir el bloque en la memoria (o más concretamente, en la memoria temporal L2 22 en esta forma de realización) (bloque 132). En cualquier caso, el agente puede invalidar la entrada de la memoria temporal desde la que el bloque ha sido desalojado. En una forma de realización adicional un sistema comprende una pluralidad de agentes configurados para almacenar datos temporalmente, en el que la pluralidad de agentes están acoplados a una interconexión; y una memoria temporal acoplada a la interconexión; en el que la memoria temporal y la pluralidad de agentes están configurados para mantener estados de coherencia de manera que, si la memoria temporal detecta un acierto para una solicitud de memoria transmitida en la interconexión, la memoria temporal es

capaz de proporcionar los datos independientemente del estado de los datos en la pluralidad de agentes, y en el que la memoria temporal está configurada para proporcionar datos antes de la fase de respuesta correspondiente a la solicitud de memoria si se detecta el acierto. En el que preferentemente, en el sistema, si un estado resultante en una de la pluralidad de agentes en respuesta a la solicitud de memoria permite a la una de la pluralidad de agentes modificar los datos, la memoria temporal es configurada para invalidar los datos en la memoria temporal.

Numerosas variaciones y modificaciones se pondrán de manifiestos para los expertos en la materia una vez que la divulgación anterior se comprenda plenamente. Se pretende que se interprete que las reivindicaciones a continuación abarcan todas estas variaciones y modificaciones.

REIVINDICACIONES

1. Un sistema que comprende:

una pluralidad de agentes configurados para almacenar temporalmente datos,
en el que la pluralidad de
agentes están acoplados a una interconexión; y
una memoria temporal acoplada a la interconexión;
en el que un primer agente de la pluralidad de agentes se configura para iniciar una transacción en la interconexión mediante la transmisión de una solicitud de memoria, y
en el que otros agentes de la pluralidad de agentes son configurados para examinar la solicitud de memoria de la interconexión y proporcionar una respuesta en una fase de respuesta de la transacción en la interconexión, **caracterizado porque** la memoria temporal se configura para detectar un acierto para la solicitud de memoria y proporcionar datos para la transacción al primer agente antes de la fase de respuesta e independientemente de la respuesta.

2. El sistema según la reivindicación 1, en el que, incluso si uno o más de los otros agentes detectan un acierto para la solicitud de memoria, el uno o más de los otros agentes no proporcionan datos si la memoria temporal detecta el acierto.

3. El sistema según la reivindicación 1 en el que la transacción resulta en un estado de memoria temporal en el primer agente que causaría que el primer agente proporcionara datos en respuesta a un examen posterior, y en el que la memoria temporal es configurada para invalidar los datos de una entrada de memoria temporal que almacena los datos en respuesta a la transacción.

4. El sistema según la reivindicación 1 en el que la pluralidad de agentes, en respuesta al desalojo de un bloque de memoria temporal que se encuentra en un estado exclusivo pero no modificado, son configurados para generar una transacción de reescritura para el bloque de memoria temporal.

5. El sistema según la reivindicación 4 en el que la memoria temporal es configurada para asignar una entrada de memoria temporal en respuesta a la transacción de reescritura y aceptar el bloque de memoria temporal en la entrada.

6. El sistema según la reivindicación 1 que comprende adicionalmente un árbitro de datos configurado para arbitrar las solicitudes para una interconexión de datos, y en el que la memoria temporal está acoplada al árbitro de datos y, en respuesta a la recepción de la solicitud de memoria de la interconexión, la memoria temporal es configurada para arbitrar de forma especulativa la interconexión de datos.

7. El sistema según la reivindicación 6 en el que el árbitro de datos es configurado para conceder la interconexión de datos a la memoria temporal, y en el que la memoria temporal es configurada para transmitir una indicación de cancelación en el bus de datos para indicar al primer agente que la especulación fue incorrecta y que el primer agente debe ignorar los datos.

8. El sistema según la reivindicación 1 en el que la memoria temporal es configurada para proporcionar una respuesta temprana al primer agente, antes de la fase de respuesta, para indicar que la memoria temporal proporcionará los datos.

9. El sistema según la reivindicación 1 en el que la memoria temporal es acoplada para recibir la solicitud de memoria de la interconexión antes que la pluralidad de agentes.

10. Un procedimiento para operar un sistema que comprende una pluralidad de agentes acoplados a una interconexión y una memoria temporal acoplada a la interconexión, en el que la pluralidad de agentes también están configurados para almacenar temporalmente datos, comprendiendo el procedimiento:

iniciar una transacción mediante la transmisión de una solicitud de memoria en la interconexión de un primer agente de la pluralidad de agentes;
examinar la solicitud de memoria de la interconexión por otros agentes de la pluralidad de agentes;
proporcionar una respuesta en una fase de respuesta de la transacción en la interconexión por los otros agentes;
detectar un acierto para la solicitud de memoria en la memoria temporal; y
proporcionar datos de la transacción al primer agente antes de la fase de
respuesta e independientemente de la respuesta, proporcionando la memoria temporal los datos en respuesta a la detección del acierto.

11. El procedimiento según la reivindicación 10 en el que, incluso si uno o más de los otros agentes detectan un acierto para la solicitud de memoria, el uno o más de los otros agentes no proporcionan datos si la memoria temporal detecta el acierto.

12. El procedimiento según la reivindicación 10 que comprende adicionalmente:

almacenar los datos de la transacción en el primer agente;
almacenar un estado de memoria temporal en el primer agente que causaría que el primer agente proporcionara datos en respuesta a un posterior examen en el primer agente; e
invalidar los datos de una entrada de memoria temporal que almacena los datos en la memoria temporal en respuesta a la transacción de memoria.

13. El procedimiento según la reivindicación 10 que comprende adicionalmente:

desalojar un bloque de memoria temporal que se encuentra en un estado exclusivo pero no modificado de uno de la pluralidad de agentes; y
generar una transacción de reescritura para el bloque de memoria temporal;
la asignación por parte de la memoria temporal de una entrada de memoria temporal en respuesta a la transacción de reescritura;
y
aceptar el bloque de memoria temporal en la entrada.

14. El procedimiento según la reivindicación 10 que comprende adicionalmente la memoria temporal solicitando de forma especulativa una interconexión de datos en respuesta a la recepción de la solicitud de memoria de la interconexión antes de detectar el acierto.

15. El procedimiento según la reivindicación 10 que comprende adicionalmente proporcionar una respuesta temprana desde la memoria temporal al primer agente, antes de la fase de respuesta, para indicar que la memoria temporal proporcionará los datos.

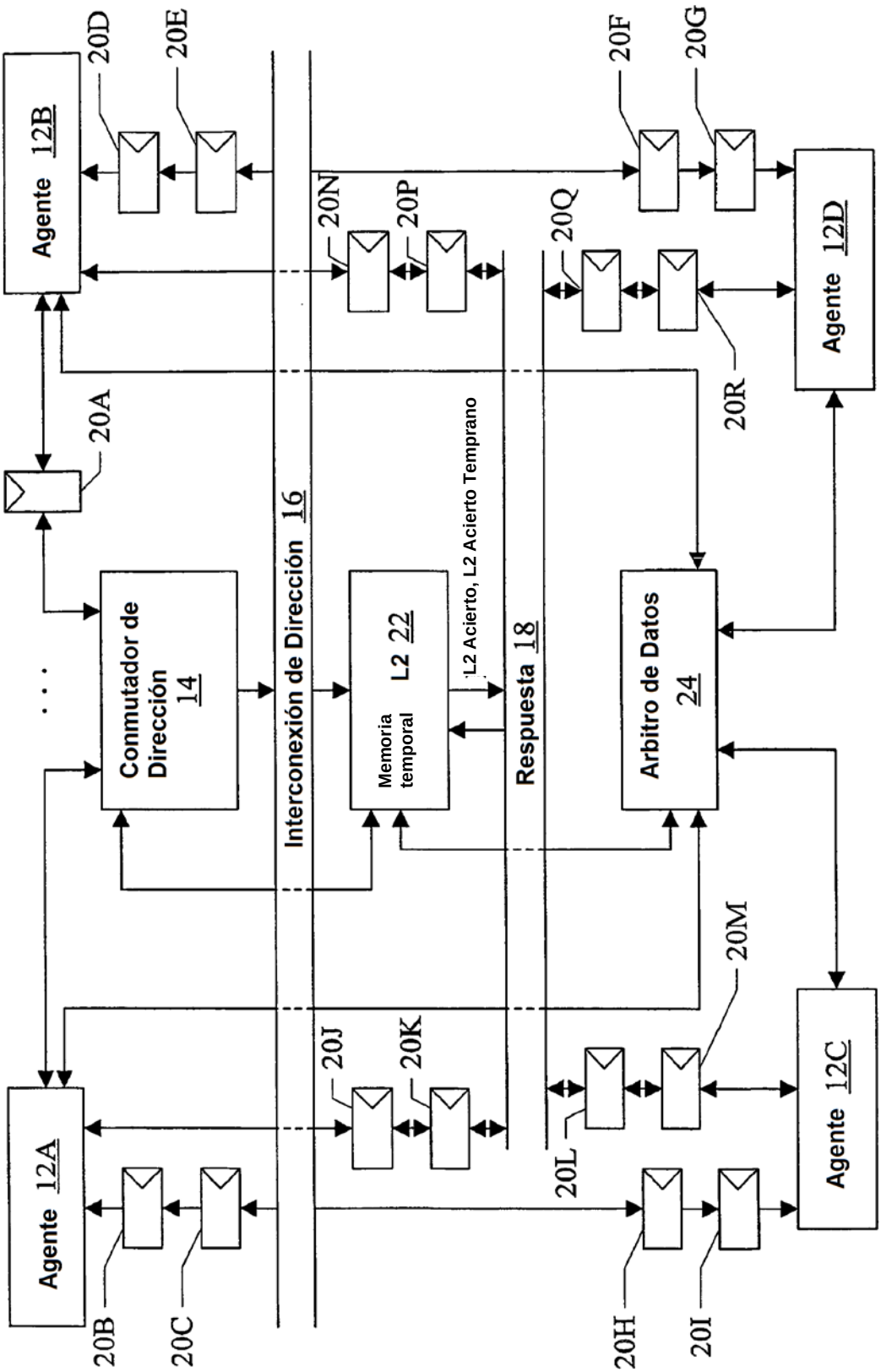
Agente 12B
20D
20E
20F
20G
20N
20P
20Q
20R
Agente 12D
20A
Conmutador de Dirección 14
Interconexión de Dirección 16
Memoria temporal L2 22
L2 Acierto, L2 Acierto Temprano
Respuesta 18
Arbitro de Datos 24
20M
20J
20K
20L
Agente 12A
Agente 12C
20B
20C
20H
20I
10

Fig. 1

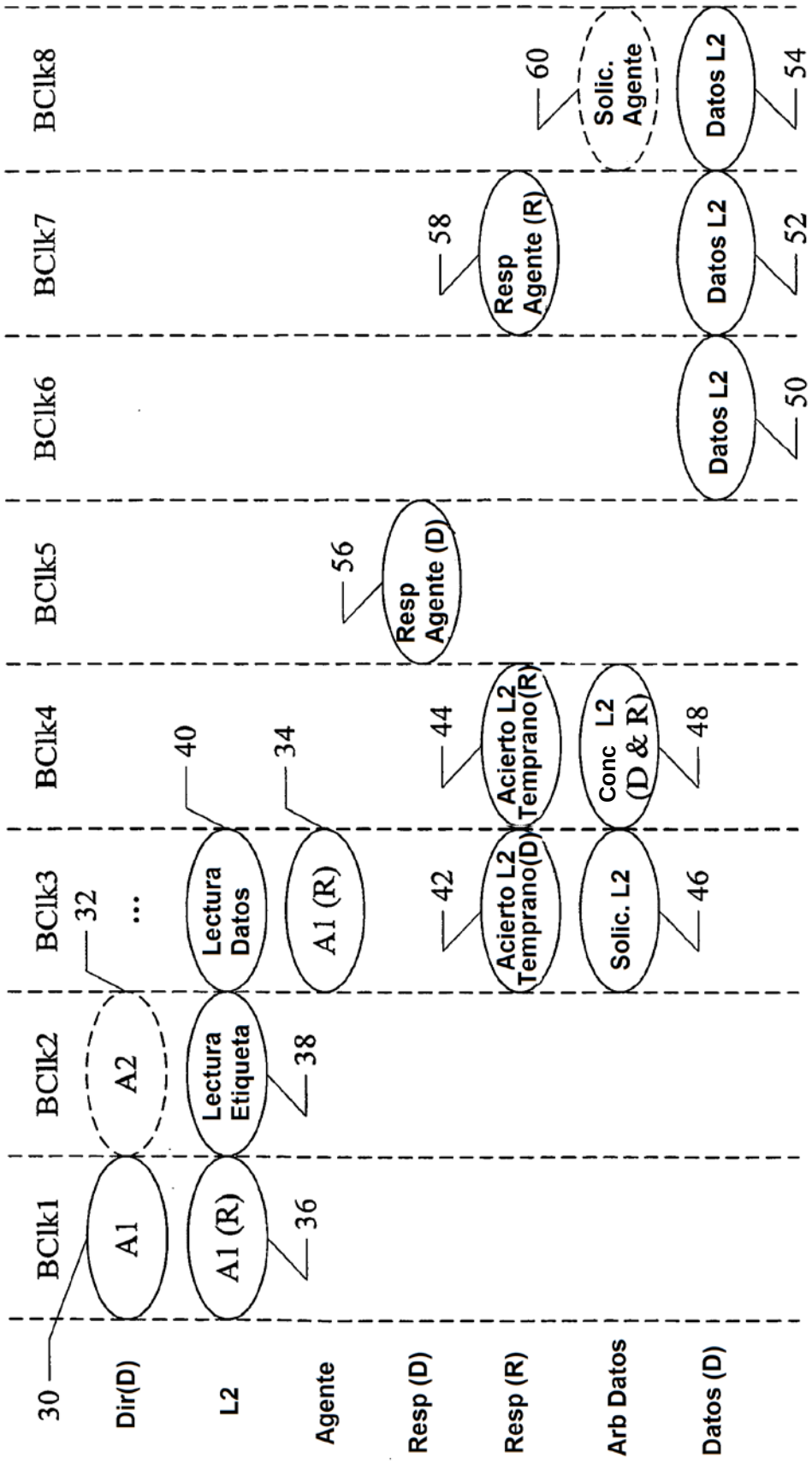
BClk1
BClk2
BClk3
BClk4
BClk5
BClk6
BClk7
BClk8
30
Dir(D)
L2
Agente
Resp (D)
Resp (R)
Arb Datos
Datos (D)
A1
A2
...
32
A1 (R)
36
Lectura Etiqueta
38
Lectura Datos
40
A1 (R)
34
Resp Agente (D)
56
Acierto L2 Temprano(D)
42
Acierto L2 Temprano(R)
44
Solic. L2
46
Conc L2 (D & R)
48
Datos L2
50
Datos L2
52
Datos L2
54
Resp Agente (R)
58
Solic. Agente
60

Fig. 2

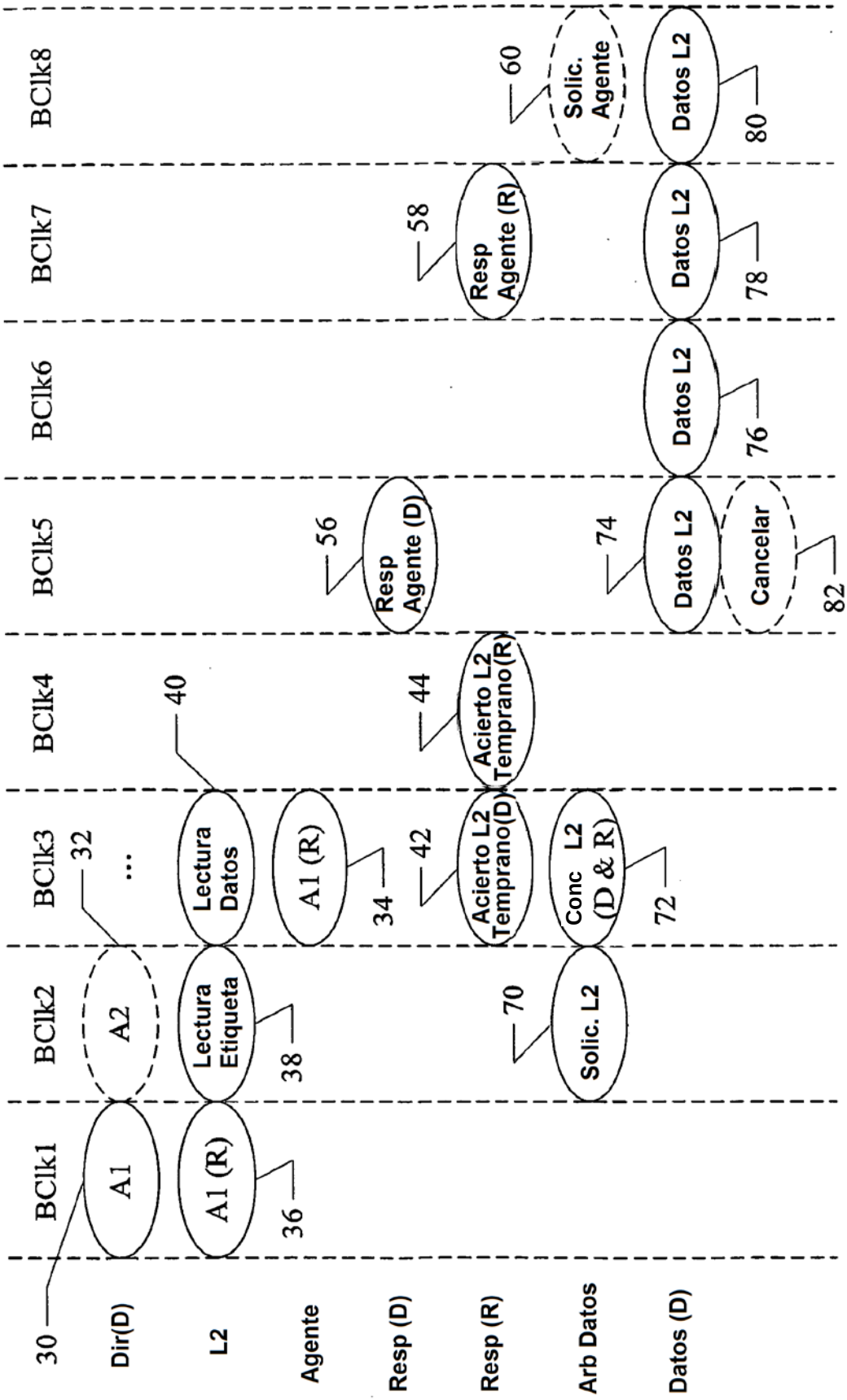
BClk1
BClk2
BClk3
BClk4
BClk5
BClk6
BClk7
BClk8
30
Dir(D)
A1
A2
...
32
L2
A1 (R)
Lectura Etiqueta
Lectura Datos
40
36
38
Agente
A1 (R)
34
Resp (D)
Resp Agente (D)
56
Resp (R)
Acierto L2 Temprano(D)
42
Acierto L2 Temprano(R)
44
58
Resp Agente (R)
60
Arb Datos
Solic. L2
70
Conc L2 (D & R)
72
Datos (D)
Datos L2
74
Datos L2
76
Datos L2
78
Datos L2
80
Cancelar
82
Solic. Agente

Fig. 3

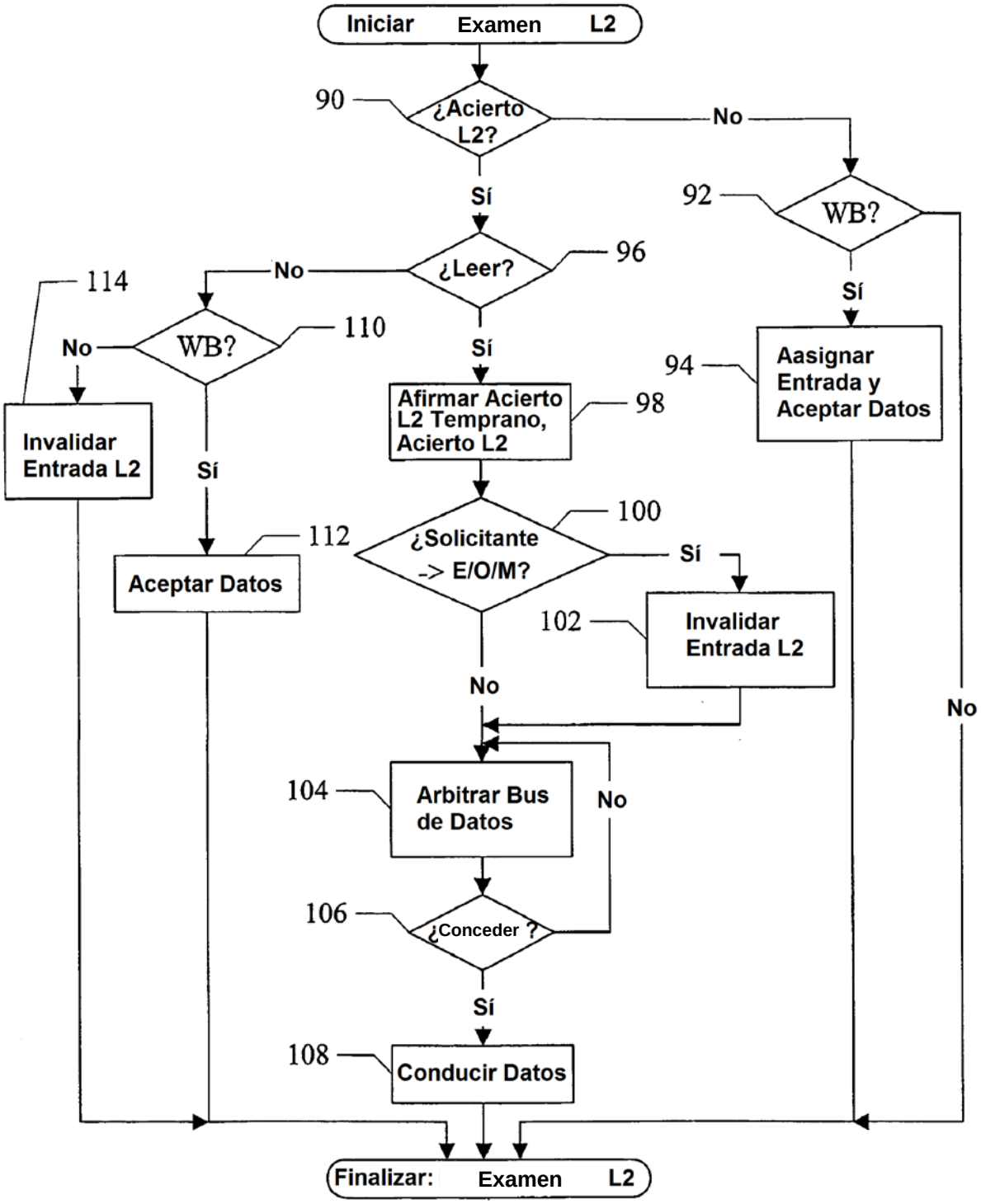
Iniciar Examen L2
90
¿Acierto L2?
No
Sí
92
WB?
Sí
94
Aasignar Entrada y Aceptar Datos
No
96
¿Leer?
No
Sí
114
No
WB?
110
Sí
Invalidar Entrada L2
Afirmar Acierto L2 Temprano, Acierto L2
98
100
¿Solicitante -> E/O/M?
Sí
102
Invalidar Entrada L2
112
Aceptar Datos
No
No
104
Arbitrar Bus de Datos
No
106
¿Conceder ?
Sí
108
Conducir Datos
Finalizar: Examen L2

Fig. 4

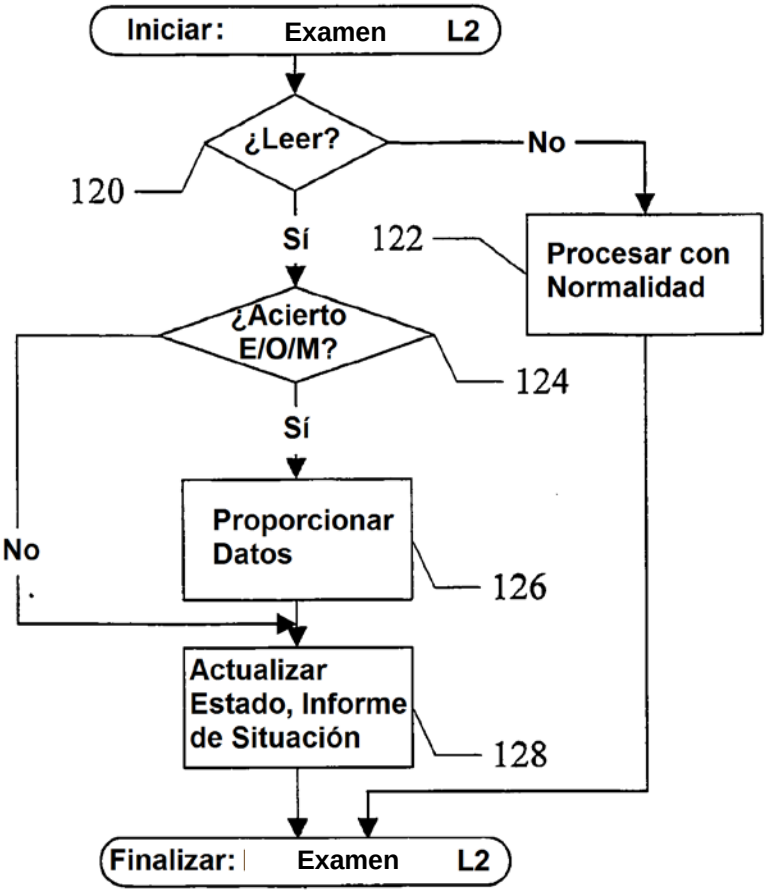
Iniciar: Examen L2
¿Leer?
No
120
Sí
122
Procesar con Normalidad
¿Acierto E/O/M?
124
Sí
Proporcionar Datos
No
126
Actualizar Estado, Informe de Situación
128
Finalizar: Examen L2

Fig. 5

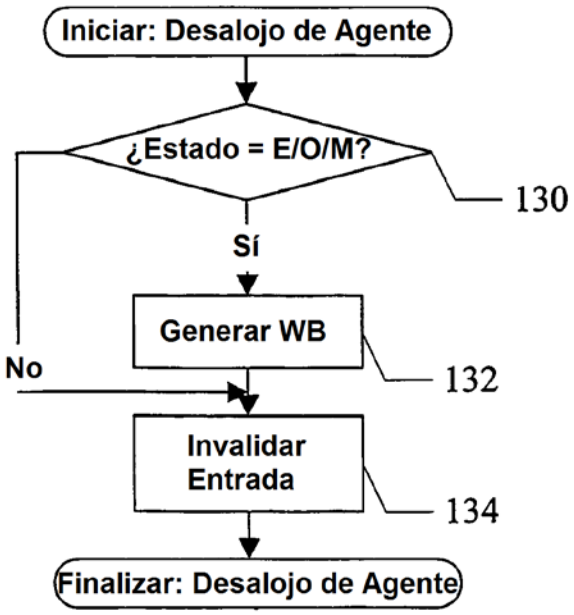
Iniciar: Desalojo de Agente
¿Estado = E/O/M?
130
Sí
Generar WB
No
132
Invalidar Entrada
134
Finalizar: Desalojo de Agente

Fig. 6