

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
24 November 2005 (24.11.2005)

PCT

(10) International Publication Number
WO 2005/111794 A1

(51) International Patent Classification⁷: **G06F 9/38**

Cupertino, CA 95014 (US). **ALSUP, Mitchell** [US/US];
2103 Kemper Cove, Austin, TX 78746 (US).

(21) International Application Number:

PCT/US2004/042683

(74) **Agent: DRAKE, Paul, S.**; Advanced Micro Devices, Inc.,
5204 East Ben White Boulevard, Mail Stop 562, Austin,
TX 78741 (US).

(22) International Filing Date:

17 December 2004 (17.12.2004)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

10/839,474

5 May 2004 (05.05.2004) US

(71) **Applicant (for all designated States except US): AD-
VANCED MICRO DEVICES, INC.** [US/US]; One
AMD Place, Mail Stop 68, P.O. Box 3453, Sunnyvale, CA
94088-3453 (US).

(81) **Designated States (unless otherwise indicated, for every
kind of national protection available):** AE, AG, AL, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ,
TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA,
ZM, ZW.

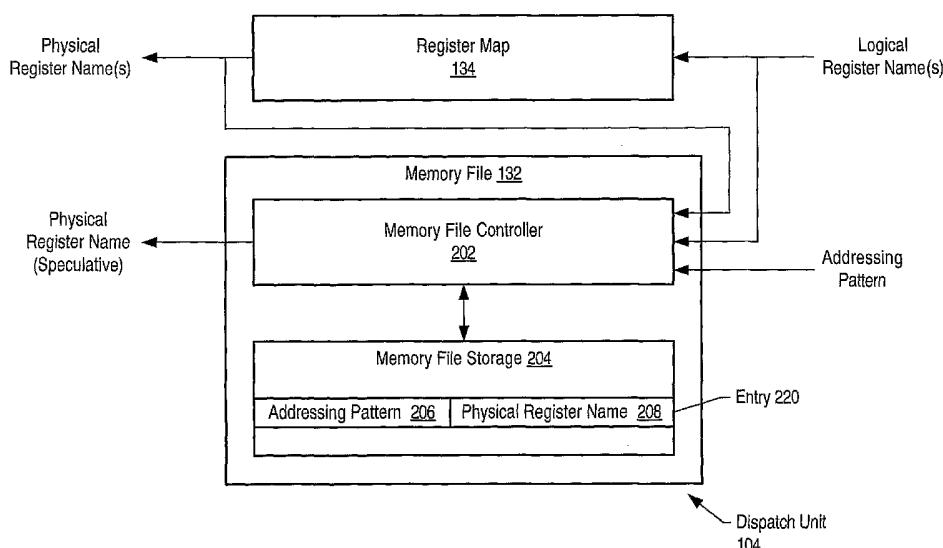
(72) **Inventors; and**

(75) **Inventors/Applicants (for US only): SANDER, Ben-
jamin, T.** [US/US]; 5701 Medicine Creek, Austin, TX
78735 (US). **RAMANI, Krishnan, V.** [IN/US]; 10215
Parkwood Drive, Apt. #2, Cupertino, CA 95014 (US).
HADDAD, Ramsey, W. [US/US]; 20728 Celeste Circle,

(84) **Designated States (unless otherwise indicated, for every
kind of regional protection available):** ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO,

[Continued on next page]

(54) **Title:** SYSTEM AND METHOD FOR VALIDATING A MEMORY FILE THAT LINKS SPECULATIVE RESULTS OF
LOAD OPERATIONS TO REGISTER VALUES



(57) **Abstract:** A system and method for linking speculative results of load operations to register values. A system (100) includes a memory file (132) including an entry (220) configured to store a first addressing pattern (206) and a first tag (208). The memory file (132) is configured to compare the first addressing pattern (206) to a second addressing pattern (206) of a load operation, and to link a data value identified by the first tag (208) to a speculative result of the load operation if there is a match. The system (100) further includes an execution core (124) coupled to the memory file (132) and configured to access the speculative result when executing a second operation that is dependent on the load operation, and a functional unit (126) coupled to the memory file (132) and configured to verify the link between the data value and the speculative result of the load operation by performing a comparison between one or more addresses.



SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

— *with international search report*

**SYSTEM AND METHOD FOR VALIDATING A MEMORY FILE THAT LINKS
SPECULATIVE RESULTS OF LOAD OPERATIONS TO REGISTER VALUES**

Technical Field

5 [0001] This invention is related to the field of processors, and more particularly, to performing data-speculative execution in a processor.

Background Art

10 [0002] Superscalar processors achieve high performance by executing multiple instructions concurrently and by using the shortest possible clock cycle consistent with their design. However, data and control flow dependencies between instructions may limit how many instructions may be issued at any given time. As a result, some processors support speculative execution in order to achieve additional performance gains.

15 [0003] One type of speculation is control flow speculation. Control flow speculation predicts the direction in which program control will proceed. For example, branch prediction may be used to predict whether a branch will be taken. Many types of branch prediction are available, ranging from methods that simply make the same prediction each time to those that maintain sophisticated histories of the previous branches in the program in order to make a history-based prediction. Branch prediction may be facilitated through hardware optimizations, compiler optimizations, or both. Based on the prediction provided by the branch prediction mechanism, instructions may be speculatively fetched and executed. When the branch instruction is finally evaluated, the branch prediction can be
20 verified. If the prediction was incorrect, any instructions that were speculatively executed based on the incorrect predication may be quashed.

[0004] Another type of speculation that has been proposed is data speculation. For example, value prediction, which predicts the value of data items, may involve observing patterns in data and basing the prediction on those patterns (e.g., an index counter variable's value may be predicted by observing how prior values of that variable are
25 incremented or decremented). Address prediction involves predicting the location of data. Yet another type of data speculation is called memory system optimism. In multi-processor systems, memory system optimism occurs when a processor speculatively executes an instruction using data from that processor's local cache before coherency checking is complete. Similarly, another type of data speculation may allow a load to speculatively execute before a store that has an uncomputed address at the time the load executes, even though the store may store data to the
30 same address that the load accesses. In all of these types of data speculation, the underlying conditions are eventually evaluated, allowing the speculation to be verified or undone. If the speculation ends up being incorrect, the instructions that executed using the speculative data may be re-executed (e.g., with updated and/or non-speculative data).

[0005] Since speculation allows execution to proceed without waiting for dependency checking to complete,
35 significant performance gains may be achieved if the performance gained from correct speculations exceeds the performance lost to incorrect speculations. Accordingly, it is desirable to be able to perform data speculation in a processor and to provide an efficient recovery mechanism for misspeculations.

DISCLOSURE OF INVENTION

40 [0006] Various embodiments of systems and methods for linking speculative results of load operations to register values are disclosed. In one embodiment, a system may include a memory file including an entry

configured to store a first addressing pattern and a first tag, where the memory file is configured to compare the first addressing pattern included in the entry to a second addressing pattern of a load operation, and where if the second addressing pattern matches the first addressing pattern stored in the entry, the memory file is configured to link a data value identified by the first tag to a speculative result of the load operation. The system may further include an execution core coupled to the memory file, where the execution core is configured to access the speculative result when executing a second operation that is dependent on the load operation, as well as a load store unit coupled to the memory file and configured to verify the link between the data value identified by the first tag and the speculative result of the load operation by performing a comparison between one or more addresses.

[0007] In another embodiment, a system may include a memory file including an entry configured to store a first addressing pattern and a first tag, where the memory file is configured to compare the first addressing pattern included in the entry to a second addressing pattern of a load operation, and where if the second addressing pattern matches the first addressing pattern stored in the entry, the memory file is configured to link a data value identified by the first tag to a speculative result of the load operation. The system may further include an execution core coupled to the memory file, where the execution core is configured to access the speculative result when executing a second operation that is dependent on the load operation, as well as a functional unit coupled to the memory file and configured to verify the link between the data value identified by the first tag and the speculative result of the load operation by performing a comparison between the speculative result and an actual result of the load operation.

BRIEF DESCRIPTION OF DRAWINGS

[0008] A better understanding of the present invention can be obtained when the following detailed description is considered in conjunction with the following drawings, in which:

[0009] FIG. 1 shows one embodiment of a processor.

[0010] FIG. 2A is a block diagram of one embodiment of a dispatch unit.

[0011] FIG. 2B shows an exemplary memory file storage entry that may be used in one embodiment.

[0012] FIG. 3 is a flowchart showing one embodiment of a method of linking the speculative result of a load operation to a data value identified by a particular tag.

[0013] FIG. 4A is a block diagram of a load store unit, according to one embodiment.

[0014] FIG. 4B is a block diagram of a load store unit, according to another embodiment.

[0015] FIG. 5A is a flowchart of a method of verifying a link between the speculative result of a load operation to a data value identified by a particular tag, according to one embodiment.

[0016] FIG. 5B is a flowchart of a method of retiring a memory operation based on whether a link involving that memory operation has been verified, according to one embodiment.

[0017] FIG. 6A shows a block diagram of a load store unit configured to verify a link between the speculative result of a load operation to a data value identified by a particular tag, according to one embodiment.

[0018] FIG. 6B shows a block diagram of an execution core configured to verify a link between the speculative result of a load operation to a data value identified by a particular tag, according to one embodiment.

[0019] FIG. 7 is a flowchart of a method of verifying a link between the speculative result of a load operation to a data value identified by a particular tag by using value comparisons, according to one embodiment.

[0020] FIG. 8 is a block diagram of one embodiment of a scheduler.

[0021] FIG. 9 is a flowchart of one embodiment of a method of issuing operations and reissuing operations in response to an indication that data speculation was incorrect.

[0022] FIG. 10 is a block diagram of a scheduler configured to selectively determine the readiness of an operation to issue based on either speculative operand sources or non-speculative operand sources, according to one embodiment.

[0023] FIG. 11 is a block diagram of another embodiment of a dispatch unit.

[0024] FIG. 12 is a block diagram of yet another embodiment of a dispatch unit.

[0025] FIG. 13 shows a speculative register map configured to maintain an indication of the color associated with a tag for a particular operand, according to one embodiment.

[0026] FIG. 14 shows one embodiment of a computer system.

[0027] FIG. 15 shows another embodiment of a computer system.

[0028] While the invention is susceptible to various modifications and alternative forms, specific embodiments thereof are shown by way of example in the drawings and will herein be described in detail. It should be understood, however, that the drawings and detailed description thereto are not intended to limit the invention to the particular form disclosed, but on the contrary, the intention is to cover all modifications, equivalents, and alternatives falling within the spirit and scope of the present invention as defined by the appended claims. Note, the headings are for organizational purposes only and are not meant to be used to limit or interpret the description or claims. Furthermore, note that the word "may" is used throughout this application in a permissive sense (i.e., having the potential to, being able to), not a mandatory sense (i.e., must). The term "include" and derivations thereof mean "including, but not limited to." The term "connected" means "directly or indirectly connected," and the term "coupled" means "directly or indirectly coupled."

MODE(S) FOR CARRYING OUT THE INVENTION

[0029] FIG. 1 is a block diagram of one embodiment of a processor 100. Processor 100 is configured to execute instructions stored in a system memory 200. Many of these instructions operate on data stored in system memory 200. Note that system memory 200 may be physically distributed throughout a computer system and may be accessed by one or more processors 100.

[0030] Processor 100 may include an instruction cache 106 and a data cache 128. Processor 100 may include a prefetch unit 108 coupled to the instruction cache 106. A dispatch unit 104 may be configured to receive instructions from instruction cache 106 and to dispatch operations to scheduler(s) 118. One or more schedulers 118 may be coupled to receive dispatched operations from dispatch unit 104 and to issue operations to one or more execution cores 124. Execution core(s) 124 may include a load/store unit 126 configured to perform accesses to data cache 128. Results generated by execution core(s) 124 may be output to a result bus 130. These results may be used as operand values for subsequently issued instructions and/or stored to register file 116. A retire queue 102 may be coupled to scheduler(s) 118 and dispatch unit 104. The retire queue may be configured to determine when each issued operation may be retired. In one embodiment, the processor 100 may be designed to be compatible with the x86 architecture. Note that processor 100 may also include many other components. For example, processor 100 may include a branch prediction unit (not shown).

[0031] Instruction cache 106 may temporarily store instructions prior to their receipt by dispatch unit 104. Instruction code may be provided to instruction cache 106 by prefetching code from the system memory 200 through prefetch unit 108. Instruction cache 106 may be implemented in various configurations (e.g., set-associative, fully-associative, or direct-mapped).

[0032] Prefetch unit 108 may prefetch instruction code from the system memory 200 for storage within instruction cache 106. In one embodiment, prefetch unit 108 may be configured to burst code from the system memory 200 into instruction cache 106. Prefetch unit 108 may employ a variety of specific code prefetching techniques and algorithms.

5 [0033] Dispatch unit 104 may output signals including bit-encoded operations executable by the execution core(s) 124 as well as operand address information, immediate data and/or displacement data. In some embodiments, dispatch unit 104 may include decoding circuitry (not shown) for decoding certain instructions into operations executable within execution core(s) 124. Simple instructions may correspond to a single operation. In some embodiments, more complex instructions may correspond to multiple operations. Upon decode of an operation that involves the update of a register, a register location within register file 116 may be reserved to store speculative register states (in an alternative embodiment, a reorder buffer may be used to store one or more speculative register states for each register). A register map 134 may translate logical register names of source and destination operands to physical register names in order to facilitate register renaming. Register map 134 may track which registers within register file 116 are currently allocated and unallocated.

10 [0034] The processor 100 of FIG. 1 supports out of order execution. A retire queue 102 may keep track of the original program sequence for register read and write operations, allow for speculative instruction execution and branch misprediction recovery, and facilitate precise exceptions. In some embodiments, retire queue 102 may also support register renaming by providing data value storage for speculative register states. In many embodiments, retire queue 102 may function similarly to a reorder buffer. However, unlike a typical reorder buffer, retire queue 102 may not provide any data value storage. In some embodiments, retire queue 102 may be implemented in a first-in-first-out configuration in which operations move to the "bottom" of the buffer as they are validated, thus making room for new entries at the "top" of the queue. As operations are retired, retire queue 102 may deallocate registers in register file 116 that are no longer needed to store speculative register states and provide signals to register map 134 indicating which registers are currently free. By maintaining speculative register states within register file 116 (or, in alternative embodiments, within a reorder buffer) until the operations that generated those states are validated, the results of speculatively-executed operations along a mispredicted path may be invalidated in the register file 116 if a branch prediction is incorrect.

15 [0035] Upon decode of a particular operation, if a required operand is a register location, register address information may be routed to register map 134 (or a reorder buffer). For example, in the x86 architecture, there are eight 32-bit logical registers (e.g., EAX, EBX, ECX, EDX, EBP, ESI, EDI and ESP). Physical register file 116 (or a reorder buffer) includes storage for results that change the contents of these logical registers, allowing out of order execution. A physical register in register file 116 may be allocated to store the result of each operation which, upon decode, is determined to modify the contents of one of the logical registers. Therefore, at various points during execution of a particular program, register file 116 (or, in alternative embodiments, a reorder buffer) may have one or more registers that contain the speculatively executed contents of a given logical register.

20 [0036] Register map 134 may assign a physical register to a particular logical register specified as a destination operand for an operation. Dispatch unit 104 may determine that register file 116 has one or more previously allocated physical registers assigned to a logical register specified as a source operand in a given operation. The register map 134 may provide a tag for the physical register most recently assigned to that logical register. This tag may be used to access the operand's data value in the register file 116 or to receive the data value via result forwarding on the result bus 130. If the operand corresponds to a memory location, the operand value

may be provided on the result bus (for result forwarding and/or storage in register file 118) through load/store unit 126. Operand data values may be provided to execution core(s) 124 when the operation is issued by one of the scheduler(s) 118. Note that in alternative embodiments, operand values may be provided to a corresponding scheduler 118 when an operation is dispatched (instead of being provided to a corresponding execution core 124 when the operation is issued).

[0037] The bit-encoded operations and immediate data provided at the outputs of dispatch unit 104 may be routed to one or more schedulers 118. Note that as used herein, a scheduler is a device that detects when operations are ready for execution and issues ready operations to one or more execution units. For example, a reservation station is a scheduler. Each scheduler 118 may be capable of holding operation information (e.g., bit encoded execution bits as well as operand values, operand tags, and/or immediate data) for several pending operations awaiting issue to an execution core 124. In some embodiments, each scheduler 118 may not provide operand value storage. Instead, each scheduler may monitor issued operations and results available in register file 116 in order to determine when operand values will be available to be read by execution core(s) 124 (from register file 116 or result bus 130). In some embodiments, each scheduler 118 may be associated with a dedicated execution core 124. In other embodiments, a single scheduler 118 may issue operations to more than one of the execution core(s) 124.

[0038] Schedulers 118 may be provided to temporarily store operation information to be executed by the execution core(s) 124. As stated previously, each scheduler 118 may store operation information for pending operations. Additionally, each scheduler may store operation information for operations that have already been executed but may still reissue. Operations are issued to execution core(s) 124 for execution in response to the values of any required operand(s) being made available in time for execution. Accordingly, the order in which operations are executed may not be the same as the order of the original program instruction sequence. Operations that involve data speculation may remain in scheduler(s) 118 until they become non-speculative so that they may be reissued if the data speculation is incorrect.

[0039] In one embodiment, each of the execution core(s) 124 may include components configured to perform integer arithmetic operations of addition and subtraction, as well as shifts, rotates, logical operations, and branch operations. A floating point unit may also be included to accommodate floating point operations. One or more of the execution core(s) 124 may be configured to perform address generation for load and store memory operations to be performed by load/store unit 126.

[0040] The execution core(s) 124 may also provide information regarding the execution of conditional branch instructions to a branch prediction unit so that if the branch was mispredicted, the branch prediction unit may flush instructions subsequent to the mispredicted branch that have entered the instruction processing pipeline and redirect prefetch unit 106. The redirected prefetch unit 106 may then begin fetching the correct set of instructions from instruction cache 106 or system memory 200. In such situations, the results of instructions in the *original program* sequence that occurred after the mispredicted branch instruction may be discarded, including those which were speculatively executed and temporarily stored in register file 116. Results produced by components within execution core(s) 124 may be output on the result bus 130 to the register file 116 if a register value is being updated. If the contents of a memory location are being changed, the results produced within execution core(s) 124 may be provided to the load/store unit 126.

[0041] Load/store unit 126 provides an interface between execution core(s) 124 and data cache 128. In one embodiment, load/store unit 126 may be configured with a load/store buffer with several storage locations for data

and address information for pending loads or stores. The load/store unit 126 may also perform dependency checking for load instructions against pending store instructions to ensure that data coherency is maintained.

[0042] Data cache 128 is a cache memory provided to temporarily store data being transferred between load/store unit 126 and the system memory 200. Like the instruction cache 106 described above, the data cache 128 may be implemented in a variety of specific memory configurations, including a set associative configuration. Additionally, data cache 106 and instruction cache 128 may be implemented in a unified cache in some embodiments.

Memory File

[0043] FIG. 2A shows one embodiment of a dispatch unit 104. In this embodiment, dispatch unit 104 includes a register map 134 and a memory file 132. As mentioned above, register map 134 may be configured to provide register renaming. Register map 134 may receive logical register names for each source and destination operand and output the physical register names of the physical registers most recently assigned to the logical registers. Memory file 132 includes a memory file controller 202 and memory file storage 204.

[0044] Memory file storage 204 includes one or more entries 220. Each entry 220 may include an addressing pattern 206 and a tag 208 associated with that entry's addressing pattern. Each tag may identify a data value by indicating where that data value will be stored (e.g., within a reorder buffer or within a register file 116) when it is generated. For example, as shown in the illustrated embodiment, each tag may identify the physical register allocated to store that data value, as indicated by register map 134. Each addressing pattern may include all or some of the information used to specify an address in an operation. For example, referring briefly to FIG. 2B, an exemplary addressing pattern 206 is shown. In this embodiment, the addressing pattern 206 includes a base logical register name 210, an index logical register name 212, and a displacement 214. Some addressing patterns 206 may include a segment logical register name of a register identifying a particular segment in memory. Memory file storage 204 may be implemented from several registers, latches, flip-flops, or other clocked storage in some embodiments. In alternative embodiments, memory file storage 204 may include one or more RAM (Random Access Memory) cells.

[0045] Note that in some embodiments, the addressing pattern 206 stored in memory file storage 204 may include less than all of the addressing information specified for an operation. For example, entries in memory file storage 204 may store fewer than all of the bits used to specify a displacement within displacement field 214. In other embodiments, the memory file controller 202 may selectively choose whether to allocate an entry in memory file storage 204 dependent on whether an entry in the memory file can store all of the addressing information for a particular operation. For example, if each entry in memory file storage 204 stores up to 16 bits of displacement information and the addressing information for a particular operation includes 24 bits of displacement information, the memory file controller 202 may not allocate a memory file entry for that operation's addressing information.

[0046] Memory file controller 202 may compare the addressing patterns specified in undischarged operations to those stored within entries in memory file storage 204. If an operation's addressing pattern does not match any of the addressing patterns currently stored within entries in memory file storage 204 (i.e., the operation's addressing pattern misses in the memory file storage), memory file controller 202 may allocate a new entry in memory file storage 204 to store all or some of that operation's addressing pattern. If there are no free entries to allocate within memory file storage 204, memory file controller 202 may select an entry to overwrite using a replacement scheme such as LRU (Least Recently Used), FIFO (First In, First Out), random replacement, etc.

[0047] In addition to storing the operation's addressing pattern within the allocated entry, the memory file controller 202 may also store a physical register name 208, which may include a tag (e.g., the name of the physical register) identifying a value being loaded from or stored into the memory location identified by that operation's addressing pattern. For example, if a load operation that loads data from memory is being handled, the entry allocated in response to that load operation may store the name of the physical register allocated to store the result of the load operation. If a store operation that writes data to a memory location is being handled, memory file controller 202 may store the physical register name of the register in which the store's source value is stored in memory file storage 204.

[0048] If an operation's addressing pattern (or a portion of that pattern) is already stored in an entry in memory file 204 (i.e., the operation's addressing pattern hits in the memory file storage), the memory file controller 202 may use or modify the entry containing the matching addressing pattern. If a load operation (an operation that loads a value from a particular address into a register) is being handled, the memory file controller 202 may output the physical register name 208 stored in the matching entry. If a store operation (an operation that stores a value from a register to a particular address) is being handled, the memory file controller 202 may overwrite the tag (e.g., physical register name 208) stored in the matching entry with the tag of the data being stored.

[0049] If a load operation is being handled and the load operation hits in the memory file storage 204, the tag output by the memory file controller 202 may be used to link the stored value identified by the tag to a speculative result of the load operation. For example, in some embodiments, when the load operation is dispatched to scheduler(s) 118, the tag output by memory file 132 may also be provided to the scheduler(s) (e.g., as a speculative source operand tag). A scheduler 118 may issue the load operation in response to the availability (e.g., in register file 116 or on result bus 130) of the value identified by that tag. An execution core 124 may execute the load operation so that the linked value is broadcast as the speculative result of the load operation on result bus 130 (note that the data value produced as the load's result may not itself be flagged or otherwise identified as a speculative value in some embodiments). In other embodiments, the data value may be linked to the speculative result of the load operation by storing the tag in a speculative map, as described below.

[0050] As a result of the link, the data value identified by the tag may be forwarded as the speculative result of the load once the data value is available (e.g., in register file 116 or on result bus 130) in order to allow dependent operations to execute using the speculative result. In many cases, this may allow dependent operations to execute using the speculative result of the load operation sooner than if their execution is delayed until the non-speculative result of the load operation becomes available. In some embodiments, the data value may be forwarded by executing the load operation so that the data value is output onto the result bus 130 as the speculative result of the load operation. For example, in one embodiment, instead of taking three cycles to perform the load non-speculatively (assuming the load hit in the data cache 128), the load may be executed in a single cycle by outputting the data value and a tag identifying the data value as the load result. In other embodiments, the data value may be forwarded in a more indirect manner by providing the tag output by memory file controller 202 directly to dependent operations (operations having an operand produced by the load operation) as a speculative operand source at dispatch. Means for forwarding the data value may include one or more of: a dispatch unit configured to modify the load operation to execute as a speculative register-to-register move operation or to provide the tag to a dependent operation as a speculative operand source tag, a scheduler configured to issue the modified load and/or the dependent operation dependent on the availability of the linked data value, and an execution core configured to

output the linked data value as the result of the load or to execute the dependent operation using the linked data value.

[0051] As shown, memory file 132 tracks the addressing patterns in operations that access data in data cache 128 (or system memory 200). As a result, register values stored in physical registers may be linked to values stored in particular addresses in memory. Note that memory file 132 allows dependencies between operations that are relatively removed from each other in the operation stream to be used to link register values to speculative load results. Thus, a memory file may provide a dependency history between operations that may be separated by several intervening operations.

[0052] FIG. 3 shows a flowchart of one embodiment of a method of linking the speculative result of a load operation to a register value. At 301, a register tag is associated with an addressing pattern, indicating that both the register and the address are likely to store the same data value. Both the tag and the addressing pattern are specified for a first load or store operation. For example, if the first operation is a load operation, the tag may identify the physical register allocated to store the result of the load and the addressing pattern may be the used to calculate the address for the load. If instead the first operation is a store operation, the addressing pattern may indicate the address of the store's destination and the tag may identify the data being stored by the store operation. In some embodiments, the tag and the addressing pattern may be associated by storing both in an entry in a memory file.

[0053] If a load operation's addressing pattern matches the addressing pattern of the first operation, the load's speculative result may be linked to the data value identified by the tag specified for the first operation, as indicated at 303-305. If the load operation's addressing pattern does not match that of the first operation, the load's addressing pattern and the tag for the load's destination may be associated (e.g., by storing both in an entry in a memory file), as indicated at 303-311. Furthermore, the load may be performed normally by accessing the data cache, as shown at 313. In one embodiment, the load operation's addressing pattern may be compared to more than one prior operation's addressing pattern. For example, the load's addressing pattern may be compared to each addressing pattern currently stored in a memory file.

[0054] If the load's speculative result is linked to a data value identified by the tag, that data value may be forwarded to one or more dependent operations as the speculative result of the load operation, as indicated at 307. The data value may be forwarded through result forwarding or by forwarding an indication that dependent operations may use the data value as a speculative operand source, as will be described in more detail below. In one embodiment, if the data value is forwarded via result forwarding, dependent operations may be executed using the forwarded data value one cycle after an operation that generates the data value completes execution. If the data value is forwarded via an indication that dependent operations may use the physical register as a speculative operand source, dependent operations may be issued as soon as the operation that generates the data value completes execution. The speculative result may be forwarded before the speculative result is verified, as indicated at 309. The speculative result may be forwarded without accessing the data cache (i.e., the speculative result may be forwarded sooner than the non-speculative result, which is generated by accessing the data cache).

[0055] If the speculative result is verified to be correct at 309, the load operation may be completed without performing a data cache access. However, if the speculative result is determined to be incorrect at 309, the data cache access may be performed in order to obtain the correct result (not shown). If so, any dependent operations that executed using the speculative result of the load may be reexecuted using the load's correct result. Note that in some situations, the verification (at 309) may be performed before the linked data value is forwarded as the

speculative result of the load (at 307). In such situations, the load may be performed normally or, if the link is determined to be correct, the data value may be forwarded as the non-speculative result of the load operation.

[0056] In order to further illustrate the operation of memory file 132, assume that a sequence of operations to be handled by dispatch unit 104 includes the following operations:

```

5      MOV EBX, [EDX + EAX – displacement A]      (LOAD 1)
      ...
      MOV ECX, [EDX + EAX – displacement A]      (LOAD 2)
      ...
      MOV [EDX + EAX – displacement A], EAX      (STORE 1).

```

10 Each of these operations may be separated by one or more intervening operations in program order. As shown, each of these three operations includes the same addressing pattern, EDX + EAX – displacement A.

[0057] When LOAD 1's addressing pattern is provided to memory file 132, memory file controller 202 may check memory file storage 204 for an addressing pattern that matches LOAD 1's addressing pattern. Assuming that the addressing pattern misses in the memory file storage 204, the memory file controller 202 may allocate an entry (either by using an unallocated entry or by overwriting an already allocated entry) to store all or some of the addressing pattern of the load operation and the physical register name of the load operation's destination physical register as provided by register map 134. Since the load operation misses in the memory file storage, the memory file controller 202 may not output a tag for that load operation.

[0058] When LOAD 2 is subsequently handled by memory file 132, its addressing pattern may match the addressing pattern in the entry allocated in response to LOAD 1 (assuming LOAD 1's entry has not been already been overwritten in response to an intervening operation). In response to LOAD 2's addressing pattern hitting in memory file storage 204, memory file controller 202 may output the physical register name of the physical register allocated to store the result of LOAD 1. This physical register name may be used to link the data value loaded by LOAD 1 to the speculative result of LOAD 2.

25 [0059] When STORE 1 is handled by memory file 132, its addressing pattern may hit in the entry allocated in response to LOAD 1 (again assuming that no intervening operation has caused this entry to be overwritten). Instead of outputting the physical register name of the physical register allocated to store the result of LOAD 1 (as was done for LOAD 2) however, memory file controller 202 may overwrite the physical register name 208 stored in that entry with the physical register name of the register containing the data being stored by STORE 1. Thus, when subsequent load operations hit in this entry, the memory file controller 202 will output the physical register name of STORE 1's source physical register instead of outputting the physical register name of LOAD 1's destination register.

[0060] Since memory file 132 is being used as a speculative structure, the accuracy of the information stored in memory file storage 204 may not be critical to the correct operation of processor 100 (e.g., mispredictions in memory file 132 may not cause errors in the output of processor 100). However, it may be desirable to improve the accuracy of memory file 132 in order to increase the benefits provided by correctly linking the speculative results of load operations to values stored in physical registers and/or to decrease any performance penalties incurred for mispredicted speculative links. In some embodiments, the accuracy of memory file 132 may be increased by invalidating entries in memory file storage 204 when updates to registers used to specify addresses are detected.

40 For example, each addressing pattern may include one or more logical register identifiers used to identify a base and an index for address calculation. If a subsequent operation modifies one of the logical registers specified as part of

an entry's addressing pattern 206, that entry may be invalidated. Thus, logical register names of the destination register of each operation may be input to memory file 132 in order to perform entry invalidations, as shown in FIG. 2A.

[0061] Additionally, in some embodiments, entries within memory file storage 204 may be invalidated in response to snooping another device gaining write access to data at a particular address. Similarly, an entry may be invalidated in response to detection of a misprediction. Generally, many conditions such as these that may affect the accuracy of the entries in the memory file storage 204 may be monitored and used to determine when to invalidate entries. However, since memory file 132 is a speculative structure, some embodiments may not implement some of these monitoring methods if the additional hardware cost to implement certain monitoring methods outweighs the potential improvement in memory file accuracy.

Memory File Validation

[0062] The load store unit 126 (or another means for verifying the link within processor 100) may verify the link of the value stored in the physical register identified by memory file 132 to the speculative result of the load operation. If the link is incorrect, the load store unit 126 may cause the load to be reissued and/or broadcast the correct result of the load operation on result bus 130. Reissuing the load may cause any dependent operations that executed using the speculative result of the load to reissue and execute using the updated, non-speculative value. Alternatively, if the link is incorrect, the processor 100 may flush the operations occurring after the misspeculation and restart execution of those operations using the correct result of the load operation.

[0063] In some embodiments, link verification may be performed in the load store unit 126 by comparing operation addresses. The load store unit 126 may compare the addresses of various memory operations to detect which operations should have been linked. In one embodiment, the load store unit 126 may verify the speculative link using store-to-load forwarding comparators. The store-to-load forwarding comparators may be used to forward the result of a younger load from an older store in addition to being used for link verification. For example, the load store unit may include comparators used to perform store-to-load forwarding. These comparators may compare the address of each pending load to the addresses of a set of older stores. The youngest of the set of older stores having an address that matches the address of a younger load may be selected. The data stored by the selected one of the set of older stores may then be forwarded as the result of the load.

[0064] FIG. 4A shows one embodiment of a load store unit 126 configured to use store-to-load forwarding (STLF) comparators 280 to perform link verification by comparing operation addresses. As shown, a load store unit 126 may include storage for outstanding loads and stores. In the illustrated embodiment, separate storage is provided for loads (in load storage 250) and stores (in store storage 260); however, in other embodiments, loads and stores may be stored in the same operation storage (e.g., an additional field may indicate whether each operation is a load, a store, or both). In some embodiments, this operation storage may be part of a scheduler 18.

[0065] The information identifying each load operation may be stored in an entry 252 in load storage 250. This information may include a tag 254 identifying the load operation, a forwarding tag 256 identifying a memory operation (another load or store) having a value to which the speculative result of the load is linked, and an address 258 indicating the address targeted by the load. Other fields (not shown) may also be included.

[0066] The forwarding tag 256 may be produced by the memory file 132 in response to a hit in the memory file 132 for a load operation (this load operation, which links its speculative result to a data value identified by a tag provided by the memory file, is referred to as the "linking load"). In addition to storing a tag 208 and addressing

pattern 206 in each memory file entry 220, the memory file 132 may store a tag identifying the memory operation (e.g., load or store) that caused that entry 220 to originally be allocated. This tag may be output as a forwarding tag 256, identifying the older memory operation whose operand is being linked to the speculative result of a younger memory operation.

5 [0067] The information identifying each store operation may be stored in an entry 262 in store storage 260. This information may include a tag 264 identifying the store operation and an address 268 indicating the address targeted by the store operation.

10 [0068] The STLTF comparators 280 may be configured to determine which store, if any, a load should forward data from. Additionally, the STLTF comparators 280 may verify links between speculative results of load operations and values. These links may be generated by a memory file 132 as described above. The STLTF comparators 280 may be configured to verify a link between the speculative result of a load operation and a store operand by searching for the store, if any, from which that load operation should have forwarded data by comparing the load's address 258 to the addresses 268 of one or more store operations. If no matching store (a store with the same address as the linking load) is found, but the forwarding tag 256 for the load indicates that the load's speculative result was linked to a data value, the STLTF comparators 280 may output an indication that the link is incorrect (assuming no other load from which that load could have forwarded data is identified, as described below). If a matching store is found, but the tag 264 of the matching store does not match the forwarding tag 256 of the load operation, the STLTF comparators 280 may similarly output an indication that the link is incorrect. If a matching store is found and the tag 264 of the matching store matches the forwarding load, the STLTF comparators 280 may output an indication verifying the link (or simply not output an indication that the link is incorrect). In some embodiments, at least part of the link verification process may take place at substantially the same time as a process by which store-to-load forwarding is performed for the load.

15 [0069] If the speculative result of a particular load has not been linked to an operand of another memory operation, the forwarding tag 256 for that load may be set to an invalid value or a value that otherwise indicates that no link verification needs to be performed for that load. For such a load, STLTF comparators 280 may perform store-to-load forwarding but not link verification.

20 [0070] In addition to verifying links between the speculative results of loads and store operands, the STLTF comparators 280 may also be configured to verify links between the speculative results of loads and other load operands. For example, the memory file 132 may link the speculative result of a younger load to the operand of an older load. Thus, the forwarding tag 256 for the younger load may identify the older load. The STLTF comparators 280 may verify such a load-load link by verifying that no stores that occur in program order between the linked loads have store addresses matching the younger load's address. Similarly to verification of links between a younger load and an older store, if the STLTF comparators 280 determine that the link is incorrect (e.g., because the address of an intervening store matches the address of the younger load), the STLTF comparators 280 may output an indication that the link is incorrect.

30 [0071] In one embodiment, the STLTF comparators may verify a link between a pair of load operations by looking for the youngest operation, out of the set of loads and stores that are older than the linking load, that has an address matching the address of the load operation whose speculative result is linked to an older operation's operand. If the youngest matching operation is the operation identified by the forwarding tag 256 stored with the linking load operation, the STLTF comparators may determine that the link is correct. If no matching operation is

40

found, or if the matching operation is not the operation identified by the forwarding tag 256, the STLF comparators 280 may output an indication that the link is incorrect.

[0072] In another embodiment, the STLF comparators 280 may look for the youngest matching load (of the set of loads older than the linking load and having addresses matching the address of the linking load) and the youngest matching store (of the set of stores older than the linking load and having addresses matching the address of the linking load). If the youngest matching load is younger (in program order) than the youngest matching store, and if the youngest matching load is the operation identified by the linking load's forwarding tag 256, the STLF comparators 280 may determine that the link is correct. If the youngest matching load is older than the youngest matching store, the STLF comparators 280 may determine that the link is incorrect (e.g., because the linking load should have been linked to the youngest matching store).

[0073] When links between a pair of loads are being considered, the STLF comparators 280 may compare multiple matching loads because the linking load may be properly linked to any matching load that is younger, in program order, than the youngest matching store that is older than the linking load (as opposed to only being properly linked to the youngest of the matching older loads). Thus, in addition to just identifying the youngest matching load, the STLF comparators may identify a set of matching loads and determine whether any one of that set is the linked-to load identified in the linking load's forwarding tag 256.

[0074] In some situations, the linked-to memory operation may have retired or been removed from the processor 100 before the link is verified. Thus, due to the removal of the older operation, the STLF comparators 280 may fail to identify a matching memory operation for a load whose speculative result is linked to an operand of the removed operation, even if the link is otherwise correct. In situations where the link cannot be verified, the STLF comparators 280 may output an indication that the link is incorrect.

[0075] In response to an indication that a link is incorrect, all or part of the memory file 132 may be invalidated. For example, in one embodiment, the particular memory file entry that was used to provide the speculative result of the load may be invalidated. In other embodiments, the entire memory file 132 may be invalidated in response to detection of an incorrect link by STLF comparators 280. Additionally, dependent operations that depend on the result of the load may execute (or re-execute, if those operations have executed already) using the non-speculative result of the load instead of using the speculative result obtained from the memory file 132.

[0076] Note that while the illustrated embodiment performs link verification using STLF comparators, other embodiments may employ dedicated comparators, not otherwise used for store-to-load forwarding, to perform link verification by comparing memory operation addresses and/or addressing patterns.

[0077] In some embodiments, each time the memory file 132 outputs a tag identifying a data value to which the speculative result of a load should be linked, the memory file 132 may also cause a flag associated with the memory operation to which that load is being linked to be set. For example, if a load's speculative result is linked to a store operand of an older store operation in response to the load's addressing pattern hitting in a memory file entry, the memory file 132 may output the tag of the data value to link to the speculative result and output an indication that the older store has been linked to a younger load. In response to the indication, the load store unit 126 may set a retire flag 269 associated with the older store, as shown in FIG. 4B. The load store unit 126 may continue to keep the retire flag 269 set until all younger load operations whose speculative results have been linked to the store's operand have been verified. It is noted that in some embodiments, a retire flag associated with an

older load operation (not shown) may also be set when a younger load's speculative result is linked to the older load's destination operand.

[0078] The load store unit 126 and/or the retire queue 102 may use the retire flag 269 associated with each store operation when determining when to deallocate an entry 262 allocated to that store operation and when to retire that store operation. If a store operation's retire flag is set, the load store unit 126 may continue to keep an entry 262 allocated to that store operation so that the link between the store's operand and the speculative result of a younger store can be verified. Likewise, the retire queue 102 may not retire the operation while its retire flag 269 is set. In embodiments including a retire flag associated with a load operation, the load store unit 126 and/or the retire queue 102 may use the retire flag associated with each load operation when determining when to deallocate an entry 252 allocated to that load operation and when to retire that load operation, in a manner similar to that described for store operations.

[0079] FIG. 5A is a flowchart of a method of verifying a link between the speculative result of a load operation to a data value identified by a particular tag, according to one embodiment. If a load's speculative result is linked an older (in program order) memory operation's operand, the tag of the linked-to older operation may be saved with the linking (younger) load, as shown at 321 and 323. For example, a memory file such as the one illustrated in FIG. 2A may be used to link the linking load's speculative result to the data value stored by an older store or to the data value loaded by an older load. The tag of the older load or store may be stored with the linking load (e.g., in a scheduler and/or load store unit) to identify which operation that load is linked to.

[0080] At some point, the load store unit may verify the link between the linking load and the linked-to older operation, as shown at 325. In some embodiments, the load store unit may use STLF comparators to perform this verification. In addition to performing link verification, the STLF comparators may also be used to perform store-to-load forwarding. Alternatively, the load store unit may have dedicated logic that only performs link verification.

[0081] If the load store unit determines that the link is correct, as determined at 325, the load store unit may determine that the speculative result of the load operation is non-speculative (at least with respect to the speculation performed by the memory file), as indicated at 327. In some embodiments, the load store unit may signal other components (e.g., the scheduler) in the processor that the link has been verified.

[0082] If the load store unit determines that the link is incorrect, as indicated at 325 and 329, the load store unit may output an indication that the link is incorrect (e.g., to the scheduler and/or dispatch unit). In response to this indication, a new result of the linking load may be generated. For example, if STLF comparators are used to verify the link, during the verification process the STLF comparators may have identified a store from which the linking load should have forwarded data. The data stored by the identified store may be forwarded as the non-speculative result of the load operation. Additionally, the memory file (or at least one entry in the memory file) may be invalidated in response to the detection of the incorrect link).

[0083] FIG. 5B is a flowchart of a method of retiring a memory operation based on whether a link involving that memory operation has been verified, according to one embodiment. If a younger load's speculative result is linked to data operated on (e.g., loaded or stored) by an older memory operation, a flag associated with the older operation (the linked-to operation) may be set, as shown at 331-333. In some embodiments, this flag may be included in a scheduler or load store unit.

[0084] If the operation's flag is set, retirement of the operation may be delayed until the link between the speculative result of the younger load and the data operated on by the older operation has been verified, as indicated at 335-337. Otherwise, the older operation may retire normally, as indicated at 339.

[0085] While the above examples perform link verification using address comparisons (i.e., by comparing addresses accessed by memory operations), other embodiments may perform link verification using value comparisons (i.e., by comparing data values resulting from memory operation accesses). For example, the link may be verified for a load operation that does not use a particular register value for address calculation by using that register value for speculative link verification. The register that is not used for address calculation may be used to store the value that is speculatively linked to the result of the load operation. When the actual result of the load operation is obtained, the load store unit (or the execution core) may compare the value in the otherwise unused register to the actual result value. If the two are equal, the link is correct. Otherwise, the link is incorrect and the correct, non-speculative result of the load operation may be broadcast in order to propagate the correction throughout the processor. Additionally, all or part of the memory file may be invalidated in response to the detection of an incorrect link.

[0086] The register used to store the speculatively linked result value may, in one embodiment, be the index register. The memory file controller may be configured to only allocate entries into the memory file storage for memory operations that do not use the index register for address calculation.

[0087] In one embodiment, value-based link verification may be performed in the load store unit 126, as shown in FIG. 6A. Here, the load store unit 126 includes dedicated value comparators 282 configured to compare the actual value 255 loaded by the load operation (either through store-to-load forwarding or by accessing the data cache and/or memory) with the speculative result 257 of the load operation. If the two values are equal, the load store unit 126 may determine that the link is correct. Otherwise, the load store unit 126 may broadcast the correct result of the load operation, loaded value 255, so that any dependent operations can execute (or re-execute) using the correct value. The load store unit 126 may also provide a signal to the dispatch unit 104 indicating that one or more entries in the memory file 132 should be invalidated in response to detection of an incorrect link.

[0088] Alternatively, value-based link verification may be performed in the execution core(s) 124, as shown in FIG. 6B. In these embodiments, a load whose speculative result is linked to a data value identified by the memory file 132 may be converted into a combination load operation and comparison operation by the dispatch unit 104. These operations may be respectively issued to the load store unit 126 and a functional unit 126. The load store unit 126 may perform the load normally. When the non-speculative result of the load is available, the comparison operation may execute by accessing the speculative result via the unused address calculation register (e.g., the index register) and comparing the speculative result to the non-speculative result generated by the load store unit 126. If the two are equal, the link is correct; otherwise, the link is incorrect.

[0089] FIG. 7 is a flowchart of a method of verifying a link between the speculative result of a load operation to a data value identified by a particular tag by using value comparisons, according to one embodiment. If a load's speculative result is linked to an operand of an older memory operation (e.g., by a memory file), an operand not needed for the load's address calculation may be used to store the speculative result of the load, as indicated at 341-343. For example, if the load does not use the index register for address calculation, the index register may be used to store the load's speculative result.

[0090] If the actual result of the load matches the speculative result of the load (e.g., as determined by performing a value comparison in a functional unit), the link is correct and the speculative result may become non-speculative, as indicated at 345 and 347. Otherwise, the link is incorrect and the actual result of the load may be broadcast on a result bus in order to propagate the correct result to any dependent operations, as indicated at 349.

Forwarding a Load's Speculative Result as a Source Operand

[0091] Various embodiments may link a load operation's speculative result to a register data value in many different ways. In some embodiments, values may be linked by identifying two sources for an operand: a speculative source and a non-speculative source. The speculative source may be the linked data value. Speculative sources may be provided for load operations to indicate the data value linked to the speculative result of the load. In some embodiments, speculative sources may also be provided for operations dependent on such a load operation. Accordingly, some operands may have two tags: one identifying the speculative source and one identifying the non-speculative source. In such embodiments, each scheduler 118 may provide tag storage for both the speculative operand and the non-speculative operand, as shown in FIG. 8.

[0092] FIG. 8 shows one embodiment of a scheduler 118 that may be included in a processor. In the illustrated embodiment, the scheduler 118 includes a scheduler controller 502 and an operation storage 504. In response to dispatch unit 104 dispatching an operation, scheduler controller 502 may allocate an entry within operation storage 504 to store information corresponding to the operation. For example, an entry 522 in operation storage 504 may include an opcode field 510, one or more operand fields, and a result field 516. The result field 516 may store a tag identifying the physical register in which the result of that entry's operation should be stored. When the operation is issued, this tag may be forwarded to each scheduler 118 on one of one or more tag buses 520. Each scheduler may compare the tags forwarded on tag buses 520 to the operand tags (both speculative and non-speculative, as described below) for pending operations in order to determine when the pending operations' operands will be available. Accordingly, an operation may be issued (or marked as being ready to issue) in response to its source operand tags appearing on tag bus 520.

[0093] Each operand field(s) may include storage for a speculative tag identifying a speculative operand source and a non-speculative tag identifying a non-speculative operand source. In the illustrated entry 522, operand 1's two sources are identified by non-speculative tag 512 and speculative tag 514. The scheduler 118 may be configured to issue an operation in response to one or more indications that the operation's operands are available. An operand is available if it is available from either a speculative source or a non-speculative source. If an operand is available from both a speculative source and a non-speculative source, the operation may be executed using the value available from the non-speculative source. In some embodiments, the scheduler 118 may prioritize issuing operations whose non-speculative operand sources are available over operations for which only speculative operand sources are available.

[0094] Note that an operation may include some operands that have speculative sources and other operands that do not have speculative sources. Also note that the same source may be a speculative source for one operation and a non-speculative source for another. In some embodiments, when an operation is executed, only one data value may be read for each operand (e.g., the execution core 124 may read either the speculative operand source or the non-speculative operand source, but not both). This may make it unnecessary to add additional ports into the register file 116. Speculative sources and physical sources may be stored in the same storage locations (e.g., within register file 116) and speculative sources may not be flagged or otherwise identified as speculative sources in some embodiments.

[0095] In many embodiments, scheduler 118 may be configured to keep entries allocated to operations after those operations have been issued to execution core(s) 124. When an operation is issued by scheduler 118 in response to the availability of one or more speculative operands, the scheduler may keep an entry 522 allocated to

that operation so that the operation can be reissued if the speculative link is incorrect. In some embodiments, the load/store unit may be configured to verify speculative links that arise when a speculative load result is linked to a data value stored in a physical register. If the link is correct, the load store unit may be configured to not broadcast a non-speculative result of the load, since the correct result is already available through the link. If so, the scheduler(s) 118 may be configured to reissue an operation if a tag identifying a non-speculative operand source for that operation is broadcast on result bus 130.

[0096] Alternatively, the load/store unit may broadcast the result of the store along with an extra status bit that masks the broadcast or indicates that the speculative link was correct and that the load should not be reissued. However, if the speculative link is incorrect, the load/store unit may perform a data cache and/or memory access in order to obtain the correct result for the load and broadcast the result of the load. In embodiments where results are always broadcast, the additional status bit may indicate that the speculative link was incorrect. Thus, in many embodiments, the same tag and result buses already available in a processor may be used to indicate that a speculative link is incorrect. In other embodiments, alternative indication mechanisms (e.g., using separate result buses 130 and/or result tag buses 520 to indicate mispredictions) may be implemented.

[0097] FIG. 9 shows a flowchart of one embodiment of a method of issuing and reissuing operations that have both speculative and non-speculative operand sources. If an operation's speculative operand source is available, the operation may be issued, as shown at 801-803. An operation's speculative operand source may become available when a data value is present in a particular register within the register file or when the data value is output on the result bus. Note that in some situations, the operation's non-speculative operand source for the same operand may be available before the speculative operand source. In those situations, the operation may be issued before the speculative operand source becomes available. Subsequent availability of the speculative source may not trigger reissue of the operation in some embodiments.

[0098] The issued operation may be executed using the data value provided by the speculative operand source, as indicated at 805, and the result of the operation may be broadcast, as indicated at 807. Broadcasting the operation's result allows dependent operations to execute.

[0099] If at some later time the speculative source is determined to be incorrect (e.g., the data value provided by the speculative source and the non-speculative source are not the same, or a speculative link used to generate the tag of the speculative source is not correct), the tag of the non-speculative source may be broadcast as an indication that the speculative source's value is incorrect. Broadcasting the tag of the non-speculative source involves broadcasting the tag in such a way that the scheduler(s) respond by reissuing the operation. For example, in some embodiments, a scheduler may respond if the tag is broadcast and a status flag associated with the tag is set to a certain value, while in other embodiments there may not be an associated status flag and a scheduler may be configured to reissue an operation any time the non-speculative tag is broadcast.

Determining Readiness for Issue based on either Speculative or Non-Speculative Sources

[0100] In some embodiments, such as the one illustrated in FIG. 10, the scheduler 118 may be configured to determine the readiness of a given operand as indicated by the availability of either the operand's speculative source or the operand's non-speculative source, but not both at the same time. This may in turn reduce potentially detrimental loading effects on the result tag bus 520.

[0101] In FIG. 10, the scheduler determines the readiness of an operation's operand by comparing either the tag 514 of the operand's speculative source or the tag 512 of the operand's non-speculative source, as selected by

multiplexer 534, to the tags currently on the result tag bus(es) 520 using comparator 532. Thus, the scheduler may determine whether that operation is ready to issue by watching for either the availability of the speculative source or the availability of the non-speculative source, but not both, on the result tag bus 520. Thus, instead of comparing each result tag broadcast on the result tag bus to both the operand's speculative source tag and the operand's non-speculative source tag, the scheduler may only compare the broadcast result tag to one or the other of the operand's source tags. If all of an operation's operands are available (either from speculative or non-speculative sources) are available, the ready flag 508 for that operation may be set, indicating that the operation is ready for issue.

[0102] Incorrect speculation detection logic 530 may control whether the scheduler determines an operation's readiness to issue based on either the operation's speculative sources or non-speculative sources. Incorrect speculation detection logic 530 may be duplicated for each pending operation in the scheduler in some embodiments such that each operation may independently watch for either speculative or non-speculative sources on result bus 520. In other embodiments, all of the pending operations may be controlled in the same way such that either all operations are watching for speculative sources (if available) or all operations are watching for non-speculative sources.

[0103] Incorrect speculation detection logic 530 may be configured to control the scheduler 118 behavior such that the scheduler determines the readiness of operations to issue based on the availability of their speculative sources so long as no incorrect speculations have been detected. When an incorrect speculation is detected, the incorrect speculation detection logic 530 may "flip" to watching for the readiness of non-speculative sources. For example, if the load store unit 126 detects an incorrect link through address or value comparisons, the incorrect speculation logic 530 may flip to using non-speculative sources to determine operation readiness to issue. The detection of the incorrect link may also cause all or part of the memory file 132 or other source of the speculation to be invalidated. The incorrect speculation detection logic 530 may continue to watch for non-speculative sources instead of speculative sources until the source of the speculation has been reset or otherwise corrected for the incorrect speculation.

[0104] When the incorrect speculation detection logic 530 "flips" from watching for the availability of speculative sources to watching for the availability of non-speculative sources, the scheduler 118 may lack information about the availability of the non-speculative sources. In some embodiments, the scheduler 118 may assume that all of the non-speculative sources are available and that the operation is thus ready to issue. If this assumption is incorrect, the operation may be reissued at a later time when the non-speculative sources are actually available (e.g., as detected when the tags of the non-speculative sources are broadcast on the result tag bus 520).

Speculative Register Map

[0105] FIG. 11 shows another embodiment of a dispatch unit 104 that may be included in a processor 100. In this embodiment, dispatch unit 104 includes a register map 134, a memory file 132, and a speculative register map 800. Like register map 134, speculative register map 800 may translate logical register names to physical register names. However, speculative register map 800 may speculatively map a logical register name to a physical register name (e.g., in response to memory file 132 linking a value stored in a physical register to a speculative result of a load operation). The speculative register map 800 may allow speculative operand values for operations that do not include addressing patterns to be linked to register data values. For example, if there is a valid speculative map for logical register EAX, an operation having EAX as a source operand may have two source tags: a non-speculative tag provided by register map 134 and a speculative tag provided by speculative register map 800. Since the

operation may issue as soon as its speculative source is available, speculative register map 800 may link data consumers directly to data producers via a speculative operand tag, bypassing any intervening loads and stores. Note that the speculative map may store tags other than physical register names in some embodiments (e.g., in embodiments having a reorder buffer that includes storage for speculative register states).

5 [0106] Speculative register map 800 includes a speculative register map controller 802 and speculative register map storage 804. Speculative register map storage may include one or more entries 820. Each entry 820 may be associated with a particular logical register and indicate a physical register identifier 812 of the physical register to which that logical register is currently speculatively mapped. Each speculative register map entry 820 may also include an indication (not shown) as to whether that entry is currently valid or not. In one embodiment, 10 speculative register map storage 804 may include an entry for each logical register. In other embodiments, speculative register map 804 may include fewer entries than there are logical registers. In such embodiments, each entry 820 may include an indication of the logical register to which that entry currently corresponds.

[0107] Speculative register map controller 802 may be configured to update an entry 820 in response to an indication that a load operation's speculative result has been linked with a data value identified by a particular 15 physical register name. In the illustrated embodiment, this indication is provided by the memory file 132. The entry 820 to update is the entry for the logical register specified as the destination of the load operation. The entry may be updated to include the physical register identifier output by the memory file 132 for that load operation. In other embodiments, speculative register map entries may be created in response to indications other than those provided by a memory file 132 (in some of these embodiments, dispatch unit 104 may not include a memory file). 20 For example, dispatch unit 104 may detect a conditional move instruction CMOV EAX, EBX if Z and responsively indicate that the speculative register map entry for EAX should identify the physical register currently mapped to EBX. Generally, speculative register map entries may be created in response to any prediction mechanism that indicates that a logical register should be speculatively mapped to a particular physical register.

[0108] In some embodiments, operations may be provided to speculative register map 800 during the same 25 cycle that they are provided to register map 134. As register map 134 performs the non-speculative register renaming for an operation, speculative register map 800 may indicate whether any of the logical registers specified as storing one of the operation's speculative source operands are linked to a particular physical register. If a valid entry exists in speculative register map storage 804 for one of the operation's logical register sources, speculative register map controller 802 may output the physical register name stored in that logical register's entry. Dispatch 30 unit 104 may output this speculative physical register name as a speculative source when the operation is dispatched to a scheduler 118. Thus, if an ADD operation is provided to speculative register map 800 and one of the ADD's sources has a valid entry in speculative register map storage 804, the tag for the physical register identified in that entry may be provided as a speculative source operand tag to scheduler 118. The scheduler may be configured to store both speculative and non-speculative operand tags, as described above, and may in some embodiments be 35 configured to reissue operations (if already issued) in response to the non-speculative tag being broadcast on a result bus.

[0109] Entries within the speculative map may be invalidated in response to an indication that a data value for a particular logical register will be modified. For example, if an operation ADD EAX, ECX is handled by the dispatch unit 104, the speculative register map controller 802 may invalidate the speculative map entry currently 40 assigned to EAX since the ADD operation will modify that register value.

[0110] Generally, speculative operand tags may be provided to scheduler 118 whenever one operation's speculative result is linked to a register data value. In some embodiments, a memory file 132 and another structure that tracks dependencies over several cycles (e.g., a speculative register map as described below) may be used to link speculative results to register values. For example, dispatch unit 104 may generate speculative tags for an operation in response to a memory file 132 identifying a link. In some embodiments, speculative tags may be generated without the use of such a speculative map. For example, a sequence of instructions may include:

```

ADD EBX, EBX                (ADD 1),
MOV [addressing pattern A], EBX    (STORE 1)
ADD [addressing pattern A], ECX    (ADD 2).

```

These instructions may be contiguous instructions (e.g., they may directly follow each other in program order). These instructions may be separated into the following component operations (shown with logical addresses translated to physical addresses) for execution within execution core(s) 124:

```

ADD PR2, PR2, PR1            (ADD 1)
MOV [addressing pattern A], PR2    (STORE 1)
MOV PR3, [addressing pattern A]    (load for ADD 2)
ADD PR4, PR3, PR5            (add for ADD 2)
MOV [addressing pattern A], PR4    (store for ADD 2)

```

Before the component load, add, and store operations of ADD 2 are dispatched, a dispatch unit 104 may detect whether there are any dependencies between any of the component operations in the sequence that would allow linking to speculative results. Additionally, the data being stored by STORE 1 may be linked to the load's speculative result (e.g., by a memory file). Since there are no intervening operations, dispatch unit may have all of the information needed to detect a dependency between the load operation and the add operation (both operations being derived from the same instruction) in the same dispatch cycle. Based on these two dependencies, the dispatch unit 104 may link the tag of the data being stored by STORE 1, PR2, to the speculative result of the load operation performed as part of ADD 2. This speculative link may in turn allow the dispatch unit to link the source of the addition operation performed as part of ADD 2 to the value stored in PR2. Accordingly, dispatch unit 104 may output an indication that PR2 may be speculatively used as the source of one of the operands for the addition but that PR3 is the non-speculative source for that operand. Thus, in one embodiment, the operations and operand identifiers output by dispatch unit may be specified as follows:

```

ADD PR2, PR2, PR1            (ADD 1)
MOV [addressing pattern A], PR2    (STORE 1)
MOV PR3, [addressing pattern A]    (load for ADD 2)
ADD PR4, PR2*, PR3, PR5        (add for ADD 2, where PR2* is a speculative source
                                for ECX and PR3 is the non-speculative source for
                                ECX)
MOV [addressing pattern A], PR4    (store for ADD 2).

```

[0111] In other embodiments, dispatch unit 104 may not be configured to identify speculative source operands for operations that depend on a load operation. Instead, as shown in FIG. 12, dispatch unit 104 may include an operation converter 180 configured to convert load operations into one or more operations that include a register-to-register move operation in order to provide speculative load results to dependent operations. The conversion of a load operation may be performed in response to an indication that a link exists between a

speculative result of the load operation and a data value identified by a particular physical register name. This indication is provided by the link detector 182, which may include a memory file 132 in some embodiments. In other embodiments, the link detector 182 may include logic configured to link data values in response to operations such as a conditional move operation, as described above.

[0112] In one embodiment, the operation converter may receive an input opcode for an operation as well as an indication as to whether a link between a register value and a speculative result of the operation is detected for that operation. If the operation is a load and a speculative link has been detected, the operation converter may output an opcode for a register-to-register move operation. The dispatch unit 104 may dispatch the register-to-register move operation, using the tag output by the link detection unit as the source operand tag for the register-to-register move.

[0113] In some embodiments, the operation converter may be configured to dispatch the resulting register-to-register move such that the scheduler stores the operand tags needed for both the register-to-register move and the original load operation in the entry allocated to the register-to-register move operation. This may allow the operation to be reissued as the original load operation if the speculative result of the register-to-register move operation is detected to be incorrect. In order to implement this, an additional source operand may be added to each register-to-register move operation that results from modifying a load operation (or, in alternative embodiments, a source operand that is already present may be modified to implement this). In some embodiments, the speculative result of the register-to-register move operation may be verified by performing the address calculation for the original load and/or comparing the linked data value to the actual load result data value. If the speculative result is incorrect, the data cache may be accessed in order to obtain the correct load result. Rebroadcast of the correct load result may cause the scheduler to reissue any dependent operations that were executed using the incorrect value.

[0114] In some embodiments, the operation converter 180 may be configured to convert a load operation into a dual-nature operation. Like a load operation, this dual-nature operation may involve both address calculation and data movement. Unlike a load, the data movement initiated by the dual-nature operation is a register-register move. Furthermore, the data movement initiated by the dual-nature operation may occur before the address calculation has completed. The address calculation may be used to verify whether the speculative link was correct. If the speculative link was incorrect, the dual-nature operation may be reissued as a normal load operation and its result may be rebroadcast to dependent operations upon completion of a data cache access.

[0115] The following examples show how different embodiments may convert this exemplary sequence of operations:

ADD PR2, PR1, PR1	(ADD 1)
...	
STORE [addressing pattern A], PR2	(STORE 1)
...	
LOAD PR3, [addressing pattern A]	(LOAD 1)
...	
ADD PR4, PR3, PR3	(ADD 2)

In this sequence, it is possible that the specified operations may be separated by one or more intervening operations. However, assuming that no intervening operations appear to modify the values used in addressing pattern A or to modify the data values stored at the address calculated from addressing pattern A and in PR2, a speculative link may be detected between the data values stored in PR2 and at the address calculated from addressing pattern A.

[0116] In one embodiment, in response to the detection of the speculative link by speculative link detector 182, operation converter 180 may convert LOAD 1 into a dual-nature move operation: MOV PR3, PR2. In addition to specifying the register source and destination, this dual-nature move operation may also specify addressing pattern A so that the address calculation for LOAD 1 may be performed. However, the move portion of the dual-nature move operation may issue as soon as ECX is available. As soon as the result of the move portion of the dual-nature operation is broadcast, ADD 2 may issue, using the speculative result of the move operation as an operand. When the address calculation is performed, the speculative link may be verified. If the speculative link is incorrect, the load/store unit may provide an indication to the scheduler that causes the scheduler to reissue the dual-nature move operation as a load operation. The result of the load operation may be broadcast, causing any dependent operations, such as ADD 2, which may have executed using the speculative result of the move to reissue. Note that this dual-nature operation may be scheduled using a single scheduler entry and that a scheduler 118 may select the dual-nature operation for issue twice: once for the load's address calculation and once for the register-to-register move.

[0117] In an alternative embodiment, operation converter 180 may convert LOAD 1 into two linked operations, LOAD PR3, [addressing pattern A] and MOV PR3, PR2, in response to detection of the speculative link. Unlike the earlier example, which involved a dual-nature operation, these linked operations may each take up an entry in a scheduler 118 instead of sharing a single scheduler entry. When the register-to-register move operation issues, dependent operations such as ADD 2 may issue using the speculative value of PR3 as an operand value. The LOAD operation may be tagged as a special type of load operation so that, if the speculative link is determined to be correct, the load's result tag is not broadcast (or is broadcast with an indication that the move's result is correct and that the dependent operations should not reissue). Also, if both the load and the move operation broadcast their results on the same cycle, an indication may be provided so that the load result is used for dependent operations instead of the move result. In some embodiments, this may occur by adding an extra bit to each tag bus within the processor 100 that indicates to which tag a scheduler should respond. If the speculative link is incorrect (e.g., as indicated by the broadcast of the load's result tag), the move operation may be cancelled (e.g., the scheduler 118 that schedules the move operation may deallocate the scheduler entry currently allocated to the register-to-register move so that the operation cannot reissue). In many embodiments, the scheduler may be configured to reissue any dependent operations that executed using the speculative load result in response to the load result tag reappearing on the result bus.

[0118] In one embodiment, if a link between a register data value identified by a particular tag and a speculative result of a load operation is detected (e.g., by a memory file), the load operation may be modified to include a register-to-register move operation having a source operand tag equal to the tag of the data value linked to the speculative load result. Execution of the register-to-register move operation may involve outputting the data value identified by the tag onto the result bus along with the tag of the load's result.

[0119] The speculative result may be verified (e.g., by performing the address calculation for the original load). If the speculative result is correct, the load's result may not be rebroadcast on the result bus (or, alternatively, the load's result may be rebroadcast along with an indication that dependent operations should not reissue in response to the rebroadcast result). If the speculative result is incorrect, the data cache may be accessed to retrieve the correct load result and the correct load result may be broadcast on the result bus. This may in turn cause any dependent operations that issued using the speculative result value to be reissued. Note that in some situations, the

speculative result may be verified before performance of the register-to-register move operation. If the speculative result is incorrect, the register-to-register move operation may not be performed.

[0120] If no link between the speculative result of the load operation and a register data value is detected (e.g., the load's addressing pattern misses in a memory file), the load may not be modified. The load may execute normally, accessing the data cache and broadcasting its result on the result bus.

Tag Coloring

[0121] In some situations, the memory file 132 may link the speculative result of a load operation to a value in a particular logical register. In some situations, that logical register may be overwritten, causing the physical register assigned to that logical register by the register map to be returned to the freelist (i.e., the list of free physical registers available to be assigned). The physical register may then be reassigned to store another value. The return of the physical register to the freelist and/or reassignment of the physical register may not affect the memory file entry specifying that physical register. Thus, subsequent links provided by the memory file 132 may not take the physical register reassignment into account. For example, consider the following sequence:

```
MOV [EBP+50], EAX
XOR EAX, EAX
MOV EBX, [EBP+50]
```

In response to this sequence, the memory file 132 may allocate an entry to the addressing pattern EBP+50 and the physical register assigned to EAX in response to the first store operation. When EAX is overwritten by the XOR operation, the current value of EAX may no longer correspond to the value that is linked to the addressing pattern in the memory file 132. However, the memory file 132 may link the speculative result of the subsequent load operation to the earlier value based on the match between the load's addressing pattern and the earlier store's addressing pattern.

[0122] In some embodiments, each physical register in register file 116 may be assigned a color. This color may be associated with the tag of each register and may be propagated throughout the processor along with the register's tag. As used herein, a color is one or more bits of information that indicate whether a register has been reassigned. In one embodiment, a color may be a single bit that is toggled each time a register is returned to or pulled from the freelist. If an entry specifying the tag of a particular register is allocated in the memory file, the memory file may also store the color associated with that register in the allocated entry. Similarly, the speculative register map may store a color 814 for each tag, as shown in FIG. 13. The color may also be stored with the tag in the register map 134 and the scheduler 118. When a register tag is output on result bus(es) 520, the color associated with that tag may also be output on the result bus(es) 520.

[0123] As an example, consider a physical register that initially has a color of '0'. This physical register, with color 0, may be assigned to a particular logical register by register map 134. Additionally, this physical register, with color 0, may be assigned to a logical register in speculative register map 800. At some point subsequent to the allocation of the speculative register map 800 entry, the physical register may be returned to the freelist and then selected for reassignment. When the physical register is pulled from the freelist, its color may be updated to '1'. At some point after the register's color is updated, the speculative register map 800 may identify a speculative source of an operation by the register's tag and the color 0, based on the entry allocated before the register was reassigned. This information identifying the speculative source may be stored in scheduler 118. Before the operation issues, another operation may produce the new value of that register, and the tag and current color, 1, may be broadcast on

the result bus 520. By comparing the current color 1 associated with the register's tag to the color 0 associated with the register's tag in the speculative register map entry, the scheduler 118 may determine that the tag no longer identifies the same value. In some embodiments, the scheduler 118 may treat this as an incorrect speculation and "flip" to watching for the non-speculative source for that operand, as described above with respect to FIG. 10. The flip may be limited to the particular operation for which the color mismatch is detected (as opposed to causing all of the pending operations in the scheduler to flip).

[0124] Typically, it is difficult to remove register to register move operations from the dependency graph without causing difficulties in the management of physical register reassignment. For example, if register to register moves are implemented by updating the register map 134 so that two logical register numbers reference the same physical register number, it may become difficult to determine when the physical register identified by that physical register number should be returned to the freelist. Register coloring may be used as an alternative to these systems.

[0125] In some embodiments, register coloring may be used when optimizing performance of register to register moves by using the speculative register map 800. For example, the speculative register map 800 may be used to remove register to register move operations from the dependency graph (i.e., operations that depend on the result of a register to register move operation will no longer have to wait for the move operation to execute if the move's source is already available). Whenever a register to register move operation is detected, the speculative register map 800 may allocate an entry linking the source and destination register specified in that move operation. The color of each physical register may be stored in the resulting speculative register map 800 entry. In this system, physical registers may be reassigned normally without the difficulties that arise in systems where the register map 134 is updated so that multiple logical registers are associated with the same physical register. When a physical register is reassigned in the system using coloring, the particular value to which the physical register is currently allocated may be differentiated from other values that may have been allocated in the same physical register before the physical register was reassigned by the color associated with each tag identifying that physical register.

[0126] The speculative register map 800 may be configured to handle both integer and floating point registers, allowing the speculative register map to remove floating point register to register move operations from the dependency graph. This may in turn allow operations such as FXCH (which exchanges a stack element with the current top of the stack) that have significant latency to be removed from the dependency graph. Note that even though the use of the speculative register map 800 may remove register to register moves from the dependency graph, the register to register move may still be executed.

Exemplary Computer Systems

[0127] FIG. 14 shows a block diagram of one embodiment of a computer system 400 that includes a processor 100 coupled to a variety of system components through a bus bridge 402. Processor 100 may include an embodiment of a dispatch unit 104, a memory file 132, a scheduler 118, and/or speculative register map 800 as described above. Other embodiments of a computer system are possible and contemplated. In the depicted system, a main memory 200 is coupled to bus bridge 402 through a memory bus 406, and a graphics controller 408 is coupled to bus bridge 402 through an AGP bus 410. Several PCI devices 412A-412B are coupled to bus bridge 402 through a PCI bus 414. A secondary bus bridge 416 may also be provided to accommodate an electrical interface to one or more EISA or ISA devices 418 through an EISA/ISA bus 420. In this example, processor 100 is coupled to

bus bridge 402 through a CPU bus 424 and to an optional L2 cache 428. In some embodiments, the processor 100 may include an integrated L1 cache (not shown).

[0128] Bus bridge 402 provides an interface between processor 100, main memory 404, graphics controller 408, and devices attached to PCI bus 414. When an operation is received from one of the devices connected to bus bridge 402, bus bridge 402 identifies the target of the operation (e.g., a particular device or, in the case of PCI bus 414, that the target is on PCI bus 414). Bus bridge 402 routes the operation to the targeted device. Bus bridge 402 generally translates an operation from the protocol used by the source device or bus to the protocol used by the target device or bus.

[0129] In addition to providing an interface to an ISA/EISA bus for PCI bus 414, secondary bus bridge 416 may incorporate additional functionality. An input/output controller (not shown), either external from or integrated with secondary bus bridge 416, may also be included within computer system 400 to provide operational support for a keyboard and mouse 422 and for various serial and parallel ports. An external cache unit (not shown) may also be coupled to CPU bus 424 between processor 100 and bus bridge 402 in other embodiments. Alternatively, the external cache may be coupled to bus bridge 402 and cache control logic for the external cache may be integrated into bus bridge 402. L2 cache 428 is shown in a backside configuration to processor 100. It is noted that L2 cache 428 may be separate from processor 100, integrated into a cartridge (e.g., slot 1 or slot A) with processor 100, or even integrated onto a semiconductor substrate with processor 100.

[0130] Main memory 200 is a memory in which application programs are stored and from which processor 100 primarily executes. A suitable main memory 200 comprises DRAM (Dynamic Random Access Memory). For example, a plurality of banks of SDRAM (Synchronous DRAM) or Rambus DRAM (RDRAM) may be suitable.

[0131] PCI devices 412A-412B are illustrative of a variety of peripheral devices such as network interface cards, video accelerators, audio cards, hard or floppy disk drives or drive controllers, SCSI (Small Computer Systems Interface) adapters and telephony cards. Similarly, ISA device 418 is illustrative of various types of peripheral devices, such as a modem, a sound card, and a variety of data acquisition cards such as GPIB or field bus interface cards.

[0132] Graphics controller 408 is provided to control the rendering of text and images on a display 426. Graphics controller 408 may embody a typical graphics accelerator generally known in the art to render three-dimensional data structures that can be effectively shifted into and from main memory 200. Graphics controller 408 may therefore be a master of AGP bus 410 in that it can request and receive access to a target interface within bus bridge 402 to thereby obtain access to main memory 200. A dedicated graphics bus accommodates rapid retrieval of data from main memory 404. For certain operations, graphics controller 408 may further be configured to generate PCI protocol transactions on AGP bus 410. The AGP interface of bus bridge 402 may thus include functionality to support both AGP protocol transactions as well as PCI protocol target and initiator transactions. Display 426 is any electronic display upon which an image or text can be presented. A suitable display 426 includes a cathode ray tube ("CRT"), a liquid crystal display ("LCD"), etc.

[0133] It is noted that, while the AGP, PCI, and ISA or EISA buses have been used as examples in the above description, any bus architectures may be substituted as desired. It is further noted that computer system 400 may be a multiprocessing computer system including additional processors (e.g., processor 100a shown as an optional component of computer system 400). Processor 100a may be similar to processor 100. More particularly, processor 100a may be an identical copy of processor 100. Processor 100a may be connected to bus bridge 402 via

an independent bus (as shown in FIG. 14) or may share CPU bus 224 with processor 100. Furthermore, processor 100a may be coupled to an optional L2 cache 428a similar to L2 cache 428.

[0134] Turning now to FIG. 15, another embodiment of a computer system 400 that may include a dispatch unit 104, a memory file 132, a scheduler 118, and/or speculative register map 800 as described above is shown.

5 Other embodiments are possible and contemplated. In the embodiment of FIG. 15, computer system 400 includes several processing nodes 612A, 612B, 612C, and 612D. Each processing node is coupled to a respective memory 614A-614D via a memory controller 616A-616D included within each respective processing node 612A-612D. Additionally, processing nodes 612A-612D include interface logic used to communicate between the processing nodes 612A-612D. For example, processing node 612A includes interface logic 618A for communicating with
10 processing node 612B, interface logic 618B for communicating with processing node 612C, and a third interface logic 618C for communicating with yet another processing node (not shown). Similarly, processing node 612B includes interface logic 618D, 618E, and 618F; processing node 612C includes interface logic 618G, 618H, and 618I; and processing node 612D includes interface logic 618J, 618K, and 618L. Processing node 612D is coupled to communicate with a plurality of input/output devices (e.g., devices 620A-620B in a daisy chain configuration)
15 via interface logic 618L. Other processing nodes may communicate with other I/O devices in a similar fashion.

[0135] Processing nodes 612A-612D implement a packet-based link for inter-processing node communication. In the present embodiment, the link is implemented as sets of unidirectional lines (e.g., lines 624A are used to transmit packets from processing node 612A to processing node 612B and lines 624B are used to transmit packets from processing node 612B to processing node 612A). Other sets of lines 624C-624H are used to
20 transmit packets between other processing nodes as illustrated in FIG. 15. Generally, each set of lines 624 may include one or more data lines, one or more clock lines corresponding to the data lines, and one or more control lines indicating the type of packet being conveyed. The link may be operated in a cache coherent fashion for communication between processing nodes or in a non-coherent fashion for communication between a processing node and an I/O device (or a bus bridge to an I/O bus of conventional construction such as the PCI bus or ISA bus).
25 Furthermore, the link may be operated in a non-coherent fashion using a daisy-chain structure between I/O devices as shown. It is noted that a packet to be transmitted from one processing node to another may pass through one or more intermediate nodes. For example, a packet transmitted by processing node 612A to processing node 612D may pass through either processing node 612B or processing node 612C as shown in FIG. 15. Any suitable routing algorithm may be used. Other embodiments of computer system 400 may include more or fewer processing nodes
30 then the embodiment shown in FIG. 15.

[0136] Generally, the packets may be transmitted as one or more bit times on the lines 624 between nodes. A bit time may be the rising or falling edge of the clock signal on the corresponding clock lines. The packets may include command packets for initiating transactions, probe packets for maintaining cache coherency, and response packets from responding to probes and commands.

[0137] Processing nodes 612A-612D, in addition to a memory controller and interface logic, may include one or more processors. Broadly speaking, a processing node comprises at least one processor and may optionally include a memory controller for communicating with a memory and other logic as desired. More particularly, each processing node 612A-612D may comprise one or more copies of processor 100. External interface unit 18 may include the interface logic 618 within the node, as well as the memory controller 616.

[0138] Memories 614A-614D may comprise any suitable memory devices. For example, a memory 614A-614D may comprise one or more RAMBUS DRAMs (RDRAMs), synchronous DRAMs (SDRAMs), static RAM,

etc. The address space of computer system 400 is divided among memories 614A-614D. Each processing node 612A-612D may include a memory map used to determine which addresses are mapped to which memories 614A-614D, and hence to which processing node 612A-612D a memory request for a particular address should be routed. In one embodiment, the coherency point for an address within computer system 400 is the memory controller 616A-616D coupled to the memory storing bytes corresponding to the address. In other words, the memory controller 616A-616D is responsible for ensuring that each memory access to the corresponding memory 614A-614D occurs in a cache coherent fashion. Memory controllers 616A-616D may comprise control circuitry for interfacing to memories 614A-614D. Additionally, memory controllers 616A-616D may include request queues for queuing memory requests.

[0139] Interface logic 618A-618L may comprise a variety of buffers for receiving packets from the link and for buffering packets to be transmitted upon the link. Computer system 400 may employ any suitable flow control mechanism for transmitting packets. For example, in one embodiment, each interface logic 618 stores a count of the number of each type of buffer within the receiver at the other end of the link to which that interface logic is connected. The interface logic does not transmit a packet unless the receiving interface logic has a free buffer to store the packet. As a receiving buffer is freed by routing a packet onward, the receiving interface logic transmits a message to the sending interface logic to indicate that the buffer has been freed. Such a mechanism may be referred to as a "coupon-based" system.

[0140] I/O devices 620A-620B may be any suitable I/O devices. For example, I/O devices 620A-620B may include devices for communicate with another computer system to which the devices may be coupled (e.g., network interface cards or modems). Furthermore, I/O devices 620A-620B may include video accelerators, audio cards, hard or floppy disk drives or drive controllers, SCSI (Small Computer Systems Interface) adapters and telephony cards, sound cards, and a variety of data acquisition cards such as GPIB or field bus interface cards. It is noted that the term "I/O device" and the term "peripheral device" are intended to be synonymous herein.

[0141] As used herein, the terms "clock cycle" or "cycle" refer to an interval of time in which the various stages of the instruction processing pipelines complete their tasks. Instructions and computed values are captured by memory elements (such as registers or arrays) according to a clock signal defining the clock cycle. For example, a memory element may capture a value according to the rising or falling edge of the clock signal.

[0142] Numerous variations and modifications will become apparent to those skilled in the art once the above disclosure is fully appreciated. It is intended that the following claims be interpreted to embrace all such variations and modifications.

Industrial Applicability

This invention may generally be applicable to the field of processors.

WHAT IS CLAIMED IS:

1. A system (100) comprising:

a memory file (132) including an entry (220) configured to store a first addressing pattern (206) and a first tag (208), wherein the memory file (132) is configured to compare the first addressing pattern (206) included in the entry (220) to a second addressing pattern (206) of a load operation, wherein if the second addressing pattern (206) matches the first addressing pattern (206) stored in the entry (220), the memory file (132) is configured to link a data value identified by the first tag (208) to a speculative result of the load operation;

an execution core (124) coupled to the memory file (132), wherein the execution core (124) is configured to access the speculative result when executing a second operation that is dependent on the load operation; and

a functional unit (126) coupled to the memory file (132) and configured to verify the link between the data value identified by the first tag (208) and the speculative result of the load operation by performing a comparison between one or more addresses.

2. The system (100) of claim 1, wherein the functional unit (126) includes store-to-load forwarding comparators (280) configured to perform the comparison and to perform non-speculative store-to-load forwarding.

3. The system (100) of claim 1, wherein the entry (220) is further configured to store a forwarding tag (256) identifying a load or store operation in response to which the entry (220) was allocated and to output said forwarding tag (256) if the second addressing pattern (206) matches the first addressing pattern (206) stored in the entry (220).

4. The system (100) of claim 3, wherein the functional unit (126) is coupled to receive the forwarding tag (256) and is further configured to perform the comparison by selecting a youngest operation that is older than the load operation and that has an address matching an address of the load operation, and by comparing the forwarding tag (256) to an identifying tag of the selected youngest operation, wherein the selected youngest operation is a load or store operation.

5. The system (100) of claim 3, wherein the functional unit (126) is coupled to receive the forwarding tag (256) and is further configured to perform the comparison by selecting a youngest load operation and a youngest store operation each older than the load operation and each having an address matching an address of the load operation, by determining that the selected youngest load operation is younger in program order than the selected youngest store operation, and by comparing the forwarding tag (256) to an identifying tag of the selected youngest load operation.

6. A method comprising:

storing an entry (220) including a first addressing pattern (206) and a first tag (208);

comparing the first addressing pattern (206) included in the entry (220) to a second addressing pattern (206) of a load operation subsequent to storing the entry (220);

if the second addressing pattern (206) matches the first addressing pattern (206) stored in the entry (220),
linking a data value identified by the first tag (208) to a speculative result of the load operation in
response to the comparing;
accessing the speculative result when executing a second operation that is dependent on the load operation
subsequent to the linking; and
verifying the link between the data value identified by the first tag (208) and the speculative result of the
load operation by performing a comparison between one or more addresses subsequent to the
linking.

7. The method of claim 6, wherein the entry (220) is allocated in response to a store operation that
occurs earlier in program order than the load operation, and further comprising setting a flag (269) associated with
the store operation in response to linking the data value identified by the tag to the speculative result of the load
operation.

8. The method of claim 7, further comprising maintaining information associated with the store
operation until the link has been verified by performing the comparison, wherein said maintaining occurs in
response to setting the flag (269).

9. The method of claim 8, further comprising not retiring the store operation until the link has been
verified, in response to setting the flag (269).

10. A method comprising:

storing an entry (220) including a first addressing pattern (206) and a first tag (208);

comparing the first addressing pattern (206) included in the entry (220) to a second addressing pattern
(206) of a load operation subsequent to storing the entry (220);

if the second addressing pattern (206) matches the first addressing pattern (206) stored in the entry (220),
linking a data value identified by the first tag (208) to a speculative result of the load operation in
response to the comparing;

accessing the speculative result when executing a second operation that is dependent on the load operation
subsequent to the linking; and

verifying the link between the data value identified by the first tag (208) and the speculative result of the
load operation by performing a comparison between the speculative result and an actual result of
the load operation subsequent to the linking.

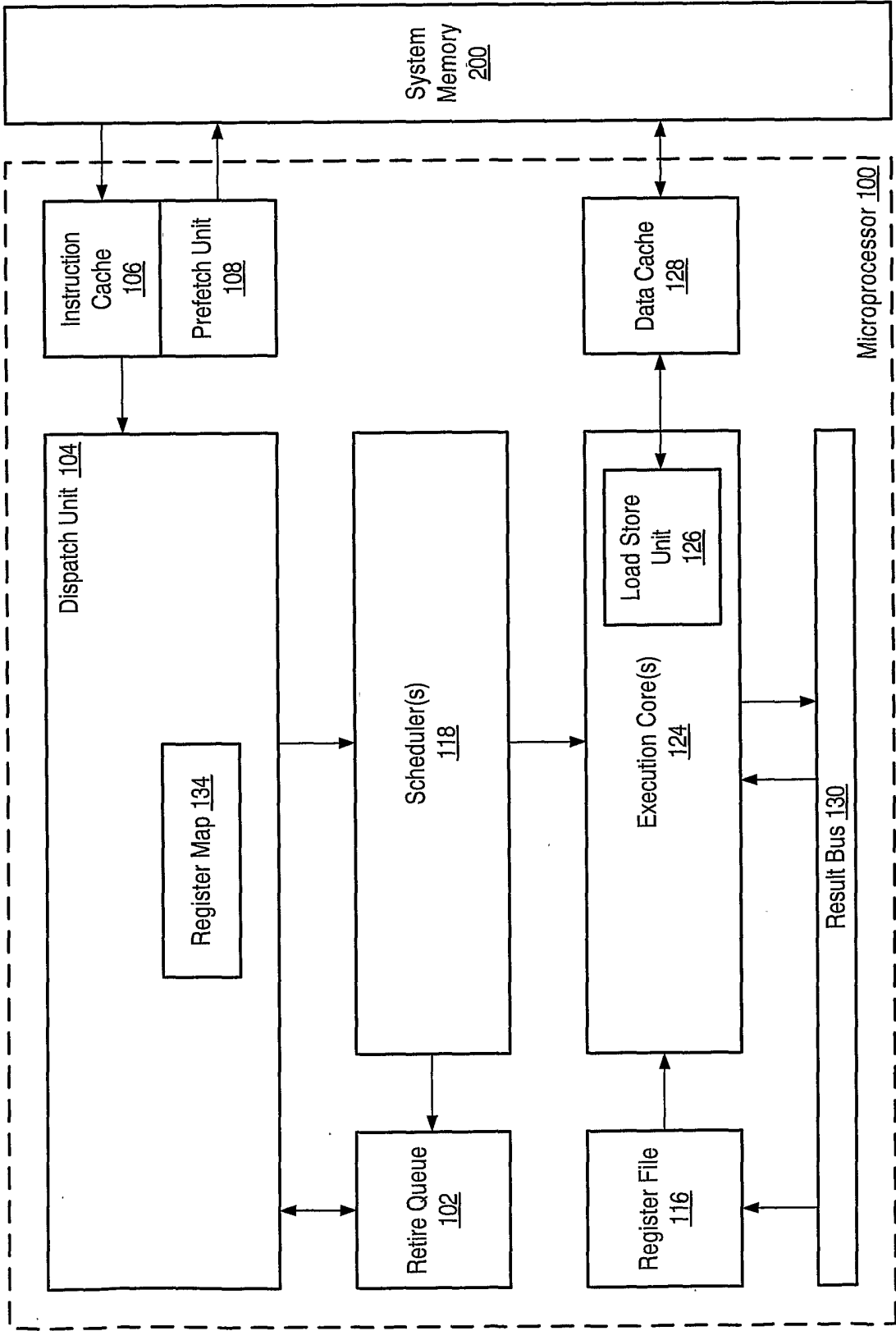


FIG. 1

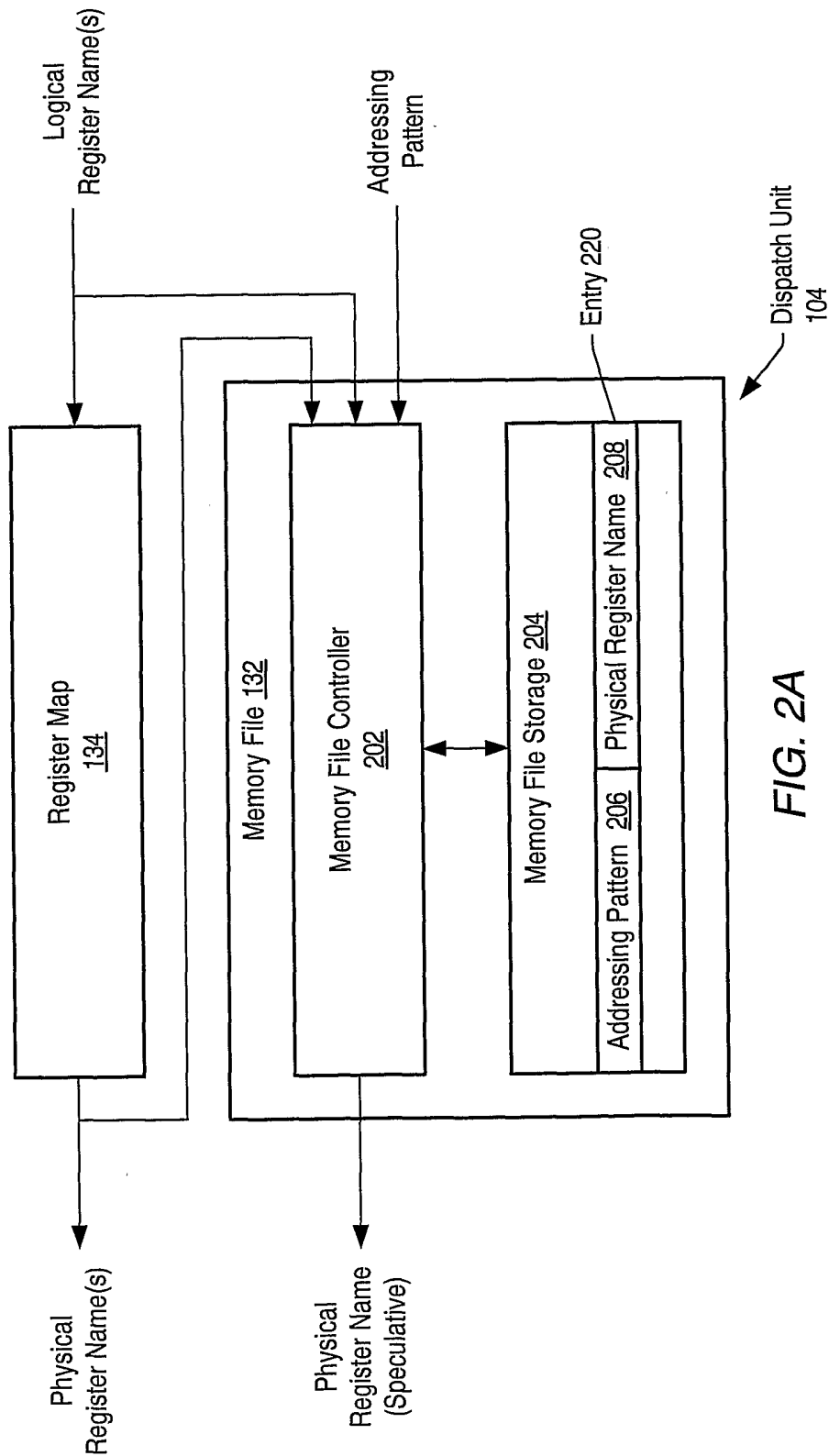


FIG. 2A

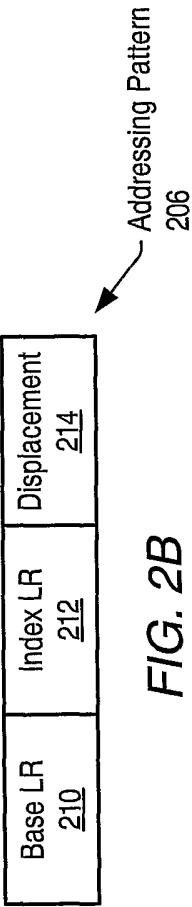


FIG. 2B

3 / 16

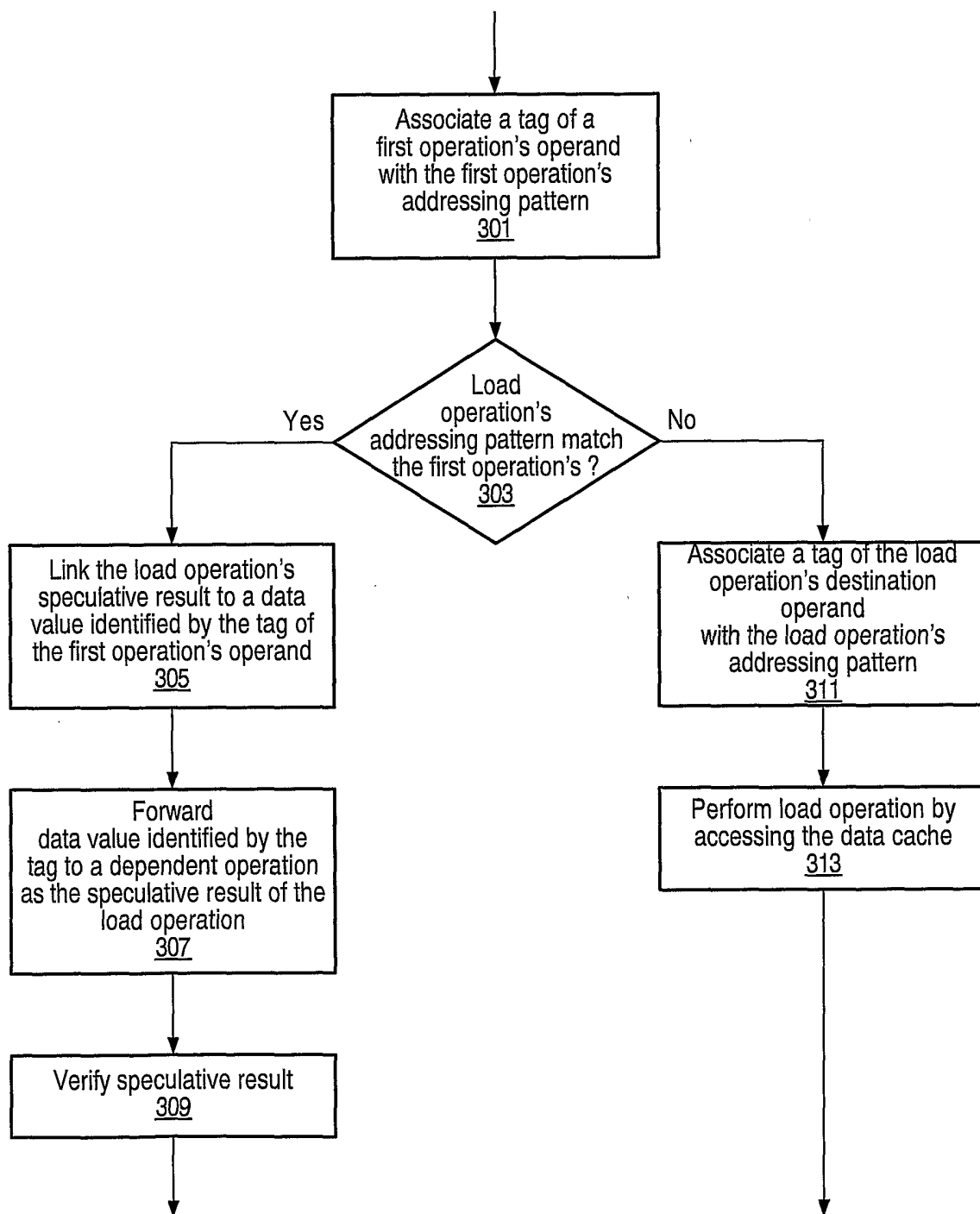


FIG. 3

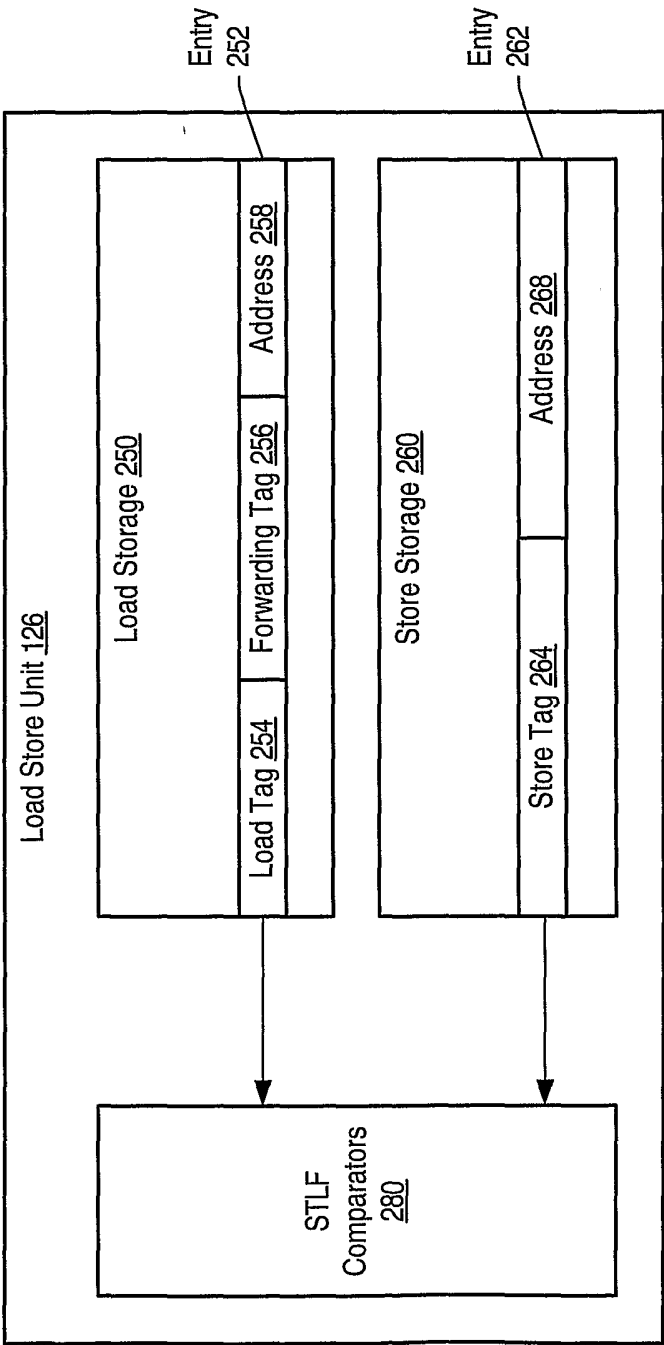


FIG. 4A

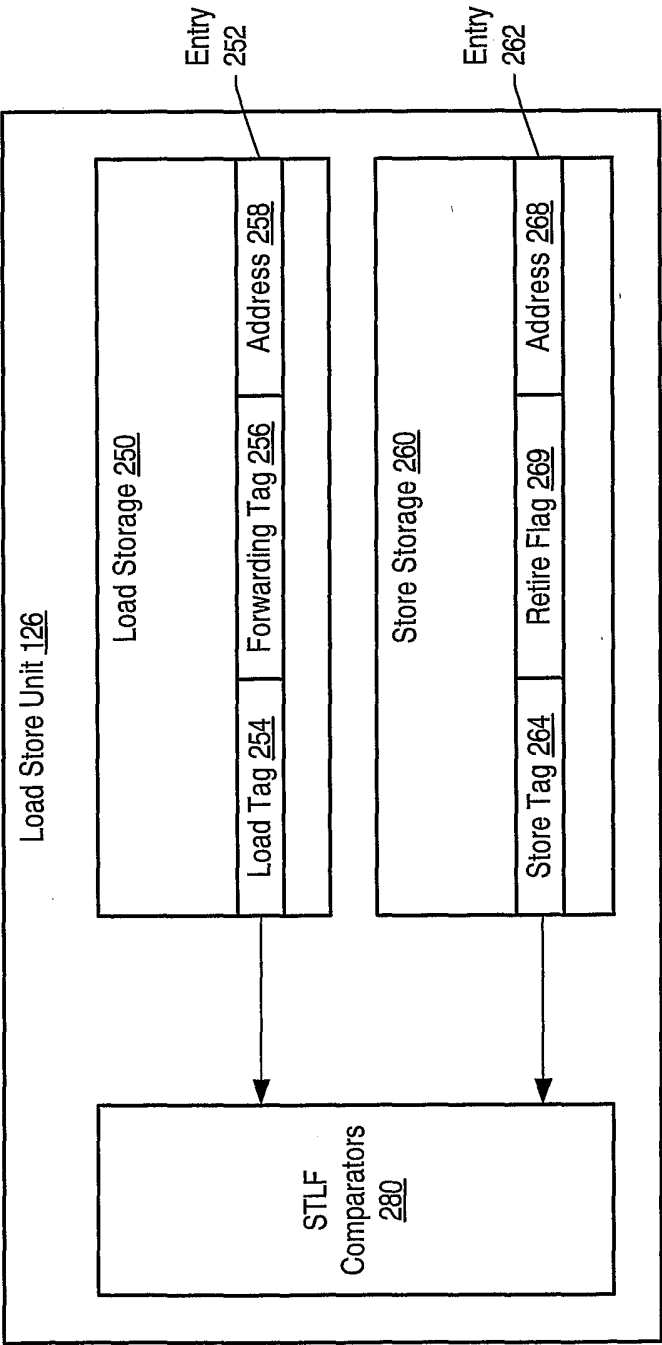


FIG. 4B

6 / 16

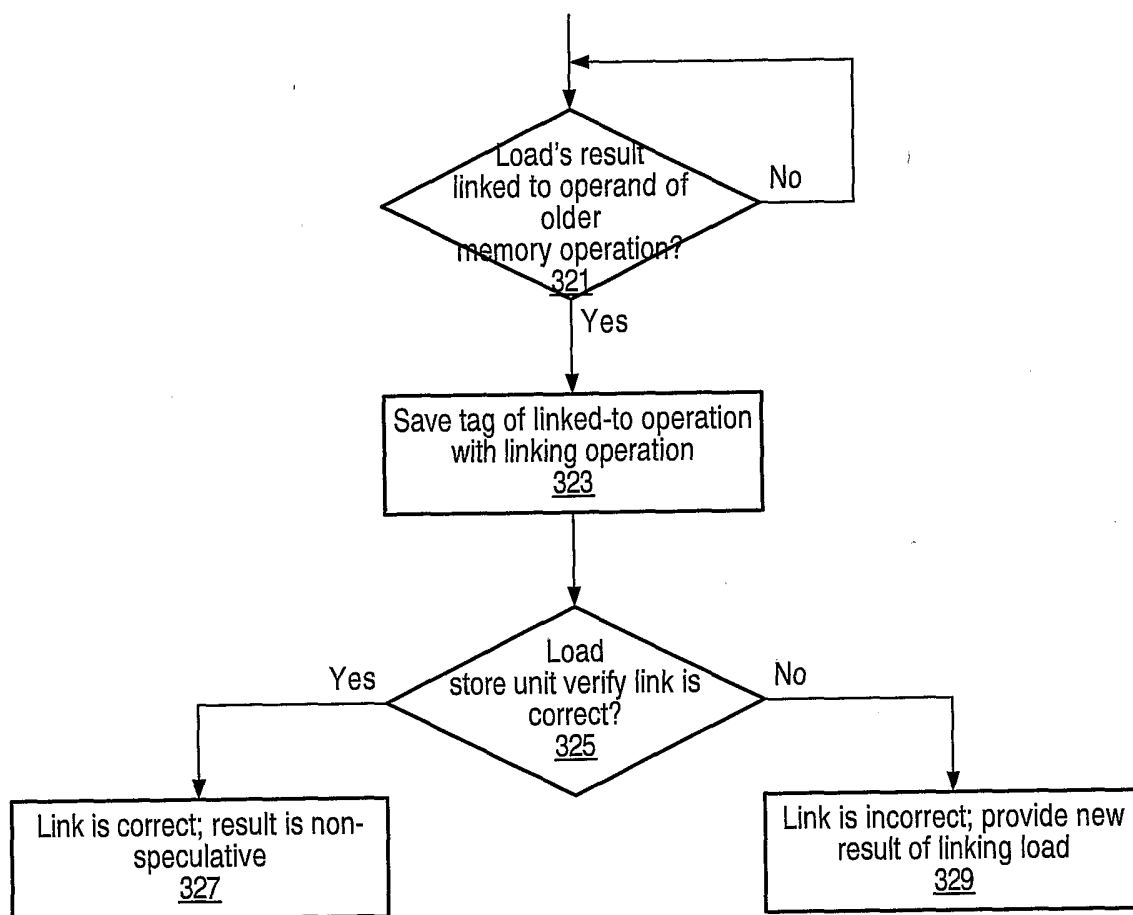


FIG. 5A

7 / 16

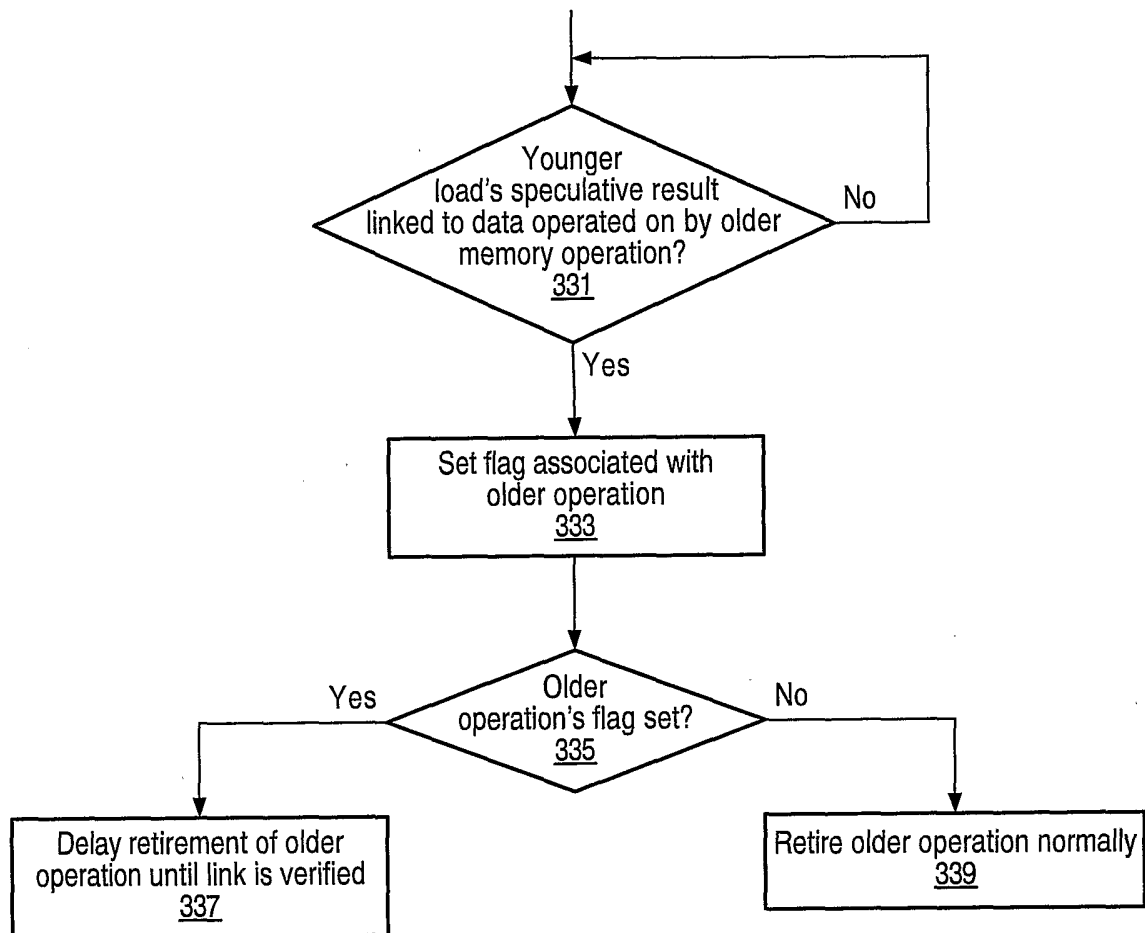


FIG. 5B

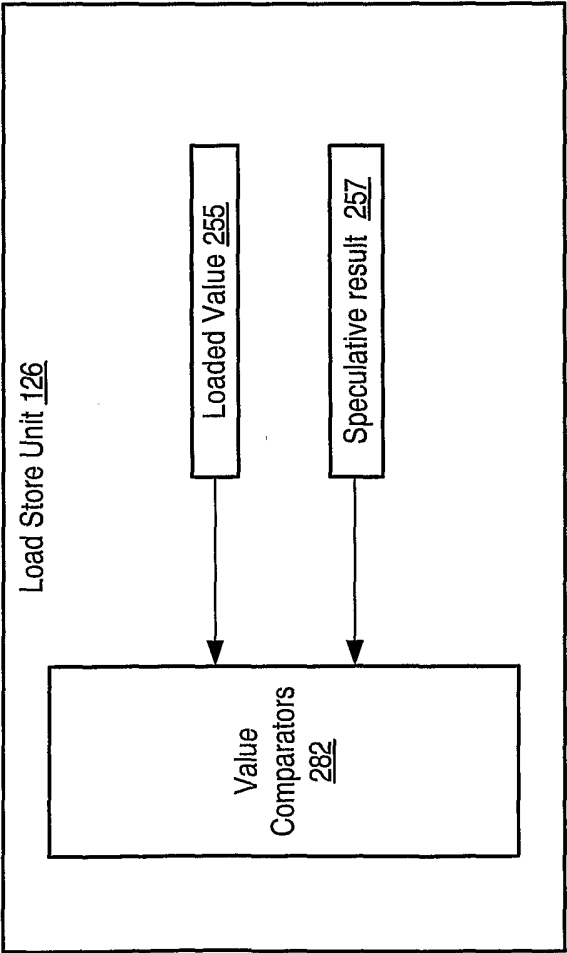


FIG. 6A

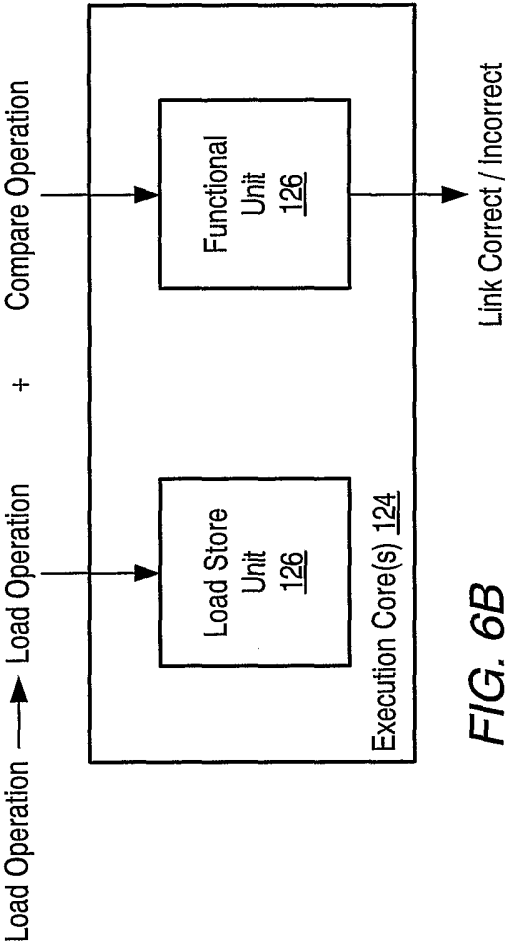


FIG. 6B

9 / 16

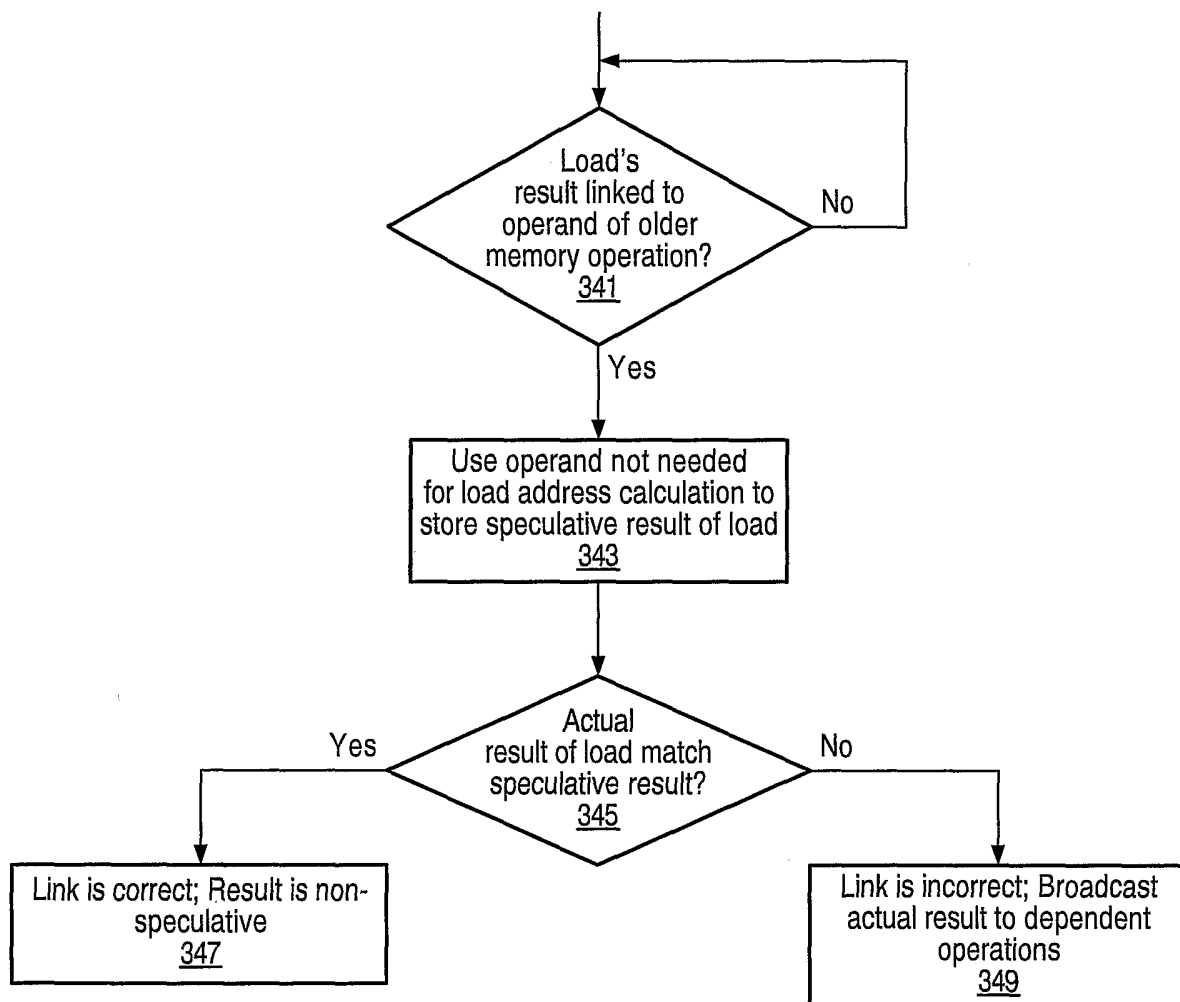


FIG. 7

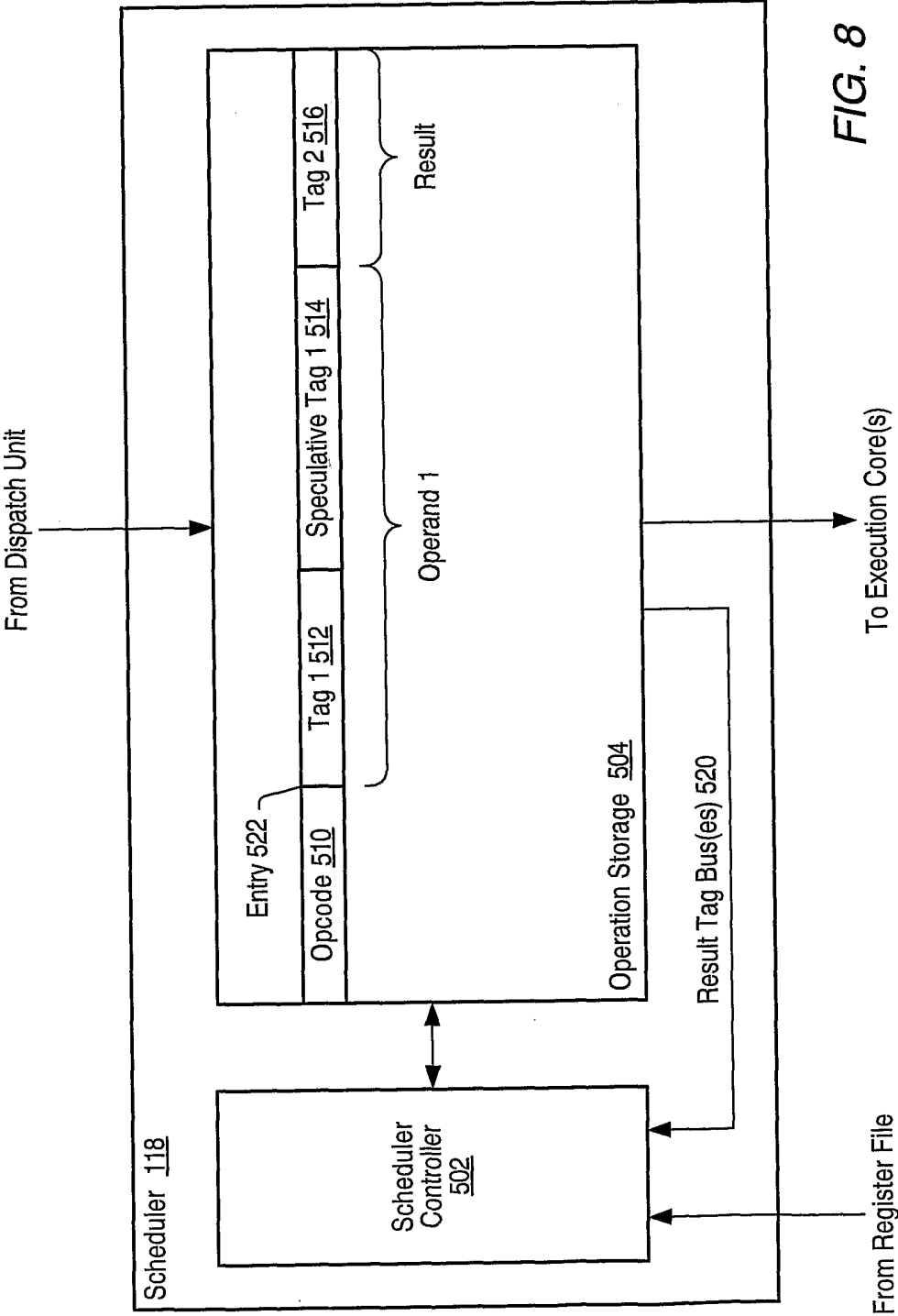


FIG. 8

11 / 16

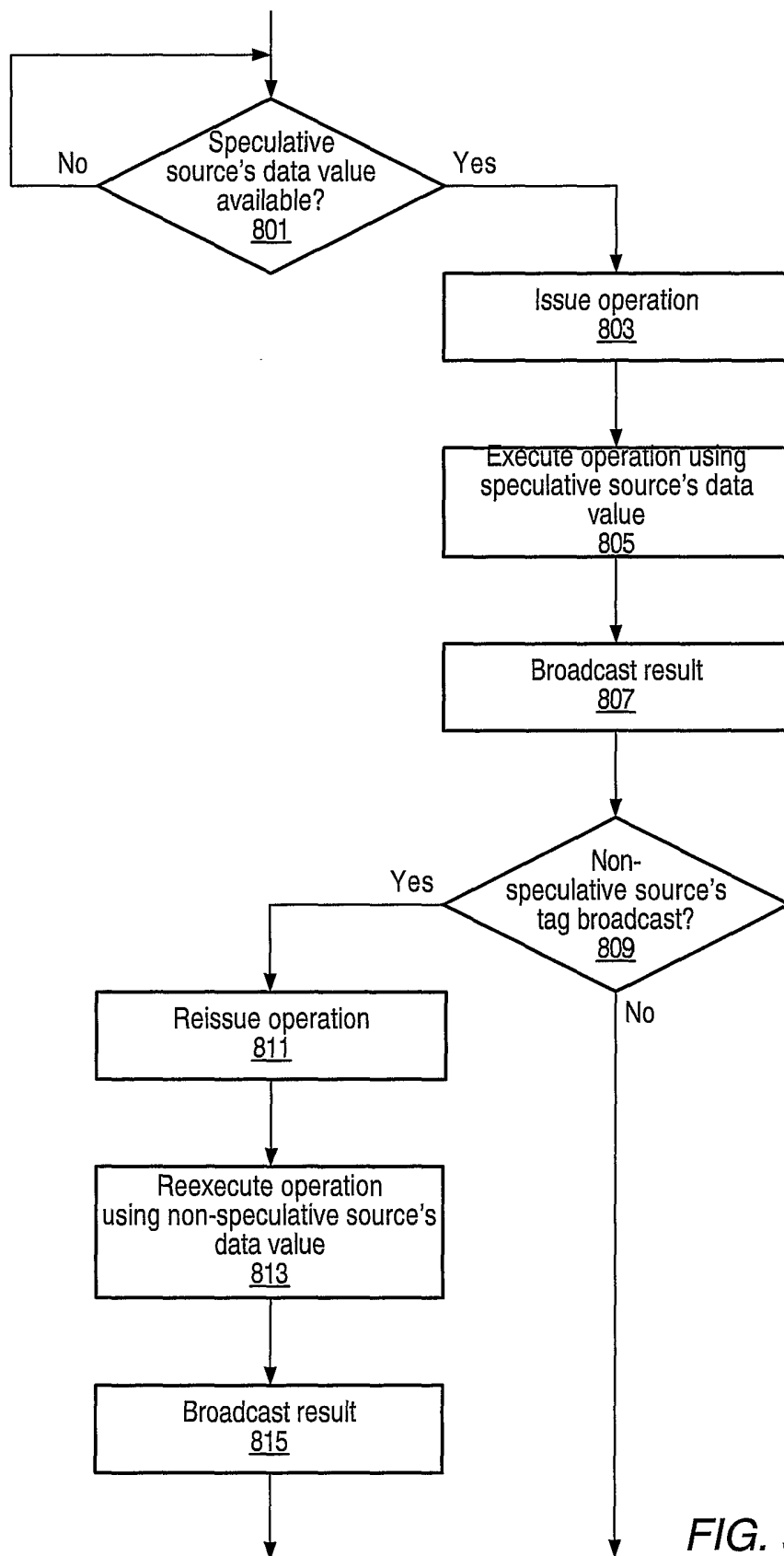


FIG. 9

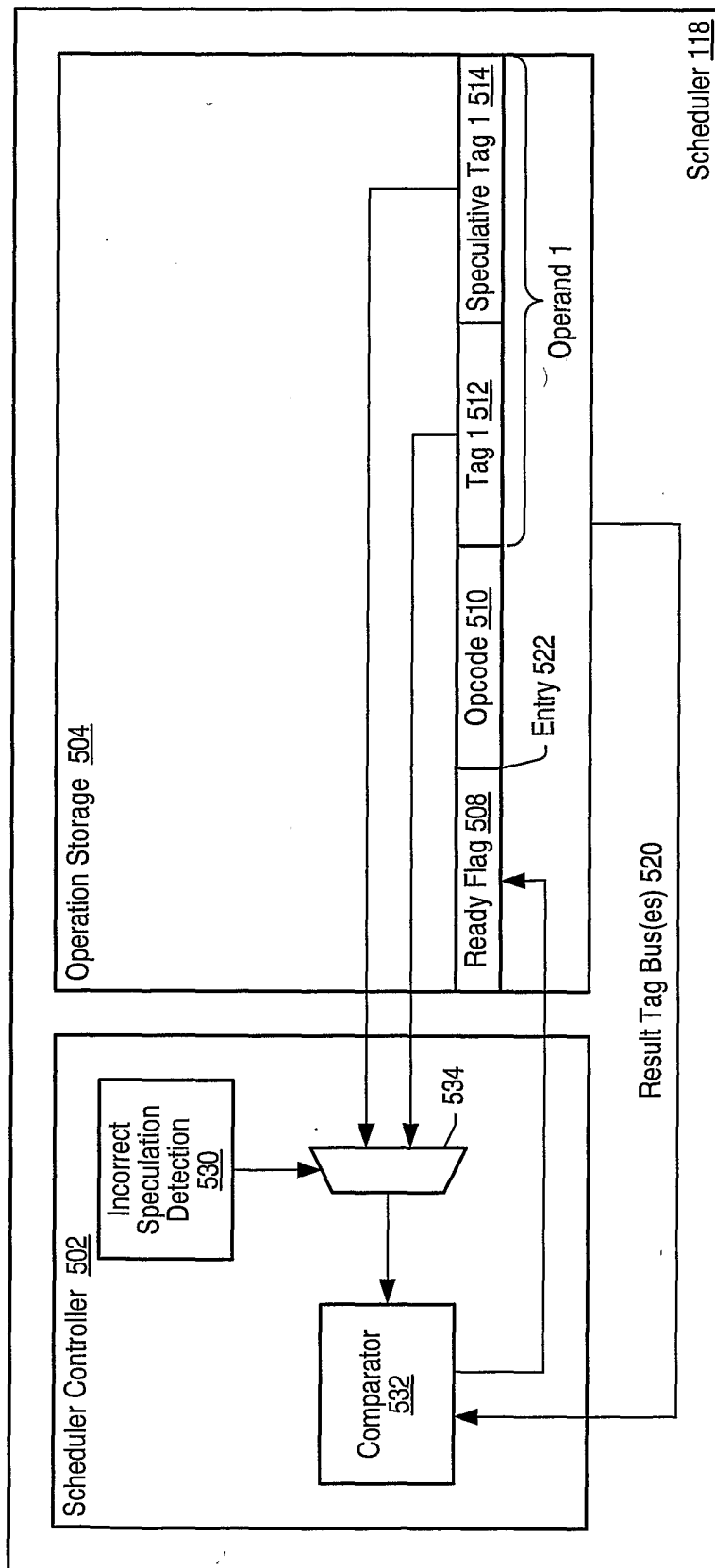


FIG. 10

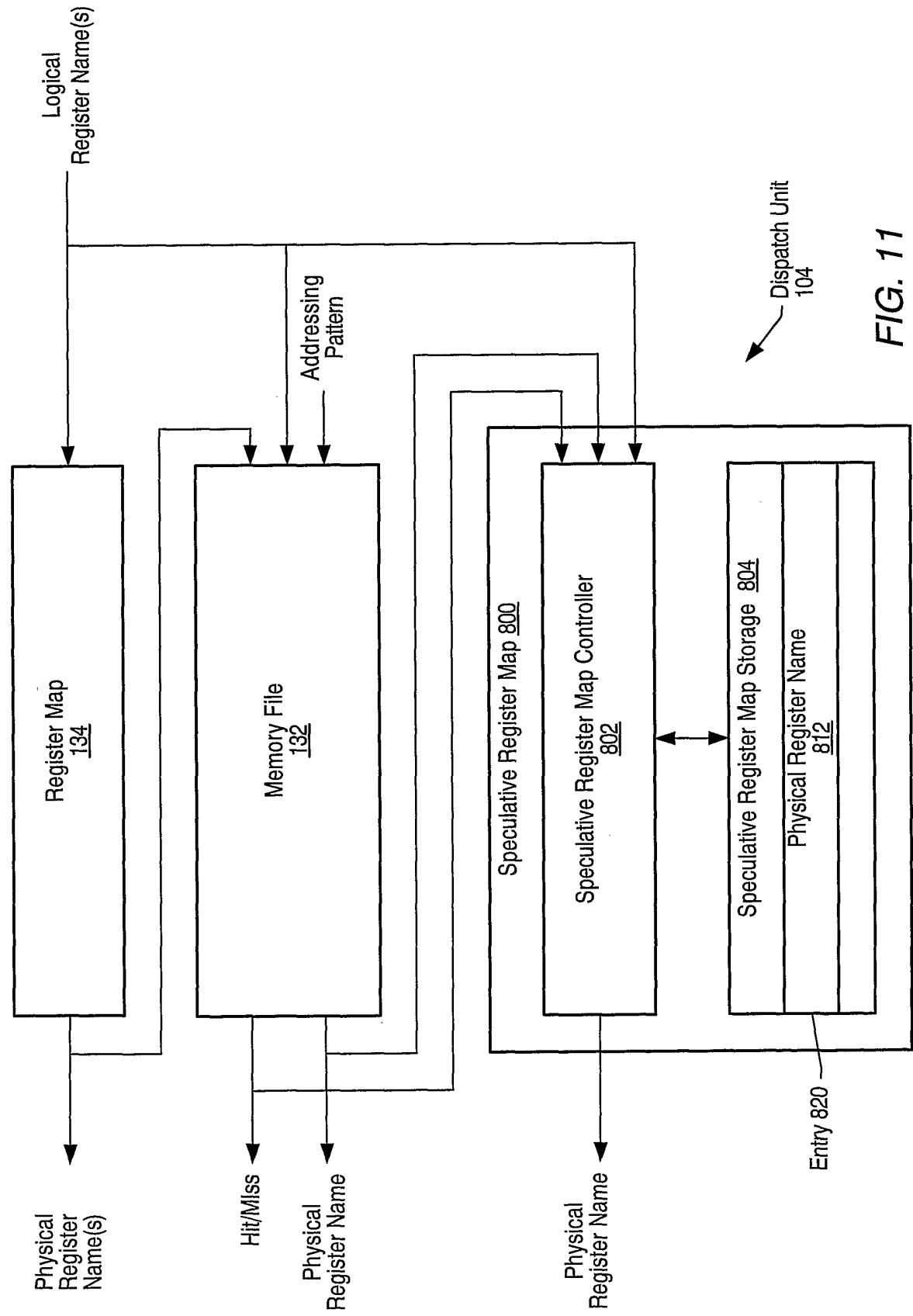


FIG. 11

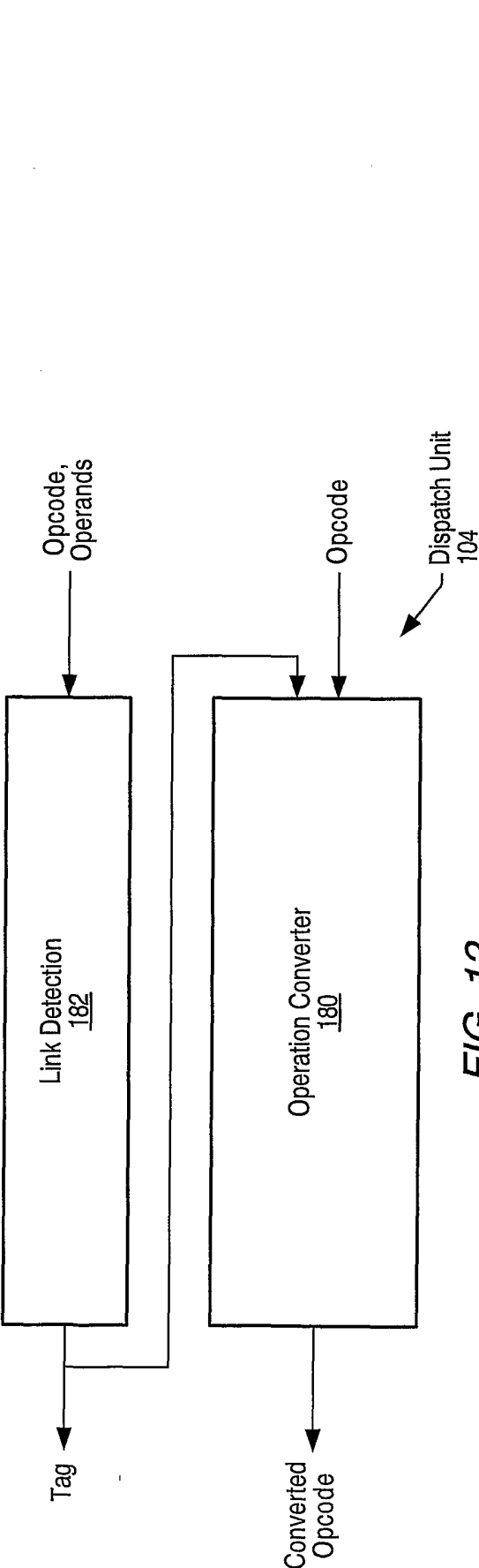


FIG. 12

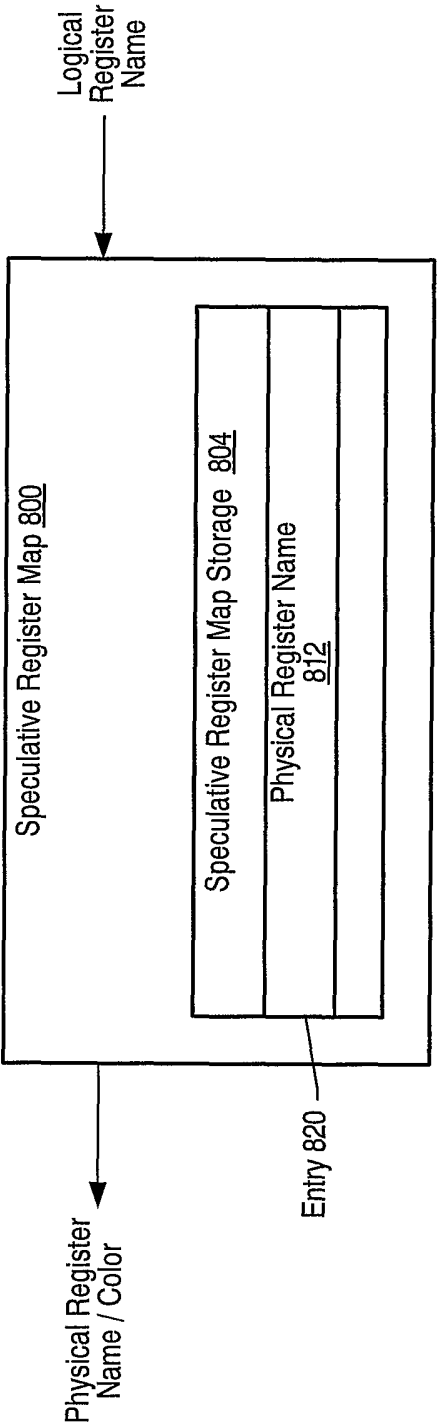


FIG. 13

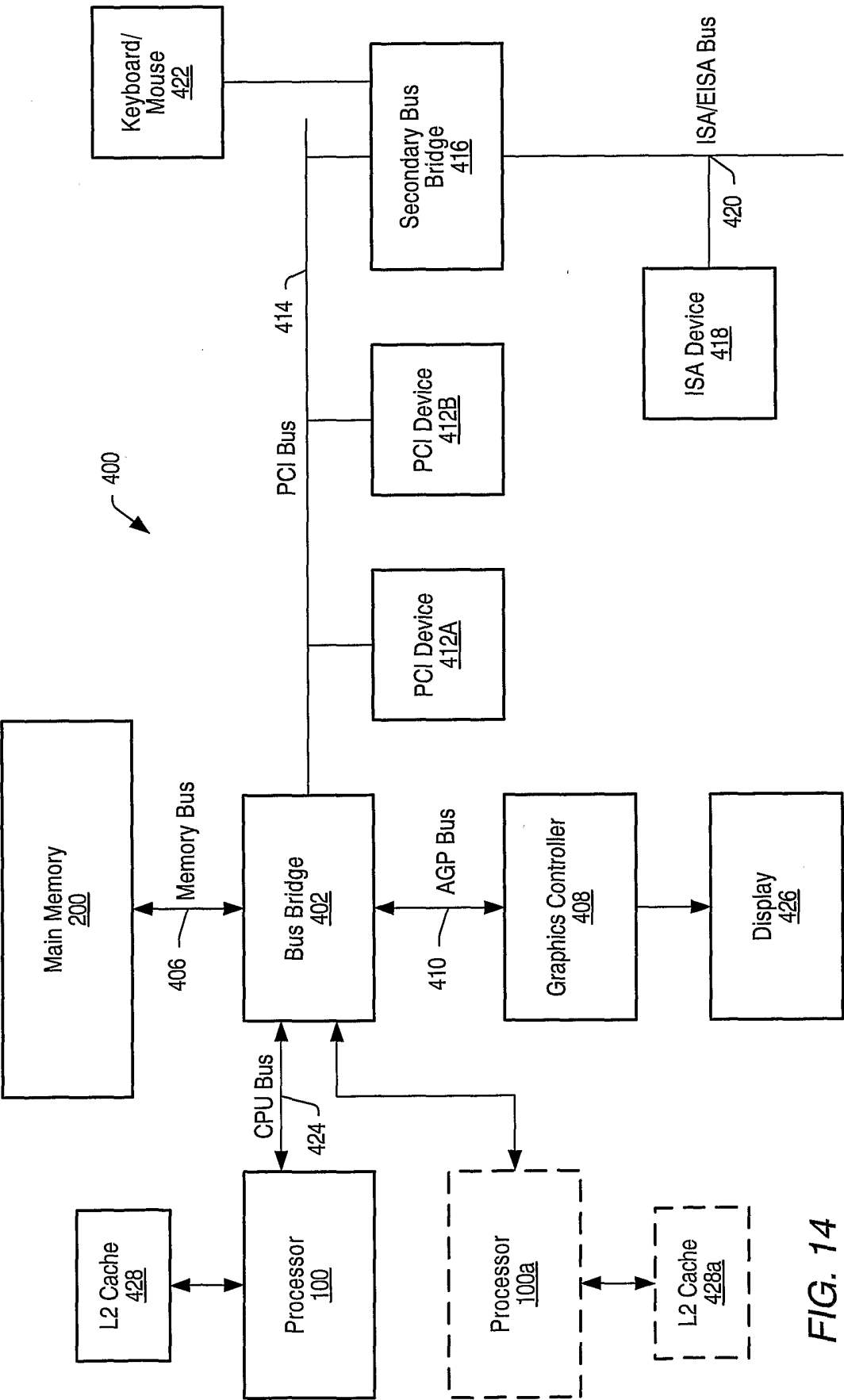


FIG. 14

16 / 16

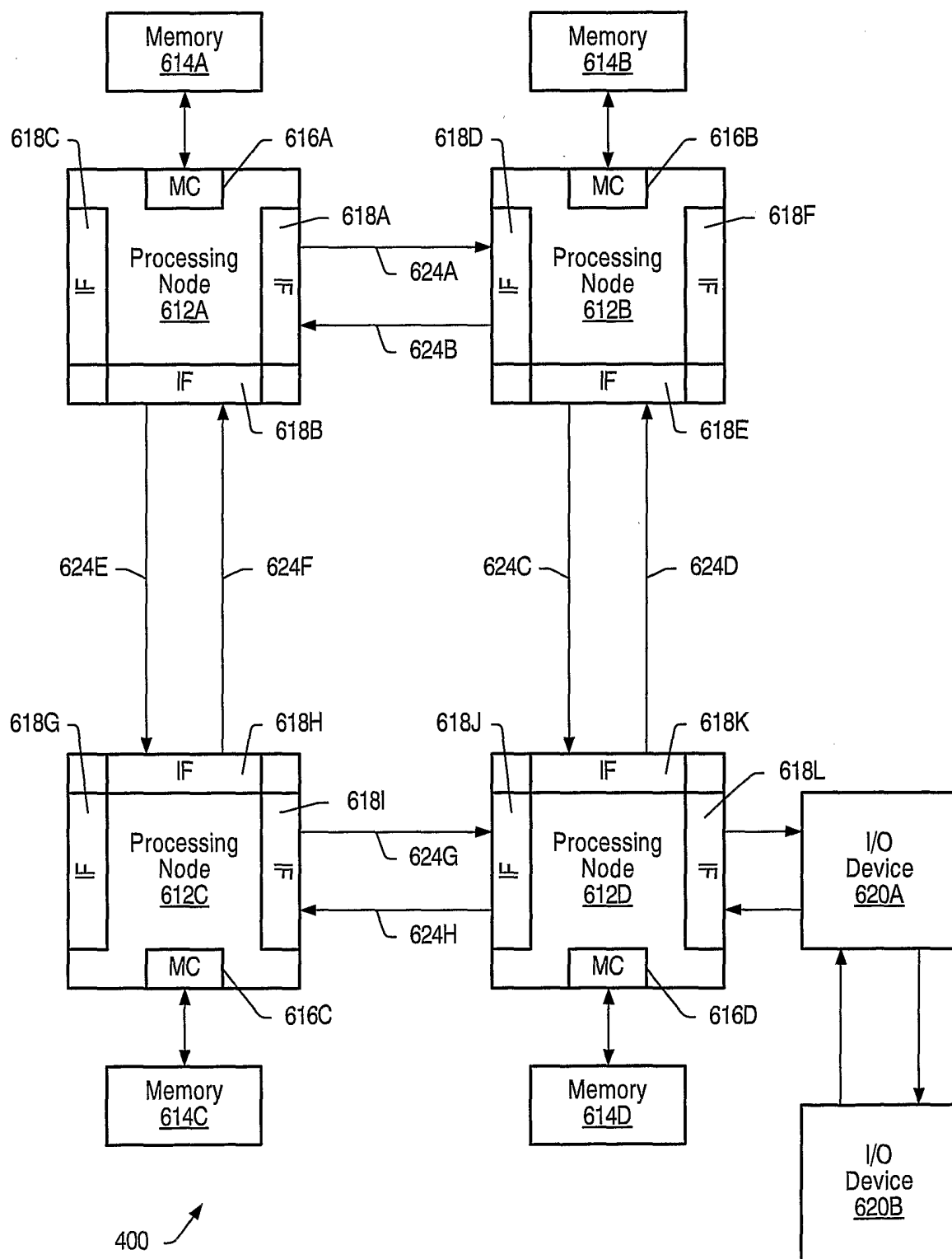


FIG. 15

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US2004/042683

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G06F9/38

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 03/093982 A (ADVANCED MICRO DEVICES, INC) 13 November 2003 (2003-11-13) page 2, line 1 - line 15; figures 1,2A,3 page 3, line 6 - page 7, line 24 page 9, line 1 - line 34 -----	1-10
X	US 6 065 103 A (TRAN ET AL) 16 May 2000 (2000-05-16) column 3, line 16 - line 35 column 9, line 4 - column 14, line 29; figures 2,3 column 16, line 8 - line 38; figure 7 -----	1-10
X	US 6 662 280 B1 (HUGHES WILLIAM A) 9 December 2003 (2003-12-09) column 2, line 17 - line 44 column 4, line 8 - column 7, line 47; figure 1 -----	1-10



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *G* document member of the same patent family

Date of the actual completion of the international search

28 February 2005

Date of mailing of the international search report

09/03/2005

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Thibaudeau, J

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US2004/042683

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 03093982	A	13-11-2003	US 2004177236 A1 AU 2002367915 A1 EP 1504340 A1 WO 03093982 A1	09-09-2004 17-11-2003 09-02-2005 13-11-2003
US 6065103	A	16-05-2000	NONE	
US 6662280	B1	09-12-2003	EP 1228426 A1 JP 2003514299 T TW 548548 B WO 0135212 A1	07-08-2002 15-04-2003 21-08-2003 17-05-2001