(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 August 2011 (04.08.2011)

PCT

(10) International Publication Number
WO 2011/093877 A1

(51) International Patent Classification:
G06F 15/16 (2006.01) *G06F 15/163* (2006.01)

(21) International Application Number:
PCT/US2010/022563

(22) International Filing Date:
29 January 2010 (29.01.2010)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US)*: **HEWLETT-PACKARD DEVELOPMENT COMPANY, L. P.** [US/US]; 11445 Compaq Center Drive W., Houston, Texas 77070 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **CEPULIS, Darren, J.** [US/US]; 11445 Compaq Center Drive W., Houston, Texas 77007 (US). **REZA, Masud, M.** [US/US]; 11445 Compaq Center Drive W., Houston, Texas 77070 (US). **STEARNS, Michael** [US/US]; 11445 Compaq Center Drive W., Houston, Texas 77070 (US). **HUA, Chanh, V.** [US/US]; 11445 Compaq Center Drive W., Houston, Texas 77070 (US).

(74) Agents: **WEBB, Steven, L.** et al.; Hewlett-Packard Company, Intellectual Property Administration, Mail Stop 35, 3404 E. Harmony Road, Fort Collins, CO 80528-9599 (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, 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, IS, JP, KE, KG, KM, 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, PE, PG, PH, PL, PT, RO, RS, RU, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i))*

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *with international search report (Art. 21(3))*

(54) **Title**: METHODS AND SYSTEMS FOR AN INTERPOSER BOARD

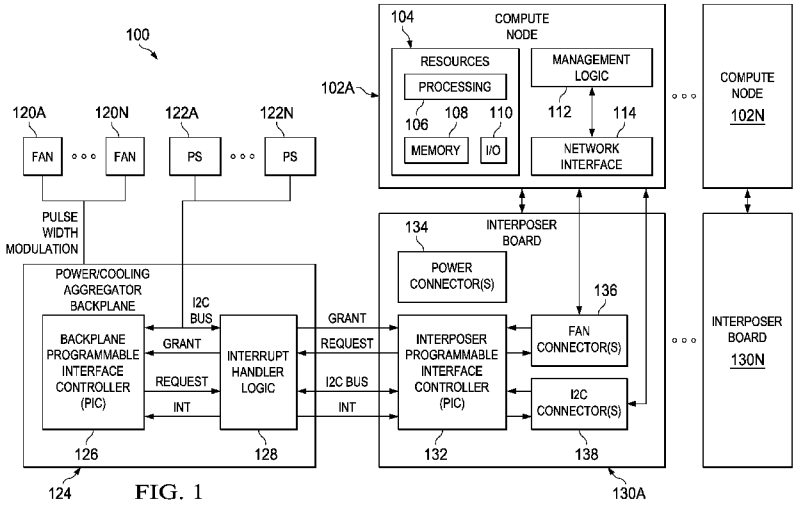


FIG. 1

(57) **Abstract**: In accordance with at least some embodiments, a system (100) includes an aggregator backplane (124) coupled to a plurality of fans (120A-120N) and power supplies (122A-122N) and configured to consolidate control and monitoring for the plurality of fans (120A-120N) and power supplies (122A-122N). The system (100) also includes a plurality of compute nodes (102A-102N) coupled to the aggregator backplane (124), wherein each compute node (102A-102N) selectively communicates with the aggregator backplane (124) via a corresponding interposer board (130A-130N). Each interposer board (130A-130N) is configured to translate information passed between its corresponding compute node (102A-102N) and the aggregator backplane (124).

WO 2011/093877 A1

METHODS AND SYSTEMS FOR AN INTERPOSER BOARD

BACKGROUND

[0001] There are many types of computer architectures. Some computer architectures combine multiple compute nodes in a shared resource enclosure. Such architectures may require customizing the power, cooling and management of each compute node being implemented in the shared resource enclosure. As an example, such customization may involve significant changes to core firmware and hardware related to the baseboard management controller (BMC) in each compute node. Such customization may have one or more of the following problems: 1) high design costs; 2) long development cycles; 3) limited choices for compute nodes due to limited resources and schedule conflicts.

BRIEF DESCRIPTION OF THE DRAWINGS

[0002] For a detailed description of exemplary embodiments of the invention, reference will now be made to the accompanying drawings in which:

[0003] Fig. 1 illustrates a system in accordance with an embodiment of the disclosure;

[0004] Fig. 2 illustrates a server enclosure in accordance with an embodiment of the disclosure;

[0005] Figs. 3 shows a table of microcontroller registers for an interposer programmable interface controller (PIC) in accordance with an embodiment of the disclosure; and

[0006] Fig. 4 illustrates a method in accordance with an embodiment of the disclosure.

NOTATION AND NOMENCLATURE

[0007] Certain terms are used throughout the following description and claims to refer to particular system components. As one skilled in the art will appreciate, computer companies may refer to a component by different names. This document does not intend to distinguish between components that differ in name but not function. In the following discussion and in the claims, the terms “including” and “comprising” are used in an open-ended fashion, and thus should be interpreted to mean “including, but not limited to… .” Also, the term “couple” or

"couples" is intended to mean either an indirect, direct, optical or wireless electrical connection. Thus, if a first device couples to a second device, that connection may be through a direct electrical connection, through an indirect electrical connection via other devices and connections, through an optical electrical connection, or through a wireless electrical connection. Also, the term "server enclosure" means a server system chassis capable of hosting multiple compute nodes with common infrastructure (*e.g.*, power supplies and fans). As used herein, a "server rack" may contain multiple of such server enclosures.

DETAILED DESCRIPTION

[0008] The following discussion is directed to various embodiments of the invention. Although one or more of these embodiments may be preferred, the embodiments disclosed should not be interpreted, or otherwise used, as limiting the scope of the disclosure, including the claims. In addition, one skilled in the art will understand that the following description has broad application, and the discussion of any embodiment is meant only to be exemplary of that embodiment, and not intended to intimate that the scope of the disclosure, including the claims, is limited to that embodiment.

[0009] In accordance with embodiments of the disclosure, an interposer board is implemented for each of a plurality of compute nodes housed in an enclosure for a computing system (*e.g.,* a server). Each interposer board interfaces its corresponding compute node to an aggregator backplane that manages the fans and power supplies for the server enclosure. Each interposer board provides predetermined functions that simplify the configuration of compute nodes in a server enclosure with an aggregator backplane. Without the interposer boards, each compute node would need to be designed or customized to communicate directly with the aggregator backplane. Such customization is possible, but is costly and/or results in undesirable latency in the release of new products (*e.g.,* multi-node server enclosure and/or racks with multiple server enclosures).

[0010] Fig. 1 illustrates a system 100 in accordance with an embodiment of the disclosure. The system 100 may represent, for example, components found in a server enclosure. As shown in Fig. 1, the system 100 comprises a plurality of compute nodes 102A-102N. Each compute node 102A-102N comprises a

motherboard or other printed circuit board (PCB) containing resources 104 such as any one or more of processing resources 106, memory resources 108 and input/output (I/O) resources 110. Further, each of the compute node 102A-102N may comprise a network interface 114 coupled to management logic 112. For example, the management logic 112 may monitor and/or log sensor data. In some embodiments, the management logic 112 is configured to transmit fan control signals and/or power supply control signals in response to detected temperature/electrical conditions. Further, the management logic 112 may handle remote computing requests received via the network interface 114. The management logic 122 may correspond to, for example, a baseboard management controller (BMC) known in the art.

[0011] In the system 100, a plurality of fans 120A-120N provide airflow/cooling for the plurality of compute nodes 102A-102N. Fig. 2 illustrates a server enclosure 200 in accordance with an embodiment of the disclosure. As shown in Fig. 2, the server enclosure 200 comprises eight compute nodes (nodes 1-8) that share a plurality of fans (*e.g.,* fans 1-8). In general, nodes (*e.g.,* nodes 1-8) are organized into zones, with each zone having at least one node and one fan assigned thereto. More specifically, the server enclosure 200 in the embodiment of Fig. 2 has four zones with two nodes and two fans in each zone. As shown, nodes 1 and 3 share two fans, nodes 2 and 4 share two fans, nodes 5 and 7 share two fans, and nodes 6 and 8 share two fans. Other zone configurations are possible and may vary depending on the size of the enclosure, the size of the compute nodes boards and components attached thereto, the heat generated by each compute node, the airflow provided by each fan and/or other characteristics.

[0012] Returning to Fig. 1, a plurality of power supplies 122A-122N provide power for the plurality of compute nodes 102A-102N. The fans 120A-120N and the power supplies 122A-122N are coupled to an aggregator backplane 124 that consolidates monitoring and control of the fans 120A-120N and the power supplies 122A-122N. In at least some embodiments, the aggregator backplane 124 comprises a backplane programmable interface controller (PIC) 126 (*e.g.,* a microcontroller) coupled to interrupt handler logic 128 (*e.g.,* a complex programmable logic device (CPLD)). The backplane PIC 126 consolidates fan

status information received from the fans 120A-120N as well as power supply status information received from the power supplies 122A-122N. The backplane PIC 126 also responds to control signals or requests for information received from the compute nodes 102A-102N.

[0013] In accordance with at least some embodiments, information and/or control signals from the aggregator backplane 124 are selectively transmitted to each of the compute nodes 102A-102N via its corresponding interposer board 130A-130N. Similarly, information and/or control signals from each of the compute nodes 102A-102N are transmitted via its corresponding interposer board 130A-130N to the aggregator backplane 124. In accordance with at least some embodiments, the interrupt handler logic 128 facilitates communications between the backplane PIC 126 and each interposer board 130A-130N using interrupt protocols.

[0014] In Fig. 1, each of the interposer boards 130A-130N comprise the same or similar components. For convenience, only the components of the interposer board 130A are shown and discussed, but it should be understood that the discussion of interposer board 130A applies to the other interposer boards (130B-130N) as well. In Fig. 1, the interposer board 130A corresponds to a printed circuit board (PCB) with an interposer programmable interface controller (PIC) 132 mounted thereon and configured to perform various functions as will be described herein. The interposer board 130A also comprises at least one power connector 134 mounted on the PCB to provide a power interface from at least one of the power supplies 122A-122N to the corresponding compute node. The power interface provided by the power connector 134 also enables the aggregator backplane 126 to monitor the power consumption of the corresponding compute node. The interposer board 130A also comprises at least one fan connector 136 mounted on the PCB and coupled to the interposer PIC 132. The fan connector 136 provides an interface for the corresponding compute node to issue fan control signals (requests) and to receive responses to such requests. The interposer board 130A also comprises a serial bus connector (*e.g.,* an I2C bus connector) 138 mounted to the PCB and coupled to the interposer PIC 132. The serial bus connector 128 provides a serial communication interface

(*e.g.,* a single-master I2C bus) between the interposer board 130A and its corresponding compute node.

[0015] In accordance with at least some embodiments, the interposer PIC 132 translates information passed between the aggregator backplane 126 and the compute node corresponding to interposer board 130A. For example, the interposer PIC 132 may translate fan control signals from pulse width modulation (PWM) to a serial bus protocol data packet (*e.g.,* I2C data packets). More specifically, the interposer PIC 132 may receive fan control PWM information from its corresponding compute node and provide a translated version of the fan control PWM information to the backplane PIC 126 via an I2C bus. Further, the interposer PIC 132 may translate fan status signals from Fault to Tach fan emulation. More specifically, the interposer PIC 132 may provide a fan tachometer signal to its corresponding compute node based on a PWM duty cycle and the actual fan status.

[0016] In at least some embodiments, the interposer PIC 132 receives compute node fan PWM input and digitizes the input in terms of duty cycle. The digitized duty cycle is then passed to the backplane PIC 126 via a multi-master I2C bus. As an example, a fan PWM value of 0 will represent a 0% duty cycle and 100 will represent a 100% duty cycle. Optionally, a compute node BMC may write the fan PWM settings directly to registers of the interposer PIC 132. In some embodiments, digitization of the fan PWM signals may be accomplished based on direct current (DC) conversion. In DC conversion, the PWM signal is converted to an analog DC signal and is digitized via analog-to-digital conversion. Alternatively, digitization of PWM signals may be accomplished using a timer and a capture/compare (CCP) technique. In the timer and CCP technique, the interposer PIC analyzes the PWM signal as a digital input and calculates the duty cycle using internal timers and CCP.

[0017] The interposer PIC 132 also reads system fan status information from the backplane PIC 126 via an I2C bus. Once the fan status information has been read, the interposer PIC 132 is able to generate a fan tach signal to drive a fan tach signal of a compute node. If any of the system fans fail, a tach signal is not generated by the interposer PIC 132 for the failed fan. Otherwise, the

interposer PIC 132 drives the fan tach signal at a frequency corresponding to the PWM of the fan using the equation FanTachCount per Sec = (PWM_DC/100) * (MaxFanRPM/60) * (FanTachPulse/Rev), where the value of FanTachPulse/Rev is normally 2 and MaxFanRPM is the Fan RPM at 100% fan PWM duty cycle.

[0018] In at least some embodiments, the interposer PIC 132 is configured to bridge a single-master serial bus topology (*e.g.,* single-master I2C) for communications between the interposer board 130A and its corresponding compute node with a multi-master serial bus topology (*e.g.,* multi-master I2C) for communications between the interposer board 130A and the aggregator backplane 126. For example, the interposer PIC 132 may provide I2C MUX arbitration based on request/grant (REQ/GNT) signals for an I2C bus. In other words, the interposer PIC 132 acts as an I2C pass thru between the BMC of a compute node and the backplane PIC 126 to avoid multi-master I2C issues.

[0019] In relation to the bridge function, the interposer PIC 132 may also manage a traffic level for the multi-master serial bus topology. For example, in at least some embodiments, the interposer PIC 132 is configured to manage consolidation and caching of data such as fan status information, power supply status information, and power consumption meter information. Using the cached information, the interposer PIC 132 is able to selectively respond to requests from a corresponding compute node without requesting updated information from the aggregator backplane 124. For example, the interposer PIC 132 may have a traffic control timing threshold (*e.g.,* 2 seconds) during which all requests (*e.g.,* fan control signals, requests for information) from the compute node corresponding to the interposer board 130A are responded to from cached information within the interposer PIC 132. Once the traffic control timing threshold is reached, the interposer PIC 132 may request updated information from the aggregator backplane 124. The request for updated information may be automated or may be in response to a related request from the compute node corresponding to the interposer board 102A. Further, in at least some embodiments, the aggregator backplane 124 is able to transmit updated information to the interposer PIC 132 without waiting for the interposer PIC 132 to

issue or forward a request. The traffic control timing threshold described previously may vary according to predetermined criteria (*e.g.,* the number of compute nodes in a given system) or ongoing traffic analysis.

[0020] In at least some embodiments, the interposer PIC 132 enables flashing firmware of the backplane PIC 126 via a single-master I2C bus. For example, during the flash process, the interposer PIC 132 receives each line of code from a compute node via a single-master I2C bus and stores the code internally. The interposer PIC 132 then obtains access to the multi-master I2C bus by generating a request (REQ) and receiving a grant (GNT) from the backplane PIC 126. Once the interposer PIC 132 has mastership of the multi-master I2C bus, the line of code is passed to the backplane PIC 126 via the multi-master I2C bus and so on.

[0021] During the flash process of the backplane PIC 126, the interposer PIC 132 has full functionality. In some embodiments, the backplane PIC 126 may set the system fans to full speed with 100% duty cycle to avoid any thermal events during flashing. During the flash, the interposer PIC 132 may return predetermined values (*e.g.,* the last known power supply status, the last known fan status, a 100% duty cycle for current fan PWM value) in response to respective requests/commands received during the flash process.

[0022] In at least some embodiments, the firmware of the interposer PIC 132 may be flashed via the single-master I2C bus as well. To perform the flash, a compute node accesses the registers of the interposer PIC 132 by writing a value to the interposer PIC Mailbox register. The compute node also writes a firmware update key to a firmware update keyhole register to put the interposer PIC in bootloader mode. In at least some embodiments, the bootloader mode of the interposer PIC 132 has various attributes. For example, the interposer PIC 132 will limit access of its registers to flash-related functions. In other words, access to non-flash related registers will be NACK'd during the bootloader mode. Once the flash process completes (*e.g.,* usually between 30 seconds to 3 minutes), the interposer PIC 132 automatically resets itself and bring itself back on-line. The reset process may be completed in less than a second and does not require the multi-node system or any of the compute nodes to cycle power or

reset. In the bootloader mode, the computer node corresponding to the interposer PIC 132 will not have access to the registers of the backplane PIC 126. However, the interposer PIC 132 continues to generate the fan tach signals at 100% PWM duty cycle for both the fans. Once the flash process completes, the interposer PIC 132 resets itself and the compute node BMC will have access to all the registers of the backplane PIC 126 again. It is expected that the BMC of the corresponding compute node will recognize the flash process of the interposer PIC 132 and avoid logging errors. If a power loss occurs during the flash process (*e.g.,* the interposer PIC 132 loses its VDD or the PIC reset pin is asserted low for a duration that causes a PIC reset), then the interposer PIC 132 returns to the bootloader mode once power is restored. The BMC of the corresponding compute node is then able to re-start the flash process.

[0023] In Fig. 1, the interrupt handler logic 128 of the aggregator backplane 124 manages communications between the backplane PIC 126 and each interposer board 130A-130N. As shown, an I2C bus (or other serial communication bus) is utilized for communications between the backplane PIC 126 and the interrupt handler logic 128. The timing of serial communications is controlled by a grant signal (GNT), a request signal (REQ) and an interrupt signal (INT). More specifically, the backplane PIC 126 may assert the REQ signal to request mastership of the I2C bus. The assertion of the REQ signal may be based on predetermined operations of the backplane PIC 126 or the INT signal being asserted by the interrupt handler logic 128 to the backplane PIC 126. In response to the REQ signal being asserted by the backplane PIC 126, the interrupt handler logic 128 selectively grants mastership of the I2C bus to the backplane PIC 126 and asserts the GNT signal to notify the backplane PIC 126 regarding the grant of mastership. In response to the GNT signal being asserted, the backplane PIC 126 is able to transmit information to the interrupt handler logic 128 via the I2C bus.

[0024] Similarly, an I2C bus (or other serial communication bus) is utilized for communications between the interrupt handler logic 128 and the interposer board 132. Again, the timing of serial communications is controlled by a grant

signal (GNT), a request signal (REQ) and an interrupt signal (INT). More specifically, the interposer PIC 132 may assert the REQ signal to request mastership of the I2C bus. The assertion of the REQ signal may be based on predetermined operations of the interposer PIC 132 or the INT signal being asserted by the interrupt handler logic 128 to the interposer PIC 132. In response to the REQ signal being asserted by the backplane PIC 126, the interrupt handler logic 128 selectively grants mastership of the I2C bus to the interposer PIC 132 and asserts the GNT signal to notify the interposer PIC 132 regarding the grant of mastership. In response to the GNT signal being asserted, the interposer PIC 132 is able to transmit information to the interrupt handler logic 128 via the I2C bus. The interrupt handler logic 128 is thus able to route information from the backplane PIC 126 to each interposer board 130A-130N. Similarly, the interrupt handler logic 128 is able to route information from each interposer board 130A-130N to the backplane PIC 126.

[0025] The system 100 described for Fig. 1 may be understood as an abstraction architecture that reduces development time of a multi-node server enclosure by enabling use of available compute node boards with few, if any, modifications. Further, the fans 120A-120N of the system 100 may be, for example, Tach or Fault. Further, the power supplies 122A-122N may be industry standard or HP's "common slot" power supplies. In operation, the aggregator backplane 124 evaluates, monitors, and controls the power supplies 122A-122N and fans 120A-120N at the initial power-on of the system 100, obviating the need for any of the compute nodes 102A-102N of performing this task. The aggregator backplane 124 also performs power metering and limiting for the entire system 100 (*e.g.,* enclosed in an enclosure) as well as for each of the compute nodes 102A-102N. The aggregator backplane 124 also steers relevant management data to and from each of the compute nodes 102A-102N via their corresponding interposer boards 130A-130N. In some embodiments, such steering of management data may be automatic and may be based on the location of each compute node 102A-102N within an enclosure. The backplane PIC 126 also controls the REQ/GNT arbitration for the multi-master I2C bus and thus is aware of which compute node is accessing its register set.

[0026] The abstraction architecture described herein greatly simplifies each compute node's BMC management support, which is normally handled through industry standard, Intelligent Platform Management Interface (IPMI) compatible Sensor Data Records (SDRs) in the BMC firmware. In accordance with at least some embodiments, each of the compute nodes 102A-102N contain an identical set of SDRs and do not have to carry multiple sets of SDRs depending on each compute node's location in the server enclosure, or in which fan zone (or power zone) a particular compute node is participating. In this manner, the complexity of designing the system 100 or another abstraction architecture system is reduced.

[0027] In the abstraction architecture of the system 100, the purpose of each interposer board 130A-130N is to simplify and adapt the power, cooling, and management signals for a single compute node into a multi-node, shared-resource architecture. For example, fan PWM outputs from a compute node are captured either as PWM signals or simple I2C bus writes by its corresponding interposer board and then converted (by an interposer PIC) to multi-master I2C bus transactions passed down to the backplane PIC 126 as fan speed requests. Further, each interposer board may retrieve and cache server management data such as power supply status (*e.g.,* AC OK, DC OK, and redundancy status), fan status, actual fan speed, compute node power consumption, and overall chassis power consumption from the backplane PIC 126 on a periodic basis. The server management data is then quickly accessible to the BMC of the compute node corresponding to the interposer board with cached server management data. In some embodiments, sets of compute nodes (*e.g.,* 8, 16 or more) are supported in a single enclosure. As the number of compute nodes increases in a multi-node shared resource architecture, the advantage of enabling each interposer PIC to control/minimize the traffic on the multi-master I2C bus increases as well.

[0028] In accordance with various industry standards, BMCs usually poll the status of sensors (driven by SDRs) as if the sensors were dedicated to a particular compute node. In the multi-node architecture described herein, the interposer PICs are able to limit the amount of traffic on the shared multi-master

I2C bus. For example, a compute node's BMC may be configured to read fan status once per second. Meanwhile, the interposer PIC 132 is able to read fan status information from the backplane PIC 126 at a fixed slower rate (*e.g.,* every 2 seconds). By having an interposer PIC return a locally cached copy of the fan status to the compute node's BMC, the amount of traffic to the backplane PIC 126 can be reduced without any design changes being required to the compute nodes.

[0029] Fig. 3 shows a table of microcontroller registers for an interposer PIC (*e.g.,* interposer PIC 132) in accordance with an embodiment of the disclosure. As shown, the table comprises an actual fan PWM duty cycle register, a fan1 tachometer count register, a fan2 tachometer count register, a fan fault register, a power supply status register ("PS status register"), a firmware update keyhole register, a fan1 speed request register, a fan2 speed request register, a mailbox register, and reserved registers. In at least some embodiments, only the mailbox register is directly accessible. The remaining registers are accessible only after the mailbox register has been written to with the correct access key. The reserved registers are read-only registers that are not used at this time.

[0030] Fig. 4 illustrates a method 400 in accordance with an embodiment of the disclosure. The method 400 is for interfacing a compute node with an aggregator backplane of a multi-node server. As shown, the method 400 comprises receiving, by an interposer board separate from the compute node, a fan control signal from the compute node via a single-master serial bus (block 402). At block 404, the interposer board translates the received fan control signal. In at least some embodiments, translating fan control signals comprises accessing registers of a programmable interface controller (PIC) such as an actual fan pulse width modulation (PWM) duty cycle register, a fan tachometer count register, and a fan fault register. If a traffic control timer has expired (determination block 406), the interposer board routes the translated fan control signal to an aggregator backplane via a multi-master serial bus (block 408). The routing step of block 408 comprises, for example, bridging a single-master I2C bus topology to a multi-master I2C bus topology while managing a traffic level for the multi-master I2C bus topology. The interposer board then caches and routes

a response to the fan control signal from the aggregator backplane to the compute node (block 410). If the traffic control timer has not expired (determination block 408), the interposer board responds to the translated fan control signal with previously cached information received from the aggregator board (block 412).

[0031] The above discussion is meant to be illustrative of the principles and various embodiments of the present invention. Numerous variations and modifications will become apparent to those skilled in the art once the above disclosure is fully appreciated. It is intended that the following claims be interpreted to embrace all such variations and modifications.

CLAIMS

What is claimed is:

1. A system (100), comprising:

an aggregator backplane (124) coupled to a plurality of fans (120A-120N) and power supplies (122A-122N) and configured to consolidate control and monitoring for the plurality of fans (120A-120N) and power supplies (122A-122N);

a plurality of compute nodes (102A-102N) coupled to the aggregator backplane (124), wherein each compute node (102A-102N) selectively communicates with the aggregator backplane (124) via a corresponding interposer board (130A-130N),

wherein each interposer board (130A-130N) is configured to translate information passed between its corresponding compute node (102A-102N) and the aggregator backplane (124).

2. The system (100) as in claim 1, wherein the aggregator backplane (124) comprises an backplane controller (126) coupled to an interrupt handler component (128), wherein the backplane controller (126) is configured to consolidate control and monitoring for the plurality of fans (120A-120N) and power supplies (122A-122N) and wherein the interrupt handler component (128) is configured to handle interrupts between the backplane controller (126) and each of the interposer boards (130A-130N).

3. The system (100) as in claims 1 or 2, wherein the backplane controller (124) comprises a backplane programmable interface controller (PIC) (126) and the interrupt handler component (128) comprises a complex programmable logic device (CPLD).

4. The system (100) as in claims 1, 2, or 3, wherein each interposer board (130A-130N) comprises an interposer controller (132) configured to translate fan control signals from pulse width modulation (PWM) to I2C data packets and to translate fan status signals from Fault to Tach fan emulation.

5. The system (100) as in claim 4, wherein the interposer controller (132) corresponds to an interposer programmable interface component (PIC) and wherein the interposer PIC is configured to manage data consolidation and caching of fan status information, power supply status information, power consumption meter information.

6. The system (100) as in claim 5, wherein the interposer PIC (132) is configured to bridge a single-master I2C bus topology to a multi-master I2C bus topology and to manage a traffic level for the multi-master I2C bus topology.

7. An interposer board (130A) for interfacing a compute node (102A) with an aggregator backplane (124) of a server enclosure (200), the interposer board (130A) comprising:

a printed circuit board (PCB); and

an interposer programmable interface controller (PIC) (132) mounted on the PCB, the interposer PIC (132) being configured to bridge a single-master serial bus topology for communications between the compute node (102A) and the interposer board (130A) with a multi-master serial bus topology for communications between the interposer board (130A) and the aggregator backplane (124).

8. The interposer board (130A) as is claim 7, wherein the interposer PIC (132) is configured to translate fan control signals from pulse width modulation (PWM) to serial bus topology data packets and to translate fan status signals from Fault to Tach fan emulation.

9. The interposer board (130A) as in claims 7 or 8, wherein the interposer PIC (132) is configured to manage traffic to the multi-master serial bus topology by caching fan status information, power supply status information, and power consumption meter information.

10. The interposer board (130A) as in claims 7, 8, or 9, wherein the interposer PIC (132) comprises a plurality of control registers (Fig. 3) including an actual fan pulse width modulation (PWM) duty cycle register, a fan tachometer count register, and a fan fault register, and a power supply status register.

11. The interposer board (130A) as in claims 7, 8, 9, or 10, further comprising a fan connector (136) mounted on the PCB and coupled to the interposer PIC (132), and a single-master serial bus connector (138) mounted to the PCB and coupled to the interposer PIC (132).

12. A method for interfacing a compute node (102A) with an aggregator backplane (124) of a multi-node server enclosure (200), the method comprising:

translating, by an interposer board (130A) separate from the compute node (102A), fan control signals received from the compute node (102A); and

selectively routing, by the interposer board (130A), translated fan control signals to the aggregator backplane (124) via a multi-master serial bus interface.

13. The method as in claim 12 further comprising caching, by the interposer board (132), server management data from the aggregator backplane (124) for access by the compute node (102A).

14. The method as in claims 12 or 13, further comprising bridging, by the interposer board (132), a single-master I2C bus topology to a multi-master I2C bus topology and managing a traffic level for the multi-master I2C bus topology.

15. The method as in claims 12, 13, or 14, further comprising selectively accessing registers (Fig. 3) of an interposer board programmable interface controller (PIC) (132) including an actual fan pulse width modulation (PWM) duty cycle register, a fan tachometer count register, and a fan fault register.

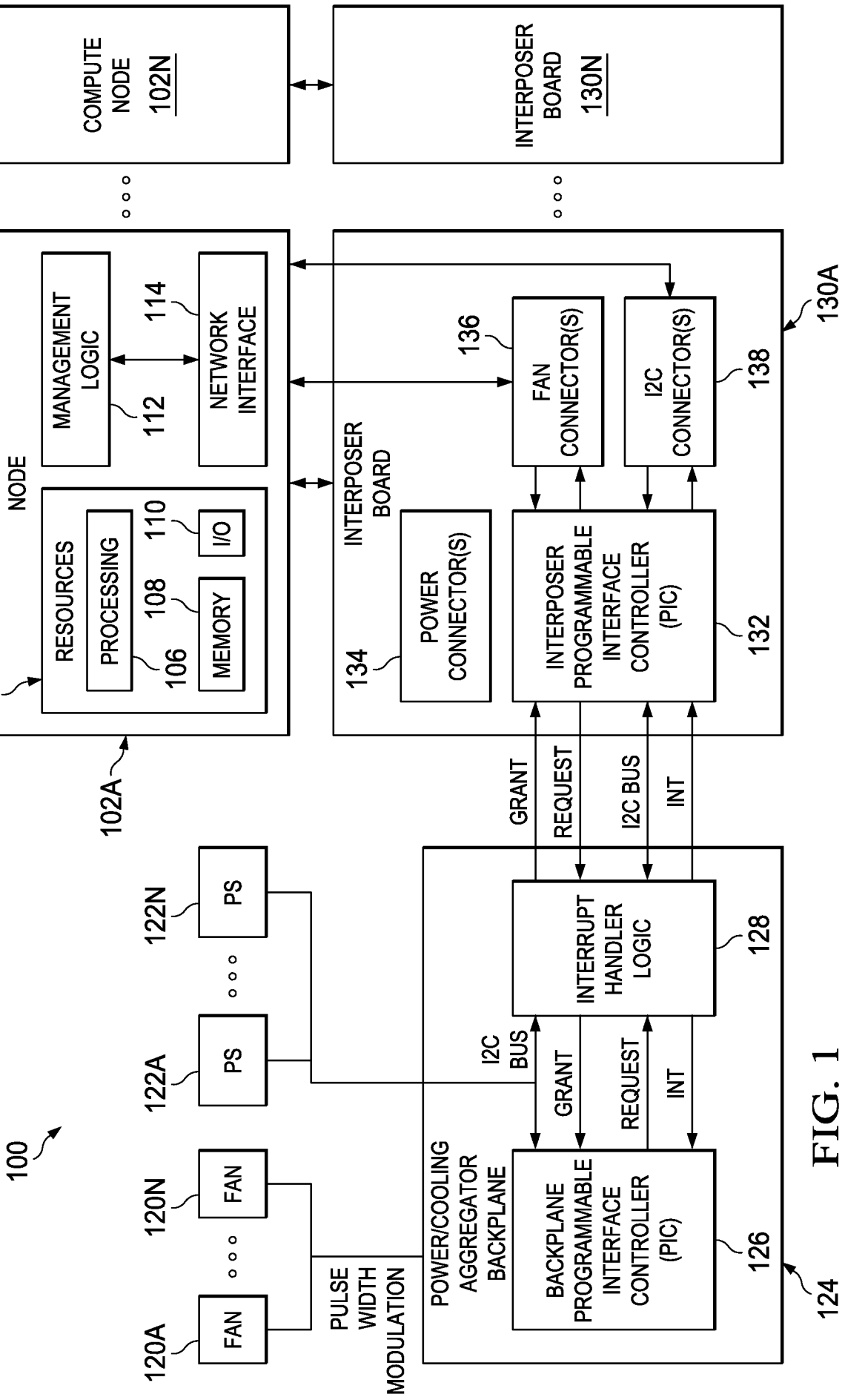

COMPUTE NODE 102N
INTERPOSER BOARD 130N
COMPUTE NODE
MANAGEMENT LOGIC
NETWORK INTERFACE
RESOURCES
PROCESSING
MEMORY
I/O
INTERPOSER BOARD
FAN CONNECTOR(S)
I2C CONNECTOR(S)
POWER CONNECTOR(S)
INTERPOSER PROGRAMMABLE INTERFACE CONTROLLER (PIC)
GRANT
REQUEST
I2C BUS
INT
INTERRUPT HANDLER LOGIC
POWER/COOLING AGGREGATOR BACKPLANE
BACKPLANE PROGRAMMABLE INTERFACE CONTROLLER (PIC)
PULSE WIDTH MODULATION
FAN
PS
100
102A
104
106
108
110
112
114
120A
120N
122A
122N
124
126
128
130A
132
134
136
138

FIG. 1

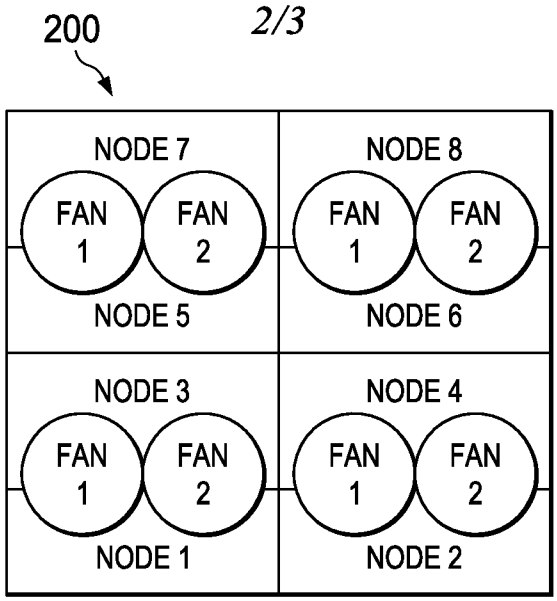

200
NODE 7
FAN 1
FAN 2
NODE 5
NODE 8
FAN 1
FAN 2
NODE 6
NODE 3
FAN 1
FAN 2
NODE 1
NODE 4
FAN 1
FAN 2
NODE 2

FIG. 2

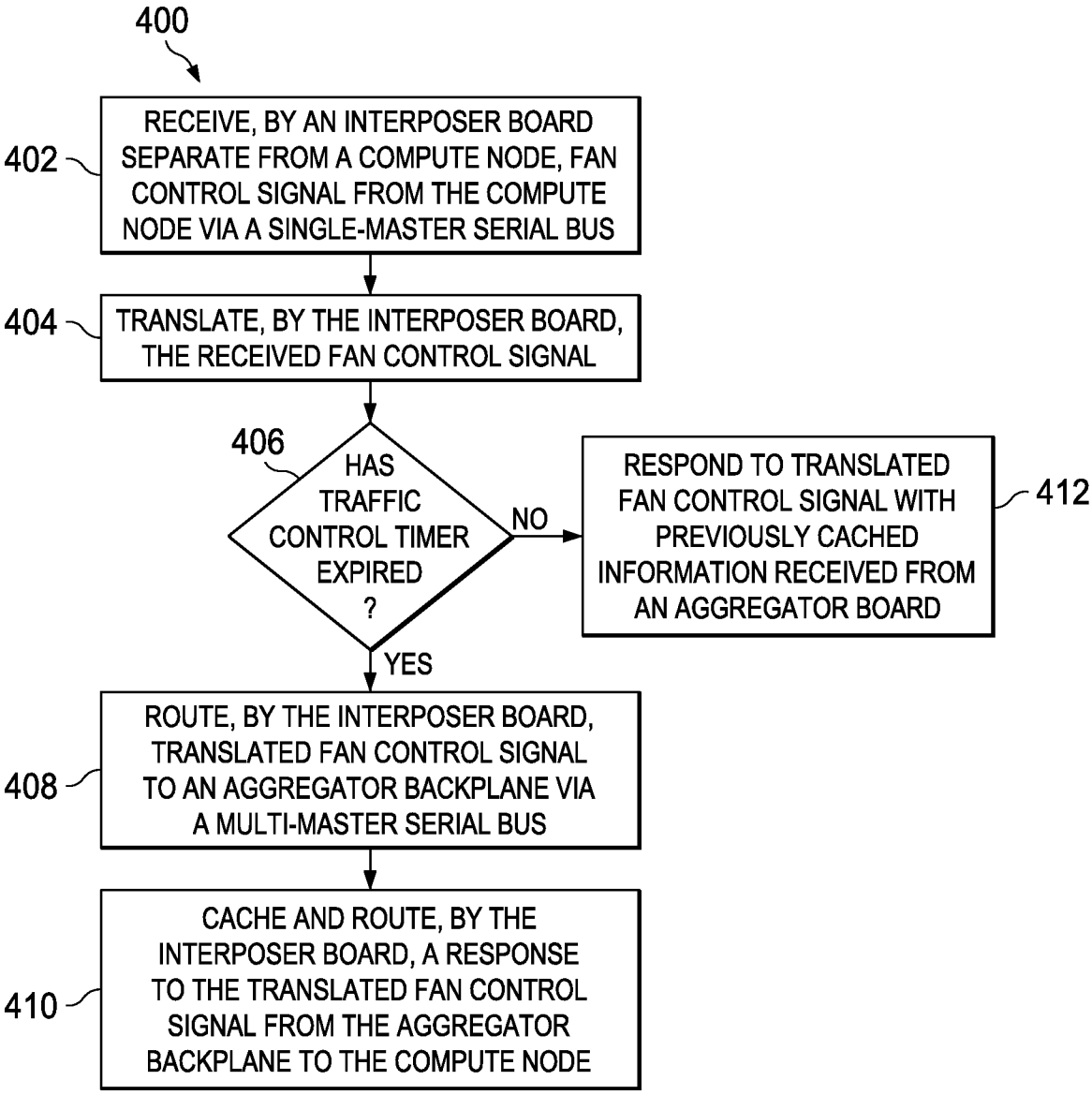

400
402
RECEIVE, BY AN INTERPOSER BOARD SEPARATE FROM A COMPUTE NODE, FAN CONTROL SIGNAL FROM THE COMPUTE NODE VIA A SINGLE-MASTER SERIAL BUS
404
TRANSLATE, BY THE INTERPOSER BOARD, THE RECEIVED FAN CONTROL SIGNAL
406
HAS TRAFFIC CONTROL TIMER EXPIRED ?
NO
412
RESPOND TO TRANSLATED FAN CONTROL SIGNAL WITH PREVIOUSLY CACHED INFORMATION RECEIVED FROM AN AGGREGATOR BOARD
YES
408
ROUTE, BY THE INTERPOSER BOARD, TRANSLATED FAN CONTROL SIGNAL TO AN AGGREGATOR BACKPLANE VIA A MULTI-MASTER SERIAL BUS
410
CACHE AND ROUTE, BY THE INTERPOSER BOARD, A RESPONSE TO THE TRANSLATED FAN CONTROL SIGNAL FROM THE AGGREGATOR BACKPLANE TO THE COMPUTE NODE

FIG. 4

MICROCONTROLLER REGISTERS

REGISTER NAME	ACCESS
ACTUAL FAN PULSE WIDTH MODULATION DUTY CYCLE	RO
FAN 1 TACHOMETER COUNT	RO
FAN 2 TACHOMETER COUNT	RO
FAN FAULT REGISTER	RO
PS STATUS REGISTER	RO
RESERVED	RO
FIRMWARE UPDATE KEYHOLE	RW
FAN 1 (PULSE WIDTH MODULATOR) SPEED REQUEST	RW
FAN 2 (PULSE WIDTH MODULATOR) SPEED REQUEST	RW
MAILBOX	RW
RESERVED	RO

RO = READ-ONLY VIA HOST I2C
RW = READ/WRITE VIA HOST I2C

FIG. 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2010/022563

A. CLASSIFICATION OF SUBJECT MATTER

G06F 15/16(2006.01)i, G06F 15/163(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 15/16; B23P 11/00; G05D 23/00; G06F 13/00; H02P 7/29

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:fan,backplane,communications

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KR 20-0302002 Y1 (KIM, MI-AE) 24 January 2003 See abstract and claims	1-15
A	US 6862644 B1 (MATTHEW A. PENDLETON et al.) 01 March 2005 See abstract and claims	1-15
A	US 2009-0299544 A1 (TANADA YUKIHIRO et al.) 03 December 2009 See abstract and claims	1-15
A	US 2006-0178786 A1 (NITIN BHAGWATH et al.) 10 August 2006 See abstract and claims	1-15

☐ 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 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	Date of mailing of the international search report
02 AUGUST 2010 (02.08.2010)	**02 AUGUST 2010 (02.08.2010)**

Name and mailing address of the ISA/KR	Authorized officer
Korean Intellectual Property Office Government Complex-Daejeon, 139 Seonsa-ro, Seo-gu, Daejeon 302-701, Republic of Korea Facsimile No. 82-42-472-7140	LEE, Jong Ick Telephone No. 82-42-481-8373

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2010/022563

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
KR 20-0302002 Y1	24.01.2003	None	
US 6862644 B1	01.03.2005	None	
US 2009-0299544 A1	03.12.2009	None	
US 2006-0178786 A1	10.08.2006	GB 0602269 D0	15.03.2006
		GB 2423167 A	16.08.2006
		JP 2006-221641 A	24.08.2006
		US 7310738 B2	18.12.2007