

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 391 733**

21 Número de solicitud: 201101381

51 Int. Cl.:

G06F 15/80

(2006.01)

12

PATENTE DE INVENCION CON EXAMEN PREVIO

B2

22 Fecha de presentación:

30.12.2011

43 Fecha de publicación de la solicitud:

29.11.2012

Fecha de modificación de las reivindicaciones:

18.12.2012

Fecha de la concesión:

26.04.2013

45 Fecha de publicación de la concesión:

10.05.2013

73 Titular/es:

**UNIVERSIDADE DE SANTIAGO DE
COMPOSTELA (100.0%)
EDIFICIO EMPRENDIA-CAMPUS VISA
15782 SANTIAGO DE COMPOSTELA (A Coruña) ES**

72 Inventor/es:

**NIETO LAREO, Alejandro Manuel;
BREA SÁNCHEZ, Victor Manuel y
LÓPEZ VILARIÑO, David**

54 Título: **ARQUITECTURA HÍBRIDA SIMD/MIMD DINÁMICAMENTE RECONFIGURABLE DE UN
COPROCESADOR PARA SISTEMAS DE VISIÓN**

57 Resumen:

La presente invención se refiere a una arquitectura híbrida Simple Instrucción-Múltiples Datos (SIMD)/Múltiples Instrucciones-Múltiples Datos (MIMD), dinámicamente reconfigurable, de un coprocesador que se emplea en sistemas altas prestaciones para tareas de visión por computador. La arquitectura comprende un conjunto de elementos de procesamiento (PE) que reciben datos a través de una red local o de una cola de entrada serie salida paralela (SIPO), ambas gestionadas por un procesado de entrada programable (PIP). Los resultados obtenidos se envían al exterior a través de una cola de entrada paralela y salida serie (PISO) o a través de la red local, estando ambos elementos gestionados por el procesador de salida programable (POP).

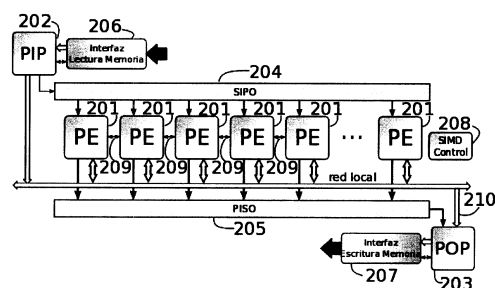


Fig. 2

ES 2 391 733 B2

DESCRIPCIÓN

Arquitectura híbrida SIMD/MIMD dinámicamente reconfigurable de un coprocesador para sistemas de visión

5 CAMPO DE LA TÉCNICA

La presente invención se refiere a una arquitectura híbrida Simple Instrucción-Múltiples Datos (SIMD)/Múltiples Instrucciones-Múltiples Datos (MIMD), dinámicamente reconfigurable, de un coprocesador que se emplea en sistemas de altas prestaciones para tareas de visión por computador.

La invención presenta una aplicación en sistemas de visión artificial para aplicaciones de vídeo-vigilancia inteligente, guiado de robots, procesamiento de imagen médica o dispositivos de seguridad en el automóvil, y en general a sistemas que se basen en el procesamiento de imágenes o vídeo digitales.

15 ANTECEDENTES DE LA INVENCION

En la actualidad existen una gran variedad de técnicas para procesar imagen y vídeo digital que son ampliamente utilizadas en la industria, algunos ejemplos de estas técnicas son la detección y el seguimiento de objetos o la monitorización de procesos. Estas técnicas suelen tener un elevado coste computacional, que implican un elevado tiempo de cómputo cuando se procesan en sistemas convencionales, como puede ser un ordenador personal, por lo que existe una gran demanda de dispositivos hardware que permitan reducir su tiempo de cómputo. La traslación de la implementación software en un ordenador a dispositivos hardware no es directa y en muchos casos hay que realizar una serie de simplificaciones que reducen la funcionalidad del sistema. Además, los elevados costes computacionales de estas técnicas van en contra de los requerimientos de bajo consumo de potencia, consumo de memoria y forma de las plataformas hardware.

La industria demanda aplicaciones de visión capaces de realizar tareas de análisis sin afectar a la precisión o al rendimiento del sistema. Los recientes avances en la industria han permitido el desarrollo de sistemas de visión por computador que poseen una gran eficiencia y un gran poder de cómputo, así como bajo consumo de potencia y un reducido coste de ingeniería no recurrente. Debido al elevado coste computacional de muchas aplicaciones industriales, tales como vídeo-vigilancia, guiado de robots o dispositivos de seguridad en el automóvil, los campos de arquitectura de computadores, computación serie y paralela y sistemas inteligentes han sufrido un gran desarrollo durante estos últimos años. Los esfuerzos realizados en estas áreas han ido encaminados al desarrollo de sistemas de procesamiento de imagen y vídeo. Además de los requerimientos de precisión y rendimiento, a este tipo de sistemas también se les exige la obtención de prototipos en un corto período de tiempo y a un bajo precio.

Los circuitos programables Field Programmable Gate Array (FPGA) se usan ampliamente para reducir el tiempo de cómputo de las aplicaciones de visión por computador. Debido a que las FPGAs ofrecen unas prestaciones menores que los circuitos integrados para aplicaciones específicas, ASIC del inglés Application Specific Integrated Circuit, al permitir reconfigurar el hardware en aplicaciones que requieran que la respuesta del sistema sea un factor crítico el producto final suele ser un ASIC. Sin embargo, en los últimos años las FPGAs han dado un gran salto al incorporar más recursos, como pueden ser bloques de Procesador de Señal Digital, del inglés Digital Signal Processor (DSP) o bloques de memoria RAM, sin un incremento significativo del coste no recurrente. Los Sistemas en Chip, del inglés System on Chip (SoC), se ven favorecidos, reduciendo el ciclo de desarrollo, por el uso de FPGAs. Así, el uso de uno u otro tipo de circuitos dependerá en gran medida del volumen de producción.

25

Los sistemas de visión por computador tienen que gestionar eficientemente una serie de operaciones, dependencias entre datos y flujo de datos. Las operaciones matemáticas que se deben realizar en los sistemas de visión por computador se pueden agrupar en:

- Operaciones de bajo nivel simples y repetitivas que trabajan sobre un conjunto de datos elevados, en las que se requiere gran paralelismo. Son operaciones con un patrón regular de acceso a los datos y de control.
- 5 • Operaciones de nivel medio en las que existe dependencias en el flujo de datos. Son operaciones que presentan un acceso regular a los datos pero en las que hay un flujo irregular de datos y control.
- Operaciones de alto nivel que trabajan sobre conjuntos complejos y reducidos de datos y que requieren una gran precisión. Debido a su complejidad este tipo de operaciones suelen ser llevadas a cabo por procesadores de propósito general.

10

Debido a la complejidad de los algoritmos de visión por computador cualquier sistema de propósito general debe ser capaz de abordar estos tres tipos de operaciones,. A causa de la variabilidad de operaciones con las que tienen que trabajar los sistemas de visión por computador se han desarrollado arquitecturas que permiten operar con la

15 diversidad de operaciones empleadas en este tipo de sistemas. Así, se han desarrollado sistemas que permiten resolver operaciones de bajo y medio nivel de forma eficiente, en las que se puede explotar el paralelismo espacial ofrecido por una implementación hardware, y que actúan como aceleradores de procesadores de propósito general, que se encargan de llevar a cabo las operaciones de alto nivel.

20 Las operaciones de bajo nivel se enmarcan dentro del campo de procesadores SIMD, que están formados por un elevado conjunto de elementos de procesamiento, que realizan la misma operación sobre distintos datos y con los que obtienen un grado de paralelismo espacial elevado. Por contra, las operaciones de nivel medio son tratadas por procesadores MIMD, que están formados por varios elementos de procesamiento,

25 en los que se ejecutan distintas instrucciones sobre distintos datos de acuerdo a flujos irregulares de datos y de control.

Los procesadores SIMD han sido ampliamente utilizados para la resolución de tareas de bajo nivel, pero el desarrollo de técnicas avanzadas de procesamiento de imágenes y vídeo han propiciado nuevas necesidades, por lo que el desarrollo de procesadores

30 MIMD se ha visto favorecido. Recientemente, se han propuesto procesadores híbridos

que se engloban en la tipología SIMD y MIMD, que combinan características de ambos tipos de procesadores.

En la [Referencia 1] se propone un sistema híbrido SIMD/MIMD cuya funcionalidad queda restringida a multiplicación de matrices, lo que limita su aplicación en algoritmos de visión por computador complejos, reduciendo su aplicación a operaciones de convolución (operación de bajo nivel). En la [Referencia 2] se propone un sistema híbrido SIMD/MIMD en el que la funcionalidad de los elementos de procesamiento se define mediante una serie de instrucciones almacenadas en un buffer. Se permite el paso de modo SIMD a MIMD y viceversa aplicando una serie de conmutadores. La principal desventaja de este sistema es que la memoria se encuentra separada de los elementos de procesamiento, con lo que se reduce la eficiencia del sistema en términos de acceso a datos. En la [Referencia 3] se propone una arquitectura híbrida SIMD/MIMD en la que los elementos de procesamiento SIMD se agrupan en bloques MIMD, lo que es ineficiente para tareas de visión por computador en las que la presencia de un gran número de elementos de procesamiento que funcionen en modo SIMD es de gran importancia. En la [Referencia 4] se propone un sistema formado por una matriz 2D de elementos de procesamiento, en el que las instrucciones se envían a los elementos de procesamiento mediante el uso de un decodificador de filas y columnas, lo que es altamente ineficiente en términos de consumo de área. En la [Referencia 5] se propone un sistema en el que la funcionalidad de los elementos de procesamiento está almacenada en memorias dentro del sistema, y por medio de conmutadores se puede elegir la funcionalidad de los elementos de procesamiento. Esta estructura presenta el inconveniente del gran espacio de memoria requerida para dotar al sistema de una funcionalidad amplia, ya que se necesitaría un gran número de memorias, lo que redundaría en el área ocupada. En la [Referencia 6] se propone un sistema híbrido SIMD/MIMD centrado en la gestión de la eficiencia de la comunicación de los elementos de procesamiento con la memoria, que acarrea una gestión complicada de la memoria y limita su aplicación en aplicaciones de visión por computador.

[Referencia 1] Xiaofang Wang and Sotirios G. Ziavras. "HERA: A Reconfigurable and Mixed-Mode Parallel Computing Engine on Platform FPGAs", en Proc. of the 16th

International Conference on Parallel and Distributed Computing and Systems (PDCS), MIT, Cambridge, MA, Nov. 9-11, 2004.

[Referencia 2] US 5,535,410. 09/1996. Watanabe T. et al. "Parallel processor having
5 decoder for selecting switch from the group of switches and concurrently inputting
MIMD instructions while performing SIMD operation".

[Referencia 3] US 7,853,775. 14/2010. Shorin Kyo. "Processing elements grouped in
MIMD sets each operating in SIMD mode by controlling memory portion as instruction
10 cache and GPR portion as TAG".

[Referencia 4] US 7,089,436. 08/2006. Ming-Hau Lee and Fadi Kurdahi. "Power saving
method and arrangement for a reconfigurable array".

15 [Referencia 5] US 7,814,296. 10/2010. Gun-Gi Luy et al. "Arithmetic units responsive
to common control signal to generate signals to selectors for selecting instructions from
among respective program memories for SIMD / MIMD processing control".

[Referencia 6] EP 2132645. 05/2011. Lieske Hanno and Kyo Shorin. "A data transfer
20 network and control apparatus for a system with an array of processing elements each
either self- or common controlled".

DESCRIPCIÓN DE LA INVENCION

25

La presente invención hace referencia a una arquitectura híbrida SIMD/MIMD de un coprocesador que se puede emplear para el procesamiento de imagen o vídeo digital en el ámbito de aplicaciones de visión por computador. La arquitectura del

coprocesador está diseñada para formar parte de un SoC. Una representación genérica de dicho sistema puede verse en la Figura 1. Un SoC genérico está formado por una unidad central de proceso, del inglés Central Process Unit (CPU), 101, un coprocesador de imagen, que realiza operaciones de cálculo intensivo, 102, una serie de elementos de

5 Entrada/Salida (E/S), como son registro de propósito general de entrada y salida, del inglés Global Purpose Input Output register, 103, un control visualizador, 103, un bus JTAG, 103, un puerto Ethernet, 103, un bus USB, 103, un controlador de memoria multipuerto, 103, y un subsistema Video Entrada/Salida (E/S), 103, y que permiten la comunicación del SoC con el exterior, ya sea con dispositivos de visualización o

10 captura de datos o con una memoria externa, una memoria flash interna, 104, un gestor de control de interrupciones, 105, un módulo para gestionar el acceso directo a una memoria exterior, del inglés Direct Memory Access (DMA), 106. En este SoC el coprocesador, cuya realización es objeto de la presente invención, actúa como un acelerador de la CPU principal llevando a cabo aquellas tareas de procesado de imagen

15 que requieran de un elevado rendimiento.

La arquitectura del coprocesador, Figura 2, comprende un conjunto de elementos de procesamiento, PE del inglés Processing Element, 201, formado por 2 o más elementos, un procesador de entrada programable, 202, PIP del inglés Programmable Input Processor, un procesador de salida programable, 203, POP del inglés

20 Programmable Output Processor, una cola de entrada serie-salida paralela, 204, SIPO del inglés Serial Input-Parallel Output, una cola de entrada paralela-salida serie, PISO del inglés Parallel Input-Serial Output, 205, una interfaz para leer la memoria externa, 206, una interfaz para escribir en la memoria externa, 207, un módulo de control SIMD global, 208, un conjunto de interconexiones directas entre los elementos de

25 procesamiento, 209, y una red de conexión local, 210.

En la presente invención, 201, se organizan pudiendo operar en dos modos diferentes, SIMD y MIMD. En el modo SIMD todos los elementos de procesamiento realizan la misma operación sobre distintos datos, mientras que en el modo MIMD cada elemento de procesamiento realiza una operación diferente sobre distintos datos. Los

30 interfaces de lectura de memoria y de escritura de memoria se encargan, respectivamente, de obtener datos y enviar datos a una memoria externa en la que se

almacenan los datos a procesar por el coprocesador (lectura) y los resultados obtenidos por el mismo (escritura). Las conexiones directas entre elementos de procesamiento, 209, permiten el intercambio de datos entre un elemento de procesamiento y sus dos vecinos adyacentes, situados a izquierda y derecha cuando el sistema funciona en modo

5 SIMD, organizándose los elementos de procesamiento de acuerdo a un array 1D. La red local de conexión, 210, posibilita el intercambio de datos entre elementos de procesamiento adyacentes cuando el sistema funciona en modo MIMD, permitiendo la distribución de los elementos de procesamiento de acuerdo a organizaciones complejas, como una estructura de Toro o de Hipercubo, tal y como se detallarán más adelante. En

10 esta arquitectura la cola SIPO, 204, se encarga de distribuir los datos provenientes de memoria, proporcionados por el PIP (202), a los elementos de procesamiento cuando estos trabajan en modo SIMD. Por otro lado, la cola PISO, 205, se encarga de enviar los datos procesados por los elementos de procesamiento, cuando estos trabajan en modo SIMD, a la memoria externa a través del POP, 203.

15 El PIP, 202, se encarga de gestionar la comunicación con la memoria externa para obtener los datos que se procesarán, mientras que el POP, 203, se encarga de recoger los datos suministrados por los elementos de procesamiento y mandarlos a la memoria externa. Aunque la funcionalidad del PIP y el POP es complementaria, encargándose el primero de la lectura de datos de memoria y el segundo de escritura de datos en la

20 memoria, su estructura es idéntica y se muestra en la Figura 3. Ambos procesadores trabajan en paralelo, posibilitando el solape de operaciones de acceso a memoria cuando sea posible. Para soportar los distintos patrones de acceso a memoria que presentan los distintos algoritmos de visión por computador, tanto el PIP como el POP calculan automáticamente direcciones de memoria usando un conjunto de registros

25 reconfigurables, 302, y una unidad aritmética y lógica, 303, ALU del inglés Arithmetic and Logic Unit. Los procesadores PIP y POP tienen una memoria de Almacenamiento de programa, 301, en la que está almacenado el programa que ejecutan el PIP y el POP, la programación de estos procesadores la realiza la CPU, 101, adaptando de esta forma dinámicamente las necesidades de cómputo. Estos procesadores permiten el

30 alineamiento automático de datos procedentes de la memoria externa, eligiendo de qué forma llegan estos a los elementos de procesamiento, facilitando así el desarrollo de

- algoritmos de visión por computador en esta arquitectura. Tanto el PIP como el POP incorporan una caché, 304, que permite el envío de ráfagas de datos entre la memoria externa y el PIP o el POP a través de la interfaz de red, 305, que permite la comunicación de la memoria caché con la cola entrada serie-salida paralela cuando los
- 5 elementos de procesamiento trabajan en modo SIMD o con la red de conexión local cuando estos trabajan en modo MIMD. Para leer datos el sistema tiene una interfaz (306 y 307) que permite la comunicación de la memoria caché con una memoria externa para leer datos de la misma cuando los elementos de procesamiento trabajan en modo SIMD (306) y MIMD (307).
- 10 La generación de las direcciones de memoria se lleva a cabo empleando aritmética entera y un conjunto de registros, organizados en forma de cuartetos, 302, que configuran el patrón de acceso a la memoria. El cuarteto de registros consiste en un registro base, que proporciona la dirección base del patrón de acceso a memoria, un registro índice, que proporciona el desplazamiento relativo dentro del conjunto de datos
- 15 a los que se accede, un registro de incremento, que proporciona el incremento del índice después de cada operación de lectura, y un registro de módulo que indica el tipo de aritmética de generación de direcciones que se puede emplear, y que puede ser:
- Lineal: es el modo de generación de direcciones más sencillo y el empleado por defecto, consiste en incrementar o disminuir el índice de acuerdo al valor
 - 20 almacenado en el registro de incremento.
 - Módulo: en este modo el valor del índice permanece dentro de unos márgenes de forma automática sin ninguna intervención de alguna instrucción de control específica, fijando el registro de módulo a un límite superior, M. Para obtener el valor del índice se incrementa o reduce su valor y se realiza una operación
 - 25 módulo M, que proporciona el nuevo valor, que queda restringido al rango 0-M. Esta aritmética de direccionamiento es útil para buffers circulares, colas FIFO, del inglés First In First Out, o para aplicar el mismo algoritmo sobre distintos bloques de datos de la memoria.
 - Acarreo inverso: en este caso el bit de acarreo se propaga en dirección inversa,
 - 30 desde el bit más significativo, del inglés Most Significant Bit (MSB), al bit

menos significativo, del inglés Less Significant Bit (LSB). Esta aritmética de direccionamiento es equivalente a cambiar el orden de los bits de los datos y del resultado y es útil para trabajar con algoritmos que empleen la transformada de Fourier rápida, del inglés Fast Fourier Transform (FFT)

- 5 El PIP y el POP operan en paralelo, por lo que cada uno de ellos puede proporcionar una dirección de memoria válida en cada ciclo. Para aumentar la flexibilidad del PIP y el POP los registros del cuarteto de registros se pueden tratar como registros de datos y realizar sobre ellos operaciones de adición o substracción para modificar sus valores. Este tipo de operación se puede realizar en tiempo de ejecución.
- 10 Así, el PIP permite el envío de datos procedentes de la memoria externa hacia la cola SIPO (modo SIMD) y a la red de conexión local (modo MIMD), mientras que el POP permite recibir datos de la cola PISO (modo SIMD) y la red de conexión local (modo MIMD) para enviarlos a la memoria externa.

- El conjunto de PEs es la parte más importante de la presente arquitectura, ya que
- 15 define la funcionalidad con la que se dota al sistema. El conjunto de PEs está formado por elementos de procesamiento conectados a través de una red de conexión local y de conexiones locales directas entre PEs adyacentes, como se muestra en la Figura 2. El PE tiene un conjunto de instrucciones reducido, del inglés Reduced Instruction Set Computer (RISC), que permite segmentar y paralelizar los algoritmos de visión por
- 20 computador de forma eficiente. El PE consiste esencialmente en una ALU, 407, que opera en punto fijo y punto flotante, un banco de registros, 402, que almacena datos, un elemento de memoria, 405, que consiste esencialmente en una memoria de acceso aleatorio, del inglés Random Access Memory (RAM), y que almacena datos con los que opera el elemento de procesamiento y varía su funcionamiento dependiendo de si el
- 25 coprocesador trabaja en modo MIMD, 403, o SIMD, 404, como se detallará más adelante, un conjunto de colas FIFO, 401, que almacena los datos que vienen de la red de conexión local, 410, a través de la interfaz de red, 409, y permite realizar la sincronización de los distintos elementos de procesamiento cuando éstos operan en modo MIMD ya que permiten realizar la escritura y lectura de forma simultánea aunque
- 30 las frecuencias de lectura y escritura sean distintas, un selector de datos, 411, que permite seleccionar los datos con los que opera la ALU de entre los datos provenientes

de la red de conexión local a través de la cola de FIFO, del banco de registros, del elemento de memoria o de los elementos de procesamiento adyacentes, 406, una unidad de control MIMD, 408, que gestiona el comportamiento del elemento de procesamiento cuando opera en modo MIMD. En modo de funcionamiento SIMD un módulo de control SIMD local (412) basado en banderas (Control Flag) permite programar la funcionalidad del elemento de procesamiento. La ALU, 407, puede operar con hasta 3 datos simultáneamente y permite la realización de 3 tipos de operaciones diferentes:

- Operaciones básicas aritméticas, que consisten en operaciones de suma, resta, multiplicación.
- 10 • Operaciones de DSP, que consisten en operaciones de multiplicación-adición, adición-multiplicación, cálculo del valor absoluto, cálculo del valor absoluto-resta o comparaciones.
- Operaciones de cálculo del máximo y mínimo de un conjunto de datos.
- Operaciones booleanas, que consisten en operaciones de desplazamiento y operaciones a nivel de bit (bitwise operations).
- 15 • Operaciones de conversión del tipo de datos: conversión a punto flotante, punto fijo y entero.

A modo ilustrativo y sin limitar el alcance de la invención se muestran algunas instrucciones soportadas en la Tabla 1.

Tabla 1. Ejemplo de juego de instrucciones.

5	Aritméticas	A+B ADD
		A-B SUB
		A*B MULT
10	Cálculo máximos y mínimos	Max(A,B) MAX
		MIN(A,B) MIN
15 20 25	Operaciones DSP	Abs(A) ABS
		Sature(A) SAT
		A=A+B*C MAC
		(A+B)*C AMUL
		Cmp = CMPEQ
		Cmp < CMPLT
		Cmp <= CMPLE
30	Operaciones booleanas	y AND
		o OR
		y no AND NOT
		o exclusivo XOR
35	Operaciones de desplazamiento	desplazamiento izquierda SLL
		desplazamiento derecha SRL
		aritmética derecha SRA

40

Se describen a continuación los dos modos de funcionamiento de la arquitectura de coprocesador: SIMD y MIMD.

Modo SIMD:

En este modo de funcionamiento todos los elementos de procesamiento realizan la misma operación sobre distintos datos y la estructura del conjunto de elementos de procesamiento adopta la forma mostrada en la Figura 5, en la que podemos identificar el PIP, 501, que suministra datos a los elementos de procesamiento, 506, a través de la cola SIPO, 503, el POP, 502, que recoge los resultados proporcionados por los elementos de procesamiento a través de la cola PISO, 504, la unidad de control SIMD global, 505, que indica la operación que realizan los elementos de procesamiento, y el conjunto de conexiones entre elementos de procesamiento adyacentes, 507, que permite el intercambio de datos entre elementos de procesamiento de forma rápida. La unidad de control SIMD global almacena el programa que ejecutan los elementos de procesamiento y realiza el cálculo de direcciones de memoria, mediante una ALU interna calcula 3 direcciones de memoria en cada ciclo. El cálculo de direcciones es idéntico al realizado por PIP y POP. Incluye un banco de registros para almacenar los valores de Base, Índice, Incremento y Módulo, aunque en este caso hacen referencia al banco de memoria interno de cada PE y no a la memoria RAM externa. El SIMD global también decodifica la instrucciones y aserta las banderas de control. La principal ventaja de este modo de operación es el elevado grado de paralelismo que se obtiene ya que todos los elementos de procesamiento realizan la misma tarea de forma conjunta. La cola PISO, que recoge los datos y los manda al POP permite que los elementos de procesamiento puedan seguir procesando datos sin tener que esperar a que los datos sean transferidos al exterior, ya que se almacenan en la cola PISO y es el POP el que gestiona el envío de los datos a la memoria externa.

El elemento de procesamiento, en este modo de funcionamiento, se organiza de acuerdo a la arquitectura mostrada en la Figura 6. En este modo de funcionamiento la ALU, 604, puede operar con datos provenientes de la cola SIPO, 607, que se almacenan en el elemento de memoria, 601, del banco de registros, 602, o de los vecinos adyacentes, 605 y 606. La selección de datos se realiza mediante un selector de datos, 603. El módulo de Banderas de Control recibe las señales de la unidad de control global SIMD, 208, que son decodificadas y definen la operación que realiza el elemento de procesamiento.

En la Fig. 7 se resume el modo de funcionamiento de la arquitectura cuando los elementos de procesamiento trabajan en modo SIMD. En primer lugar se cargan los programas a ejecutar en el PIP, POP y SIMD control, quedando estos en modo de espera hasta que reciben una señal de inicio. A continuación el PIP busca el bloque de

5 datos en la memoria externa indicado en los cuartetos de registros que proporcionan las direcciones de memoria. Los datos se copian en la cola SIPO hasta que ésta se llena, mientras tanto el resto de los módulos se encuentran en modo de espera, según se indica en la Fig. 7. Una vez que la cola SIPO se ha llenado los datos se envían a las memorias locales de los elementos de procesamiento. Una vez que los elementos de

10 procesamiento han recibido los datos comienzan a operar sobre ellos de acuerdo a la funcionalidad con la que han sido programados. Simultáneamente con este proceso, la cola SIPO carga nuevos datos, que no se pueden mandar a los elementos de procesamiento ya que estos están ocupados procesando los datos recibidos anteriormente. Una vez que los elementos de procesamiento han finalizado envían los

15 resultados a la cola PISO y al mismo tiempo la cola SIPO vuelca nuevos datos a los elementos de procesamiento. El POP extrae los datos de la cola PISO y los envía a la memoria externa de acuerdo a las direcciones proporcionadas por los cuartetos de registros de direcciones del POP. El proceso se repite hasta que el PIP, SIMD control y el POP generan una señal de finalización de programa.

20

Modo MIMD:

25 En este modo de funcionamiento los elementos de procesamiento trabajan de forma coordinada pero cada uno ejecutando su propio conjunto de instrucciones sobre datos diferentes. En modo MIMD los elementos de procesamiento trabajan de acuerdo a una Red de Procesos de Kahn, del inglés Kahn Process Network (KPN). Así, el algoritmo que ejecuta el coprocesador se divide en varias partes, una de las partes es la

30 encargada de proporcionar datos a los elementos de procesamiento y está almacenada en el procesador PIP, otra parte es la encargada de extraer los resultados, que está

almacenada en el procesador POP y finalmente otra parte proporciona un programa por elemento de procesamiento que incluye la funcionalidad, control de flujo de datos y acceso a la red de conexión local por parte de los distintos elementos de procesamiento. En este modo de funcionamiento la arquitectura del coprocesador se puede configurar

5 como una estructura Toro 2-D, hipercubo u otra disposición de los elementos de procesado que se adapte mejor al conjunto de algoritmos que se pretenden desarrollar. En la Fig.8 se ilustra una realización de la arquitectura trabajando en modo MIMD en la que se ha adoptado una topología de conexiones de Toro 2-D. En cualquier caso, la red física quedará fijada y no puede modificarse. El elemento de procesamiento, en este

10 modo de funcionamiento, adopta la estructura de la Fig. 9, que consiste esencialmente en un elemento de memoria, 901, que almacena tanto datos como el programa que ejecuta el elemento de procesamiento, una unidad de control MIMD, 902, que gestiona el funcionamiento del elemento de procesamiento, un interfaz de comunicación con la red de conexión local de entrada, 903, y de salida, 904, que permite el intercambio de

15 datos entre los distintos elementos de procesamiento, un banco de registros, 909, que permite trabajar con datos provenientes del elemento de memoria evitando su latencia de acceso a los datos, un conjunto de colas FIFO, 907, tantas como vecinos estén conectados al elementos de procesamiento, que almacenan los datos provenientes de la red de conexión local, un valor inmediato, 909, que permite definir una constante

20 predefinida con la que se puede operar, y que viene determinada en la propia instrucción que se carga en el elemento de procesamiento, una ALU, 905, que permite operar hasta con 3 elementos, un selector de datos, 906, que proporciona datos provenientes del banco de registros, de la red de conexión local o del valor inmediato a la ALU y un selector de datos, 908, que permite seleccionar qué cadena del flujo de datos

25 proveniente de la red de conexión local se envía a la ALU.

En la Fig. 10 se puede ver el modo de funcionamiento de la arquitectura cuando los elementos de procesamiento trabajan en modo MIMD. En este caso se envían datos de distinto tamaño a los distintos elementos de procesamiento y cada uno de los elementos de procesamiento ejecuta sus propias operaciones sobre los datos recibidos.

30 Así, el proceso se inicia con la carga de los programas en el PIP, POP y en los elementos de procesamiento. El PIP suministra los datos a los procesadores de acuerdo a la programación realizada. Una vez que los elementos de procesamiento reciben los

datos comienzan a procesarlos. Finalmente, el POP extrae los resultados de los elementos de procesamiento y los envía a la memoria externa de acuerdo a las direcciones proporcionadas por los cuartetos de registros de direcciones del POP.

5

La descripción de la arquitectura se realiza mediante un lenguaje de descripción de hardware y permite su implementación directa sobre una FPGA o ASIC definiendo previamente la configuración de los parámetros del sistema, como son:

- 10 1. Número de elementos de procesamiento.
2. Longitud de las palabras de datos.
3. Capacidad de las memorias de los elementos de procesamiento.
4. Capacidad de los bancos de registros de los elementos de procesamiento.
5. Longitud de las colas.
- 15 6. Topología de interconexión entre los elementos de procesamiento.
7. Tipo de operaciones que implementan las ALUs de los elementos de procesamiento.

Este grado de configuración ofrece una funcionalidad muy elevada al coprocesador
20 al permitir adaptar el funcionamiento a la aplicación específica.

BREVE DESCRIPCIÓN DE LAS FIGURAS

Las realizaciones de la invención son ilustradas a modo de ejemplo y no actúan de
25 modo limitante, en las figuras que se acompañan:

La Fig. 1 muestra la arquitectura de un sistema para el procesado de imágenes y/o vídeo en sistemas de visión por computador.

La Fig. 2 muestra la arquitectura del coprocesador objeto de la presente invención, en la que se distinguen las distintas partes del mismo.

La Fig. 3 muestra la arquitectura de los procesadores que gestionan el intercambio de información con la memoria, el PIP y el POP.

La Fig. 4 muestra la arquitectura del elemento de procesamiento del coprocesador objeto de la presente invención.

- 5 La Fig. 5 muestra la arquitectura del coprocesador objeto de la presente invención cuando funciona en modo SIMD.

La Fig. 6 muestra la arquitectura de un elemento de procesamiento cuando trabaja en modo SIMD.

- 10 La Fig. 7 muestra el cronograma de operaciones de la arquitectura cuando los elementos de procesamiento trabajan en modo SIMD.

La Fig. 8 muestra una realización de la topología de conexiones en forma de Toro 2-D entre los elementos de procesamiento cuando estos trabajan en modo MIMD.

La Fig. 9 muestra la configuración de un elemento de procesamiento en modo MIMD.

- 15 La Fig. 10 muestra el cronograma de operaciones de la arquitectura cuando los elementos de procesamiento trabajan en modo MIMD.

REIVINDICACIONES

1. Una arquitectura de un coprocesador empleado para el procesamiento de imágenes
5 que comprende :
 - a. un conjunto de elementos de procesamiento formado por dos o más
elementos (201) que pueden trabajar en modo SIMD (Single Instruction
Multiple Data) o MIMD (Multiple Instruction Multiple Data);
 - b. un procesador de entrada programable (202), PIP (Programmable Input
10 Processor), que suministra datos al conjunto de elementos de
procesamiento;
 - c. un procesador de salida programable (203), POP (Programmable Output
Processor), que extrae los datos obtenidos por el conjunto de elementos
de procesamiento hacia el exterior del coprocesador;
 - 15 d. una red de conexión local (210) que permite la comunicación entre los
elementos de procesamiento cuando estos trabajan en modo MIMD;
 - e. un conjunto de interconexiones directas entre los elementos de
procesamiento (209) que permite la conexión de uno de estos elementos
de procesamiento con sus 2 vecinos adyacentes, es decir, los situados a
20 izquierda y derecha cuando estos trabajan en modo SIMD;
 - f. una cola de entrada serie-salida paralela (204), SIPO (Serial Input-
Parallel Output), que distribuye los datos proporcionados por el
procesador de entrada programable a los elementos de procesamiento
cuando estos trabajan en modo SIMD;
 - 25 g. una cola de entrada paralela-salida serie (205), PISO (Parallel Input-
Serial Output), que extrae los datos de los elementos de procesamiento y
los distribuye al procesador de salida programable cuando estos trabajan
en modo SIMD;
 - h. un módulo de control SIMD global (208) que permite programar la
30 funcionalidad del conjunto de elementos de procesamiento cuando estos
trabajan en modo SIMD;
 - i. una interfaz de lectura de memoria (206), que permite la comunicación

- del PIP con la memoria externa para realizar la lectura de datos; y
- j. una interfaz de escritura de memoria (207), que permite la comunicación del POP con la memoria externa para realizar la escritura de datos.
2. La arquitectura según la reivindicación 1, apartado a, en la que los elementos de procesamiento (201) consisten esencialmente en:
- 5 a. una Unidad Aritmética y Lógica (407), ALU (Arithmetic and Logic Unit), que opera con aritmética en punto fijo o aritmética en punto flotante y permite la representación de datos con signo o sin signo;
- 10 b. un elemento de memoria (405) que proporciona datos a la Unidad Aritmética y Lógica cuando el elemento de procesamiento trabaja en los modos SIMD o MIMD, y que almacena el conjunto de instrucciones que el elemento de procesamiento debe ejecutar cuando trabaja en modo MIMD (404);
- 15 c. un banco de registros (402) que permite el almacenamiento de datos;
- d. una interfaz de lectura y escritura que comunica el elemento de procesamiento con la red de conexión local y con los vecinos adyacentes, que permite el intercambio de datos entre los elementos de procesamiento;
- 20 e. un módulo de control SIMD local (412) basado en banderas (Control Flag) que permite programar la funcionalidad del elemento de procesamiento cuando este trabaja en modo SIMD;
- f. un módulo de control MIMD (408) que permite programar la funcionalidad de los elementos de procesamiento cuando estos trabajan en modo MIMD;
- 25 g. un selector de datos (411) que proporciona los datos a la Unidad Aritmética y Lógica;
- h. una cola FIFO (401), del inglés First In-First Out, que almacena datos provenientes de la red de acceso local y permite realizar la sincronización del flujo de datos de los distintos elementos de procesamiento;
- 30 i. un interfaz de red (409) que permite en envío y recepción de datos de la red de conexión local (410); y

- j. una conexión directa a los elementos de procesamiento adyacentes (406) que permite intercambiar información cuando los elementos de procesamiento trabajan en modo MIMD.
- 3. La arquitectura según la reivindicación 2, apartado a, en la que la Unidad Aritmética y Lógica puede operar con hasta 3 datos y permite realizar las siguientes operaciones:
 - a. operaciones básicas aritméticas;
 - b. operaciones de procesador digital de señal, DSP (Digital Signal Processor);
 - 10 c. operaciones para el cálculo del máximo o el mínimo de un conjunto de datos;
 - d. operaciones booleanas: desplazamiento (shift) y operaciones a nivel de bit (bitwise operations); y
 - e. operaciones de conversión del tipo de datos: conversión a punto flotante, punto fijo y entero
- 15 4. La arquitectura según la reivindicación 2, apartado b, en la que el elemento de memoria consiste esencialmente en una Memoria de Acceso Aleatorio, RAM (Random Access Memory), de doble puerto que almacena datos, cuando los elementos de procesamiento trabajan en modo SIMD, y datos e instrucciones cuando los elementos de procesamiento trabajan en modo MIMD.
- 20 5. La arquitectura según la reivindicación 2, apartado e, en la que el módulo de control SIMD local basado en banderas recibe instrucciones del módulo de control SIMD global y las demultiplexa para que sean ejecutadas por el elemento de procesamiento.
- 25 6. La arquitectura según la reivindicación 2, apartado f, en la que el módulo de control MIMD local recibe instrucciones del Procesador de Entrada Programable, que establecen la funcionalidad del elemento de procesamiento.
- 7. La arquitectura según la reivindicación 2, apartado g, en la que el selector de datos puede seleccionar datos del banco de registros, del elemento de memoria, de la red de conexión local y de la conexión directa del elemento de procesamiento con sus vecinos adyacentes.
- 30 8. La arquitectura según la reivindicación 1, apartado b, en la que el procesador de

entrada programable (202) consiste esencialmente en:

- a. una memoria (301) que almacena las instrucciones que va a ejecutar el procesador de entrada programable con el fin de enviar información a los elementos de procesamiento;
 - 5 b. un conjunto de registros (302) organizados en cuartetos que controlan la aritmética de generación de direcciones de memoria;
 - c. una unidad aritmética y lógica (303) de enteros que opera con datos con signo y sin signo, que calcula direcciones de memoria de acuerdo a la aritmética de generación de direcciones de memoria;
 - 10 d. una memoria caché (304) que almacena direcciones de memoria y datos que se envían a los elementos de procesamiento;
 - e. una interfaz (305) que permite la comunicación de la memoria caché con la cola entrada serie-salida paralela cuando los elementos de procesamiento trabajan en modo SIMD o con la red de conexión local cuando estos trabajan en modo MIMD; y
 - 15 f. una interfaz (306 y 307) que permite la comunicación de la memoria caché con una memoria externa para leer datos de la misma cuando los elementos de procesamiento trabajan en modo SIMD (306) y MIMD (307).
- 20 9. La arquitectura según la reivindicación 8, apartado b, en la que cada cuarteto de registros consiste en:
- a. un registro, denominado registro base, que almacena la dirección base del conjunto de datos a los que se accede;
 - b. un registro, denominado registro índice, que almacena el desplazamiento relativo dentro del conjunto de datos a los que se accede;
 - 25 c. un registro, denominado registro de incremento, que almacena el incremento del registro índice que se realiza después de cada operación de lectura; y
 - d. un registro, denominado registro de módulo, que almacena el tipo de
 - 30 aritmética de generación de direcciones que se usa para la obtención de direcciones de memoria.
10. La arquitectura según la reivindicación 1, apartado c, en la que el procesador

de salida programable (203) consiste esencialmente en:

- a. una memoria (301) que almacena las instrucciones que va a ejecutar el procesador de salida programable para gestionar los datos proporcionados por los elementos de procesamiento;
 - 5 b. un conjunto de registros (302) organizados en cuartetos que controlan la aritmética de generación de direcciones de memoria;
 - c. una unidad aritmética y lógica (303) de enteros que opera con datos con signo y sin signo, que calcula y proporciona direcciones de memoria;
 - d. una memoria caché (304) que almacena direcciones de memoria y datos
10 enviados por los elementos de procesamiento;
 - e. una interfaz (305) que permite la comunicación de la memoria caché con la cola entrada paralelo-salida serie cuando los elementos de procesamiento trabajan en modo SIMD o con la red de conexión local cuando estos trabajan en modo MIMD; y
 - 15 f. una interfaz (306 y 307) que permite la comunicación de la memoria caché con una memoria externa para leer datos de la misma cuando los elementos de procesamiento trabajan en modo SIMD (306) y MIMD (307).
11. La arquitectura según la reivindicación 10, apartado b, en la que cada cuarteto
20 de registros comprende:
- a. un registro, denominado registro base, que almacena la dirección base del conjunto de datos a los que se accede;
 - b. un registro, denominado registro índice, que almacena el desplazamiento relativo dentro del conjunto de datos;
 - 25 c. un registro, denominado registro de incremento, que almacena el incremento del registro índice que se realiza después de cada operación de lectura; y
 - d. un registro, denominado registro de módulo, que almacena el tipo de aritmética que se usa para la generación de direcciones de memoria.
- 30 12. La arquitectura según la reivindicación 1, apartado d, en la que la red de conexión local puede ser configurada con distintas topologías de red.
13. La arquitectura según la reivindicación 1, en la que los siguientes parámetros

son configurables:

- a. número de elementos de procesamiento;
 - b. longitud de las palabras de datos;
 - c. capacidad de las memorias de los elementos de procesamiento;
 - 5 d. capacidad de los bancos de registros de los elementos de procesamiento;
 - e. longitud de las colas;
 - f. topología de interconexión entre los elementos de procesamiento; y
 - g. tipo de operaciones que implementan las ALUs de los elementos de procesamiento;
- 10 14. La arquitectura según la reivindicación 1 sintetizable en una FPGA (Field Programmable Gate Array) y ASIC mediante el uso de un lenguaje de descripción de hardware.

15

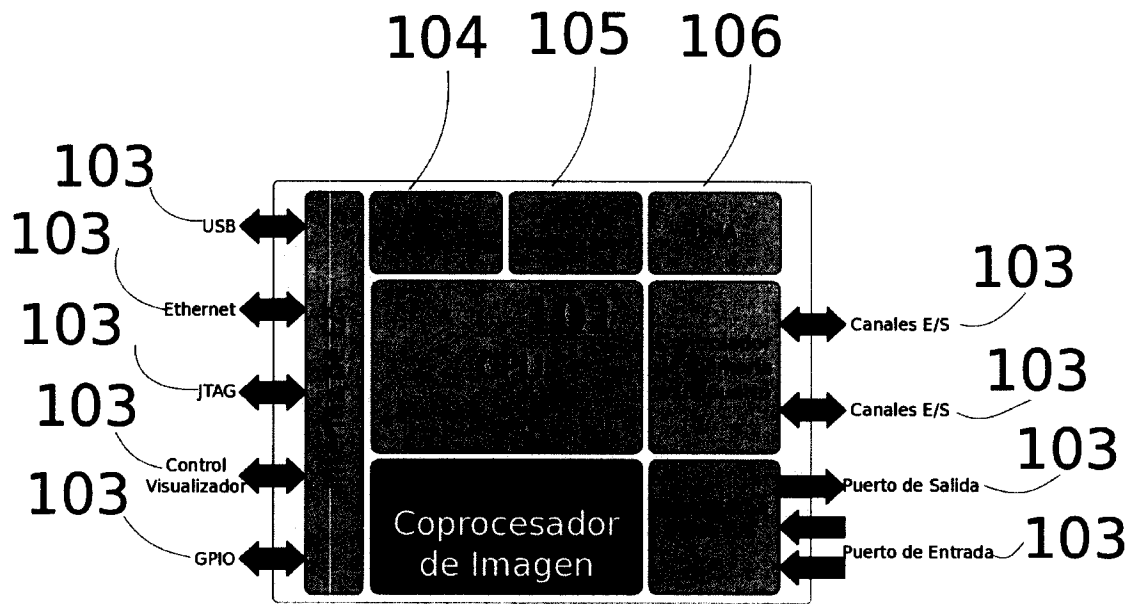


Fig. 1

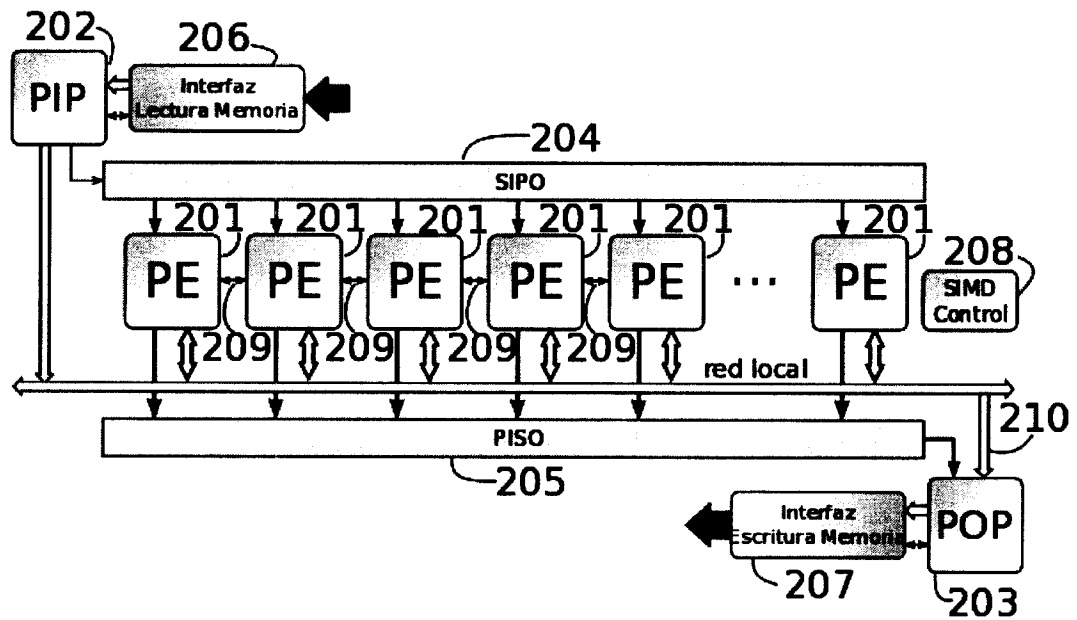


Fig. 2

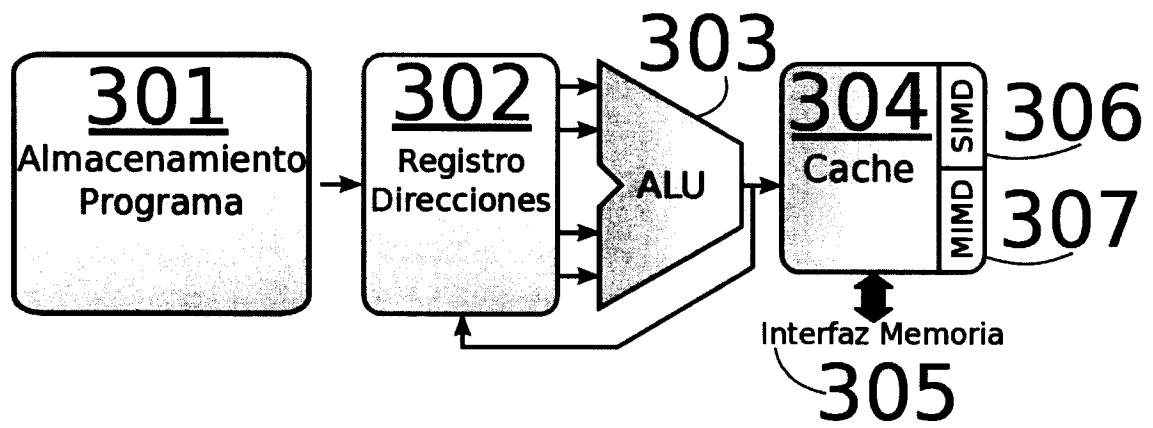


Fig. 3

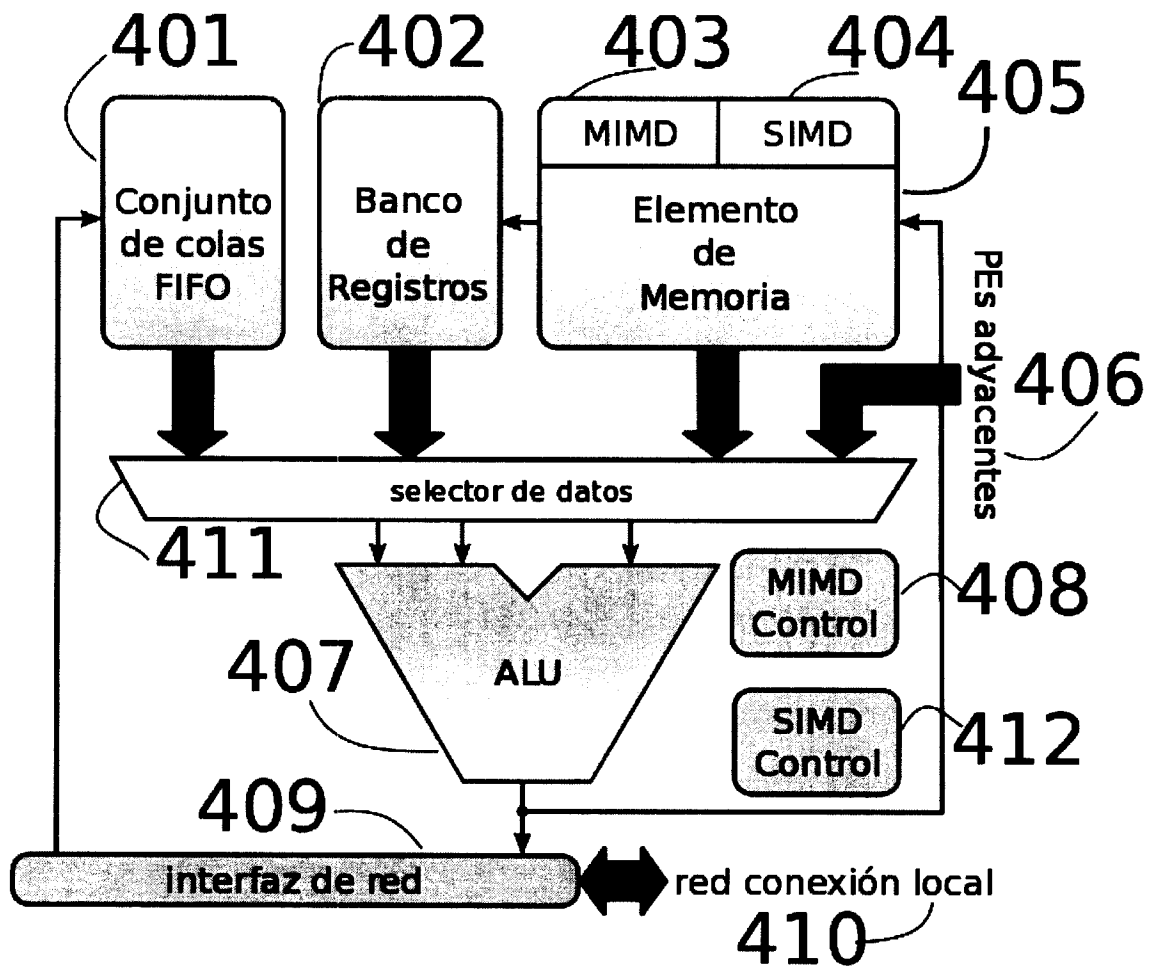


Fig. 4

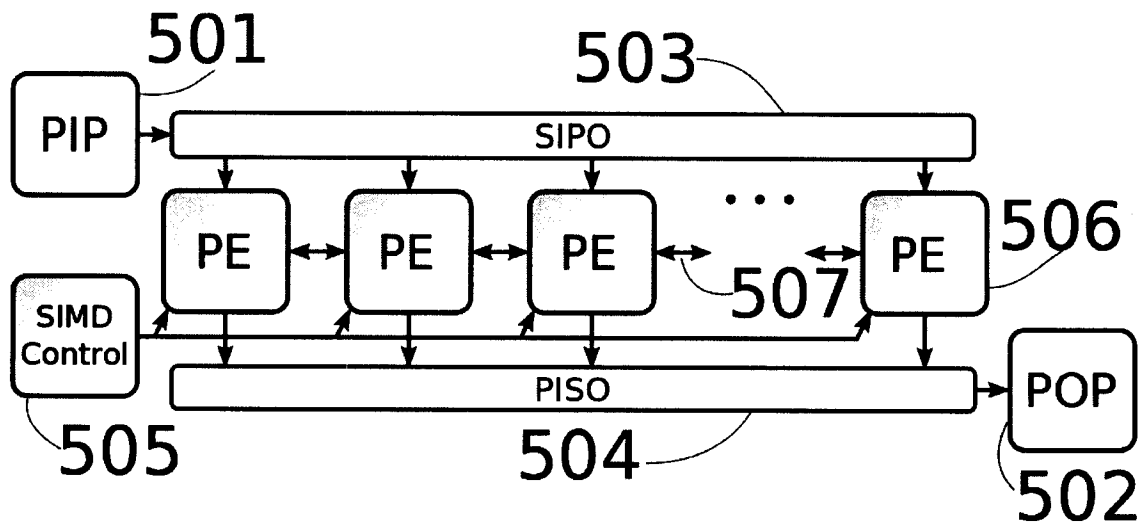


Fig. 5

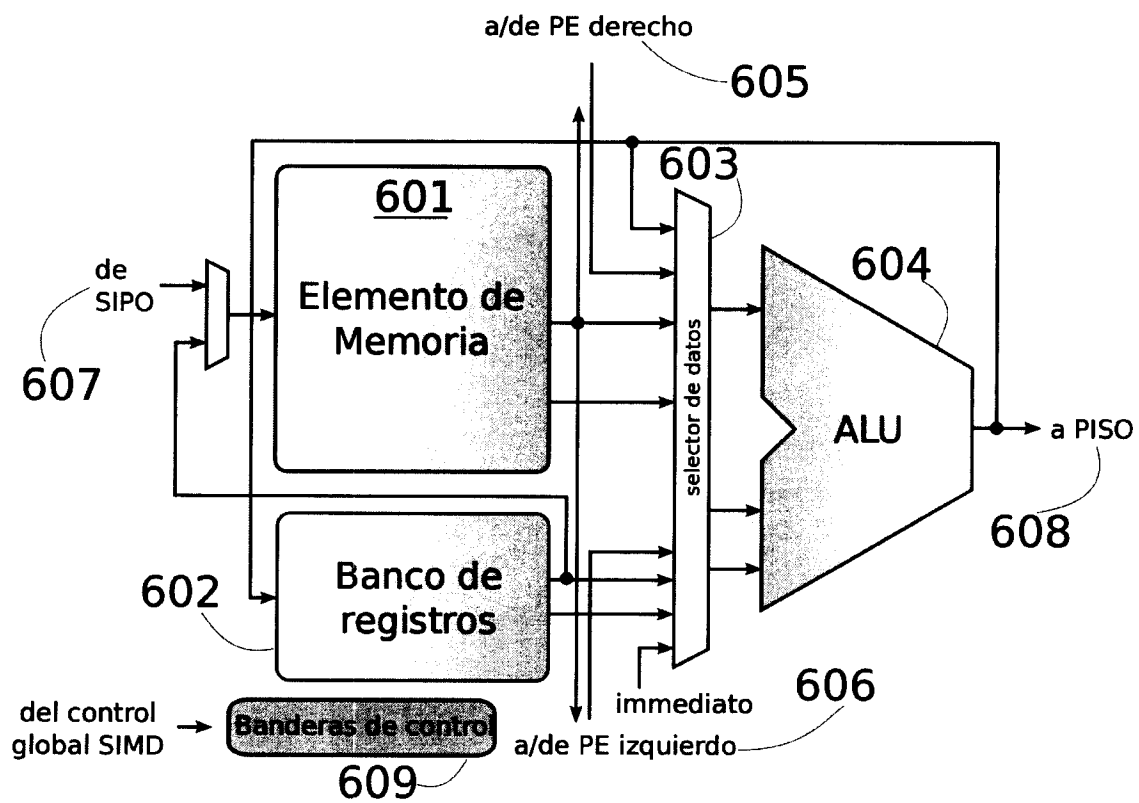


Fig. 6

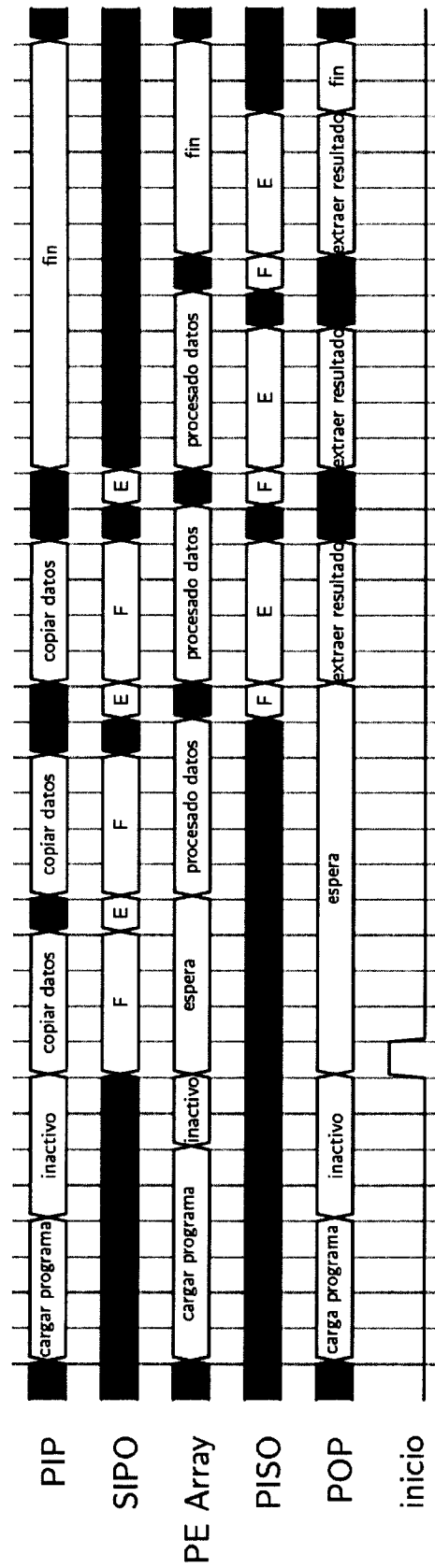


Fig. 7

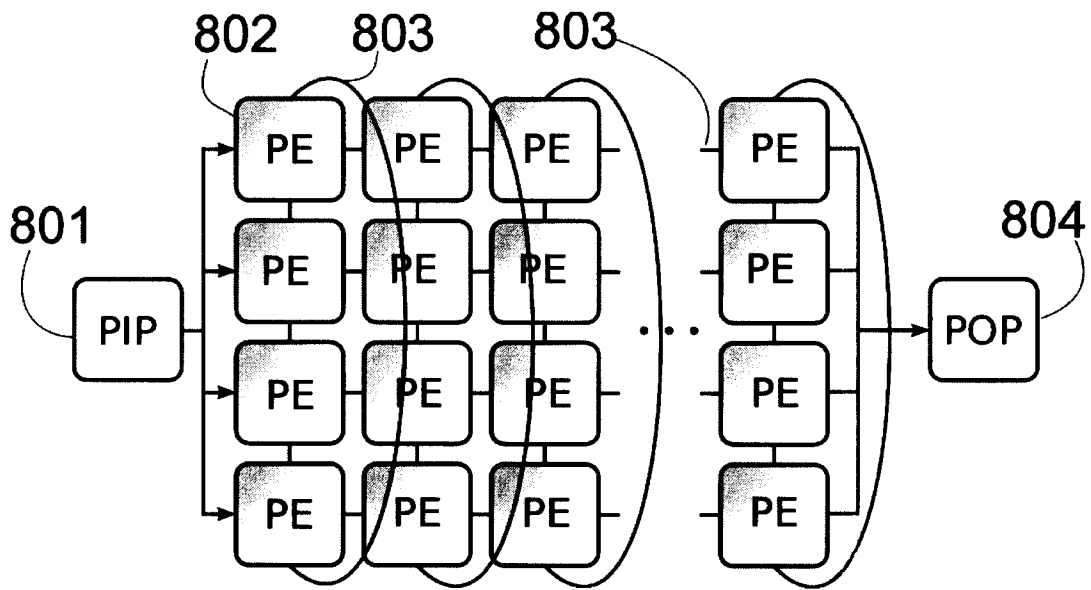


FIG. 8

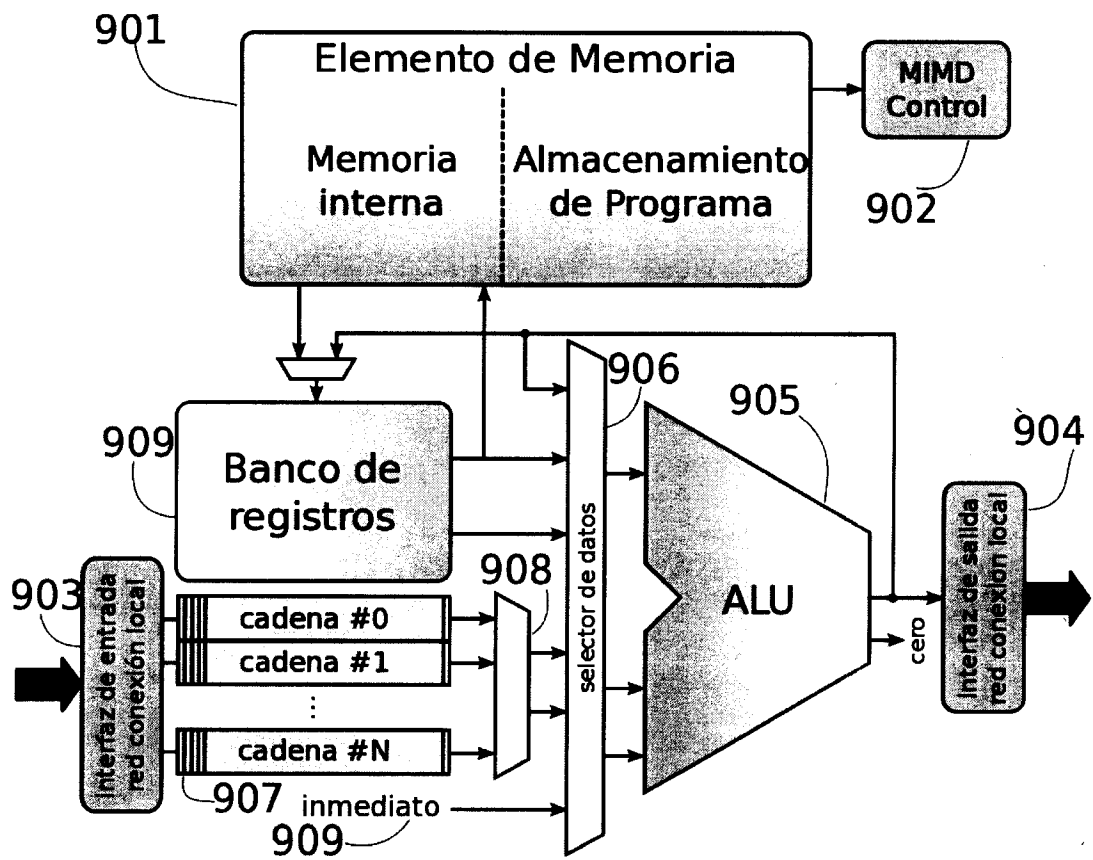


Fig. 9

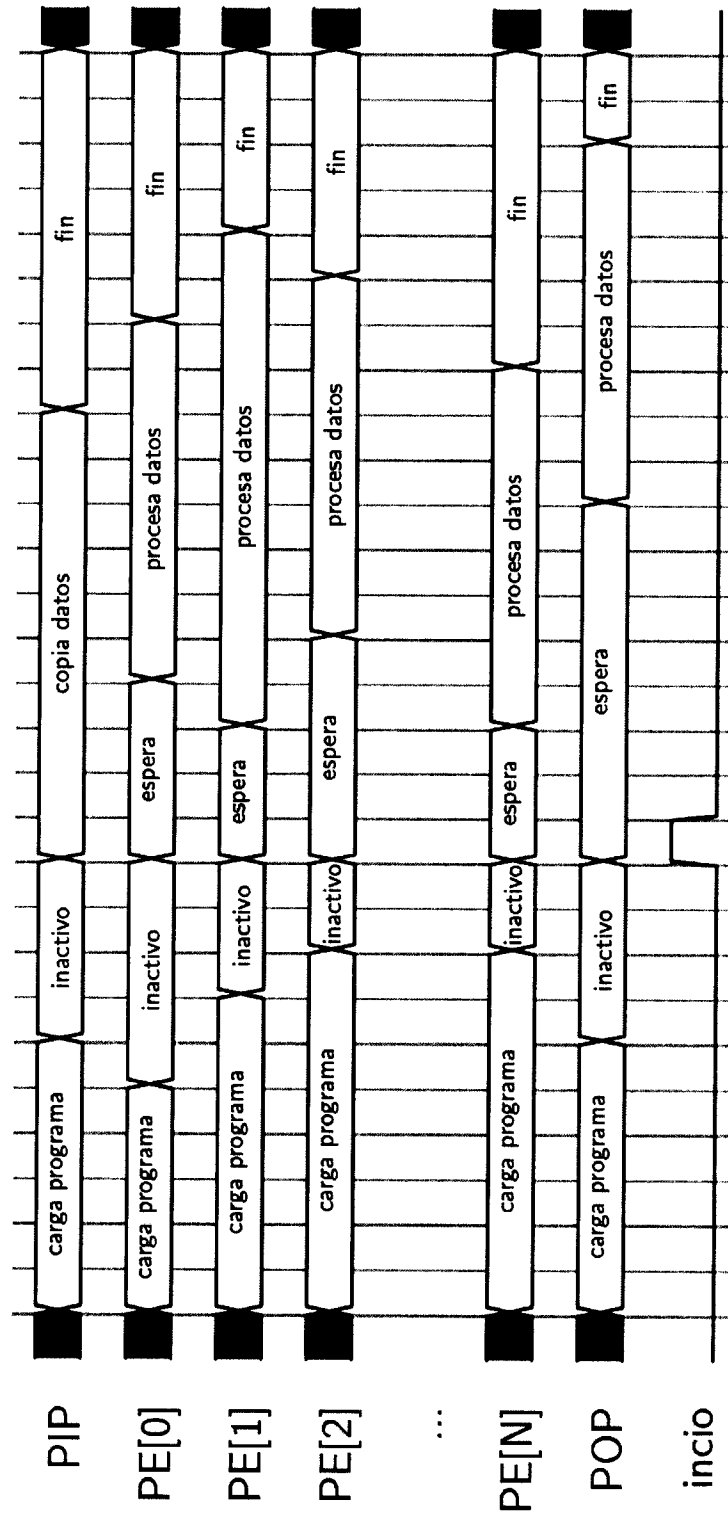


Fig. 10



OFICINA ESPAÑOLA
DE PATENTES Y MARCAS

ESPAÑA

②① N.º solicitud: 201101381

②② Fecha de presentación de la solicitud: 30.12.2011

③② Fecha de prioridad:

INFORME SOBRE EL ESTADO DE LA TÉCNICA

⑤① Int. Cl.: **G06F15/80** (2006.01)

DOCUMENTOS RELEVANTES

Categoría	⑤⑥ Documentos citados	Reivindicaciones afectadas
A	US 2010088489 A1 (LIESKE ET AL.) 08.04.2010, todo el documento.	1
A	US 6948050 B1 (GOVE ET AL.) 20.09.2005, todo el documento.	1
A	NOMOTO, S.; KYO, S.; OKAZAKI, S.; , "A dynamic SIMD/MIMD mode switching processor for embedded real-time image recognition systems," Solid State Circuits Conference (A-SSCC), 2011 IEEE Asian , vol., no., pp.17-20, 14-16 Nov. 2011. doi: 10.1109/ASSCC.2011.6123653 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6123653&isnumber=6123555 . Todo el documento.	1

Categoría de los documentos citados

X: de particular relevancia

Y: de particular relevancia combinado con otro/s de la misma categoría

A: refleja el estado de la técnica

O: referido a divulgación no escrita

P: publicado entre la fecha de prioridad y la de presentación de la solicitud

E: documento anterior, pero publicado después de la fecha de presentación de la solicitud

El presente informe ha sido realizado

☒ para todas las reivindicaciones

☐ para las reivindicaciones nº:

Fecha de realización del informe
25.07.2012

Examinador
M. L. Alvarez Moreno

Página
1/4

Documentación mínima buscada (sistema de clasificación seguido de los símbolos de clasificación)

G06F

Bases de datos electrónicas consultadas durante la búsqueda (nombre de la base de datos y, si es posible, términos de búsqueda utilizados)

INVENES, EPODOC, WPI, INSPEC

Fecha de Realización de la Opinión Escrita: 25.07.2012

Declaración

Novedad (Art. 6.1 LP 11/1986)	Reivindicaciones 1-14	SI
	Reivindicaciones	NO
Actividad inventiva (Art. 8.1 LP11/1986)	Reivindicaciones 1-14	SI
	Reivindicaciones	NO

Se considera que la solicitud cumple con el requisito de aplicación industrial. Este requisito fue evaluado durante la fase de examen formal y técnico de la solicitud (Artículo 31.2 Ley 11/1986).

Base de la Opinión.-

La presente opinión se ha realizado sobre la base de la solicitud de patente tal y como se publica.

1. Documentos considerados.-

A continuación se relacionan los documentos pertenecientes al estado de la técnica tomados en consideración para la realización de esta opinión.

Documento	Número Publicación o Identificación	Fecha Publicación
D01	US 2010088489 A1 (LIESKE et al.)	08.04.2010
D02	US 6948050 B1 (GOVE et al.)	20.09.2005
D03	NOMOTO, S.; KYO, S.; OKAZAKI, S.; , "A dynamic SIMD/MIMD mode switching processor for embedded real-time image recognition systems," Solid State Circuits Conference (A-SSCC), 2011 IEEE Asian , vol., no., pp.17-20, 14-16 Nov. 2011. doi: 10.1109/ASSCC.2011.6123653 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6123653&isnumber=6123555 . Todo el documento.	

2. Declaración motivada según los artículos 29.6 y 29.7 del Reglamento de ejecución de la Ley 11/1986, de 20 de marzo, de Patentes sobre la novedad y la actividad inventiva; citas y explicaciones en apoyo de esta declaración

El documento D01 muestra una arquitectura de procesadores configurable para trabajar en modo SIMD o MIMD. La configuración de los elementos PE [párrafos 0021, 0065-0066; figura 1] consiste en una estructura en anillo que los conecta secuencialmente. Cada PE puede operar en modo auto-controlado o en modo de control común, existiendo una unidad de control global (GCU) que permite el control general en este último modo. La arquitectura propuesta en D01 difiere completamente de la propuesta en la solicitud, siendo alguna una de sus diferencias más relevantes la inexistencia de una red local común a todos los PE que facilite la operación en modo MIMD.

El documento D02 también se encuentra orientado al procesamiento de imágenes, y propone una arquitectura en la que los elementos de procesamiento pueden operar de forma reconfigurable en modo SIMD o MIMD. La figura 10 muestra un esquema de la topología propuesta donde todos los elementos de procesamiento se encuentran conectados al conjunto de memorias a través de un elemento conmutador (cross-bar switch) común. También difiere totalmente de la arquitectura propuesta en la reivindicación 1.

El documento D03 muestra otra arquitectura dinámicamente reconfigurable orientada al procesamiento de imágenes. Se elige una estructura en anillo que conecta los diversos PE y se reutiliza el mismo hardware mediante agrupaciones de PEs para operar en ambos modos (SIMD y MIMD). Los apartados A y B muestran las diversas configuraciones, pero al igual que en el caso anterior difiere totalmente de la propuesta realizada en la reivindicación 1.

Los documentos D01 a D03 se citan únicamente a modo de ejemplo de otro tipo de arquitecturas de procesadores que puedan trabajar en modo SIMD y MIMD. No se han encontrado documentos que muestren arquitecturas similares a la definida en la reivindicación independiente 1. A la vista de los documentos anteriores las reivindicaciones 1 a 14 tienen actividad inventiva según el artículo 8 de la Ley de Patentes.