



- (51) International Patent Classification:
G06F 13/14 (2006.01) G06F 12/00 (2006.01)
- (21) International Application Number:
PCT/US2014/010275
- (22) International Filing Date:
6 January 2014 (06.01.2014)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
13/743,795 17 January 2013 (17.01.2013) US
- (71) Applicant: INTEL CORPORATION [US/US]; 2200 Mission College Boulevard, Santa Clara, California 95054 (US).
- (72) Inventors: CUTTER, Daniel F.; 9 Charles Street, Maynard, Massachusetts 01754 (US). FANNING, Blaise; 989 Smith Way, Folsom, California 95630 (US). NAGARAJAN, Ramadass; 10568 SW Washington St., Portland, Oregon 97225 (US). NIELL, Jose S.; 4 Pasture Way, Franklin, Massachusetts 02038 (US). BERNSTEIN, Debra; 321 Old Lancaster Road, Sudbury, Massachusetts 01776 (US). LIMAYE, Deepak; 11609 Larch Valley Drive, Austin, Texas 78754 (US). SCHOINAS, Ioannis T.; 1125 NW 12th Av #801, Portland, Oregon 97209 (US). IYER, Ravishankar; 15934 NW Andalusian Way, Portland, Oregon 97229 (US).
- (74) Agents: ROZMAN, Mark J. et al.; Trop, Pruner & Hu, P.C. 1616 S. Voss Rd., Ste. 750, Houston, Texas 77057-2631 (US).

- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to the identity of the inventor (Rule 4.17(i))

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(54) Title: ARBITRATING MEMORY ACCESSES VIA A SHARED MEMORY FABRIC

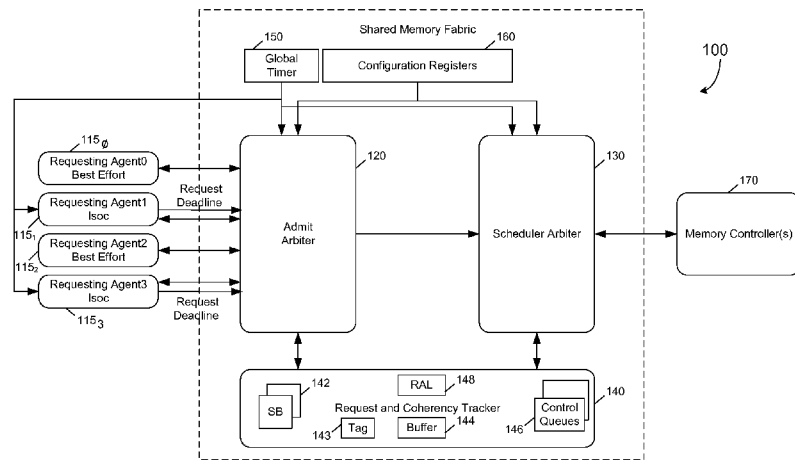


FIG. 1

(57) Abstract: In an embodiment, a shared memory fabric is configured to receive memory requests from multiple agents, where at least some of the requests have an associated deadline value to indicate a maximum latency prior to completion of the memory request. Responsive to the requests, the fabric is to arbitrate between the requests based at least in part on the deadline values. Other embodiments are described and claimed.

WO 2014/113225 A1

ARBITRATING MEMORY ACCESSES VIA A SHARED MEMORY FABRIC

Background

[0001] In many computer systems with multiple devices, an arbitration is performed to provide access to a shared resource such as a shared memory. Different types of arbitration mechanisms are provided to enable arbitration between the different agents or requestors. Some systems use a fixed priority arbitration system in which different agents are allocated a particular priority. However, this can lead to unfairness in usage and starvation of one or more agent's ability to obtain access to the shared resource. Other arbitration systems provide for a round robin-based approach to allocating access to the shared resource.

[0002] In either event, the arbitration that is performed does not take into account factors regarding the shared resource such as its power state. As such, requests granted access to the shared resource cause the resource to exit a low power state, although the device does not require such immediate access to the shared resource. Accordingly, current arbitration systems suffer from various drawbacks.

Brief Description of the Drawings

[0003] FIG. 1 is a block diagram of a portion of a shared memory fabric in accordance with an embodiment of the present invention.

[0004] FIG. 2 is a block diagram of further details of an admit arbiter in accordance with an embodiment of the present invention.

[0005] FIG. 3 is a flow diagram of a method for updating age values for an agent upon a determination of an arbitration winner in accordance with one embodiment of the present invention.

[0006] FIG. 4 is a block diagram of an admit arbiter state machine in accordance with an embodiment of the present invention.

[0007] FIG. 5 is a flow diagram of a method for performing first level arbitration in an admit arbiter in accordance with an embodiment of the present invention.

[0008] FIG. 6 is a block diagram of a portion of a resource allocation logic in accordance with an embodiment of the present invention.

[0009] FIG. 7 is a block diagram of a scoreboard index generation logic in accordance with an embodiment of the present invention.

[0010] FIG. 8 is a block diagram of a state machine for a scheduler arbiter in accordance with an embodiment of the present invention.

[0011] FIG. 9 is a flow diagram of a method for performing memory scheduling in accordance with an embodiment of the present invention.

[0012] FIG. 10 is a block diagram of an SoC in accordance with an embodiment of the present invention.

[0013] FIG. 11 is a block diagram of components present in a computer system in accordance with an embodiment of the present invention.

Detailed Description

[0014] In various embodiments, a shared memory fabric couples multiple independent devices, also referred to herein as agents, to a shared memory (e.g., via an intervening memory controller). In many embodiments, the shared memory fabric is an interconnect structure of a single die semiconductor device that includes intellectual property (IP) logic blocks of different types. The shared memory fabric is configured to enable compliance with quality of service (QoS) requirements for time critical isochronous devices while also providing memory bandwidth proportioning for non-isochronous devices, also referred to herein as best effort devices.

[0015] In this way, reliable and predictable allocation and scheduling of memory bandwidth occurs to support multiple devices and device types connected to the shared memory fabric. By including QoS functionality in a common shared memory fabric (rather than a memory controller or other non-fabric circuitry), the design may be more easily reused across multiple semiconductor devices such as system on chips (SoCs) since the design is independent of memory technology.

[0016] Embodiments thus perform resource allocation, bandwidth apportioning and time-aware QoS properties in a shared memory fabric to provide predictable and

reliable memory bandwidth and latencies to meet the requirements of the devices connected to the fabric.

[0017] A class of service category is assigned to each device coupled to the shared memory fabric. In an embodiment, this assignment can be identified using configuration registers of the fabric. Multiple classes of service may be supported by the fabric. Although the scope of the present invention is not limited in this regard, in one embodiment devices of two classes of service categories may be present, including an isochronous class of service category used for latency sensitive devices and a best effort class of service category used for devices that that can tolerate longer latencies to service their requests to memory. Although the scope of the present invention is not limited in this regard, in some embodiments latency sensitive devices include content rendering devices such as audio or video players, camera devices and so forth, while lower priority devices include processor cores, graphics processing units and so forth.

[0018] Time, in the form of a request deadline, is communicated from the isochronous devices to the fabric to indicate to the fabric the required latency to complete a request to memory. To enable synchronization, the fabric broadcasts a global timer to all isochronous requesting agents. This global timer is continuously driven on outputs from the fabric so that is available for sampling by the isochronous devices. Responsive to this time value, the agents determine a latency requirement for completion of a request, and add this latency value to the global timer value to form a deadline for the request. As an example, the latency for a read can be determined by the amount of data in the agent's data buffer and the drain rate of the buffer by the agent. If the agent consumes 1 cache line of data every 250 nanoseconds (ns) and has 8 cache lines of data in the buffer, the required deadline for a new request would $8 \times 250\text{ns}$ or 2 microseconds (us) before the buffer is empty. Based on this communicated latency or deadline value, the fabric may make better scheduling decisions based on knowledge of the current power state of the memories and the required latencies for other unscheduled memory requests pending in the fabric. This deadline communication may improve memory bandwidth and also save system power.

[0019] The use of request deadlines provides the fabric with latency information for each request from an isochronous device. Configuration registers programmed within the fabric provide the fabric with information about the memory configuration such as the latency required for the memories to exit a low power, e.g., self-refresh, state. The fabric also controls when the memory controller causes the attached memory to enter and exit the self-refresh state by sending an indication to the memory controller, e.g., in the form of a status channel. The fabric determines when the memories should enter and exit self-refresh by evaluating the latency requirements for all pending memory requests. Because the fabric has knowledge of the required latency for all pending memory requests and required latency to exit self-refresh, greater management of power state transitions of the memories may result in additional power savings versus the prior art.

[0020] Embodiments may also provide for efficiency in memory bandwidth by allowing memory requests to be scheduled out of order; however this may result in long scheduling latencies for some requests. To resolve such concern, the fabric assigns a priority level to each isochronous memory request, e.g., a high or low priority. When scheduling high priority isochronous requests the amount of out of order scheduling allowed is less than what is acceptable when scheduling best effort or low priority isochronous requests. Limiting the amount of out-of-order scheduling for high priority requests is done in order to ensure that the request latency requirement is met. Because request priority is determined from the deadline of the request the fabric can determine immediately after a request is scheduled what the priority levels of other pending requests are for an isochronous device. Using the deadline method the priority level of all pending requests only change when the global timer increments.

[0021] Embodiments may also improve portability and reuse of the sophisticated QoS memory scheduling algorithms across multiple SoC implementations, in that intelligent memory scheduling logic is incorporated in the fabric, while technology specific memory controller logic may be implemented within the memory controller.

[0022] Embodiments may also incorporate anti-starvation algorithms into multiple arbitration points of the fabric. In one embodiment, these anti-starvation algorithms

include a weighted age-based arbitration method used by an admit arbiter and an oldest of available scheduling queues used in a memory scheduler and request tracker. In addition, request weights may be used to switch between different priority levels at the arbitration points in the fabric and for switching from scheduling read requests to write requests, in contrast to a fixed priority arbitration in which requests from high priority isochronous devices always win.

[0023] In an embodiment, the shared memory fabric includes two arbitration points that are used for scheduling requests being sent to the memory controller. The first arbitration point is used to admit requests from the devices into the shared memory fabric and is referred to as an admit arbiter. The second arbitration point is used to schedule the requests sent to the memory controller from the shared memory fabric and is referred to as a scheduler arbiter.

[0024] Each device connected to the shared memory fabric has a request interface that is connected between the device and fabric. The request interface supplies information about the request that can be used for QoS memory scheduling. In an embodiment, this information includes a memory address, order ID field and an opcode field. For isochronous devices an additional field called a request deadline field is provided to indicate the required latency needed to complete the request. Note that in some implementations of SoCs the memory fabric interface may be connected to other fabrics or switches which allows multiple devices to share a common request interface.

[0025] Referring now to FIG. 1, shown is a block diagram of a portion of a shared memory fabric in accordance with an embodiment of the present invention. As shown in FIG. 1, a shared memory fabric 100 is coupled between a plurality of agents 115₀-115₃ (generically agent 115) and a memory controller 170. Note that in some embodiments more than one memory controller is present. Of course while not shown for ease of illustration understand that the memory controller is coupled to a system memory such as a dynamic random access memory (DRAM) or other system memory.

[0026] In the embodiment shown in FIG. 1, different types of agents are coupled to shared memory fabric 100. Specifically, the different agents include a first class of service (COS) agent type, namely so-called isochronous agents and a second class of service agent type, namely so-called best effort COS agents. As seen, each of the agents 115 may communicate request information to an admit arbiter 120. In turn, admit arbiter 120 can communicate corresponding control type information back to the agents. In addition, the isochronous agents (namely agents 115₁ and 115₃ in the embodiment of FIG. 1) further include an additional link to communicate request deadline information to admit arbiter 120. To this end, these agents are further configured to receive global timing information from a global timer 150, also coupled to both admit arbiter 120 and a scheduler arbiter 130.

[0027] In the embodiment of FIG. 1, admit arbiter 120 may be configured to receive incoming requests from agents 115 (and request deadline information from isochronous agents) and to select appropriate requests to admit to scheduler arbiter 130. To aid in its arbitration process, admit arbiter 120 receives configuration information from a set of configuration registers 160, further coupled to scheduler arbiter 130. In addition, a request and coherency tracker 140 may be coupled to arbiters 120 and 130. In general, tracker 140 may include multiple scoreboards 142, a data buffer 144 and corresponding address tag storage 145, control queues 146 and other resources such as various buffers, logic such as resource allocation logic 148, and so forth. In some implementations, the tag array and data buffer may be located elsewhere than the tracker. Although shown at this high level in the embodiment of FIG. 1, understand the scope of the present invention is not limited in this regard.

[0028] The shared memory fabric may include certain finite resources that are first allocated before a request from a requesting agent can be granted by the admit arbiter. These resources include available entries in the internal data buffer and address tag storage. Other finite resources include available entries in the memory scheduler and request tracker scoreboards. There is a one-to-one correspondence in resources for the fabric's internal data buffer, tag array and memory scheduler scoreboard. In an embodiment, these resources are allocated to a predetermined

region (e.g., a cache line width such as 64 bytes) of memory. Each active request is also allocated its own entry in the request and coherency tracker, but multiple requests to the same region in memory share the same entry in the data buffer, tag array and memory scheduler scoreboard. Although it is possible for more than 1 request to be allocated to the same data buffer, tag array, and scheduler scoreboard entry, only 1 read request is scheduled to the memory controller for all outstanding read requests in the request and coherency tracker.

[0029] The request interface for all devices connects to the admit arbiter of the fabric. Isochronous devices use the deadline field of the request bus to indicate to the fabric the required latency to complete the request. The fabric sends a global timer value to all isochronous devices that are attached to the fabric. For each request to be sent to the fabric, the isochronous device, e.g., in a deadline logic, determines the required latency needed for the request to complete and adds the value to the current value of the global timer in order to create the request deadline. Different methods may be used by different isochronous devices to determine the required latency for the request, but all isochronous devices indicate to the fabric the request latency using a deadline field of the request interface.

[0030] In an embodiment, the admit arbiter has two levels of priority. There is a high priority path in the arbiter that is used for urgent isochronous requests. A request is considered urgent if the requesting agent is configured as an isochronous agent and the deadline field of the request is less than a value stored in a configuration register specifying a threshold value, referred to an urgency threshold value. The admit arbiter also has a low priority path used for best effort requests and for isochronous requests that are not considered urgent. The final level of arbitration is done using a priority selector that selects between the winner of the high priority arbitration and the winner of the low priority arbitration.

[0031] In one embodiment, the admit arbiter final selector has two modes that can be selected using a configuration register. The first mode is a fixed priority mode in which, assuming at least one high priority request is present at the input of the admit arbiter, the selector chooses the winner of the high priority arbitration path before choosing the winner of the low priority arbitration path. The second mode of the final

selector is a weighted round robin mode in which the final selector switches between granting the high priority path to granting the low priority path after N number of high priority requests are granted. The selector then grants M number of low priority requests from the winner of the low priority path before switching back to granting requests from the high priority path. In an embodiment, the values for N and M are specified using configuration registers.

[0032] Referring now to FIG. 2, shown is a block diagram of further details of an admit arbiter in accordance with an embodiment of the present invention. As shown in FIG. 2, arbiter 120 receives incoming requests from the requesting agents. In this illustration, requesting agents 115₀ and 115₁ are non-isochronous or best effort agents, while agents 115₂ and 115₃ are isochronous agents. Note that the isochronous agents may include or be coupled to deadline determination logic 118 that is used to calculate required latency for requests. In an embodiment in which at least some of the agents are third party IP blocks, this logic can be implemented in wrapper or interface logic that couples the agent to the shared memory fabric.

[0033] In the embodiment shown, admit arbiter 120 includes a first age-based arbiter 122 and a second age-based arbiter 124, which correspond to low and high priority age-based arbiters, respectively. Thus as seen, requests from all of agents 115 are provided to first arbiter 122, while only requests from isochronous agents 115₂ and 115₃ are provided to second arbiter 124. To determine whether a particular request from one of the isochronous agents is of an urgent status, a pair of deadline checker logics 121₀ and 121_n are each coupled to receive requests from a corresponding one of these isochronous agents, as well as global timing information from global timer 150. Based on a comparison of the deadline information provided by the agent and the global timing information, an indication of an urgent status for a corresponding request can be provided to second arbiter 124.

[0034] In operation, arbiters 122 and 124 operate to select an arbitration winner from a set of incoming requests. In the embodiment shown, this determination is based in part on information from an age storage 126 that stores an age value for each of the agents. The corresponding winners from each of the arbiters may be coupled to a priority arbiter selector 125 that selects based on mode of operation a

corresponding request to provide to scheduler arbiter 130 (not shown for ease of illustration in FIG. 2). To this end, selector 125 may select a request for admission to the scheduler arbiter based at least in part on information in a priority storage 129. Although shown at this high level in the embodiment of FIG. 2, understand the scope of the present invention is not limited in this regard.

[0035] Weighted Age-Based Arbitration Details

[0036] The age-based algorithm implemented by the admit arbiter is such that the requesting agent which has waited the longest since last being granted by the arbiter will be given the highest priority level. Once an agent has received the highest priority level, the priority level for that agent will not change unless that agent has been granted by the arbiter. In this way, starvation issues commonly found using round robin arbitration may be avoided by ensuring that the priority level for a requesting agent can only increase in priority level until that requesting agent has been granted by the arbiter.

[0037] The admit arbiter also allows for agent weights to be assigned to all requesting agents. Weights are used to allocate a percentage of the request bandwidth for each requesting agent. In an embodiment, a weight value is specified for each agent via a value stored in an agent weight configuration register. The percentage of request bandwidth that is allocated to an agent is equal to the agent weight value divided by the sum of weights for all agents, in an embodiment.

[0038] Weighted Age-Based Algorithm

[0039] The admit arbiter weighted age-based algorithm is based on the relative age of when a requesting agent was last granted by the arbiter. For each requesting agent that connects to the admit arbiter, there is one age counter instantiated and one weight counter instantiated. Both the high priority and low priority arbitration paths in the admit arbiter share common age and weight counters for the agents connected to the admit arbiter. The updating of the requesting agent's age and weight registers is determined by the final selector (namely the priority arbiter selector 125) after choosing the final arbitration winner.

[0040] The age registers (e.g., of age storage 126) for all requesting agents are first initialized responsive to receiving a reset input to the admit arbiter. When reset asserts, the age registers are initialized to unique values in a range starting at 0 and ending at a value of N-1, where the value of N equals the number of request interfaces connected to the admit arbiter.

[0041] Prior to any requests being asserted by the requesting agents, the agent weight counters (e.g., of weight storage 128) are initialized from programmed values in the agent weight configuration registers of the fabric. Once the weight counters are initialized the counter for an agent decrements by one for each request granted for that agent. Once an agent's weight counter reaches zero and if the agent is granted again by the admit arbiter, the counter is reloaded with the value programmed in the configuration register for that agent's weight.

[0042] In one embodiment, the age-based arbitration method performed in first and second arbiters 122 and 124 uses a request bit vector (each arbiter having its own vector) to determine the winner of the arbitration. When a request is asserted for an agent the arbiter uses the age value for the requesting agent as the priority level of the request. The priority levels for the arbiter and thus the range of the bit vector width is from 0 to N-1. The age-based algorithm guarantees that the age values for all requesting agents are always unique and therefore there is always only one winner per arbitration.

[0043] The arbiter updates the age registers for all agents when the weight counter for the winner of the request arbitration has reached zero. In one embodiment, the age registers for all agents are updated according to the following rules that guarantee the age values for the agents are always a unique value.

[0044] Rule 1: when the agent's age equals the age of the winner of the arbitration, the age register for that agent is set to zero to indicate youngest request age or lowest priority.

[0045] Rule 2: when the agent's age is less than the winner of the arbitration, the agent's age register is incremented by 1.

[0046] Rule 3: when the agent's age is greater than the winner of the arbitration, the agent's age register does not change.

[0047] Referring now to FIG. 3, shown is a flow diagram of a method for updating age values for an agent upon a determination of an arbitration winner, which may be performed to update age values when the winner's weight value equals zero. As seen, method 200, which may be performed by the priority arbiter selector, begins by determining whether the age value of an agent equals the winner value (diamond 210). If so, control passes to block 215 where the age value for this winning agent can be updated to the lowest priority level, which in an embodiment may be equal to zero. From both of block 215 and diamond 210, control passes to diamond 220 where it can be determined whether the age value is less than the winner value (namely corresponding to the age of the agent). If so, control passes to block 225 where the agent's age value can be updated, e.g., incremented. If none of these conditions occur, the agent's age is greater than the winner of the arbitration, and as such the age value for this particular agent does not change. Note that method 200 can be performed for each agent at the conclusion of each arbitration round when a winner is selected. Although shown at this high level in the embodiment of FIG. 3, understand the scope of the present invention is not limited in this regard.

[0048] Referring now to FIG. 4, shown is a block diagram of an admit arbiter state machine in accordance with an embodiment of the present invention. As shown in FIG. 4, state machine 250, which may be present within admit arbiter 120 of FIG. 1, first enters into an initialization (INIT) state 255 from a reset assertion. From this state, control passes into an active state 260 in which it remains so long as no requests are received. When a request is received and a granted agent has a weight of zero, control passes to an update age state 270 in which age storages are updated and a weight counter for an arbitration winner is reloaded to a predetermined value, e.g., obtained from a configuration register. Control then passes to one of active state 260, decrement agent weight state 280 or remains at update age state 270, depending upon whether an additional request is present and a value of the granted agent's weight. Similarly at decrement agent weight state 280 a winner arbitration weight counter is decremented. But here no weight counter

reloads are performed. Although shown at this high level in FIG. 4, understand the scope of the present invention is not limited in this regard.

[0049] Table 1 provides a description of each state of an admit architecture in accordance with one embodiment of the present invention.

Table 1

State	Description
Init	Reset is asserted: Agent Weights reloaded to values in configuration registers Agent Age registers set to unique Agent ID values
Active	No Agent Requests: Agent Age and Weight registers retain same state
Decrement Agent Weights	Requests asserted from 1 or more Agent. Winner of arbitration weight counter is non-zero. Weight counter of Winner is decremented
Update Age	Requests asserted from 1 or more Agent. Winner of arbitration weight counter is zero. Agent Age registers updated Weight counters for winner of arbitration reload to value in configuration registers

[0050] Referring now to FIG. 5, shown is a flow diagram of a method for performing first level arbitration in an admit arbiter in accordance with an embodiment of the present invention. As shown in FIG. 5, method 300 may be performed within the admit arbiter both for purposes of performing arbitration amongst incoming memory requests, as well as updating various age and weight values based upon an arbitration. As seen in FIG. 5, method 300 may begin by receiving a memory request from a device coupled to the fabric (block 310). More specifically to illustrate operation with regard to deadline-based requests from a latency sensitive device, assume that this memory request includes or is associated with a deadline value and is thus provided from an isochronous or latency sensitive device. As one such example this latency sensitive device is a media player. As seen, control passes to diamond 315 where it can be determined whether the deadline value is greater than a latency threshold. In an embodiment, this latency threshold is a minimum latency from the time a request is received until it is completed (e.g., by provision of requested data back to the requesting device provision of a write completion for a

write request). Note that the deadline value is in one embodiment a maximum latency that the requesting device can tolerate for handling the memory request.

[0051] Still referring to FIG. 5, if it is determined that the deadline value is greater than the latency threshold, control passes to block 320_b where the memory request is forwarded to a low priority arbiter. Otherwise control passes to block 320_a where the memory request is forwarded to a high priority arbiter.

[0052] Still referring to FIG. 5, note the presence of parallel paths such that at block 325 (blocks 325_a and 325_b) an arbitration is performed in the corresponding arbiter that is based on a bit vector associated with the age values for the devices that provide requests to the corresponding arbiter. Next at block 330 (blocks 330_a and 330_b), the winning memory requests are forwarded to a final arbiter. Then at block 335 a final arbitration is performed to select the winner memory request. Depending upon a mode of configuration for this final arbiter, the winner request can be selected from the high priority arbiter only, or a weighting between high priority and low priority paths may occur. Thus at this point the winning memory request is forwarded to a memory scheduler scoreboard where it can be stored in an entry to thus enable arbitration in the memory scheduler arbiter to consider this memory request.

[0053] Still referring to FIG. 5, various updating operations may further be performed responsive to selection of a winner by the final arbiter. Specifically, at diamond 340 it can be determined whether the weight value of the winner agent equals zero. If so, control passes to block 345 where this weight value can be updated to its configured value, e.g., stored in a configuration register of the shared memory fabric. Control next passes to block 350 where the age values for all agents can be updated (block 350). To this end all non-winning agents may have their age value incremented, while the winning agent may have its age value set to a lowest priority value, e.g., zero. If instead at diamond 340 it is determined that the weight value of the winner agent is not zero, control passes to block 355 where the weight value of the winner agent is decremented. Although shown at this high level in the embodiment of FIG. 5, understand the scope of the present invention is not limited in this regard.

[0054] Shared Memory Fabric Shared Resource Allocation

[0055] The memory fabric includes logic to allow for fair allocation of the shared resources within the fabric, e.g., resource allocation logic 148 of FIG. 1. In one embodiment, these shared resources are the fabric's internal data buffer, address tag storage and request tracker scoreboards. Since there are no dedicated resources for any of the requesting agents, mechanisms may limit the number of outstanding requests that are pending in the fabric for each of the agents, while also allowing entries to be reserved for an agent, e.g., by reserving virtual entries in these shared resources. The fabric allows for the specification of agent limits to prevent any one requesting agent from using up all the available shared resources of the fabric.

[0056] A portion of the memory scheduling algorithm deals with minimizing the performance impact of read-to-write turnaround times for memory technologies. In order to minimize the number of times the memory scheduler switches between scheduling read requests to scheduling write requests, a flush pool is used for queuing write requests. The flush pool allows write requests targeting memory to be accumulated in the memory fabric until enough write requests have been received to allow the fabric's memory scheduler to send the write requests to the memory controller as a burst of back-to-back requests. In order to prevent all available resource in the fabric to be used up by the flush pool, a flush limit can be specified. When specified, the flush limit causes the fabric to block new write requests from all agents at the admit arbiter until the number of entries in the flush pool is less than the value programmed for the flush pool.

[0057] Memory Fabric Flush Pool for Write Requests

[0058] When a write request is received from a requesting agent, the fabric transfers the write data from the requesting agent to an internal data buffer. Once the new data is written to the fabric's internal data buffer and the request is retired from the agent's point of view, the buffer entry is considered to be in the "flush pool". For coherent memory traffic the fabric may receive snooped requests from the requesting agents. Snooped requests can be either read or write requests to

memory. When the fabric receives a snooped read or write request from a requesting agent, it sends a snoop request to all caching agents coupled to the fabric. The caching agents will respond to a snooped request that hits in their cache and will return the write back (WB) data for a cache line that has been modified by the caching agent. The WB data is then written into the fabric's internal data buffer and is then considered to be included in the flush pool of write requests targeting memory. When the number of entries in the flush pool reaches the value programmed for the flush limit, new write requests, e.g., as determined by decoding of the request opcode field, are blocked at the admit arbiter.

[0059] Memory Fabric Reservations and Limits

[0060] The memory fabric allows reservations to be specified for any agent using agent reservation configuration registers. Using these configuration registers the user can specify the number of entries in the memory fabric to reserve for each agent. The reserved entries for an agent are the first entries allocated to the agent and the last entries to be retired for the agent. In order to determine if an agent's reserved entries are being allocated or retired, each agent has a request counter that is compared against the value specified in the configuration register. If the value in the request counter is less than or equal to the value in the configuration register, the agent's reserved entries are being used.

[0061] The mechanism used to provide agents with reserved entries varies the full threshold limit as reserved entries are allocated or freed for requesting agents. Initially, the full threshold for all agents is calculated by subtracting the total number of reserved entries for all agents (e.g., as specified by configuration registers) from the total number of entries in the scoreboards. As reserved entries are allocated to an agent, an accumulator is used to adjust the full threshold based on the total number of reserved entries that have been used. Agents that have used their reserved entries or do not have reserved entries specified are blocked when the total number of pending requests in the memory fabric reaches this adjusted full threshold. Agents that have not used their reserved entries are not blocked by the admit arbiter until they have used all their reserved entries and the total number of pending requests reaches the adjusted full threshold limit.

[0062] Agent limits may also be specified in configuration registers of the memory fabric. These agent limits may be disabled by setting the request limit for an agent to zero, in an embodiment. When agent limits are disabled any agent may be allocated all existing entries of the request tracker. In order to prevent a single agent from using all request tracker entries, a request limit can be specified for the agent. When the agent's request counter reaches the request limit specified for the agent the request input to the admit arbiter for that agent is disabled. When the request tracker retires requests for the agent and the agent's request counter becomes less than the agent's request limit, the request input to the admit arbiter for that agent is enabled.

[0063] Referring now to FIG. 6, shown is a block diagram of a portion of a resource allocation logic in accordance with an embodiment of the present invention. As shown in FIG. 6, logic 360 may be used to control allocation of various resources shared between all of the agents. As seen, an adder 368 determines a total number of reserved entries based on agent reserve values received from a configuration storage 365. From this total reserve entry value, a number of tag entries is subtracted at subtracter 370. The resulting value is provided through a flop 372 to an adder 375 which combines this value with a number of reserved entries used, received from a flop 374 that is alternately incremented and decremented based on increment and decrement reserve count values, described further below.

[0064] As such, the sum generated by adder 375 corresponds to an adjusted full threshold value that is provided to one input of a comparator 382 that further receives a number of allocated tag entries from a flop 376. If it is determined that the adjusted full threshold value is less than or equal to this number of allocated tag entries, a full flag is generated and used to mask requests of agents that have no reserve entries or have used their reserve entries.

[0065] As further seen, another comparator 380 is configured to receive a given requestor's reserve configuration value and a request counter value for that requestor (from a flop 378). The comparator thus generates an indication as to whether that requester has any free reserved entries, which is provided as an input to a pair of AND gates 384 and 385 that further receive indications of a channel grant

and a retirement of an entry for that channel. As such, these AND gates thus generate, respectively the increment and decrement values for the corresponding requestor. Similar logic and operations are performed for the other requestors, with all increment and decrement reserve values being provided to corresponding OR gates 386 and 387 that respectively generate the increment reserve count value and the decrement reserve count value.

[0066] Finally, the request counter value for a requestor is provided to another comparator 390 along with a configured limit value for that requestor to thus determine whether this requestor has reached its limit. If so, an indication of this limit is used to mask off the requests from this agent for further arbitration. Although shown at this high level in the embodiment of FIG. 6, understand the scope of the present invention is not limited in this regard.

[0067] Shared Memory Fabric Scheduler Arbitration Details

[0068] Embodiments may incorporate multiple scheduling algorithms to enhance reuse across multiple SoCs that support different memory technologies. The fabric's memory scheduler logic contains advanced QoS scheduling algorithms, and is also optimized to minimize performance bottlenecks that are commonly found in most memory technologies. The typical performance bottlenecks that occur using, e.g., DRAM, memories include entering and exiting of low power memory states, read-to-write turnaround times, consecutive memory accesses to the same DRAM bank but to different rows of memory, and consecutive memory accesses to different DRAM memory ranks. By including complex out of order scheduling algorithms in the shared memory fabric's scheduling logic, the fabric can be adapted to many different SoCs by attaching simplified technology-specific constraint solvers to the fabric to support their unique requirements for memory technologies or configurations.

[0069] In addition to improving the portability of the memory scheduling logic, embodiments also provide predictability of memory request latency in that the combination of advanced out-of-order scheduling algorithm with QoS scheduling logic results in improved predictability of the maximum request latency, in that the memory controller has much less flexibility to reorder memory requests.

[0070] Once a request is granted by the admit arbiter, it is enqueued into the scheduler scoreboard. The scheduler scoreboard stores information about the request that it uses to forward the request to the memory controller in order to perform a read or write to memory. In one embodiment, the information includes request address, request length, command type (read or write), class of service category, memory channel, memory bank, memory rank, and page hit/miss status.

[0071] Memory Scheduler Oldest of Available Queue

[0072] Embodiments provide for out of order page aware scheduling that is based on a history of requests sent to the memory controller, although the fabric has no direct knowledge of the true state of the memory bank. More specifically, the fabric's memory scheduler uses the scheduler scoreboard as a history buffer of requests that have been sent to memory. Because the scheduler scoreboard is used to reflect the history of requests, it seeks to retain the status information for a request in the scoreboard as long as possible. The memory scheduler uses a structure called the oldest of available queue to determine the oldest scoreboard entry that is available to be reallocated.

[0073] The oldest of available queue is also used by the memory scheduler to avoid starvation issues that can arise due to the out-of-order scheduling of the requests to memory. The fabric's memory scheduler uses the oldest of available queue to determine how many requests of the same class of service category and type, read or write, have bypassed the oldest pending request to memory. Once the number of requests that have bypassed the oldest request reaches a preprogrammed limit (e.g., set by software) the fabric's memory scheduler disables out-of-order scheduling of requests and grants the oldest pending request.

[0074] As mentioned above, the scheduler keeps track of the relative age of all requests in its scoreboard using the oldest of available queue. When a request targeting a new memory address is granted by the admit arbiter an index pointer into the scheduler scoreboard is enqueued into the tail entry of the oldest of available queue which is then considered to be the newest request. When all pending requests have completed transferring data to/from the requesting agents and to/from

the memory controllers, a scoreboard entry is available to be reallocated and can be reallocated for a new request granted by the admit arbiter. Due to the out-of-order scheduling, the oldest entry in the oldest of available queue may not always be available for reallocation.

[0075] To select the scoreboard entry to be reallocated to a new request, the scheduler detects whether all outstanding requests to a scoreboard entry have completed. In one embodiment, the scheduler uses a request bit vector having a length equal to the number of scoreboard entries to indicate which entries are available for reallocation. A bit set to 1 in the request bit vector indicates the entry corresponding to that bit position is available for reallocation. The request bit vector is then sent to the oldest of available queue. The oldest of available queue uses the indexes stored in the queue to select the bit in the request vector corresponding to the request for that entry of the queue. Each entry of the queue is associated with a unique bit in the request vector and a "find first" function is performed starting from the oldest entry in the queue to determine the oldest available request to be reallocated. After determining the oldest available entry to be reallocated, the scoreboard index for that entry is output from the oldest of available queue.

[0076] Referring now to FIG. 7, shown is a block diagram of a scoreboard index generation logic in accordance with an embodiment of the present invention. As shown in FIG. 7, logic 400 includes a plurality of flip flops 410₀-410_n coupled in a serial configuration to store a corresponding scoreboard index. As seen, flip-flops 410 are configured to receive a scoreboard index corresponding to an index pointer into a scoreboard of the scheduler which is also the index to the tag array and data buffer. Flip-flops 410 may be configured in an order from newest (namely flip flop 410₀) to an oldest (namely flip flop 410_n). In an embodiment, each flip flop may be a D-type flip-flop, although the scope of the present invention is not limited in this regard, and further understand that other storage elements may be used.

[0077] As seen, an output of each flip-flops 410 is coupled to one of a corresponding plurality of multiplexer 420₀-420_n, each of which is further configured to receive a bit of a scoreboard request vector. As such, this bit vector provides an indication, e.g., via a set bit to indicate that a corresponding scoreboard entry is

available for reallocation. Using the outputs from multiplexers 420_0 - 420_n , a grant signal can be generated either directly from the comparator output (as from comparator 420_n) or via a corresponding one of logic gates 430_0 - 430_n (which in the embodiment shown are configured as AND gates having a first input received from a corresponding multiplexer 420 and a second input corresponding to an inverted output of a corresponding OR gate 425_0 - 425_{n-2}). In this way only a single one of the grant signals may be active at a time.

[0078] As further seen in FIG. 7, the grant output signals may be coupled to a corresponding one of a plurality of AND gates 435_0 - 435_n also configured to receive an incoming index signal. In turn the outputs from AND gates 435_n may be coupled to an OR gate 440 to thus output a scoreboard index corresponding to the oldest available entry such that a 1-hot multiplexer function is performed that thus provides a one hot multiplexing of the scoreboard index of the granted request.

[0079] Shared Memory Fabric Memory Scheduling Details

[0080] The fabric memory scheduler contains three state machines that work together to schedule requests sent to the memory controller.

[0081] Referring now to FIG. 8, shown is a block diagram of a state machine for a scheduler arbiter in accordance with an embodiment of the present invention. As shown in FIG. 8, state machine 500 , which may be performed in hardware, software and/or firmware such as scheduler arbiter 130 of FIG. 1, may begin by entering into an initialization state INIT upon reset of the system. Control next passes into a self-refresh state machine 510 that includes an enter self-refresh state 512 , a request self-refresh state 513 , and an exit self-refresh state 516 .

[0082] As seen in FIG. 8 from exit self-refresh state 516 , control passes into a read/write grant state machine 520 that in turn includes a grant read request state 522 and a grant write request state 524 . From these states control in turn passes into a read state machine 530 that includes a plurality of states, namely a bypass grant state 532 , a high priority read request grant state 534 , a best effort grant read request state 536 , and a low priority isochronous grant read request state 538 . Although shown at this high level with particular conditions causing exit and entry

into the various states (as described further below in Table 2 and 4), understand the scope of the present invention is not limited in this regard and various modifications and alternatives are possible.

[0083] Self-Refresh State Machine

[0084] Embodiments may control when the memories are allowed to enter and exit the low power memory state, also referred to as the self-refresh state. The self-refresh state machine is responsible for controlling when to send an indication to the memory controller to enter or exit self-refresh. For best effort read requests, the self-refresh state machine transitions immediately to the exit self-refresh state. For isochronous read requests, the memory scheduler checks the request deadline to determine if it is to exit self-refresh in order to satisfy the required read latency for the request. To determine if exiting self-refresh is required for meeting the isochronous read requirement, the memory scheduler subtracts the deadline of the request from the current value of the global timer. The result of the subtraction is checked against a configuration register in the fabric that is programmed to reflect the worst case latency needed for the memory controller to exit self-refresh and the fabric to return data to the request agent.

[0085] For write requests, the fabric counts the number of dirty entries in the flush pool and checks the result against a programmable threshold value, termed the flush high water mark. If the number of dirty entries exceeds the value of the flush high water mark, the self-refresh state machine passes control to the exit self-refresh state. In addition, the fabric checks for read/write conflicts to the same tag address in which the request is blocked by the admit arbiter. When the fabric determines that a request is blocked by an address conflict, agent limit or if the request tracker or memory scheduler scoreboards are full, control passes from the self-refresh state machine to the exit self-refresh state. The fabric also contains a configuration register that can be programmed to disable entering self-refresh, in an embodiment.

[0086] When the memory scheduler sends an indication to the memory controller to exit self-refresh, requests may begin to be sent to the memory controller. The memory scheduler continues to send an indication to the memory controller to

remain out of self-refresh while it is actively sending memory requests to the memory controller. When the memory scheduler completes sending all read requests to the memory controller and the number of write requests in the flush pool is below the casual high water mark limit, the memory scheduler transitions to the request self-refresh state.

[0087] In the request self-refresh state if no new requests are granted by the admit arbiter the state machine transitions to the enter self-refresh state after a programmable delay value called the enter self-refresh delay is met. In an embodiment, this delay is programmed in configuration registers in the fabric. If new requests are granted by the admit arbiter, the self-refresh state machine may transition to the exit self-refresh state under certain conditions. If a new best effort read request is received or if a write request is received that results in the number of entries in the flush pool exceeding the number programmed in the flush high water mark configuration register, the self-refresh state machine transitions from the request self-refresh state back to the exit self-refresh state. If an isochronous read request is received when the state machine is in the request self-refresh state, the deadline value of the request is checked against a programmed value called the enter self-refresh threshold. If the deadline latency is greater than the enter self-refresh threshold, the state machine continues in request self-refresh state if the deadline latency for a request is below the enter self-refresh threshold the state machine will transition to the exit self-refresh state.

[0088] The self-refresh state machine drives status to the memory controller to remain out of self-refresh until the state machine transitions to the enter self-refresh state. Once in the enter self-refresh state the state machine sends an indication to the memory controller to enter self-refresh.

[0089] Table 2 below is a description of a self-refresh state machine in accordance with an embodiment of the present invention.

TABLE 2

Current State	Condition	Description	Next State	Outputs
Unknown	Reset	Reset Pin Asserted	Enter Self Refresh	Fabric Drives indication to memory controller to Enter Self Refresh
Enter Self Refresh	Memory Scheduler Idle	Number of Flush Entries less than Flush HWM and No Best Effort Read Requests and No ISOC read requests with deadline times less than Exit Self Refresh Threshold	Enter Self Refresh	Fabric Drives indication to memory controller to Enter Self Refresh
Enter Self Refresh	Exit Self Refresh 1	Number of Flush Entries greater than Flush HWM or Best Effort Read Requests or ISOC read requests with deadline times less than Exit Self Refresh Threshold or ISOC read request blocked by Agent Limit or Fabric Scoreboard full indications	Exit Self Refresh	Fabric Drives indication to memory controller to Exit Self Refresh
Exit Self Refresh	Memory Scheduler Active	Isochronous or Best Effort Read requests pending or number of Flush Pool entries above Casual HWM	Exit Self Refresh	Fabric Drives indication to memory controller to Exit Self Refresh
Exit Self Refresh	Request Self Refresh	No Isochronous or Best Effort Read requests pending and number of Flush Pool entries is below Casual HWM	Request Self Refresh	Fabric Drives indication to memory controller to Exit Self Refresh
Request Self Refresh	Request Self Refresh	No Best Effort Read requests received and number of Flush Pool entries is below Flush HWM and Enter Self Refresh Timer less than Enter Self Refresh Delay value	Request Self Refresh	Fabric Drives indication to memory controller to Exit Self Refresh

Request Self Refresh	Exit Self Refresh 2	Received Isochronous Read Request with deadline less than Enter Self Refresh Threshold or Received Best Effort Read request spending or number of Flush Pool entries is now above Flush HWM	Exit Self Refresh	Fabric Drives indication to memory controller to Exit Self Refresh
Request Self Refresh	Request Self Refresh	No Best Effort Read requests received and number of Flush Pool entries is below Flush HWM and Enter Self Refresh Timer greater than Enter Self Refresh Delay value	Enter Self Refresh	Fabric Drives indication to memory controller to Enter Self Refresh

[0090] Read/Write Grant State Machine

[0091] In an embodiment, the memory scheduler uses configurable threshold values to specify when to start and stop transferring a burst of write requests to the memory controller. The memory scheduler may perform different types of transfers of write data to memory, e.g., a high priority transfer and a low priority transfer, also termed herein as a high priority flush of write requests and casual flush of write requests to memory, respectively. When the number of entries in the flush pool reaches or exceeds a threshold value (the flush high water mark), the memory scheduler begins scheduling a high priority write flush to memory and begins sending write requests to the memory controller. The memory scheduler continues to schedule write requests using the high priority flush mechanism until the number of entries in the flush pool reaches or is less than a threshold value (the flush low water mark).

[0092] A casual flush may also be performed by the fabric memory scheduler. A casual flush is triggered when the memory scheduler has completed sending all read requests to the memory controller and the number of entries in the flush pool exceeds a threshold value (the casual flush limit. In an embodiment, the casual flush limit can be typically set lower than the high water mark, but greater than or equal to the low water mark, for performance reasons. In some cases this casual flush limit can be set to 0 to flush all write data to memory. Once the last read request is sent

to the memory controller, if the number of entries in the flush pool is above the casual flush limit, a counter called the casual flush timer starts incrementing every clock cycle. If no new read requests to memory are received by the fabric and the casual flush timer reaches the value specified by the casual flush delay, which is a threshold stored in a configuration register, the memory scheduler begins sending write requests to the memory controller. This casual flush continues until the number of entries in the flush pool is less than the casual flush limit or until a new read request is received by the fabric.

[0093] The read/write grant state machine is responsible for switching from granting read requests to granting write requests. In an embodiment, the memory scheduler is configurable to allow write requests to have priority over read requests or to use weights when switching between read requests and write requests (in order to prevent starvation of reads when the system is saturated by write requests). When weights are enabled, the memory fabric uses configuration registers to specify the read and write weights independently.

[0094] Table 3 below is a description of a read/write grant state machine in accordance with an embodiment of the present invention.

TABLE 3

Current State	Condition	Description	Next State	Outputs
Unknown	Reset	Reset Pin Asserted	Grant Read Requests	Memory Scheduler sends Read Requests to Memory controller
Grant Read Request	Grant Read Request	Number of Flush Entries less than Flush HWM and Read/Write weights disabled OR Number of Flush Entries is greater than HWM and Read/Write weights enabled and read weight count is greater than 0	Grant Read Request	Memory Scheduler sends Read Requests to Memory controller

Grant Read Request	Grant Write Request	Number of Flush Entries greater than Flush HWM and Read/Write weights disabled OR Number of Flush Entries is greater than HWM and Read/Write weights enabled and read weight count is equal to 0 OR No read requests pending and Number of Flush Entries is greater than Casual HWM and Casual Timer has expired	Grant Write Request	Memory Scheduler sends Write Requests to Memory controller
Grant Write Request	Grant Write Request	Number of Flush Entries greater than Flush HWM and Read/Write weights disabled OR Number of Flush Entries is greater than LWM and Read/Write weights enabled and write weight count is greater than 0	Grant Write Request	Memory Scheduler sends Write Requests to Memory controller
Grant Write Request	Grant Read Request	Pending Read Requests and Number of Flush Entries less than Flush LWM OR Pending Read Requests and Number of Flush Entries is greater than LWM and Read/Write weights enabled and write weight count is equal 0	Grant Read Request	Memory Scheduler sends Read Requests to Memory controller

[0095] Read State Machine

[0096] The read state machine is responsible for switching between high priority isochronous read requests, best effort read requests and low priority isochronous read requests. The read state machine can be configured to operate in one of

multiple modes. In one embodiment, two such modes are provided. A first mode is a fixed priority mode where the read state machine gives high priority isochronous reads highest priority, best effort read requests medium priority, and low priority isochronous read requests receive the lowest priority. A second mode is to enable the use of weights for switching between high priority isochronous reads and best effort read requests. In this mode, low priority isochronous requests are only granted when there are no longer any high priority isochronous or best effort read requests.

[0097] Table 4 is a description of a read state machine in accordance with an embodiment of the present invention.

TABLE 4

Current State	Condition	Description	Next State	Outputs
Unknown	Reset	Reset Pin Asserted	Bypass Grant	Enable Bypass path from output of Admit Arbiter to Memory controller
Bypass Grant	No Read Request	No Read Requests Pending In Scheduler	Bypass Grant	Enable Bypass path from output of Admit Arbiter to Memory controller
Bypass Grant	High Priority ISOC Requests	Out of Self Refresh and High Priority ISOC Requests Pending	Grant High Priority ISOC Requests	Memory Scheduler Sends High Priority Read requests to Memory controller
Bypass Grant	Best Effort Requests	Out of Self Refresh and No High Priority ISOC Requests and Best Effort Requests pending	Grant Best Effort Requests	Memory Scheduler Sends Best Effort Read requests to Memory controller
Bypass Grant	Low Priority ISOC Requests	Out of Self Refresh and No High Priority ISOC Requests and No Best Effort Requests and Low Priority ISOC Requests Pending	Gran Low Priority ISOC Requests	Memory Scheduler Sends Low Priority Read requests to Memory controller

Grant High Priority ISOC Requests	High Priority ISOC Requests	Out of Self Refresh and High Priority ISOC Requests Pending and ISOC Weights not equal 0	Grant High Priority ISOC Requests	Memory Scheduler Sends High Priority Read requests to Memory controller
Grant High Priority ISOC Requests	Best Effort Requests	Out of Self Refresh and No High Priority ISOC Requests Pending and ISOC Weights equal 0 and Best Effort Requests pending	Grant Best Effort Requests	Memory Scheduler Sends Best Effort Read requests to Memory controller
Grant High Priority ISOC Requests	Low Priority ISOC Requests	Out of Self Refresh and No High Priority ISOC Requests and No Best Effort Requests and Low Priority ISOC Requests Pending	Grant Low Priority ISOC Requests	Memory Scheduler Sends Low Priority Read requests to Memory controller
Grant High Priority ISOC Requests	No Read Requests Pending	Out of Self Refresh and No High Priority ISOC Requests and No Best Effort Requests and No Low Priority ISOC Requests	Bypass Grant	Enable Bypass path from output of Admit Arbiter to Memory controller
Grant Best Effort Requests	Best Effort Requests	Out of Self Refresh and No High Priority ISOC Requests or ISOC Weights equal 0 and Best Effort Requests Pending	Grant Best Effort Requests	Memory Scheduler Sends Best Effort Read requests to Memory controller
Grant Best Effort Requests	High Priority ISOC Requests	Out of Self Refresh and High Priority ISOC Requests Pending and ISOC Weights not equal 0 or BE weights equal 0	Grant High Priority ISOC Requests	Memory Scheduler Sends High Priority Read requests to Memory controller
Grant Best Effort Requests	Low Priority ISOC Requests	Out of Self Refresh and No High Priority ISOC Requests and No Best Effort Requests and Low Priority ISOC Requests Pending	Grant Low Priority ISOC Requests	Memory Scheduler Sends Low Priority Read requests to Memory controller

Grant Best Effort Requests	No Read Requests Pending	Out of Self Refresh and No High Priority ISOC Requests and No Best Effort Requests and No Low Priority ISOC Requests	Bypass Grant	Enable Bypass path from output of Admit Arbiter to Memory controller
Grant Low Priority ISOC Requests	High Priority ISOC Requests	Out of Self Refresh and High Priority ISOC Requests Pending	Grant High Priority ISOC Requests	Memory Scheduler Sends High Priority Read requests to Memory controller
Grant Low Priority ISOC Requests	Best Effort Requests	Out of Self Refresh and No High Priority ISOC Requests and Best Effort Requests pending	Grant Best Effort Requests	Memory Scheduler Sends Best Effort Read requests to Memory controller
Grant Low Priority ISOC Requests	Low Priority ISOC Requests	Out of Self Refresh and No High Priority ISOC Requests and No Best Effort Requests and Low Priority ISOC Requests Pending	Grant Low Priority ISOC Requests	Memory Scheduler Sends Low Priority Read requests to Memory controller
Grant Low Priority ISOC Requests	No Read Requests Pending	Out of Self Refresh and No High Priority ISOC Requests and No Best Effort Requests and No Low Priority ISOC Requests	Bypass Grant	Enable Bypass path from output of Admit Arbiter to Memory controller

[0098] Scheduler Agent Weights

[0099] The memory scheduler uses agent weights for proportioning memory bandwidth between agents within the same class of service category. In an embodiment, configuration registers specify the weight value for each requesting agent, and a weight counter is provided for each agent. The agent weight configuration registers are common between the admit arbiter and the memory scheduler.

[0100] When there are no requests pending in the memory scheduler for any of the agents connected to the fabric, the agent weight counters are loaded with values specified in the agent weight configuration registers. When requests are granted by the admit arbiter and enqueued into the memory scheduler scoreboard, an agent ID field is stored in the memory scheduler scoreboard along with the request information. When the memory scheduler grants a request in its scoreboard, the agent ID field is used to determine the source of the request and the weight counter for that agent is decremented by 1. Once an agent's weight counter has reached zero, the remaining requests for that agent are masked and no longer take part in the scheduler arbitration. When an agent is masked from arbitration due to its weight counter reaching zero, the memory scheduler continues to schedule requests from the remaining agents. Once the weight counters for all agents have reached zero or if an agent's weight counter is non-zero but there are no remaining requests for that agent, all agent weight counters are reloaded with the values from agent weight configuration registers.

[0101] Referring now to FIG. 9, shown is a flow diagram of a method for performing memory scheduling in accordance with an embodiment of the present invention. As shown in FIG. 9, method 600 may be performed by a scheduler arbiter of the shared memory fabric. As seen, method 600 may be performed by selecting a memory request from the memory scheduler scoreboard for delivery to a memory controller (block 610). Various considerations may be taken into account in determining the appropriate entry including state of the memory, state of the various requests, relationship between address locations of the pending requests and so forth. Next at block 620 the weight value for the selected agent is updated. In an embodiment a decrementing of the weight value is performed. Note that while the initial value for the weight value for the agents is the same as obtained from the configuration register also used by the admit arbiter, understand that different weight counters are provided for each arbiter to enable independent control of these weight values.

[0102] Still referring to FIG. 9, next at diamond 630 it can be determined whether the weight value of the selected agent is equal to zero. Note that this determination may be in an embodiment in which zero is the lowest priority value. Of course the

scope of the present invention is not limited in this regard and this determination may in other embodiments be based on comparison to another predetermined value. If it is determined that the weight value is zero, control passes to block 640 where this selected agent is masked from further arbitration within the memory scheduler.

[0103] From both of diamonds 630 and block 640 control passes to diamond 650 where it can be determined whether the weight value of all agents equals zero. If so, control passes to block 660 where the weight values for all the agents can be updated to their configured values, e.g., obtained from a configuration register of the fabric. Otherwise, control passes from diamond 650 to diamond 670 to determine whether there are any remaining requests in the memory scheduler for agents having a non-zero weight value. If so, those requests can be handled, e.g., via another iteration of method 600. Otherwise if no additional requests remain, control passes to block 660 where the weight values can be updated as described. Although shown with this particular implementation in the embodiment of FIG. 9, understand the scope of the present invention is not limited in this regard.

[0104] Table 5 below provides example operation of memory scheduling for plurality of clock cycles, based on initial weight values for three agents as follows:

Agent0 Weight = 4
 Agent 1 Weight = 2
 Agent 2 Weight = 1

TABLE 5

Clock Cycle	Agent 0 Req	Agent 0 Req Mask	Agent 0 Weight Counter	Agent 1 Req	Agent 1 Req Mask	Agent 1 Weight Counter	Agent 2 Req	Agent 2 Req Mask	Agent 2 Weight Counter	Reload Agent Weights	Agent Grant
1	False	False	4	False	False	2	False	False	1	True	No Grant
2	True	False	4	True	False	2	True	False	1	False	Grant Agent 1
3	True	False	4	True	False	1	True	False	1	False	Grant Agent 2
4	True	False	4	True	False	1	True	True	0	False	Grant Agent 0
5	True	False	3	True	False	1	True	True	0	False	Grant Agent 0
6	True	False	2	True	False	1	True	True	0	False	Grant Agent 1
7	True	False	2	True	True	0	True	True	0	False	Grant Agent 0

8	True	False	1	True	True	0	True	True	0	True	Grant Agent 0
9	True	False	4	True	False	2	True	False	1	False	Grant Agent 0
10	True	False	3	True	False	2	True	False	1	False	Grant Agent 0
11	True	False	2	True	False	2	True	False	1	False	Grant Agent 1
12	True	False	2	True	False	1	True	False	1	False	Grant Agent 2
13	True	False	2	True	False	1	True	True	0	False	Grant Agent 0
14	True	False	1	True	False	1	True	True	0	False	Grant Agent 0
15	True	True	0	True	False	1	True	True	0	True	Grant Agent 1
16	True	False	4	True	False	2	True	False	1	False	Grant Agent 0
17	True	False	3	True	False	2	True	False	1	False	Grant Agent 1
18	True	False	3	True	False	2	True	False	1	False	Grant Agent 0

[0105] Out of Order Page Aware Scheduling

[0106] The memory scheduler reorders requests sent to the memory controller and seeks to optimize the stream of requests for the maximum memory bandwidth possible. The memory scheduler contains configuration registers programmed to provide the scheduler with information about the memory controller to which it is attached. In one embodiment, these configuration registers include information about what address bits are used for the memory channel, bank, rank and row addresses. Using the memory configuration information programmed in the configuration registers the memory scheduler determines the bank, rank, row and channel of each request in the scheduler scoreboard. The memory scheduler scoreboard also contains a page hit status bit for each request that is used to optimize requests sent to the memory controller so that requests to the same page in memory are sent to the memory controller before sending request to a different page.

[0107] After initialization and before any requests are sent to the memory controller, the memory scheduler clears all page hit status bits in its scoreboard. As requests are sent to the memory controller the memory scheduler updates the page hit status bits in the scoreboard to indicate whether other requests are to the same page or to

a different page in memory. Although the scheduler is not aware of the actual state of the page in a given memory bank, these page hit status bits may be used as a hint as to which requests are the best candidates to send to the memory controller for optimal memory bandwidth.

[0108] When a request is sent to the memory controller, the memory scheduler compares the channel, rank and bank information for all other requests pending in the scoreboard. If the channel, rank and bank information of a scoreboard entry matches a request that is sent to the memory controller the row address of the entry is compared against the row address of the request sent to the memory controller. If the row address of a scoreboard entry matches for the request the page hit status bit is set to 1; if the row address does not match the request the page hit status bit is set to 0 indicating a page miss. For scoreboard entries where the channel, rank or bank bits are different than the request sent to the memory controller, no update of the page hit status occurs.

[0109] As new requests are granted by the admit arbiter and enqueued into the scheduler scoreboard, the row address information is compared against all entries currently in the scoreboard. If the row address of the new request matches one or more entries in the scheduler scoreboard and the page hit status bit of any matching entries is set, the page hit status for the new request is also set. If the row address does not match any entries in the scoreboard or all entries it matches have the page hit status set to zero, the page hit status for the new request is also set to zero.

[0110] Using the page hit and rank status information stored in the scheduler scoreboard, the memory scheduler reorders requests sent to the memory controller based on a priority encoded scheduling scheme that has been determined to provide optimal bandwidth for most DRAM-based memory technologies. The memory scheduler grants higher priority requests before granting requests with lower priority levels.

[0111] Table 6 below shows the different priority levels used by a memory scheduler in accordance with one embodiment of the present invention.

TABLE 6

Memory Scheduler Page Aware Scheduling Priority

Pagehit Status	Rank Status	Priority Level
Pagehit	Same Rank	Priority Level 3 (Highest)
Pagehit	Different Rank	Priority Level 2
Pagemiss	Same Rank	Priority Level 1
Pagemiss	Different Rank	Priority Level 0 (Lowest)

[0112] Age Based Memory Scheduling and Starvation Prevention

[0113] In order to prevent starvation of requests due to the out-of-order page aware scheduling algorithm, the concept of age is used at least in part to schedule requests. For each class of service (COS) category, the memory scheduler contains a configuration register to specify an out-of-order (OOO) scheduling limit. To provide a shorter maximum read latency for the isochronous COS category, the OOO scheduling limit is typically set to a smaller value than the OOO scheduling limit of the best effort COS category. The memory scheduler creates a request bit vector for all pending requests in its scoreboard for the best effort and isochronous COS categories. These request bit vectors are sent to the oldest of available queue, which determines the oldest request that is still pending. The oldest of available queue outputs a one hot encoded bit vector with the bit set to 1 to indicate the oldest request. As the memory scheduler grants requests OOO based on its page aware scheduling algorithm, the memory scheduler counts how many requests were granted that were not the oldest pending request for each COS category. Once the counter reaches the OOO scheduling limit for the COS category which may be determined by performance analysis done for worst case acceptable latency for a COS category, the page aware scheduling logic is disabled and the oldest request for the COS category is granted by the memory scheduler. Anytime that the oldest request for a COS category is granted, the counter for that COS category is reset to zero. To provide the lowest possible latency for a COS category the OOO scheduling limit can be programmed to zero, essentially disabling the page aware scheduling logic for that COS category. When the OOO scheduling limit is set to zero for a COS category, requests to memory may be scheduled using request age, which is determined by the oldest of available queue.

[0114] Best Effort Maximum Latency Starvation Prevention

[0115] For best effort read requests, the fabric utilizes the deadline storage information in the scheduler scoreboard to store a value that is used to specify a maximum latency value for scheduling best effort requests. The scoreboard is a pool of entries and a request stored in the scoreboard may be either a best effort or isochronous request determined by the request's class of service category, also stored in the scoreboard for each request. In the case a request in the scoreboard is a best effort read request, a maximum allowable latency, e.g., a preprogrammed value stored in a configuration register, is used to schedule the request. When the request is enqueued in the scoreboard and is a best effort read request the maximum latency value is added to the current value of the global timer. Once the global timer reaches the value stored for the best effort requests' maximum latency, page aware scheduling is ignored for the request and results in the request being scheduled when it is the oldest request pending, e.g., as determined by the oldest of available queue.

[0116] Request Tracker Write Priority and Weights

[0117] The request tracker is responsible for the transfer of data from the requesting agents to the internal memory buffer of the fabric. The write protocol used by the shared memory fabric causes all write data to be transferred in request order from the requesting agent to the internal memory buffer in the fabric. In one embodiment, the request tracker uses separate linked lists per agent to preserve the ordering of the write requests. The request tracker may perform coherency checks for a write request prior to transferring data from the requesting agent to the internal data buffer.

[0118] For write requests, the request tracker may be configured to support one or more priority levels. When a request is granted by the admit arbiter the deadline information for the request is stored in an array having a length corresponding to the number of entries in the request tracker. The fabric uses a threshold value, e.g., stored in a configuration register, to specify when a request deadline value is considered to be high priority. Each deadline value for a request is compared

against the threshold value programmed in the configuration register. When the deadline latency is less than the value in the configuration register, a bit is set in the tracker's scoreboard entry for the request indicating the request is a high priority request.

[0119] When enabled for two priority level operation, if a write request for an agent reaches the head of the linked list and the high priority bit is set for the request the write request is considered to be high priority. If any write requests at the head of any of the agent linked lists indicate the write request is a high priority request, all low priority write requests at the head of the other linked list for other agents are masked before being input to the write request arbiter. If multiple requests of the same priority level are present at the head of the agent linked lists, an arbitration is performed to select which agent to choose to transfer the write data.

[0120] Request Tracker Write Request Arbiter

[0121] The write request arbiter uses a weighted priority based fair arbiter to select which agent to transfer write data. The weights for the write request arbiter are programmed in configuration registers in the request tracker. The write arbiter assigns each agent a unique priority at reset. On each cycle, the arbiter only considers request candidates with data that is ready to transfer, and grants to the requester with the highest priority. When granted, a request candidate's weight is decremented by 1. If the granted candidate already had a weight of zero, then the arbiter also updates request candidate priorities as follows: the granted candidate's priority is set to the lowest priority (e.g., zero); all candidates with priorities lower than the granted candidate increment their priority, and all candidates with priorities higher than the granted candidate leave their priority unchanged.

[0122] Request Tracker Read Data Return

[0123] Requesting agents either support in order data return or out-of-order data return. To support out-of-order data return, an order ID field is used. An order ID is sent from the agent with each request and is stored in the request tracker scoreboard. Requests from the same agent that have the same order ID are returned in request order. Data for requests from the same agent having different

order IDs do not need to be returned in request order. In an embodiment, the request tracker uses linked lists for ensuring read data is properly ordered when it is returned to the requesting agent.

[0124] The entry of the internal data buffer where data is to be written is chosen prior to a request being granted by the admit arbiter. When a request is granted by the admit arbiter, request information including the index into the internal data buffer is forwarded to the request tracker. As data is returned from the memory controller, the memory scheduler forwards a read completion indication to the request tracker, which includes the index field into the internal data buffer where the data is being written and an indication of which chunks of the memory address have completed a read of memory. When the request tracker receives a read completion, it compares the index field with the index fields for all requests stored in the request tracker scoreboard. If a scoreboard entries' index field matches a read completion for a request and all chunk bits for the request are set for the read completion, a bit is set in the request tracker scoreboard indicating the read request has completed.

[0125] If a read request has reached the head of the linked list and the read completion status bit in the request tracker is set and all coherency checks for the request have completed, the request is available to return read data to the agent. Similar to write requests, the request tracker uses the request deadline information for a scoreboard entry to indicate request priority. In one embodiment, the request tracker creates two request bit vectors for scoreboard entries that have data ready to return to the requesting agents. One bit vector is for low priority read requests and the other bit vector is for high priority read requests. The request bit vectors are input to the request tracker oldest of available queue. The oldest of available queue determines which request is the oldest for both request bit vectors. The request tracker has a configuration mode which when enabled will cause a return of data from the oldest high priority request selected by the oldest of available queue before returning data for any low priority requests. When support of the high priority data return is not enabled, the request tracker treats all scoreboard entries that are ready to return read data as having the same priority level. In this mode, only the low priority bit vector is used as an input to the oldest of available queue that in turn

determines the oldest read request in the scoreboard. Read data for the scoreboard entry determined to be the oldest is then returned to the requesting agent.

[0126] Embodiments may be used in many different SoCs or other semiconductor devices that integrate various IPs onto a single die to connect these IPs to memory via a memory fabric. Still further a memory fabric in accordance with an embodiment of the present invention may be used to provide a QoS level for meeting isochronous requirements of at least some of these IPs.

[0127] Referring now to FIG. 10, shown is a block diagram of an SoC in accordance with an embodiment of the present invention. As shown in FIG. 10, SoC 700 is a single die semiconductor device including multiple IP blocks along with a shared memory arbiter as described above. In the embodiment of FIG. 10 a plurality of cores 710₀-710_n are provided, each of which can independently execute instructions. In one embodiment, all of these cores are of a single design such as an in-order core design, e.g., of an Intel Architecture™ such as an Atom™-based core. In other embodiments, the cores may be out-of-order processors such as an Intel Architecture™ (IA) 32 core such as an Intel Core™-based design. In other embodiments, a mix of heterogeneous cores may be provided. In addition, a plurality of graphics engines, namely independent graphics units 720₀-720_n may be provided each to independently perform graphics operations. As seen, the multiple cores are coupled to a shared cache memory 715 such as a level 2 (L2) cache and similarly, the graphics engines are coupled to another shared cache memory 725.

[0128] A system agent 730 is coupled to these cores and graphics engines via corresponding in-die interconnects 728 and 729. As seen, system agent 730 includes a shared memory fabric 735 which may be configured as described herein. Of course various other logic, controllers and other units such as a power management unit may be present within system agent 730. As seen, shared memory fabric 735 communicates with a memory controller 740 that in turn couples to an off-chip memory such as a system memory configured as DRAM. In addition, system agent 730 is coupled via a set of interconnects 744 to one or more internal agents 750 such as various peripheral devices. In an embodiment, interconnect 744 may include a priority channel interconnect, a sideband channel interconnect, and a

memory channel interconnect. A similarly configured interconnect 746 provides for communication between system agent 730 and one or more off-chip agents (not shown for ease of illustration in the embodiment of FIG. 10). Although shown at this high level in FIG. 10, understand the scope of the present invention is not limited in this regard.

[0129] Referring now to FIG. 11, shown is a block diagram of components present in a computer system in accordance with an embodiment of the present invention. As shown in FIG. 11, system 800 can include many different components. These components can be implemented as ICs, portions thereof, discrete electronic devices, or other modules adapted to a circuit board such as a motherboard or add-in card of the computer system, or as components otherwise incorporated within a chassis of the computer system. Note also that the block diagram of FIG. 11 is intended to show a high level view of many components of the computer system. However, it is to be understood that additional components may be present in certain implementations and furthermore, different arrangement of the components shown may occur in other implementations.

[0130] As seen in FIG. 11, a processor 810, which may be a low power multicore processor socket such as an ultra low voltage processor, may act as a main processing unit and central hub for communication with the various components of the system. Such processor can be implemented as a SoC as described herein. In one embodiment, processor 810 may be an Intel® Architecture Core™-based processor such as an i3, i5, i7 or another such processor available from Intel Corporation, Santa Clara, CA, such as a processor that combines one or more Core™-based cores and one or more Intel® ATOM™-based cores to thus realize high power and low power cores in a single SoC. However, understand that other low power processors such as available from Advanced Micro Devices, Inc. (AMD) of Sunnyvale, CA, an ARM-based design from ARM Holdings, Ltd. or a MIPS-based design from MIPS Technologies, Inc. of Sunnyvale, CA, or their licensees or adopters may instead be present in other embodiments such as an Apple A5 or A6 processor.

[0131] Processor 810 may communicate with a system memory 815, which in an embodiment can be implemented via multiple memory devices to provide for a given amount of system memory. To provide for persistent storage of information such as data, applications, one or more operating systems and so forth, a mass storage 820 may also couple to processor 810. Also shown in FIG. 11, a flash device 822 may be coupled to processor 810, e.g., via a serial peripheral interface (SPI). This flash device may provide for non-volatile storage of system software, including a basic input/output software (BIOS) as well as other firmware of the system.

[0132] Various input/output (IO) devices may be present within system 800. Specifically shown in the embodiment of FIG. 11 is a display 824 which may be a high definition LCD or LED panel configured within a lid portion of the chassis. This display panel may also provide for a touch screen 825, e.g., adapted externally over the display panel such that via a user's interaction with this touch screen, user inputs can be provided to the system to enable desired operations, e.g., with regard to the display of information, accessing of information and so forth. In one embodiment, display 824 may be coupled to processor 810 via a display interconnect that can be implemented as a high performance graphics interconnect. Touch screen 825 may be coupled to processor 810 via another interconnect, which in an embodiment can be an I²C interconnect. As further shown in FIG. 11, in addition to touch screen 825, user input by way of touch can also occur via a touch pad 830 which may be configured within the chassis and may also be coupled to the same I²C interconnect as touch screen 825.

[0133] For perceptual computing and other purposes, various sensors may be present within the system and can be coupled to processor 810 in different manners. Certain inertial and environmental sensors may couple to processor 810 through a sensor hub 840, e.g., via an I²C interconnect. In the embodiment shown in FIG. 11, these sensors may include an accelerometer 841, an ambient light sensor (ALS) 842, a compass 843 and a gyroscope 844. Other environmental sensors may include one or more thermal sensors 846 which may couple to processor 810 via a system management bus (SMBus) bus, in one embodiment.

[0134] Also seen in FIG. 11, various peripheral devices may couple to processor 810 via a low pin count (LPC) interconnect. In the embodiment shown, various components can be coupled through an embedded controller 835. Such components can include a keyboard 836 (e.g., coupled via a PS2 interface), a fan 837, and a thermal sensor 839. In some embodiments, touch pad 830 may also couple to EC 835 via a PS2 interface. In addition, a security processor such as a trusted platform module (TPM) 838 in accordance with the Trusted Computing Group (TCG) TPM Specification Version 1.2, dated Oct. 2, 2003, may also couple to processor 810 via this LPC interconnect.

[0135] System 800 can communicate with external devices in a variety of manners, including wirelessly. In the embodiment shown in FIG. 11, various wireless modules, each of which can correspond to a radio configured for a particular wireless communication protocol, are present. One manner for wireless communication in a short range such as a near field may be via a near field communication (NFC) unit 845 which may communicate, in one embodiment with processor 810 via an SMBus. Note that via this NFC unit 845, devices in close proximity to each other can communicate. For example, a user can enable system 800 to communicate with another (e.g.,) portable device such as a smartphone of the user via adapting the two devices together in close relation and enabling transfer of information such as identification information payment information, data such as image data or so forth. Wireless power transfer may also be performed using a NFC system.

[0136] As further seen in FIG. 11, additional wireless units can include other short range wireless engines including a WLAN unit 850 and a Bluetooth unit 852. Using WLAN unit 850, Wi-Fi™ communications in accordance with a given Institute of Electrical and Electronics Engineers (IEEE) 802.11 standard can be realized, while via Bluetooth unit 852, short range communications via a Bluetooth protocol can occur. These units may communicate with processor 810 via, e.g., a USB link or a universal asynchronous receiver transmitter (UART) link. Or these units may couple to processor 810 via an interconnect via a Peripheral Component Interconnect Express™ (PCIe™) protocol in accordance with the PCI Express™ Specification Base Specification version 3.0 (published January 17, 2007), or another such

protocol such as a serial data input/output (SDIO) standard. Of course, the actual physical connection between these peripheral devices, which may be configured on one or more add-in cards, can be by way of the next generation form factor (NGFF) connectors adapted to a motherboard.

[0137] In addition, wireless wide area communications, e.g., according to a cellular or other wireless wide area protocol, can occur via a WWAN unit 856 which in turn may couple to a subscriber identity module (SIM) 857. In addition, to enable receipt and use of location information, a GPS module 855 may also be present. Note that in the embodiment shown in FIG. 11, WWAN unit 856 and an integrated capture device such as a camera module 854 may communicate via a given USB protocol such as a USB 2.0 or 3.0 link, or a UART or I²C protocol. Again the actual physical connection of these units can be via adaptation of a NGFF add-in card to an NGFF connector configured on the motherboard.

[0138] To provide for audio inputs and outputs, an audio processor can be implemented via a digital signal processor (DSP) 860, which may couple to processor 810 via a high definition audio (HDA) link. Similarly, DSP 860 may communicate with an integrated coder/decoder (CODEC) and amplifier 862 that in turn may couple to output speakers 863 which may be implemented within the chassis. Similarly, amplifier and CODEC 862 can be coupled to receive audio inputs from a microphone 865 which in an embodiment can be implemented via dual array microphones to provide for high quality audio inputs to enable voice-activated control of various operations within the system. Note also that audio outputs can be provided from amplifier/CODEC 862 to a headphone jack 864. Although shown with these particular components in the embodiment of FIG. 11, understand the scope of the present invention is not limited in this regard.

[0139] The following examples pertain to further embodiments. In an embodiment, an apparatus includes a fabric coupled between IP blocks of a semiconductor device and a memory controller, where the fabric is to receive memory requests from the IP blocks, at least some of which have a deadline value associated therewith to indicate a maximum latency prior to completion of the memory request. The fabric is to

arbitrate between the memory requests based at least in part on the deadline value of the at least some memory requests.

[0140] The fabric may include an admit arbiter to receive the memory requests and to select one of the memory requests to provide to a memory scheduler arbiter, and the memory scheduler arbiter may be coupled to the admit arbiter to select a first memory request from a plurality of memory requests to provide to the memory controller.

[0141] The admit arbiter includes a first arbiter to receive the memory requests from a first subset of the plurality of IP blocks and a first portion of the memory requests from a second subset of the plurality of IP blocks, and a second arbiter to receive a second portion of the memory requests from the second subset of the plurality of IP blocks. The second subset of the plurality of IP blocks include latency sensitive devices. The second arbiter is further to receive a deadline value with each of the second portion of the memory requests received from the second subset of the plurality of IP blocks.

[0142] In an embodiment, the admit arbiter is controllable to operate in a first mode in which memory requests from the second arbiter are selected over memory requests from the first arbiter and a second mode in which M memory requests from the first arbiter are selected after N memory requests from the second arbiter are selected, where M and N are obtained from a configuration register.

[0143] In an embodiment, the admit arbiter further includes a third arbiter coupled to an output of the first arbiter and the second arbiter and to arbitrate between the outputs of the first and second arbiters.

[0144] The memory scheduler arbiter may reorder the memory requests out of order based at least in part on address information of the memory requests, and based at least in part on page hit status information associated with a memory request stored in an entry of a scoreboard of the fabric.

[0145] In an embodiment, an interface logic is coupled to a first IP block which includes a deadline logic to generate the deadline value based on a latency requirement for a first memory request of the first IP block and a global timer value

provided from the fabric to the first IP block. A deadline checker may be coupled to receive the deadline value from the first IP block and to route the first memory request to a first arbiter or to a second arbiter based on the deadline value and a value of a global timer.

[0146] In an embodiment, resource allocation logic reserves virtual entries in a plurality of shared resources of the fabric for each of the plurality of IP blocks. This resource allocation logic is further to limit each of the plurality of IP blocks to a predetermined number of the plurality of shared resources, and the plurality of shared resources includes a flush pool having a plurality of entries each to store a write request, and the resource allocation logic is to limit the plurality of entries to a flush limit, where the admit arbiter is to block incoming write requests when a size of the flush pool equals the flush limit.

[0147] In an embodiment, a method includes: receiving a memory request from a device of a plurality of devices coupled to a shared memory fabric of a SoC, where the memory request is associated with a deadline value; forwarding the memory request to a first arbiter of the shared memory fabric responsive to the deadline value being less than a latency threshold; performing an arbitration in the first arbiter between a first plurality of memory requests including the memory request based at least in part on age values of the plurality of devices and forwarding a winner memory request of the first plurality of memory requests to a final arbiter of the shared memory fabric; and performing a final arbitration in the final arbiter to select a final memory request for storage in a scoreboard of the shared memory fabric.

[0148] In an embodiment, the method further includes forwarding the memory request to a second arbiter responsive to the deadline value being greater than the latency threshold, performing an arbitration in the second arbiter between a second plurality of memory requests including the memory request based at least in part on the age values of the plurality of devices, and forwarding a winner memory request of the second plurality of memory requests to the final arbiter.

[0149] In an embodiment, the method further includes determining whether a weight value of a device associated with the final memory request equals a

predetermined value, and if so updating the weight value to a configured value and updating the age values for others of the plurality of devices. And, when the weight value of the device associated with the final memory request does not equal the predetermined value, the weight value of the device associated with the final memory request is decremented.

[0150] In an embodiment, the method further includes: selecting via a memory scheduler of the shared memory fabric a memory request from the scoreboard for delivery to a memory controller; and if a weight value of a device associated with the selected memory request equals a predetermined value, masking the device from further arbitration within the memory scheduler.

[0151] In an embodiment, the method further includes determining whether a weight value of each of the plurality of devices equals the predetermined value and if so, updating the weight value for each of the plurality of devices to a configured value.

[0152] In an embodiment, a system on chip includes a plurality of cores each to independently execute instructions, a plurality of graphics engines each to independently execute graphics operations, a first content rendering device corresponding to a first latency sensitive device, a second content rendering device corresponding to a second latency sensitive device, and a fabric coupled to the plurality of cores, the plurality of graphics engines, the first and second content rendering devices and a memory controller, the fabric to receive memory requests from the plurality of cores, the plurality of graphics engines and the first and second content rendering devices. The memory requests from the first and second content rendering devices have a deadline value associated therewith to indicate a maximum latency prior to completion of the memory request, and the fabric is to arbitrate between the memory requests based at least in part on the deadline value of the memory requests from the first and second content rendering devices.

[0153] In an embodiment, the fabric further includes: an admit arbiter to receive the memory requests and to select one of the memory requests to provide to a memory scheduler arbiter; and the memory scheduler arbiter coupled to the admit arbiter to

select a first memory request from a plurality of memory requests to provide to the memory controller. The admit arbiter may include a first arbiter to receive the memory requests from the plurality of cores and the plurality of graphics engines and to conditionally receive at least some memory requests from the first and second content rendering devices, and a second arbiter to receive the memory requests from the first and second content rendering devices. The memory scheduler arbiter may reorder at least some of the memory requests out of order based at least in part on address information of the memory requests and page hit status information associated with a memory request stored in an entry of a scoreboard of the fabric.

[0154] A deadline checker may receive a deadline value associated with a first memory request from the first content rendering device and to route the first memory request to the first arbiter or to the second arbiter based on the deadline value and a value of a global timer.

[0155] Embodiments may be implemented in code and may be stored on a non-transitory storage medium having stored thereon instructions which can be used to program a system to perform the instructions. The storage medium may include, but is not limited to, any type of disk including floppy disks, optical disks, solid state drives (SSDs), compact disk read-only memories (CD-ROMs), compact disk rewritables (CD-RWs), and magneto-optical disks, semiconductor devices such as read-only memories (ROMs), random access memories (RAMs) such as dynamic random access memories (DRAMs), static random access memories (SRAMs), erasable programmable read-only memories (EPROMs), flash memories, electrically erasable programmable read-only memories (EEPROMs), magnetic or optical cards, or any other type of media suitable for storing electronic instructions.

[0156] While the present invention has been described with respect to a limited number of embodiments, those skilled in the art will appreciate numerous modifications and variations therefrom. It is intended that the appended claims cover all such modifications and variations as fall within the true spirit and scope of this present invention.

What is claimed is:

- 1 1. An apparatus comprising:
2 a fabric means coupled between a plurality of intellectual property (IP) blocks
3 of a semiconductor device and a memory controller, the fabric means for receiving
4 memory requests from the plurality of IP blocks, at least some of the memory
5 requests having a deadline value associated therewith to indicate a maximum
6 latency prior to completion of the memory request, wherein the fabric means for
7 arbitrating between the plurality of memory requests based at least in part on the
8 deadline value of the at least some memory requests.

- 1 2. The apparatus of claim 1, wherein the fabric means comprises:
2 an admit arbiter means for receiving the memory requests and selecting one
3 of the memory requests to provide to a memory scheduler arbiter means; and
4 the memory scheduler arbiter means coupled to the admit arbiter means for
5 selecting a first memory request from a plurality of memory requests to provide to the
6 memory controller.

- 1 3. The apparatus of claim 2, wherein the admit arbiter means includes a first
2 arbiter to receive the memory requests from a first subset of the plurality of IP blocks
3 and a first portion of the memory requests from a second subset of the plurality of IP
4 blocks, and a second arbiter to receive a second portion of the memory requests
5 from the second subset of the plurality of IP blocks.

- 1 4. The apparatus of claim 3, wherein the second subset of the plurality of IP
2 blocks comprise latency sensitive devices.

- 1 5. The apparatus of claim 4, wherein the second arbiter is further to receive a
2 deadline value with each of the second portion of the memory requests received
3 from the second subset of the plurality of IP blocks.

1 6. The apparatus of claim 2, wherein the admit arbiter means is controllable to
2 operate in a first mode in which memory requests from the second arbiter are
3 selected over memory requests from the first arbiter, and a second mode in which M
4 memory requests from the first arbiter are selected after N memory requests from
5 the second arbiter are selected, wherein M and N are obtained from a configuration
6 register.

1 7. The apparatus of claim 3, wherein the admit arbiter means further includes a
2 third arbiter coupled to an output of the first arbiter and the second arbiter and to
3 arbitrate between the outputs of the first and second arbiters.

1 8. The apparatus of claim 2, wherein the memory scheduler arbiter means for
2 reordering the memory requests out of order based at least in part on address
3 information of the memory requests.

1 9. The apparatus of claim 8, wherein the memory scheduler arbiter means for
2 reordering the memory requests out of order based at least in part on page hit status
3 information associated with a memory request stored in an entry of a scoreboard of
4 the fabric.

1 10. The apparatus of claim 1, further comprising an interface means coupled to a
2 first IP block, the interface means including a deadline means for generating the
3 deadline value based on a latency requirement for a first memory request of the first
4 IP block and a global timer value provided from the fabric to the first IP block.

1 11. The apparatus of claim 8, further comprising a deadline checker to receive the
2 deadline value from the first IP block and to route the first memory request to a first
3 arbiter or to a second arbiter based on the deadline value and a value of a global
4 timer.

1 12. The apparatus of claim 1, further comprising resource allocation means for
2 reserving virtual entries in a plurality of shared resources of the fabric means for
3 each of the plurality of IP blocks.

1 13. The apparatus of claim 12, wherein the resource allocation means is further to
2 limit each of the plurality of IP blocks to a predetermined number of the plurality of
3 shared resources.

1 14. The apparatus of claim 12, wherein the plurality of shared resources includes
2 a flush pool having a plurality of entries each to store a write request, wherein the
3 resource allocation means for limiting the plurality of entries to a flush limit, and
4 wherein the admit arbiter means is to block incoming write requests when a size of
5 the flush pool equals the flush limit.

1 15. A method comprising:
2 receiving a memory request from a device of a plurality of devices coupled to
3 a shared memory fabric of a system on chip (SoC), the memory request associated
4 with a deadline value;
5 forwarding the memory request to a first arbiter of the shared memory fabric
6 responsive to the deadline value being less than a latency threshold;
7 performing an arbitration in the first arbiter between a first plurality of memory
8 requests including the memory request based at least in part on age values of the
9 plurality of devices and forwarding a winner memory request of the first plurality of
10 memory requests to a final arbiter of the shared memory fabric; and
11 performing a final arbitration in the final arbiter to select a final memory
12 request for storage in a scoreboard of the shared memory fabric.

1 16. The method of claim 15, further comprising forwarding the memory request to
2 a second arbiter responsive to the deadline value being greater than the latency
3 threshold, performing an arbitration in the second arbiter between a second plurality
4 of memory requests including the memory request based at least in part on the age

5 values of the plurality of devices, and forwarding a winner memory request of the
6 second plurality of memory requests to the final arbiter.

1 17. The method of claim 15, further comprising determining whether a weight
2 value of a device associated with the final memory request equals a predetermined
3 value, and if so updating the weight value to a configured value and updating the age
4 values for others of the plurality of devices.

1 18. The method of claim 17, further comprising when the weight value of the
2 device associated with the final memory request does not equal the predetermined
3 value, decrementing the weight value of the device associated with the final memory
4 request.

1 19. The method of claim 15, further comprising:
2 selecting via a memory scheduler of the shared memory fabric a memory
3 request from the scoreboard for delivery to a memory controller; and
4 if a weight value of a device associated with the selected memory request
5 equals a predetermined value, masking the device from further arbitration within the
6 memory scheduler.

1 20. The method of claim 19, further comprising determining whether a weight
2 value of each of the plurality of devices equals the predetermined value and if so,
3 updating the weight value for each of the plurality of devices to a configured value.

1 21. A system on chip comprising:
2 a plurality of cores each to independently execute instructions;
3 a plurality of graphics engines each to independently execute graphics
4 operations;
5 a first content rendering device corresponding to a first latency sensitive
6 device;
7 a second content rendering device corresponding to a second latency
8 sensitive device; and

9 a fabric means coupled to the plurality of cores, the plurality of graphics
10 engines, the first and second content rendering devices and a memory controller
11 means, the fabric means for receiving memory requests from the plurality of cores,
12 the plurality of graphics engines and the first and second content rendering devices,
13 wherein the memory requests from the first and second content rendering devices
14 have a deadline value associated therewith to indicate a maximum latency prior to
15 completion of the memory request, the fabric means for arbitrating between the
16 memory requests based at least in part on the deadline value of the memory
17 requests from the first and second content rendering devices.

1 22. The system of claim 21, wherein the fabric means further comprises:
2 an admit arbiter means for receiving the memory requests and selecting one
3 of the memory requests to provide to a memory scheduler arbiter means; and
4 the memory scheduler arbiter means coupled to the admit arbiter means for
5 selecting a first memory request from a plurality of memory requests to provide to the
6 memory controller.

1 23. The system of claim 22, wherein the admit arbiter means includes a first
2 arbiter to receive the memory requests from the plurality of cores and the plurality of
3 graphics engines and to conditionally receive at least some memory requests from
4 the first and second content rendering devices, and a second arbiter to receive the
5 memory requests from the first and second content rendering devices.

1 24. The system of claim 22, wherein the memory scheduler arbiter means for
2 reordering at least some of the memory requests out of order based at least in part
3 on address information of the memory requests and page hit status information
4 associated with a memory request stored in an entry of a scoreboard of the fabric.

1 25. The system of claim 23, further comprising a deadline checker to receive a
2 deadline value associated with a first memory request from the first content
3 rendering device and to route the first memory request to the first arbiter or to the
4 second arbiter based on the deadline value and a value of a global timer.

1 26. A communication device arranged to perform the method of any one of claims
2 15 to 20.

1 27. At least one machine readable medium comprising a plurality of instructions
2 that in response to being executed on a computing device, cause the computing
3 device to carry out a method according to any one of claims 15 to 20.

1 28. An apparatus for processing instructions, configured to perform the method of
2 any one of claims 15 to 20.

1 29. An apparatus comprising means for performing the method of any one of
2 claims 15 to 20.

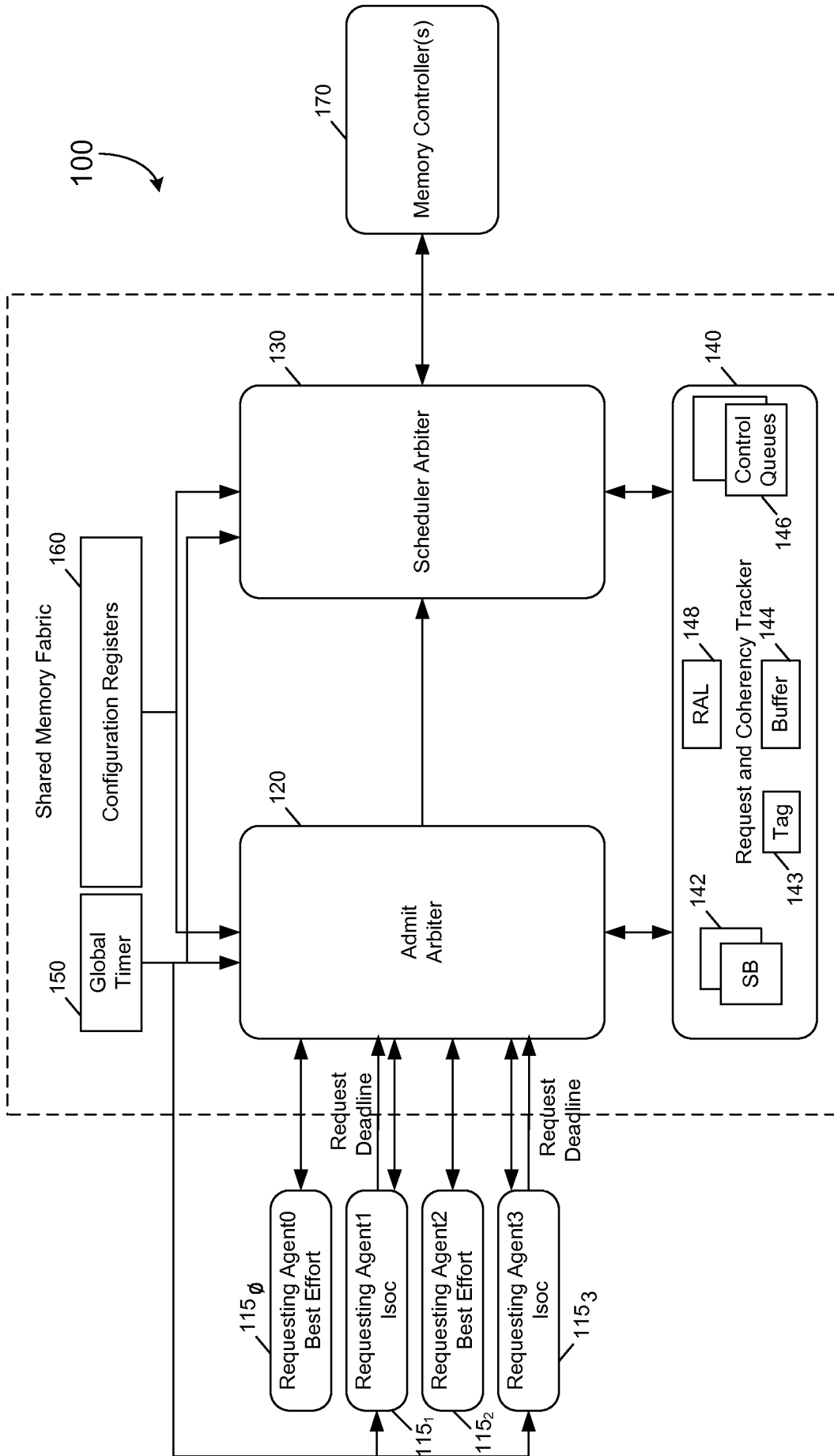


FIG. 1

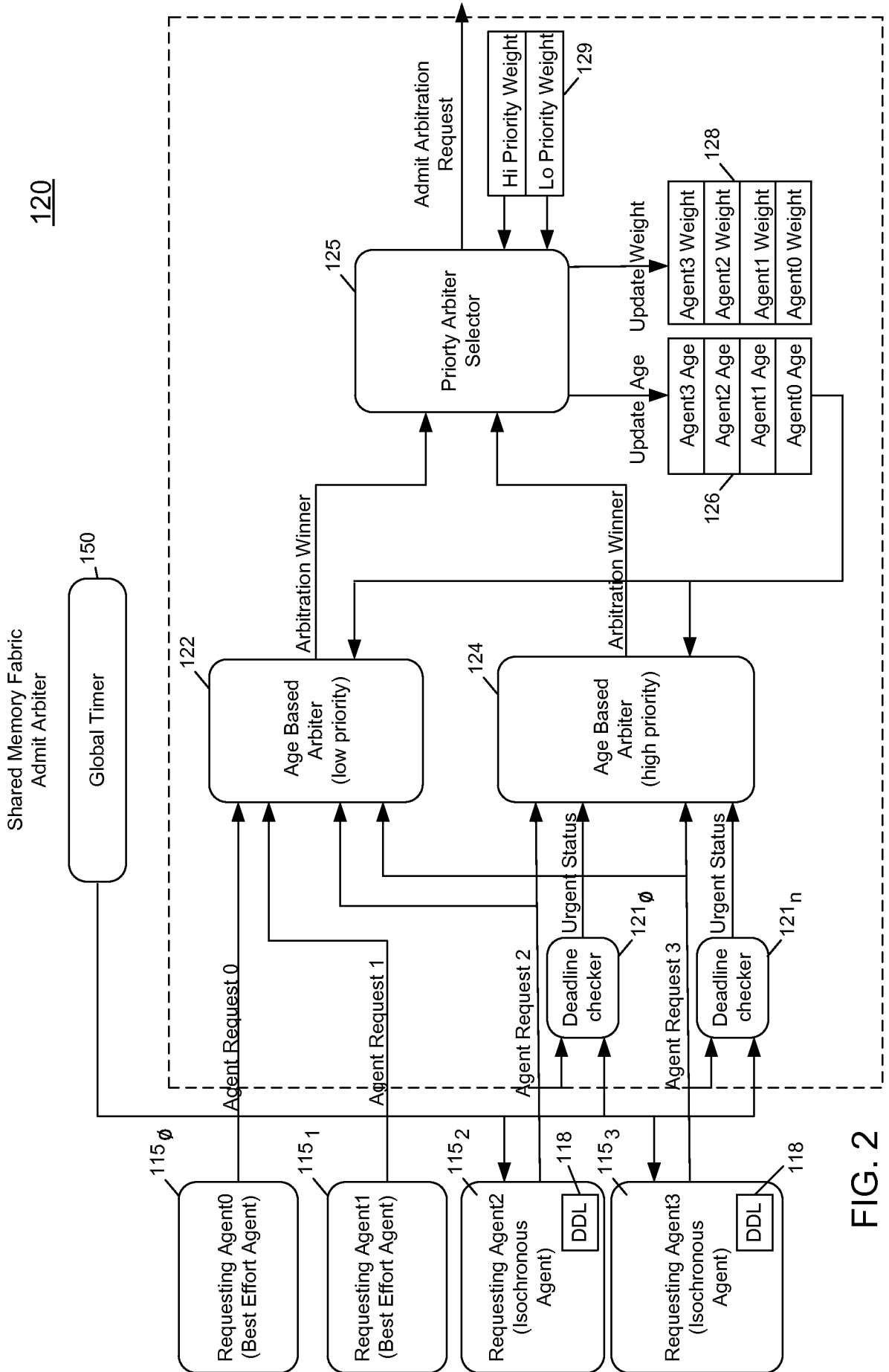


FIG. 2

3/11

200

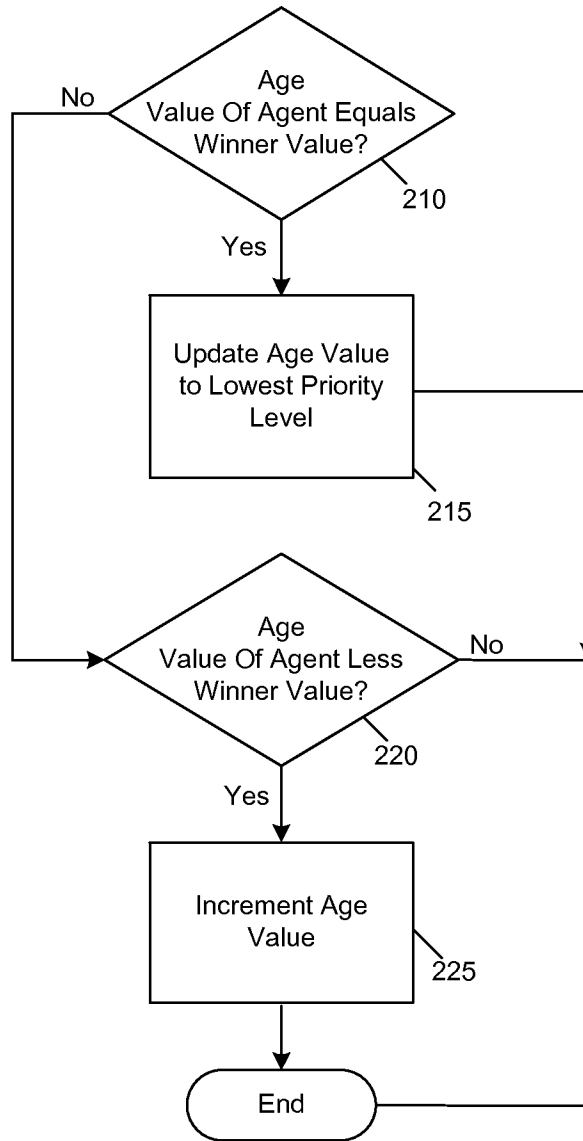


FIG. 3

250

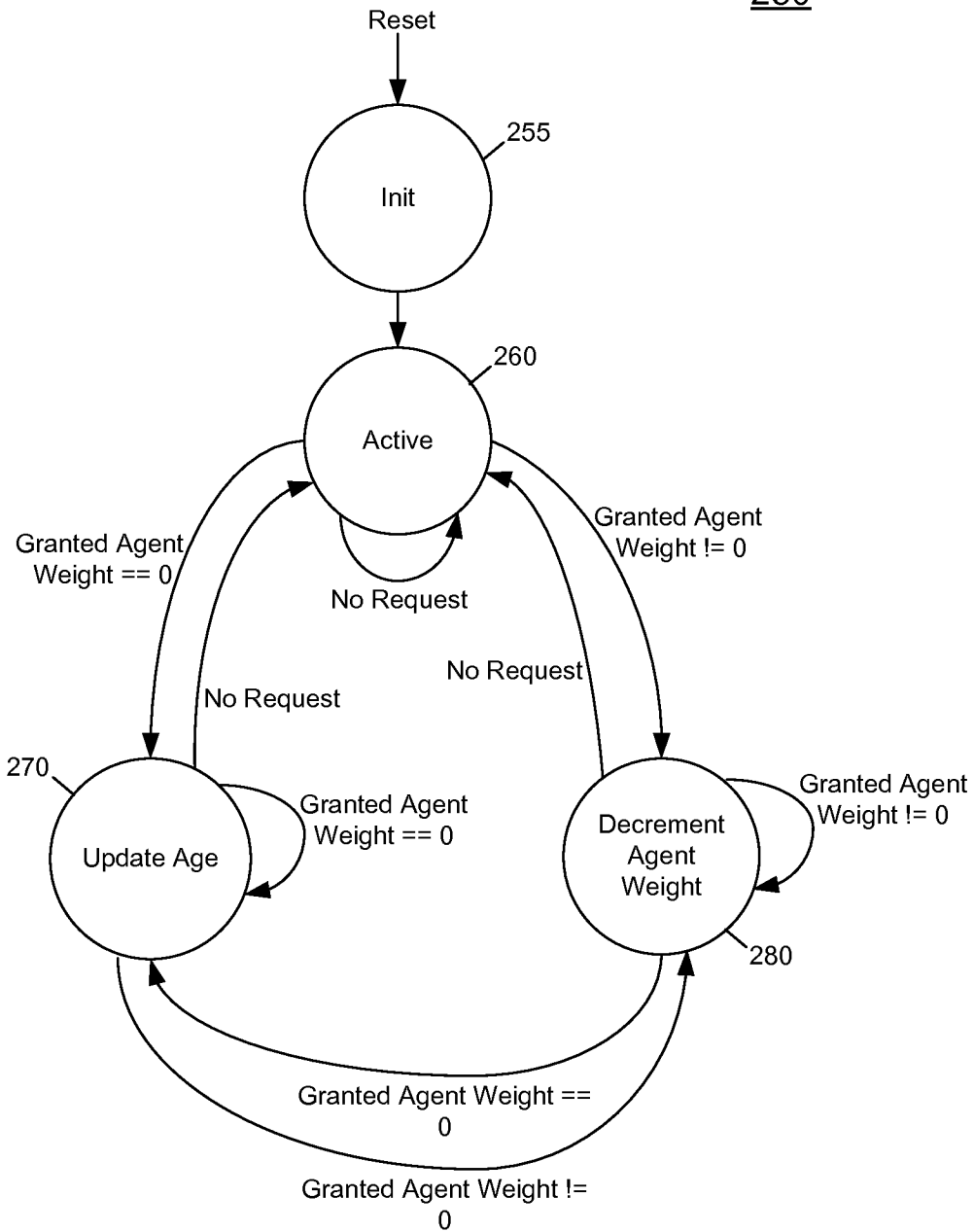


FIG. 4

300

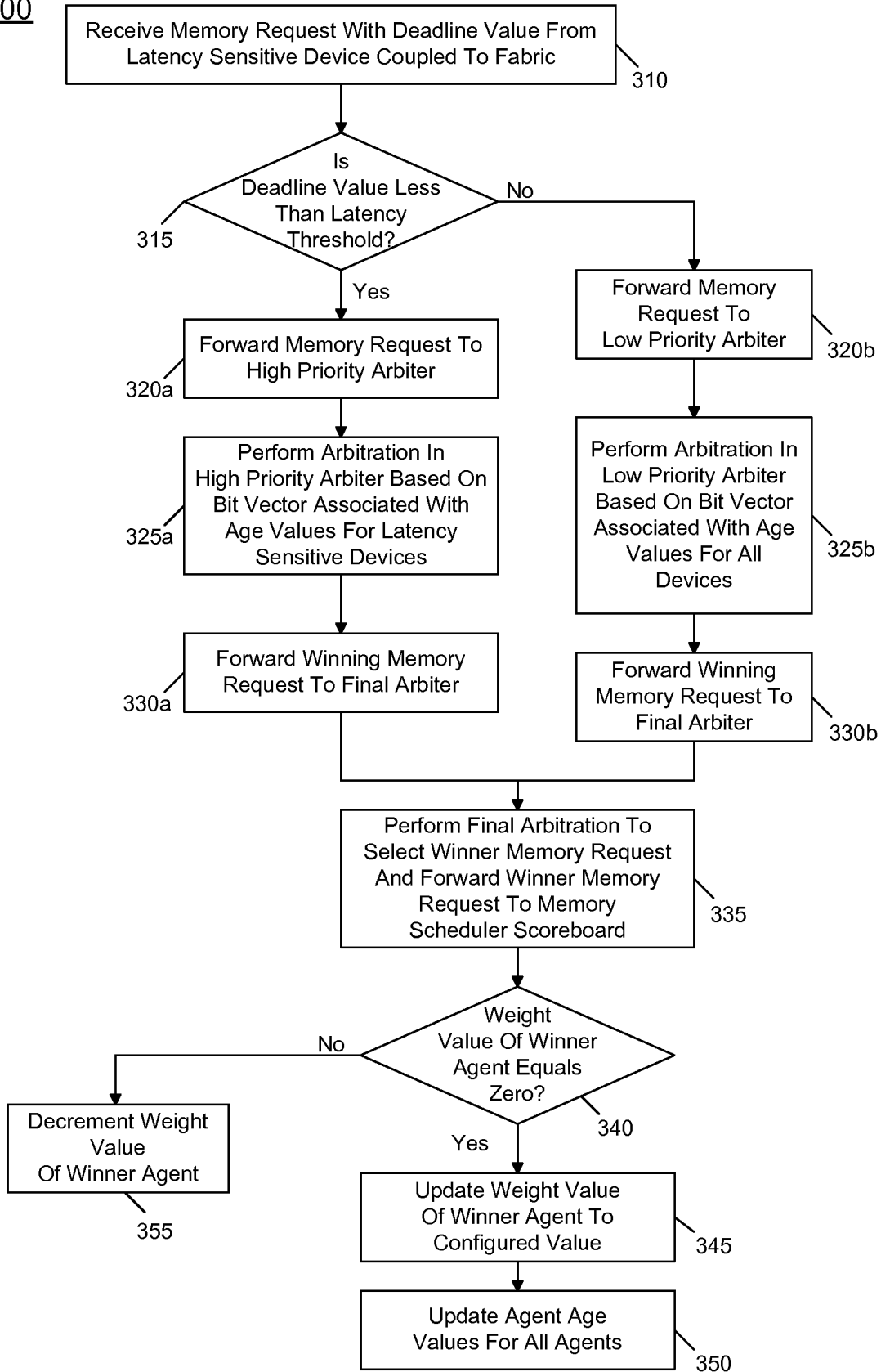


FIG. 5

360

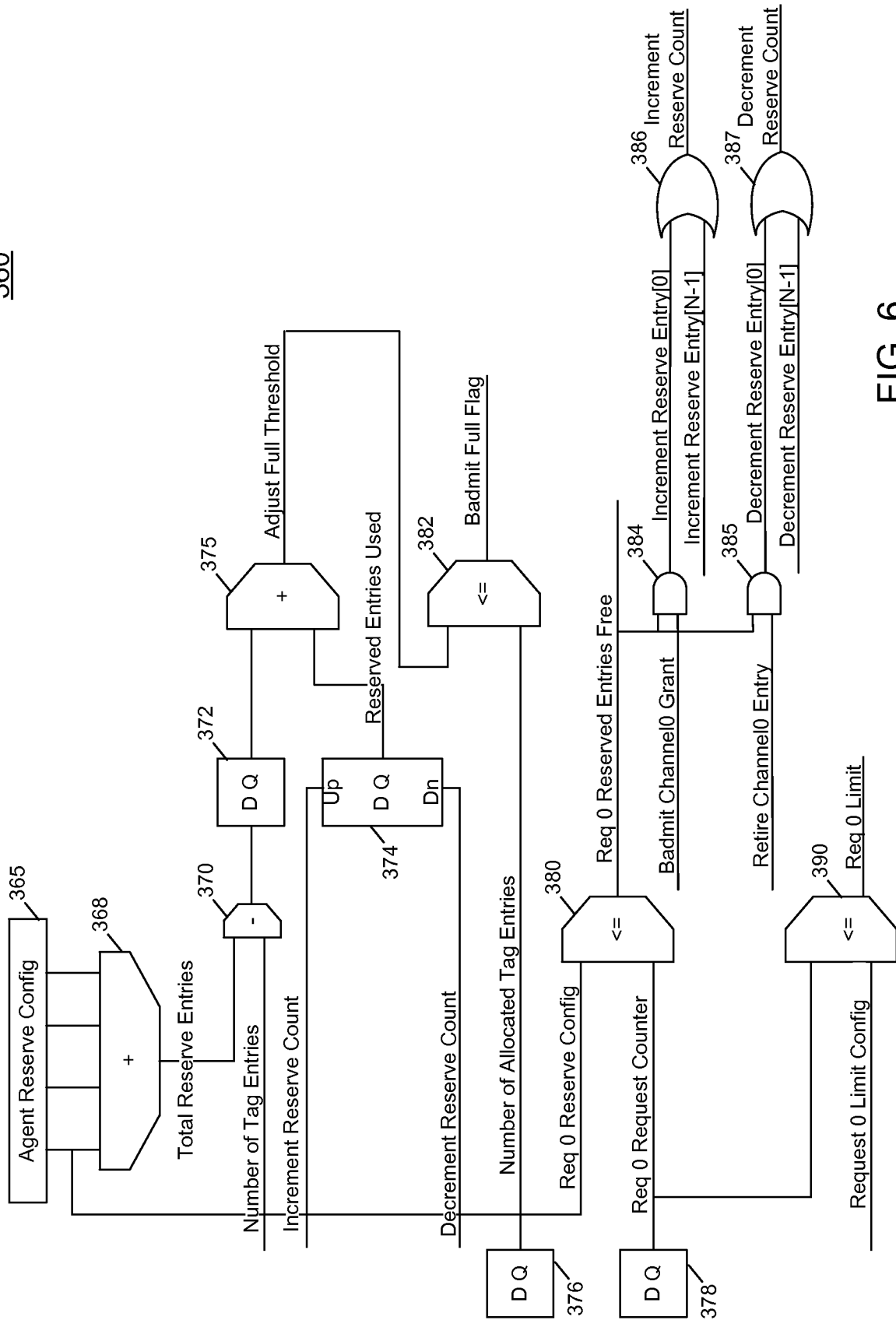
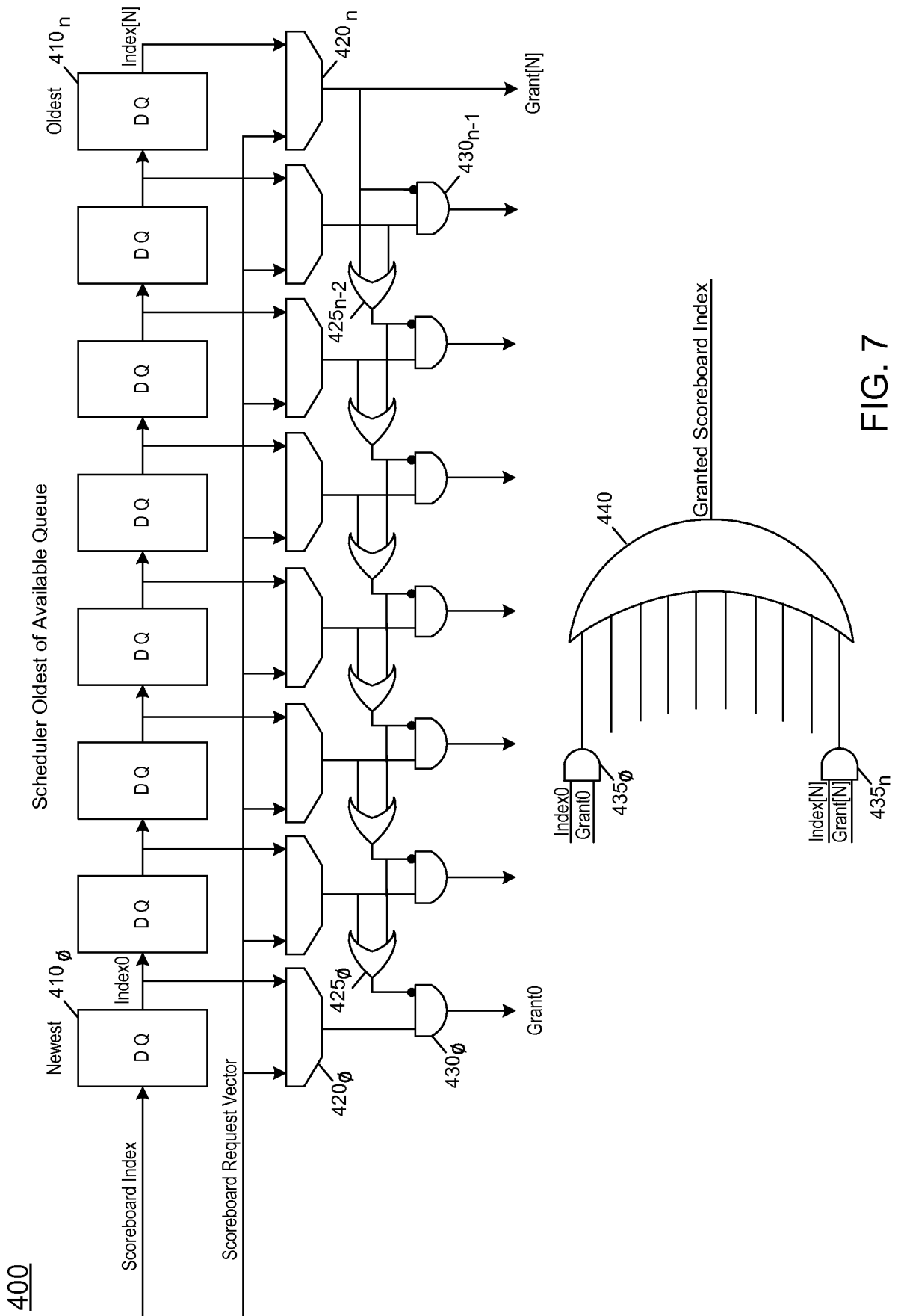


FIG. 6



500

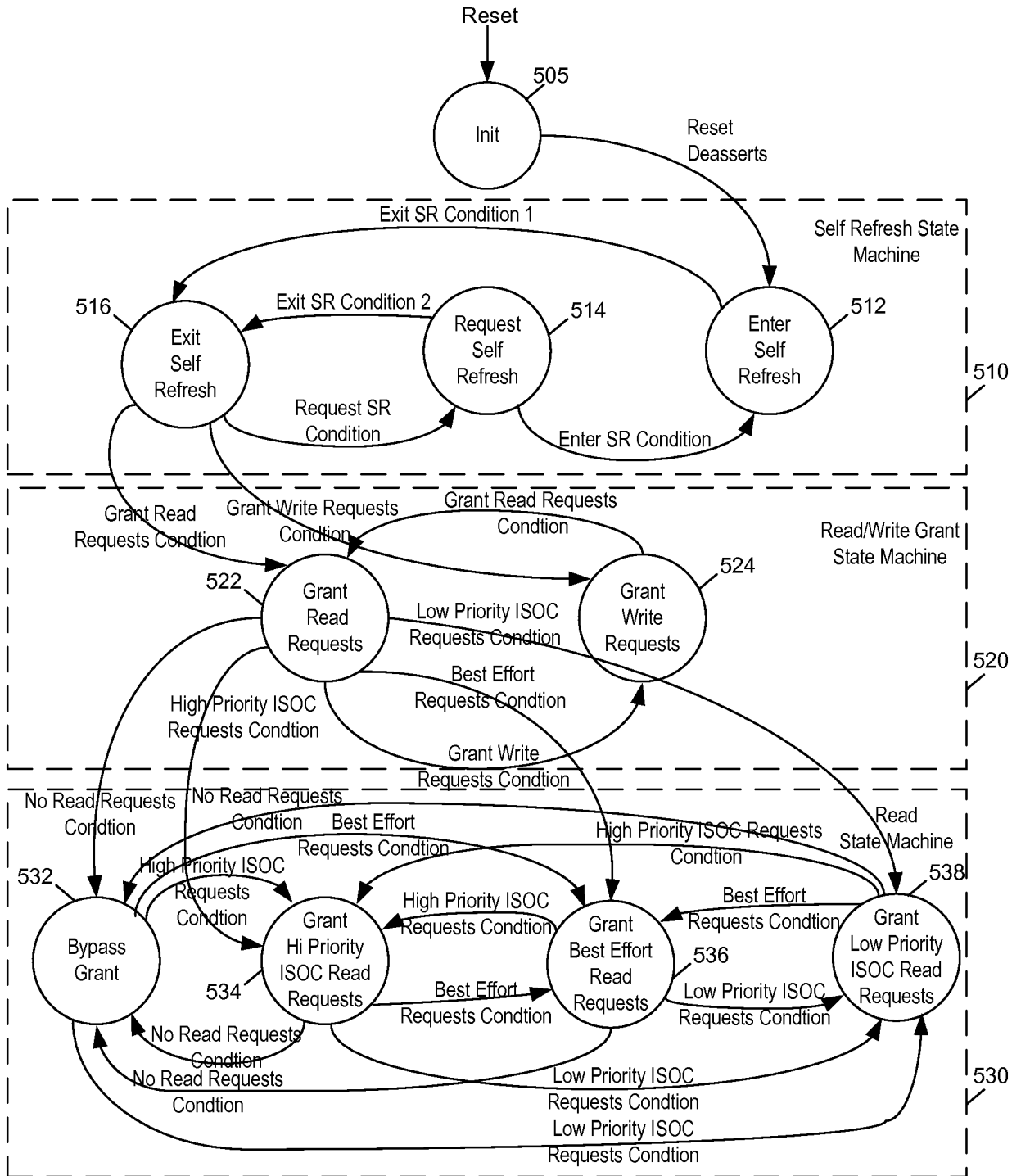


FIG. 8

9/11

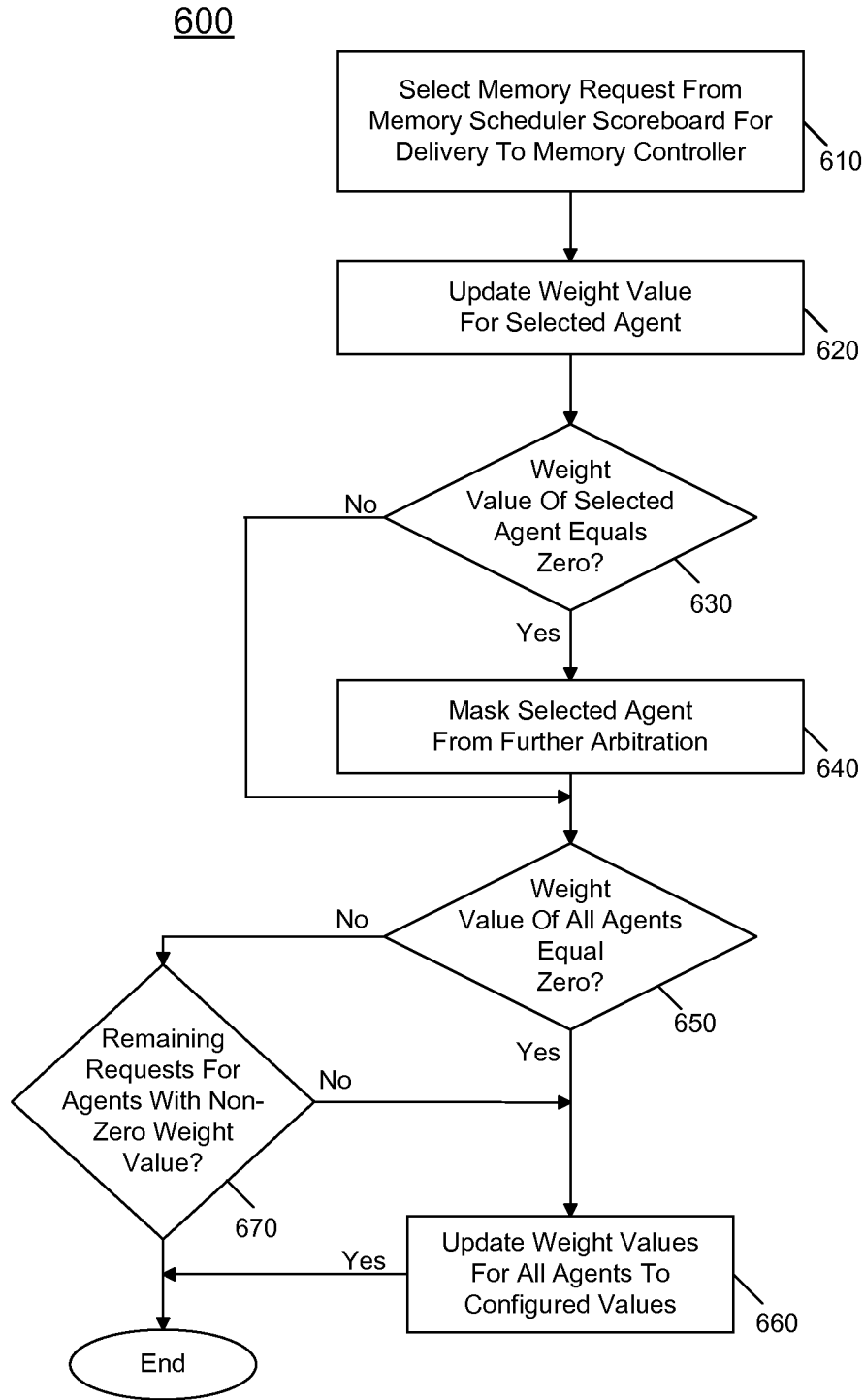


FIG. 9

700

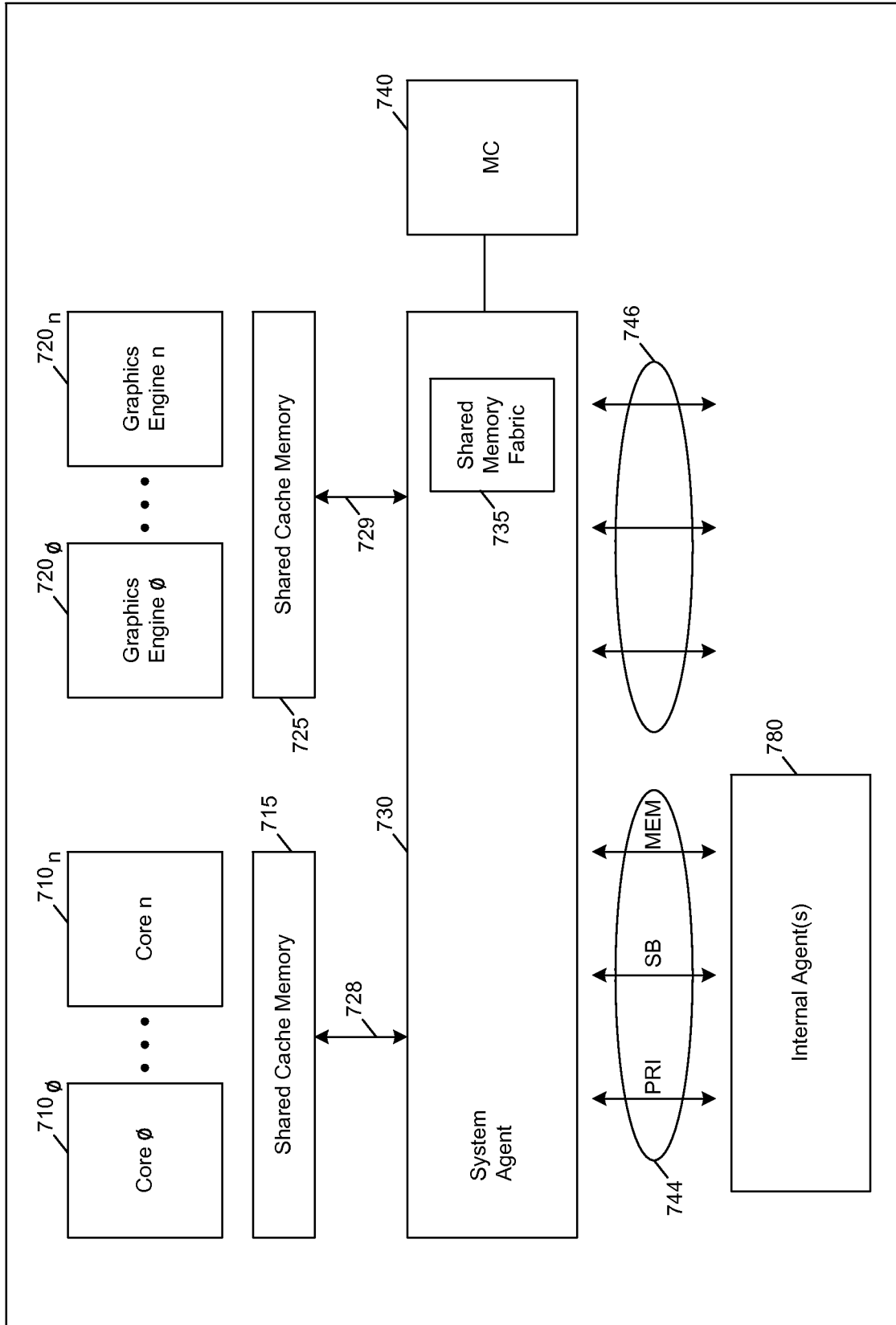


FIG. 10

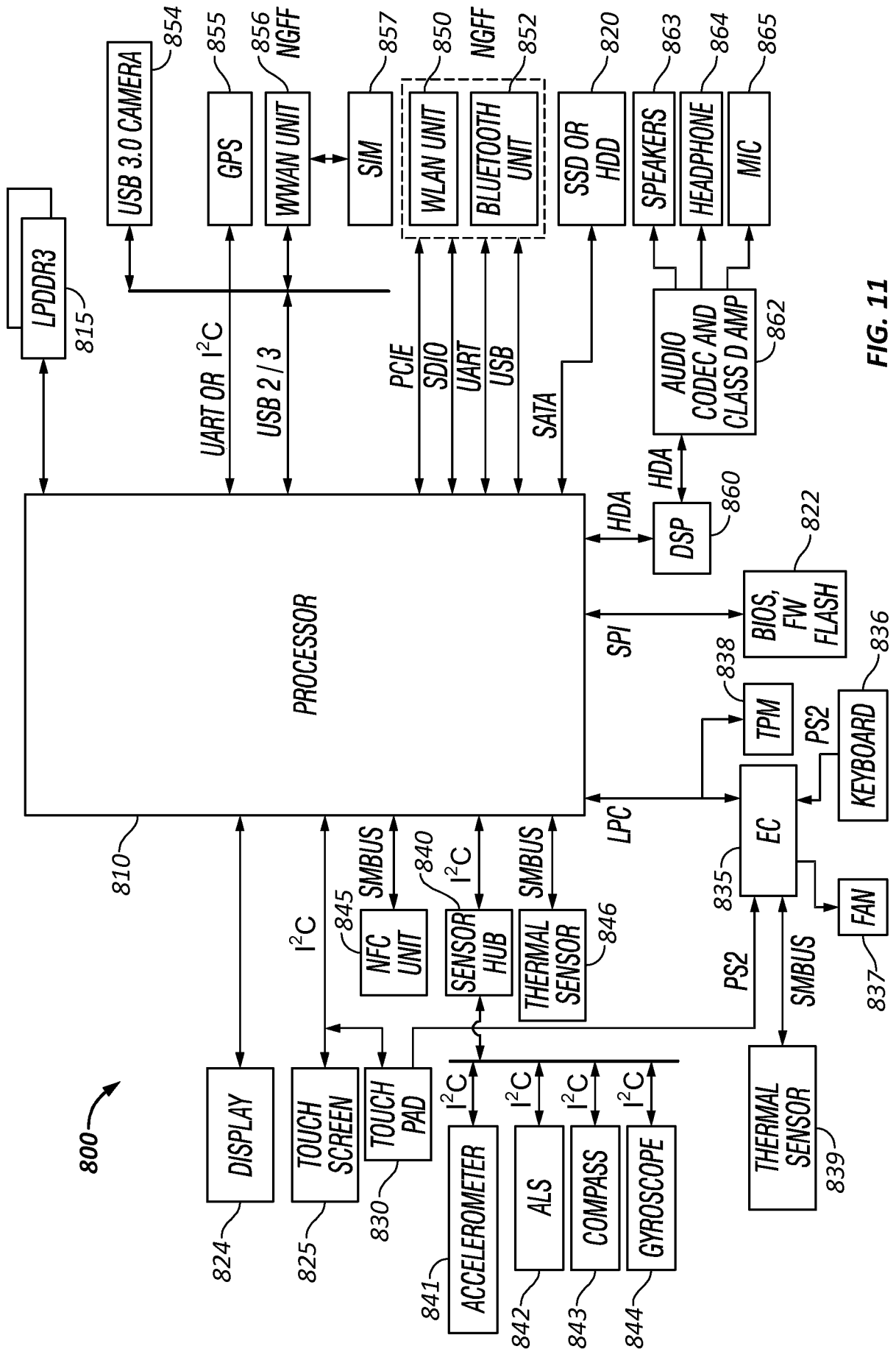


FIG. 11

A. CLASSIFICATION OF SUBJECT MATTER**G06F 13/14(2006.01)i, G06F 12/00(2006.01)i**

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)

G06F 13/14; G06F 12/00; G06F 13/36; G06F 13/28; G06F 13/00

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

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

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

eKOMPASS(KIPO internal) & Keywords: arbitration, intellectual property, core, block, request

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2012-0173786 A1 (MOSHE B. SIMON et al.) 05 July 2012 See paragraphs [0002]-[0012], [0014]-[0020], [0034], [0043]; and figure 1.	1, 10
Y		2
A		3-9, 11-29
Y	US 8190804 B1 (KRISHNAN SRINIVASAN et al.) 29 May 2012 See column 1, lines 34-55; column 2, line 60 - column 5, line 21; and figures 1, 4.	2
A		1, 3-29
A	US 2010-0115167 A1 (JEAN-LOUIS TARDIEUX et al.) 06 May 2010 See paragraphs [0002]-[0013], [0016]-[0020], [0032]-[0033], [0035], [0039], [0063]-[0066], [0068]-[0077]; and figures 1A-1B, 5-6.	1-29
A	US 2011-0238941 A1 (KUN XU et al.) 29 September 2011 See paragraphs [0011]-[0016], [0022]-[0024], [0039], [0042]-[0049]; and figures 1-2, 6.	1-29

 Further documents are listed in the continuation of Box C. See patent family 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 application or patent 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

"&" document member of the same patent family

Date of the actual completion of the international search

22 May 2014 (22.05.2014)

Date of mailing of the international search report

23 May 2014 (23.05.2014)

Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon Metropolitan City, 302-701,
Republic of Korea

Facsimile No. +82-42-472-7140

Authorized officer

JANG, Ho Keun

Telephone No. +82-42-481-8187



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2014/010275

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2012-0173786 A1	05/07/2012	US 2010-005470 A1 US 2014-019664 A1 US 8151008 B2	07/01/2010 16/01/2014 03/04/2012
US 8190804 B1	29/05/2012	US 8484397 B1	09/07/2013
US 2010-0115167 A1	06/05/2010	CN 101739372 A CN 101739372 B EP 2194461 A2 EP 2194461 A3 EP 2194461 B1 EP 2538338 A1 EP 2538338 B1 EP 2538339 A1 EP 2538339 B1 EP 2538340 A1 EP 2544095 A2 EP 2544095 A3 TW 201019129 A US 8032678 B2	16/06/2010 07/08/2013 09/06/2010 01/09/2010 19/12/2012 26/12/2012 12/02/2014 26/12/2012 12/02/2014 26/12/2012 09/01/2013 01/05/2013 16/05/2010 04/10/2011
US 2011-0238941 A1	29/09/2011	US 8560796 B2 WO 2011-126625 A2 WO 2011-126625 A3	15/10/2013 13/10/2011 22/12/2011