

(12) STANDARD PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 2010355801 B2**

- (54) Title
Store/store block instructions for communicating with adapters
- (51) International Patent Classification(s)
G06F 9/312 (2006.01) **G06F 13/38** (2006.01)
- (21) Application No: **2010355801** (22) Date of Filing: **2010.11.08**
- (87) WIPO No: **WO11/160710**
- (30) Priority Data
- | | | |
|-------------------|-------------------|--------------|
| (31) Number | (32) Date | (33) Country |
| 12/821,194 | 2010.06.23 | US |
- (43) Publication Date: **2011.12.29**
(44) Accepted Journal Date: **2014.04.24**
- (71) Applicant(s)
International Business Machines Corporation
- (72) Inventor(s)
Greiner, Dan;Craddock, David;Gregg, Thomas;Farrell, Mark
- (74) Agent / Attorney
Spruson & Ferguson, L 35 St Martins Tower 31 Market St, Sydney, NSW, 2000
- (56) Related Art
WO 2002/041157 A2 (SUN MICROSYSTEMS INC) 23 May 2002

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
29 December 2011 (29.12.2011)

(10) International Publication Number
WO 2011/160710 A1

(51) International Patent Classification:
G06F 9/312 (2006.01) *G06F 13/38* (2006.01)

(21) International Application Number:
PCT/EP2010/067028

(22) International Filing Date:
8 November 2010 (08.11.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
12/821,194 23 June 2010 (23.06.2010) US

(71) Applicant (for all designated States except US): **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, New York 10504 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **GREINER, Dan** [US/US]; IBM Corporation, M/P SVL/090/F374, 555 Bailey Avenue, Santa Teresa Lab, San Jose, California 95141-1003 (US). **CRADDOCK, David** [GB/US]; IBM

Corporation, M/P P318, 2455 South Road, Poughkeepsie, New York 12601-5400 (US). **GREGG, Thomas** [US/US]; IBM Corporation, M/P P314, 2455 South Road, Poughkeepsie, New York 12601-5400 (US). **FARRELL, Mark** [US/US]; IBM Corporation, M/P P310, 2455 South Road, Poughkeepsie, New York 12601-5400 (US).

(74) Agent: **STRETTON, Peter, John**; IBM United Kingdom Limited, Intellectual Property Law, Hursley Park, Winchester Hampshire SO21 2JN (GB).

(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.

[Continued on next page]

(54) Title: STORE/STORE BLOCK INSTRUCTIONS FOR COMMUNICATING WITH ADAPTERS

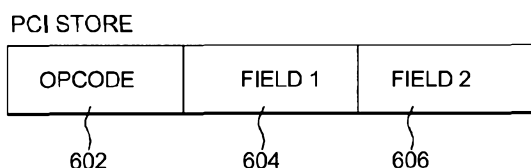


FIG. 6A

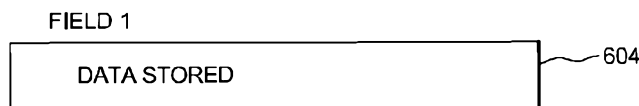


FIG. 6B



FIG. 6C

(57) Abstract: Communication with adapters of a computing environment is facilitated. Instructions are provided that explicitly target the adapters. Information provided in an instruction is used to steer the instruction to an appropriate location within the adapter.



(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

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

STORE/STORE BLOCK INSTRUCTIONS FOR COMMUNICATING WITH ADAPTERS

Technical Field

This invention relates, in general, to input/output processing of a computing environment, and in particular, to facilitating communication with adapters of the computing environment.

Background

A computing environment may include one or more types of input/output devices, including various types of adapters. One type of adapter is a Peripheral Component Interconnect (PCI) or Peripheral Component Interconnect Express (PCie) adapter. This adapter includes one or more address spaces used in communicating data between the adapter and the system to which the adapter is attached. The PCI specification is available from the World Wide Web at www.pcisig.com/home.

US Patent No. 6,704,831, issued March 9, 2004, Avery, "Method and Apparatus for Converting Address Information Between PCI Bus Protocol and a Message-Passing Queue- Oriented Bus Protocol," describes PCI load/store operations and DMA operations are implemented via work queue pairs in a message-passing, queue-oriented bus architecture. PCI address space is divided into segments and, each segment, in turn, is divided into regions. A separate work queue is assigned to each segment. A first portion of a PCI address is matched against the address ranges represented by the segments and used to select a memory segment and its corresponding work queue. An entry in the work queue holds a second portion of the PCI address which specifies a region within the selected segment that is assigned to a specific PCI device. In one embodiment, PIO load/store operations are implemented by selecting a work queue assigned to PIO operations and creating a work queue entry with the PCI address of a register on a PCI device and a pointer to the PIO data. The work queue entry is sent to a PCI bridge where the PCI address is extracted and used to program the appropriate device register with the data using the data pointer. DMA transfers are also implemented by selecting a work queue by means comparing a portion of the PCI address generated by the PCI device to an address range table and selecting a work queue that services the address range. A work queue entry is created with the remainder

of the PCI address and a pointer to the DMA data. An RDMA operation is used to perform the DMA transfer. The page and region data is used in connection with a translation protection table in the host channel adapter to access physical memory and perform the DMA transfer.

US Patent No. 7,613,847, issued November 3, 2009, Kjos et al, "Partially Virtualizing an I/O Device for Use by Virtual Machines," describes a computer system comprises a physical computer and a virtual machine monitor executable on the physical computer and configured to create an emulation of at least one guest operating system adapted to control the physical computer. The computer system further comprises a host executable on the physical computer that manages physical resources coupled to the physical computer on behalf of the virtual machine monitor and the at least one guest operating system. The host is adapted to virtualize a Peripheral Component Interconnect (PCI) configuration address space whereby the at least one guest operating system controls PCI input/output (I/O) devices directly and in absence of I/O emulation.

In some systems, a portion of an address space of the central processing unit (CPU) coupled to the adapter is mapped to an address space of the adapter enabling CPU instructions that access storage to directly manipulate the data in the adapter's address space.

The background discussion (including any potential prior art) is not to be taken as an admission of the common general knowledge in the art in any country. Any references discussed state the assertions of the author of those references and not the assertions of the applicant of this application. As such, the applicant reserves the right to challenge the accuracy and relevance of the references discussed.

An object of the present invention is to at least provide the public with a useful choice.

BRIEF SUMMARY

It is acknowledged that the terms "comprise", "comprises" and "comprising" may, under varying jurisdictions, be attributed with either an exclusive or an inclusive meaning. For the purpose of this specification, and unless otherwise noted, these terms are intended to have an inclusive meaning - i.e. they will be taken to mean an inclusion of

the listed components that the use directly references, but optionally also the inclusion of other non-specified components or elements. It will be understood that this intended meaning also similarly applies to the terms mentioned when used to define steps in a method or process.

In accordance with an aspect of the present invention, a capability is provided for facilitating communication with adapters, such as PCI or PCIe adapters. Control instructions specifically designed for communicating data to and from adapters are provided and used for communication.

The shortcomings of the prior art are overcome and advantages are provided through the provision of a computer program product for executing a store instruction for storing data in an adapter. The computer program product includes a computer readable storage medium readable by a processing circuit and storing instructions for execution by the processing circuit for performing a method. The method includes, for instance, obtaining a machine instruction for execution, the machine instruction being defined for computer execution

according to a computer architecture, the machine instruction including, for instance, an opcode field identifying a store to adapter instruction; a first field identifying a first location that includes data to be stored in an adapter; a second field identifying a second location, the contents of which include a function handle identifying the adapter, a designation of an address space within the adapter in which data is to be stored, and an offset within the address space; and executing the machine instruction, the executing including using the function handle to obtain a function table entry associated with the adapter; obtaining a data address of the adapter using at least one of information in the function table entry and the offset; and storing data from the first location in a specific location in the address space identified by the designation of the address space, the specific location identified by the data address of the adapter.

Further, a computer program product for executing a store block instruction for storing data in an adapter is provided. The computer program product includes a computer readable storage medium readable by a processing circuit and storing instructions for execution by the processing circuit for performing a method. The method includes, for instance, obtaining a machine instruction for execution, the machine instruction being defined for computer execution according to a computer architecture, the machine instruction including, for instance, an opcode field identifying a store block to adapter instruction; a first field identifying a first location, the contents of which include a function handle identifying an adapter, and a designation of an address space within the adapter in which data is to be stored; a second field identifying a second location that includes an offset within the address space; a third field identifying a third location that includes an address in memory that includes data to be stored in the adapter; and executing the machine instruction, the executing including using the function handle to obtain a function table entry associated with the adapter; obtaining a data address of the adapter using information in the function table entry and the offset; and storing data obtained from memory at the address in the third field into a specific location in the address space identified by the designation of the address space, the specific location identified by the data address of the adapter.

Methods and systems relating to one or more aspects of the present invention are also described and claimed herein.

Additional features and advantages are realized through the techniques of the present invention. Other embodiments and aspects of the invention are described in detail herein and are considered a part of the claimed invention.

5 BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

One or more aspects of the present invention are particularly pointed out and distinctly claimed as examples in the claims at the conclusion of the specification. The foregoing and other objects, features, and advantages of the invention are apparent from the following
10 detailed description taken in conjunction with the accompanying drawings in which:

FIG. 1A depicts one embodiment of a computing environment to incorporate and use one or more aspects of the present invention;

15 FIG. 1B depicts one embodiment of a device table entry located in the I/O hub of FIG. 1A and used in accordance with an aspect of the present invention;

FIG. 1C depicts another embodiment of a computing environment to incorporate and use one or more aspects of the present invention;

20 FIG. 2 depicts one example of address spaces of an adapter function, in accordance with an aspect of the present invention;

25 FIG. 3A depicts one example of a function table entry used in accordance with an aspect of the present invention;

FIG. 3B depicts one embodiment of a function handle used in accordance with an aspect of the present invention;

30 FIG. 4A depicts one embodiment of a PCI Load instruction used in accordance with an aspect of the present invention;

FIG. 4B depicts one embodiment of a field used by the PCI Load instruction of FIG. 4A, in accordance with an aspect of the present invention;

FIG. 4C depicts one embodiment of another field used by the PCI Load instruction of FIG. 4A, in accordance with an aspect of the present invention;

FIGs. 5A-5B depict one embodiment of the logic to perform a PCI Load operation, in accordance with an aspect of the present invention;

FIG. 6A depicts one embodiment of a PCI Store instruction used in accordance with an aspect of the present invention;

FIG. 6B depicts one embodiment of a field used by the PCI Store instruction of FIG. 6A, in accordance with an aspect of the present invention;

FIG. 6C depicts one embodiment of another field used by the PCI Store instruction of FIG. 6A, in accordance with an aspect of the present invention;

FIGs. 7A-7B depict one embodiment of the logic to perform a PCI Store operation, in accordance with an aspect of the present invention;

FIG. 8A depicts one embodiment of a PCI Store Block instruction used in accordance with an aspect of the present invention;

FIG. 8B depicts one embodiment of a field used by the PCI Store Block instruction of FIG. 8A, in accordance with an aspect of the present invention;

FIG. 8C depicts one embodiment of another field used by the PCI Store Block instruction of FIG. 8A, in accordance with an aspect of the present invention;

FIG. 8D depicts one embodiment of yet another field used by the PCI Store Block instruction of FIG. 8A, in accordance with an aspect of the present invention;

FIGs. 9A-9B depict one embodiment of the logic to perform a PCI Store Block operation, in accordance with an aspect of the present invention;

5 FIG. 10 depicts one embodiment of a computer program product incorporating one or more aspects of the present invention;

FIG. 11 depicts one embodiment of a host computer system to incorporate and use one or more aspects of the present invention;

10 FIG. 12 depicts a further example of a computer system to incorporate and use one or more aspects of the present invention;

FIG. 13 depicts another example of a computer system comprising a computer network to incorporate and use one or more aspects of the present invention;

15 FIG. 14 depicts one embodiment of various elements of a computer system to incorporate and use one or more aspects of the present invention;

20 FIG. 15A depicts one embodiment of the execution unit of the computer system of FIG. 14 to incorporate and use one or more aspects of the present invention;

FIG. 15B depicts one embodiment of the branch unit of the computer system of FIG. 14 to incorporate and use one or more aspects of the present invention;

25 FIG. 15C depicts one embodiment of the load/store unit of the computer system of FIG. 14 to incorporate and use one or more aspects of the present invention; and

FIG. 16 depicts one embodiment of an emulated host computer system to incorporate and use one or more aspects of the present invention.

30

DETAILED DESCRIPTION

In accordance with an aspect of the present invention, one or more control instructions are provided to facilitate communication with adapters of a computing environment. The control instructions are specifically designed for communicating data to and from the address spaces of the adapters.

As used herein, the term adapter includes any type of adapter (e.g., storage adapter, network adapter, processing adapter, PCI adapter, cryptographic adapter, other type of input/output adapters, etc.). In one embodiment, an adapter includes one adapter function. However, in other embodiments, an adapter may include a plurality of adapter functions. One or more aspects of the present invention are applicable whether an adapter includes one adapter function or a plurality of adapter functions. In one embodiment, if an adapter includes a plurality of adapter functions, then each function may be communicated with in accordance with an aspect of the present invention. Moreover, in the examples presented herein, adapter is used interchangeably with adapter function (e.g., PCI function) unless otherwise noted.

One embodiment of a computing environment to incorporate and use one or more aspects of the present invention is described with reference to FIG. 1A. In one example, a computing environment 100 is a System z® server offered by International Business Machines Corporation. System z® is based on the z/Architecture® offered by International Business Machines Corporation. Details regarding the z/Architecture® are described in an IBM® publication entitled, "z/Architecture Principles of Operation," IBM Publication No. SA22-7832-07, February 2009. IBM®, System z® and z/Architecture® are registered trademarks of International Business Machines Corporation, Armonk, New York. Other names used herein may be registered trademarks, trademarks or product names of International Business Machines Corporation or other companies.

In one example, computing environment 100 includes one or more central processing units (CPUs) 102 coupled to a system memory 104 (a.k.a., main memory) via a memory controller 106. To access system memory 104, a central processing unit 102 issues a read or write request that includes an address used to access system memory. The address included in the

request is typically not directly usable to access system memory, and therefore, it is translated to an address that is directly usable in accessing system memory. The address is translated via a translation mechanism (XLATE) 108. For example, the address is translated from a virtual address to a real or absolute address using, for instance, dynamic address translation (DAT).

The request, including the address (translated, if necessary), is received by memory controller 106. In one example, memory controller 106 is comprised of hardware and is used to arbitrate for access to the system memory and to maintain the memory's consistency. This arbitration is performed for requests received from CPUs 102, as well as for requests received from one or more adapters 110. Like the central processing units, the adapters issue requests to system memory 104 to gain access to the system memory.

In one example, adapter 110 is a Peripheral Component Interconnect (PCI) or PCI Express (PCIe) adapter that includes one or more PCI functions. A PCI function issues a request that is routed to an input/output hub 112 (e.g., a PCI hub) via one or more switches (e.g., PCIe switches) 114. In one example, the input/output hub is comprised of hardware, including one or more state machines, and is coupled to memory controller 106 via an I/O-to-memory bus 120.

The input/output hub includes, for instance, a root complex 116 that receives the request from a switch. The request includes an input/output address that is provided to an address translation and protection unit 118 which accesses information used for the request. As examples, the request may include an input/output address used to perform a direct memory access (DMA) operation or to request a message signaled interruption (MSI). Address translation and protection unit 118 accesses information used for the DMA or MSI request. As a particular example, for a DMA operation, information may be obtained to translate the address. The translated address is then forwarded to the memory controller to access system memory.

In one example, as described with reference to FIG. 1B, information used for the DMA or MSI request issued by an adapter is obtained from a device table entry 130 of a device table

132 located in the I/O hub (c.g., in the address translation and protection unit). The device table entry includes information for the adapter, and each adapter has at least one device table entry associated therewith. For instance, there is one device table entry per address space (in system memory) assigned to the adapter. For a request issued from an adapter
5 (e.g., PCI function 138), a device table entry is located using a requestor ID provided in the request.

Referring now to FIG. 1C, in a further embodiment of a computing environment, in addition to or instead of one or more CPUs 102, a central processing complex is coupled to memory
10 controller 106. In this example, a central processing complex 150 includes, for instance, one or more partitions or zones 152 (e.g., logical partitions LP1-LPn), one or more central processors (c.g., CP1-CPm) 154, and a hypervisor 156 (e.g., a logical partition manager), each of which is described below.

Each logical partition 152 is capable of functioning as a separate system. That is, each logical partition can be independently reset, initially loaded with an operating system or a hypervisor (such as z/VM® offered by International Business Machines Corporation, Armonk, New York), if desired, and operate with different programs. An operating system, a hypervisor, or an application program running in a logical partition appears to have access
20 to a full and complete system, but only a portion of it is available. A combination of hardware and Licensed Internal Code (also referred to as microcode or millicode) keeps a program in a logical partition from interfering with the program in a different logical partition. This allows several different logical partitions to operate on a single or multiple physical processor in a time slice manner. In this particular example, each logical partition
25 has a resident operating system 158, which may differ for one or more logical partitions. In one embodiment, operating system 158 is a z/OS® or zLinux operating system, offered by International Business Machines Corporation, Armonk, New York. z/OS® and z/VM® are registered trademarks of International Business Machines Corporation, Armonk, New York.

30 Central processors 154 are physical processor resources that are allocated to the logical partitions. For instance, a logical partition 152 includes one or more logical processors, each of which represents all or a share of the physical processor resource 154 allocated to the

partition. The underlying processor resource may either be dedicated to that partition or shared with another partition.

Logical partitions 152 are managed by hypervisor 156 implemented by firmware running on processors 154. Logical partitions 152 and hypervisor 156 each comprise one or more programs residing in respective portions of central storage associated with the central processors. One example of hypervisor 156 is the Processor Resource/Systems Manager (PR/SM), offered by International Business Machines Corporation, Armonk, New York.

As used herein, firmware includes, e.g., the microcode, millicode and/or macrocode of the processor. It includes, for instance, the hardware-level instructions and/or data structures used in implementation of higher-level machine code. In one embodiment, it includes, for instance, proprietary code that is typically delivered as microcode that includes trusted software or microcode specific to the underlying hardware and controls operating system access to the system hardware.

Although, in this example, a central processing complex having logical partitions is described, one or more aspects of the present invention may be incorporated in and used by other processing units, including single or multi-processor processing units that are not partitioned, among others. The central processing complex described herein is only one example.

As described above, adapters can issue requests to the processors requesting various operations, such as direct memory accesses, message signaled interrupts, etc. Further, the processors can issue requests to the adapters. For instance, returning to FIG. 1B, a processor 102 can issue a request to access an adapter function 138. The request is routed from the processor to the adapter function via I/O hub 112 and one or more switches 114. In this embodiment, the memory controller is not shown. However, the I/O hub may be coupled to the processor directly or via a memory controller.

As an example, an operating system 140 executing within the processor issues an instruction to the adapter function requesting a particular operation. In this example, instructions issued

by the operating system are specific to the I/O infrastructure. That is, since the I/O infrastructure is based on PCI or PCIe (both of which are referred to herein as PCI, unless otherwise noted), the instructions are PCI instructions. Example PCI instructions include PCI Load, PCI Store and PCI Store Block, to name a few. Although, in this example, the I/O infrastructure and instructions are based on PCI, in other embodiments, other infrastructures and corresponding instructions may be used.

In one particular example, the instructions are directed to a specific location within an address space of the adapter function. For instance, as shown in FIG. 2, an adapter function 138 includes storage 200, which is defined as a plurality of address spaces, including, for instance: a configuration space 202 (e.g., PCI configuration space for a PCI function); an I/O space 204 (e.g., PCI I/O space); and one or more memory spaces 206 (e.g., PCI memory space). In other embodiments, more, less or different address spaces may be provided. The instructions are targeted to a particular address space and a particular location within the address space. This ensures that the configuration (e.g., operating system, LPAR, processor, guest, etc.) issuing the instruction is authorized to access the adapter function.

To facilitate processing of the instructions, information stored in one or more data structures is used. One such data structure that includes information regarding adapters is a function table 300 stored, for instance, in secure memory. As shown in FIG. 3A, in one example, a function table 300 includes one or more function table entries (FTEs) 302. In one example, there is one function table entry per adapter function. Each function table entry 302 includes information to be used in processing associated with its adapter function. In one example, function table entry 302 includes, for instance:

Instance Number 308: This field indicates a particular instance of the function handle associated with the function table entry;

Device Table Entry (DTE) Index 1...n 310: There may be one or more device table indices, and each index is an index into a device table to locate a device table entry (DTE). There are one or more device table entries per adapter function, and each entry includes information associated with its adapter function, including information used to process requests of the

adapter function (e.g., DMA requests, MSI requests) and information relating to requests to the adapter function (e.g., PCI instructions). Each device table entry is associated with one address space within system memory assigned to the adapter function. An adapter function may have one or more address spaces within system memory assigned to the adapter function.

Busy Indicator 312: This field indicates whether the adapter function is busy;

Permanent Error State Indicator 314: This field indicates whether the adapter function is in a permanent error state;

Recovery Initiated Indicator 316: This field indicates whether recovery has been initiated for the adapter function;

Permission Indicator 318: This field indicates whether the operating system trying to enable the adapter function has authority to do so;

Enable Indicator 320: This field indicates whether the adapter function is enabled (e.g., 1=enabled, 0=disabled);

Requestor Identifier (RID) 322: This is an identifier of the adapter function and may include, for instance, a bus number, device number and function number. This field is used, for instance, for accesses of a configuration space of the adapter function.

For instance, the configuration space may be accessed by specifying the configuration space in an instruction issued by the operating system (or other configuration) to the adapter function. Specified in the instruction is an offset into the configuration space and a function handle used to locate the appropriate function table entry that includes the RID. The firmware receives the instruction and determines it is for a configuration space. Therefore, it uses the RID to generate a request to the I/O hub, and the I/O hub creates a request to access the adapter. The location of the adapter function is based on the RID, and the offset

specifies an offset into the configuration space of the adapter function. For instance, the offset specifies a register number in the configuration space.

Base Address Register (BAR) (1 to n) 324: This field includes a plurality of unsigned integers, designated as BAR0 – BAR_n, which are associated with the originally specified adapter function, and whose values are also stored in the base address registers associated with the adapter function. Each BAR specifies the starting address of a memory space or I/O space within the adapter function, and also indicates the type of address space, that is whether it is a 64 or 32 bit memory space, or a 32 bit I/O space, as examples;

In one example, it is used for accesses to memory space and/or I/O space of the adapter function. For instance, an offset provided in an instruction to access the adapter function is added to the value in the base address register associated with the address space designated in the instruction to obtain the address to be used to access the adapter function. The address space identifier provided in the instruction identifies the address space within the adapter function to be accessed and the corresponding BAR to be used;

Size 1....n 326: This field includes a plurality of unsigned integers, designated as SIZE0 – SIZE_n. The value of a Size field, when non-zero, represents the size of each address space with each entry corresponding to a previously described BAR.

Further details regarding BAR and Size are described below.

1. When a BAR is not implemented for an adapter function, the BAR field and its corresponding size field are both stored as zeros.
2. When a BAR field represents either an I/O address space or a 32-bit memory address space, the corresponding size field is non-zero and represents the size of the address space.
3. When a BAR field represents a 64-bit memory address space,
 - a. The BAR_n field represents the least significant address bits.
 - b. The next consecutive BAR_{n+1} field represents the most significant address bits.

- c. The corresponding SIZE_n field is non-zero and represents the size of the address space.
- d. The corresponding SIZE_{n+1} field is not meaningful and is stored as zero.

5 Internal Routing Information 328: This information is used to perform particular routing to the adapter. It includes, for instance, node, processor chip, and I/O hub addressing information, as examples.

10 Status Indicator 330: This provides an indication of whether load/store operations are blocked, as well as other indications.

In one example, the busy indicator, permanent error state indicator, and recovery initiated indicator are set based on monitoring performed by the firmware. Further, the permission indicator is set, for instance, based on policy. The BAR information is based on
15 configuration information discovered during a bus walk by the processor (e.g., firmware of the processor). Other fields may be set based on configuration, initialization and/or events. In other embodiments, the function table entry may include more, less or different information. The information included may depend on the operations supported by or enabled for the adapter function.

20 To locate a function table entry in a function table that includes one or more entries, in one embodiment, a function handle is used. For instance, one or more bits of the function handle are used as an index into the function table to locate a particular function table entry.

25 Referring to FIG. 3B, additional details regarding a function handle are described. In one example, a function handle 350 includes an enable indicator 352 that indicates whether the PCI function handle is enabled; a PCI function number 354 that identifies the function (this is a static identifier, and in one embodiment, is an index into the function table); and an instance number 356 which indicates the particular instance of this function handle. For
30 example, each time the function is enabled, the instance number is incremented to provide a new instance number.

In accordance with an aspect of the present invention, to access an adapter function, a configuration issues a request to the adapter function, which is executed by a processor. In the examples herein, the configuration is an operating system, but in other examples, it may be a system, processor, logical partition, guest, etc. These requests are via specific
5 instructions, which access the adapter. Example instructions include PCI Load, PCI Store, and PCI Store Block instructions. These instructions are specific to the adapter architecture (e.g., PCI). Further details regarding these instructions are described below. For instance, one embodiment of a PCI Load instruction is described with reference to FIGs. 4A-5B; one embodiment of a PCI Store instruction is described with reference to FIGs. 6A-7B; and one
10 embodiment of a PCI Store Block instruction is described with reference to FIGs. 8A-9B.

Referring initially to FIG. 4A, one embodiment of a PCI Load instruction is depicted. As shown, a PCI Load instruction 400 includes, for instance, an opcode 402 indicating the PCI Load instruction; a first field 404 specifying a location at which data fetched from an adapter
15 function will be loaded; and a second field 406 specifying a location at which various information is included regarding the adapter function from which data is to be loaded. The contents of the locations designated by Fields 1 and 2 are further described below.

In one example, Field 1 designates a general register, and as depicted in FIG. 4B, the
20 contents 404 of that register include a contiguous range of one or more bytes loaded from the location of the adapter function specified in the instruction. In one example, the data is loaded into the rightmost byte positions of the register.

In one embodiment, Field 2 designates a pair of general registers that include various
25 information. As shown in FIG. 4B, the contents of the registers include, for instance:

Enabled Handle 410: This field is an enabled function handle of the adapter function from which the data is to be loaded;

30 Address Space 412: This field identifies the address space within the adapter function from which the data is to be loaded;

Offset Within Address Space 414: This field specifies the offset within the specified address space from which the data is to be loaded;

Length field 416: This field specifies the length of the load operation (e.g., the number of bytes to be loaded); and

Status field 418: This field provides a status code which is applicable when the instruction completes with a predefined condition code.

In one embodiment, the bytes loaded from the adapter function are to be contained within an integral boundary in the adapter function's designated PCI address space. When the address space field designates a memory address space, the integral boundary size is, for instance, a double word. When the address space field designates an I/O address space or a configuration address space, the integral boundary size is, for instance, a word.

One embodiment of the logic associated with a PCI Load instruction is described with reference to FIGs. 5A-5B. In one example, the instruction is issued by an operating system (or other configuration) and executed by the processor (e.g., firmware) executing the operating system. In the examples herein, the instructions and adapter functions are PCI-based. However, in other examples, a different adapter architecture and corresponding instructions may be used.

To issue the instruction, the operating system provides the following operands to the instruction (e.g., in one or more registers designated by the instruction): the PCI function handle, the PCI address space (PCIAS), the offset into the PCI address space, and the length of the data to be loaded. Upon successful completion of the PCI Load instruction, the data is loaded in the location (e.g., register) designated by the instruction.

Referring to FIG. 5A, initially, a determination is made as to whether the facility allowing for a PCI Load instruction is installed, INQUIRY 500. This determination is made by, for instance, checking an indicator stored in, for instance, a control block. If the facility is not installed, an exception condition is provided, STEP 502. Otherwise, a determination is made

as to whether the operands are aligned, INQUIRY 504. For instance, if certain operands need to be in even/odd register pairs, a determination is made as to whether those requirements are met. If the operands are not aligned, then an exception is provided, STEP 506. Otherwise, if the facility is installed and the operands are aligned, a determination is made as to whether the handle provided in the operands of the PCI Load instruction is enabled, INQUIRY 508. In one example, this determination is made by checking an enable indicator in the handle. If the handle is not enabled, then an exception condition is provided, STEP 510.

If the handle is enabled, then the handle is used to locate a function table entry, STEP 512. That is, at least a portion of the handle is used as an index into the function table to locate the function table entry corresponding to the adapter function from which data is to be loaded.

Thereafter, if the configuration issuing the instruction is a guest, a determination is made as to whether the function is configured for use by a guest, INQUIRY 514. If it is not authorized, then an exception condition is provided, STEP 516. This inquiry may be ignored if the configuration is not a guest or other authorizations may be checked, if designated. (In one example, in the z/Architecture®, a pageable guest is interpretively executed via the Start Interpretive Execution (SIE) instruction, at level 2 of interpretation. For instance, the logical partition (LPAR) hypervisor executes the SIE instruction to begin the logical partition in physical, fixed memory. If z/VM® is the operating system in that logical partition, it issues the SIE instruction to execute its guests (virtual) machines in its V=V (virtual) storage. Therefore, the LPAR hypervisor uses level-1 SIE, and the z/VM® hypervisor uses level-2 SIE.)

A determination is then made as to whether the function is enabled, INQUIRY 518. In one example, this determination is made by checking an enable indicator in the function table entry. If it is not enabled, then an exception condition is provided, STEP 520.

If the function is enabled, then a determination is made as to whether the address space is valid, INQUIRY 522. For instance, is the specified address space a designated address space of the adapter function and one that is appropriate for this instruction. If the address space is

invalid, then an exception condition is provided, STEP 524. Otherwise, a determination is made as to whether load/store is blocked, INQUIRY 526. In one example, this determination is made by checking the status indicator in the function table entry. If load/store is blocked, then an exception condition is provided, STEP 528.

5

However, if load/store is not blocked, a determination is made as to whether recovery is active, INQUIRY 530. In one example, this determination is made by checking the recovery initiated indicator in the function table entry. If recovery is active, then an exception condition is provided, STEP 532. Otherwise, a determination is made as to whether the function is busy, INQUIRY 534. This determination is made by checking the busy indicator in the function table entry. If the function is busy, then a busy condition is provided, STEP 536. With a busy condition, the instruction can be retried, instead of dropped.

10

If the function is not busy, then a further determination is made as to whether the offset specified in the instruction is valid, INQUIRY 538. That is, is the offset in combination with the length of the operation within the base and length of the address space, as specified in the function table entry. If not, then an exception condition is provided, STEP 540. However, if the offset is valid, then a determination is made as to whether the length is valid, INQUIRY 542. That is, subject to the address space type, offset within the address space, and an integral boundary size is the length valid. If not, then an exception condition is provided, STEP 544. Otherwise, processing continues with the load instruction. (In one embodiment, the firmware performs the above checks.)

15

20

Continuing with FIG. 5B, a determination is made by the firmware as to whether the load is for a configuration address space of the adapter function, INQUIRY 550. That is, based on the configuration of the adapter function's memory, is the specified address space provided in the instruction a configuration space. If so, then the firmware performs various processing to provide the request to a hub coupled to the adapter function; the hub then routes the request to the function, STEP 552.

25

30

For example, the firmware obtains the requestor ID from the function table entry pointed to by the function handle provided in the instruction operands. Further, the firmware

determines based on information in the function table entry (e.g., the internal routing information) the hub to receive this request. That is, an environment may have one or more hubs and the firmware determines the hub coupled to the adapter function. It then forwards the request to the hub. The hub generates a configuration read request packet that flows out
5 on the PCI bus to the adapter function identified by the RID in the function table entry. The configuration read request includes the RID and offset (i.e., data address) that are used to fetch the data, as described below.

Returning to INQUIRY 550, if the designated address space is not a configuration space,
10 then once again the firmware performs various processing to provide the request to the hub, STEP 554. The firmware uses the handle to select a function table entry and from that entry it obtains information to locate the appropriate hub. It also calculates a data address to be used in the load operation. This address is calculated by adding the BAR starting address (with the BAR being that associated with the address space identifier provided in the
15 instruction) obtained from the function table entry to the offset provided in the instruction. This calculated data address is provided to the hub. The hub then takes that address and includes it in a request packet, such as a DMA read request packet, that flows out over the PCI bus to the adapter function.

Responsive to receiving the request either via STEP 552 or STEP 554, the adapter function fetches the requested data from the specified location (i.e., at the data address) and returns that data in a response to the request, STEP 556. The response is forwarded from the adapter function to the I/O hub. Responsive to receiving the response, the hub forwards the response to the initiating processor. The initiating processor then takes the data from the response
20 packet and loads it in the designated location specified in the instruction (e.g., field 1 404). The PCI Load operation concludes with an indication of success (e.g., setting a condition code of zero).

In addition to a load instruction that retrieves data from an adapter function and stores it in a
30 designated location, another instruction that may be executed is a store instruction. The store instruction stores data at a specified location in the adapter function. One embodiment of a PCI Store instruction is described with reference to FIG. 6A. As shown, a PCI Store

instruction 600 includes, for instance, an opcode 602 indicating the PCI Store instruction; a first field 604 specifying a location that includes data to be stored in an adapter function; and a second field 606 specifying a location at which various information is included regarding the adapter function to which data is to be stored. The contents of the locations designated by Fields 1 and 2 are further described below.

In one example, Field 1 designates a general register, and as depicted in FIG. 6B, the contents 604 of that register include a contiguous range of one or more bytes of data to be stored into the specified location of an adapter function. In one example, the data in the rightmost byte positions of the register are stored.

In one embodiment, Field 2 designates a pair of general registers that include various information. As shown in FIG. 6B, the contents of the register include, for instance:

Enabled Handle 610: This field is an enabled function handle of the adapter function to which the data is to be stored;

Address Space 612: This field identifies the address space within the adapter function to which the data is to be stored;

Offset Within Address Space 614: This field specifies the offset within the specified address space to which the data is to be stored;

Length field 616: This field specifies the length of the store operation (e.g., the number of bytes to be stored); and

Status field 618: This field provides a status code which is applicable when the instruction completes with a predefined condition code.

On embodiment of the logic associated with a PCI Store instruction is described with reference to FIGs. 7A-7B. In one example, the instruction is issued by an operating system, and executed by the processor (e.g., firmware) executing the operating system.

To issue the instruction, the operating system provides the following operands to the instruction (e.g., in one or more registers designated by the instruction): the PCI function handle, the PCI address space (PCIAS), the offset into the PCI address space, the length of the data to be stored, and a pointer to the data to be stored. Upon successful completion of the PCI Store instruction, the data is stored in the location designated by the instruction.

Referring to FIG. 7A, initially, a determination is made as to whether the facility allowing for a PCI Store instruction is installed, INQUIRY 700. This determination is made by, for instance, checking an indicator stored in, for instance, a control block. If the facility is not installed, an exception condition is provided, STEP 702. Otherwise, a determination is made as to whether the operands are aligned, INQUIRY 704. For instance, if certain operands need to be in even/odd register pairs, a determination is made as to whether those requirements are met. If the operands are not aligned, then an exception is provided, STEP 706. Otherwise, if the facility is installed and the operands are aligned, a determination is made as to whether the handle provided in the operands of the PCI Store instruction is enabled, INQUIRY 708. In one example, this determination is made by checking an enable indicator in the handle. If the handle is not enabled, then an exception condition is provided, STEP 710.

If the handle is enabled, then the handle is used to locate a function table entry, STEP 712. That is, at least a portion of the handle is used as an index into the function table to locate the function table entry corresponding to the adapter function at which data is to be stored.

Thereafter, if the configuration issuing the instruction is a guest, a determination is made as to whether the function is configured for use by a guest, INQUIRY 714. If it is not authorized, then an exception condition is provided, STEP 716. This inquiry may be ignored if the configuration is not a guest or other authorizations may be checked, if designated.

A determination is then made as to whether the function is enabled, INQUIRY 718. In one example, this determination is made by checking an enable indicator in the function table entry. If it is not enabled, then an exception condition is provided, STEP 720.

If the function is enabled, then a determination is made as to whether the address space is valid, INQUIRY 722. For instance, is the specified address space a designated address space of the adapter function and one that is appropriate for this instruction. If the address space is invalid, then an exception condition is provided, STEP 724. Otherwise, a determination is made as to whether load/store is blocked, INQUIRY 726. In one example, this determination is made by checking the status indicator in the function table entry. If load/store is blocked, then an exception condition is provided, STEP 728.

However, if the load/store is not blocked, a determination is made as to whether recovery is active, INQUIRY 730. In one example, this determination is made by checking the recovery initiated indicator in the function table entry. If recovery is active, then an exception condition is provided, STEP 732. Otherwise, a determination is made as to whether the function is busy, INQUIRY 734. This determination is made by checking the busy indicator in the function table entry. If the function is busy, then a busy condition is provided, STEP 736. With a busy condition, the instruction can be retried, instead of dropped.

If the function is not busy, then a further determination is made as to whether the offset specified in the instruction is valid, INQUIRY 738. That is, is the offset in combination with the length of the operation within the base and length of the address space, as specified in the function table entry. If not, then an exception condition is provided, STEP 740. However, if the offset is valid, then a determination is made as to whether the length is valid, INQUIRY 742. That is, subject to the address space type, offset within the address space, and an integral boundary size is the length valid. If not, then an exception condition is provided, STEP 744. Otherwise, processing continues with the store instruction. (In one embodiment, the firmware performs the above checks.)

Continuing with FIG. 7B, a determination is made by the firmware as to whether the store is for a configuration address space of the adapter function, INQUIRY 750. That is, based on the configuration of the adapter function's memory, is the specified address space provided in the instruction a configuration space. If so, then the firmware performs various processing to provide the request to a hub coupled to the adapter function; the hub then routes the request to the function, STEP 752.

For example, the firmware obtains the requestor ID from the function table entry pointed to by the function handle provided in the instruction operands. Further, the firmware determines based on information in the function table entry (e.g., the internal routing information) the hub to receive this request. That is, an environment may have one or more
5 hubs and the firmware determines the hub coupled to the adapter function. It then forwards the request to the hub. The hub generates a configuration write request packet that flows out on the PCI bus to the adapter function identified by the RID in the function table entry. The configuration write request includes the RID and offset (i.e., data address) that are used to store the data, as described below.

10 Returning to INQUIRY 750, if the designated address space is not a configuration space, then once again the firmware performs various processing to provide the request to the hub, STEP 754. The firmware uses the handle to select a function table entry and from that entry it obtains information to locate the appropriate hub. It also calculates a data address to be
15 used in the store operation. This address is calculated by adding the BAR starting address obtained from the function table entry to the offset provided in the instruction. This calculated data address is provided to the hub. The hub then takes that address and includes it in a request packet, such as a DMA write request packet, that flows out over the PCI bus to the adapter function.

20 Responsive to receiving the request either via STEP 752 or STEP 754, the adapter function stores the requested data at the specified location (i.e., at the data address), STEP 756. The PCI Store operation concludes with an indication of success (e.g., setting a condition code of zero).

25 In addition to the load and store instructions, which typically load or store a maximum of, e.g., 8 bytes, another instruction that may be executed is a store block instruction. The store block instruction stores larger blocks of data (e.g., 16, 32, 64, 128 or 256 bytes) at a specified location in the adapter function; the block sizes are not necessarily limited to
30 powers of two in size. In one example, the specified location is in a memory space of the adapter function (not an I/O or configuration space).

One embodiment of a PCI Store Block instruction is described with reference to FIG. 8A. As shown, a PCI Store Block instruction 800 includes, for instance, an opcode 802 indicating the PCI Store Block instruction; a first field 804 specifying a location at which various information is included regarding the adapter function to which data is to be stored; a
5 second field 806 specifying a location that includes an offset within the specified address space into which the data is to be stored; and a third field 808 specifying a location that includes an address in system memory of data to be stored in the adapter function. The contents of the locations designated by Fields 1, 2 and 3 are further described below.

10 In one embodiment, Field 1 designates a general register that includes various information. As shown in FIG. 8B, the contents of the register include, for instance:

Enabled Handle 810: This field is an enabled function handle of the adapter function to which the data is to be stored;

15 Address Space 812: This field identifies the address space within the adapter function to which the data is to be stored;

Length field 814: This field specifies the length of the store operation (e.g., the number of
20 bytes to be stored); and

Status field 816: This field provides a status code which is applicable when the instruction completes with a predefined condition code.

25 In one example, Field 2 designates a general register, and as depicted in FIG. 8C, the contents of the register include a value (e.g., 64-bit unsigned integer) that specifies the offset within the specified address space into which the data is to be stored.

30 In one example, Field 3, as depicted in FIG. 8D, includes the logical address in system memory of the first byte of data 822 to be stored in the adapter function.

One embodiment of the logic associated with a PCI Store Block instruction is described with reference to FIGs. 9A-9B. In one example, the instruction is issued by an operating system, and executed by the processor (e.g., firmware) executing the operating system.

5 To issue the instruction, the operating system provides the following operands to the instruction (e.g., in one or more registers designated by the instruction): the PCI function handle, the PCI address space (PCIAS), the offset into the PCI address space, the length of the data to be stored, and a pointer to the data to be stored. The pointer operand may
10 comprise both a register and a signed or unsigned displacement. Upon successful completion of the PCI Store Block instruction, the data is stored in the location in the adapter designated by the instruction.

Referring to FIG. 9A, initially, a determination is made as to whether the facility allowing for a PCI Store Block instruction is installed, INQUIRY 900. This determination is made
15 by, for instance, checking an indicator stored in, for instance, a control block. If the facility is not installed, an exception condition is provided, STEP 902. Otherwise, if the facility is installed, a determination is made as to whether the handle provided in the operands of the PCI Store Block instruction is enabled, INQUIRY 904. In one example, this determination is made by checking an enable indicator in the handle. If the handle is not enabled, then an
20 exception condition is provided, STEP 906.

If the handle is enabled, then the handle is used to locate a function table entry, STEP 912. That is, at least a portion of the handle is used as an index into the function table to locate the function table entry corresponding to the adapter function at which data is to be stored.

25 Thereafter, if the configuration issuing the instruction is a guest, a determination is made as to whether the function is configured for use by a guest, INQUIRY 914. If it is not authorized, then an exception condition is provided, STEP 916. This inquiry may be ignored if the configuration is not a guest or other authorizations may be checked, if designated.

A determination is then made as to whether the function is enabled, INQUIRY 918. In one example, this determination is made by checking an enable indicator in the function table entry. If it is not enabled, then an exception condition is provided, STEP 920.

5 If the function is enabled, then a determination is made as to whether the address space is valid, INQUIRY 922. For instance, is the specified address space a designated address space of the adapter function and one that is appropriate for this instruction (i.e., a memory space). If the address space is invalid, then an exception condition is provided, STEP 924. Otherwise, a determination is made as to whether load/store is blocked, INQUIRY 926. In
10 one example, this determination is made by checking the status indicator in the function table entry. If load/store is blocked, then an exception condition is provided, STEP 928.

However, if the load/store is not blocked, a determination is made as to whether recovery is active, INQUIRY 930. In one example, this determination is made by checking the recovery
15 initiated indicator in the function table entry. If recovery is active, then an exception condition is provided, STEP 932. Otherwise, a determination is made as to whether the function is busy, INQUIRY 934. This determination is made by checking the busy indicator in the function table entry. If the function is busy, then a busy condition is provided, STEP 936. With a busy condition, the instruction can be retried, instead of dropped.

20 If the function is not busy, then a further determination is made as to whether the offset specified in the instruction is valid, INQUIRY 938. That is, is the offset in combination with the length of the operation within the base and length of the address space, as specified in the function table entry. If not, then an exception condition is provided, STEP 940. However, if
25 the offset is valid, then a determination is made as to whether the length is valid, INQUIRY 942. That is, subject to the address space type, offset within the address space, and an integral boundary size is the length valid. If not, then an exception condition is provided, STEP 944. Otherwise, processing continues with the store block instruction. (In one embodiment, the firmware performs the above checks.)

30 Continuing with FIG. 9B, a determination is made by the firmware as to whether the storage that includes the data to be stored is accessible, INQUIRY 950. If not, an exception

condition is provided, STEP 952. If so, then the firmware performs various processing to provide the request to a hub coupled to the adapter function; the hub then routes the request to the function, STEP 954.

5 For example, the firmware uses the handle to select a function table entry and from that entry it obtains information to locate the appropriate hub. It also calculates a data address to be used in the store block operation. This address is calculated by adding the BAR starting address (with the BAR being identified by the address space identifier) obtained from the function table entry to the offset provided in the instruction. This calculated data address is
10 provided to the hub. In addition, the data referenced by the address provided in the instruction is fetched from system memory and provided to the I/O hub. The hub then takes that address and data and includes it in a request packet, such as a DMA write request packet, that flows out over the PCI bus to the adapter function.

15 Responsive to receiving the request, the adapter function stores the requested data at the specified location (i.e., at the data address), STEP 956. The PCI Store Block operation concludes with an indication of success (e.g., setting a condition code of zero).

20 Described in detail above is a capability for communicating with adapters of a computing environment using control instructions specifically designed for such communication. The communication is performed without using memory mapped I/O and is not limited to control registers in the adapter function. The instructions ensure that the configuration that issues the instruction is authorized to access the adapter function. Further, for the Store Block instruction, it ensures that the specified main storage location is within the configuration's
25 memory.

In the embodiments described herein, the adapters are PCI adapters. PCI, as used herein, refers to any adapters implemented according to a PCI-based specification as defined by the Peripheral Component Interconnect Special Interest Group (PCI-SIG), including but not
30 limited to, PCI or PCIe. In one particular example, the Peripheral Component Interconnect Express (PCIe) is a component level interconnect standard that defines a bi-directional communication protocol for transactions between I/O adapters and host systems. PCIe

communications are encapsulated in packets according to the PCIe standard for transmission on a PCIe bus. Transactions originating at I/O adapters and ending at host systems are referred to as upbound transactions. Transactions originating at host systems and terminating at I/O adapters are referred to as downbound transactions. The PCIe topology is based on point-to-point unidirectional links that are paired (e.g., one upbound link, one downbound link) to form the PCIe bus. The PCIe standard is maintained and published by the PCI-SIG, as noted above in the Background section.

As will be appreciated by one skilled in the art, aspects of the present invention may be embodied as a system, method or computer program product. Accordingly, aspects of the present invention may take the form of an entirely hardware embodiment, an entirely software embodiment (including firmware, resident software, micro-code, etc.) or an embodiment combining software and hardware aspects that may all generally be referred to herein as a "circuit," "module" or "system". Furthermore, aspects of the present invention may take the form of a computer program product embodied in one or more computer readable medium(s) having computer readable program code embodied thereon.

Any combination of one or more computer readable medium(s) may be utilized. The computer readable medium may be a computer readable storage medium. A computer readable storage medium may be, for example, but not limited to, an electronic, magnetic, optical, electromagnetic, infrared or semiconductor system, apparatus, or device, or any suitable combination of the foregoing. More specific examples (a non-exhaustive list) of the computer readable storage medium include the following: an electrical connection having one or more wires, a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), an optical fiber, a portable compact disc read-only memory (CD-ROM), an optical storage device, a magnetic storage device, or any suitable combination of the foregoing. In the context of this document, a computer readable storage medium may be any tangible medium that can contain or store a program for use by or in connection with an instruction execution system, apparatus, or device.

Referring now to FIG. 10, in one example, a computer program product 1000 includes, for instance, one or more computer readable storage media 1002 to store computer readable program code means or logic 1004 thereon to provide and facilitate one or more aspects of the present invention.

5

Program code embodied on a computer readable medium may be transmitted using an appropriate medium, including but not limited to wireless, wireline, optical fiber cable, RF, etc., or any suitable combination of the foregoing.

10

Computer program code for carrying out operations for aspects of the present invention may be written in any combination of one or more programming languages, including an object oriented programming language, such as Java, Smalltalk, C++ or the like, and conventional procedural programming languages, such as the "C" programming language, assembler or similar programming languages. The program code may execute entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network (LAN) or a wide area network (WAN), or the connection may be made to an external computer (for example, through the Internet using an Internet Service Provider).

15

20

25

30

Aspects of the present invention are described herein with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems) and computer program products according to embodiments of the invention. It will be understood that each block of the flowchart illustrations and/or block diagrams, and combinations of blocks in the flowchart illustrations and/or block diagrams, can be implemented by computer program instructions. These computer program instructions may be provided to a processor of a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which execute via the processor of the computer or other programmable data processing apparatus, create means for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

These computer program instructions may also be stored in a computer readable medium that can direct a computer, other programmable data processing apparatus, or other devices to function in a particular manner, such that the instructions stored in the computer readable medium produce an article of manufacture including instructions which implement the function/act specified in the flowchart and/or block diagram block or blocks.

The computer program instructions may also be loaded onto a computer, other programmable data processing apparatus, or other devices to cause a series of operational steps to be performed on the computer, other programmable apparatus or other devices to produce a computer implemented process such that the instructions which execute on the computer or other programmable apparatus provide processes for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

The flowchart and block diagrams in the figures illustrate the architecture, functionality, and operation of possible implementations of systems, methods and computer program products according to various embodiments of the present invention. In this regard, each block in the flowchart or block diagrams may represent a module, segment, or portion of code, which comprises one or more executable instructions for implementing the specified logical function(s). It should also be noted that, in some alternative implementations, the functions noted in the block may occur out of the order noted in the figures. For example, two blocks shown in succession may, in fact, be executed substantially concurrently, or the blocks may sometimes be executed in the reverse order, depending upon the functionality involved. It will also be noted that each block of the block diagrams and/or flowchart illustration, and combinations of blocks in the block diagrams and/or flowchart illustration, can be implemented by special purpose hardware-based systems that perform the specified functions or acts, or combinations of special purpose hardware and computer instructions.

In addition to the above, one or more aspects of the present invention may be provided, offered, deployed, managed, serviced, etc. by a service provider who offers management of customer environments. For instance, the service provider can create, maintain, support, etc. computer code and/or a computer infrastructure that performs one or more aspects of the present invention for one or more customers. In return, the service provider may receive

payment from the customer under a subscription and/or fee agreement, as examples. Additionally or alternatively, the service provider may receive payment from the sale of advertising content to one or more third parties.

5 In one aspect of the present invention, an application may be deployed for performing one or more aspects of the present invention. As one example, the deploying of an application comprises providing computer infrastructure operable to perform one or more aspects of the present invention.

10 As a further aspect of the present invention, a computing infrastructure may be deployed comprising integrating computer readable code into a computing system, in which the code in combination with the computing system is capable of performing one or more aspects of the present invention.

15 As yet a further aspect of the present invention, a process for integrating computing infrastructure comprising integrating computer readable code into a computer system may be provided. The computer system comprises a computer readable medium, in which the computer medium comprises one or more aspects of the present invention. The code in combination with the computer system is capable of performing one or more aspects of the
20 present invention.

Although various embodiments are described above, these are only examples. For example, computing environments of other architectures can incorporate and use one or more aspects of the present invention. As examples, servers other than System z® servers, such as Power
25 Systems servers or other servers offered by International Business Machines Corporation, or servers of other companies can include, use and/or benefit from one or more aspects of the present invention. Further, although in the examples herein, the adapters and PCI hub are considered a part of the server, in other embodiments, they do not have to necessarily be considered a part of the server, but can simply be considered as being coupled to system
30 memory and/or other components of a computing environment. The computing environment need not be a server. Further, although tables are described, any data structure can be used and the term table is to include all such data structures. Yet further, although the adapters

are PCI based, one or more aspects of the present invention are usable with other adapters or other I/O components. Adapter and PCI adapter are just examples. Moreover, the FTE or the parameters of the FTE can be located and maintained in other than secure memory, including, for instance, in hardware (e.g., PCI function hardware). The DTE, FTE and/or
5 handle may include more, less or different information, as well as any of the instructions or instruction fields. Many other variations are possible.

Further, other types of computing environments can benefit from one or more aspects of the present invention. As an example, a data processing system suitable for storing and/or
10 executing program code is usable that includes at least two processors coupled directly or indirectly to memory elements through a system bus. The memory elements include, for instance, local memory employed during actual execution of the program code, bulk storage, and cache memory which provide temporary storage of at least some program code in order to reduce the number of times code must be retrieved from bulk storage during execution.

15 Input/Output or I/O devices (including, but not limited to, keyboards, displays, pointing devices, DASD, tape, CDs, DVDs, thumb drives and other memory media, etc.) can be coupled to the system either directly or through intervening I/O controllers. Network adapters may also be coupled to the system to enable the data processing system to become
20 coupled to other data processing systems or remote printers or storage devices through intervening private or public networks. Modems, cable modems, and Ethernet cards are just a few of the available types of network adapters.

Referring to FIG. 11, representative components of a Host Computer system 5000 to
25 implement one or more aspects of the present invention are portrayed. The representative host computer 5000 comprises one or more CPUs 5001 in communication with computer memory (i.e., central storage) 5002, as well as I/O interfaces to storage media devices 5011 and networks 5010 for communicating with other computers or SANs and the like. The CPU 5001 is compliant with an architecture having an architected instruction set and
30 architected functionality. The CPU 5001 may have dynamic address translation (DAT) 5003 for transforming program addresses (virtual addresses) into real addresses of memory. A DAT typically includes a translation lookaside buffer (TLB) 5007 for caching translations so

that later accesses to the block of computer memory 5002 do not require the delay of address translation. Typically, a cache 5009 is employed between computer memory 5002 and the processor 5001. The cache 5009 may be hierarchical having a large cache available to more than one CPU and smaller, faster (lower level) caches between the large cache and each CPU. In some implementations, the lower level caches are split to provide separate low level caches for instruction fetching and data accesses. In one embodiment, an instruction is fetched from memory 5002 by an instruction fetch unit 5004 via a cache 5009. The instruction is decoded in an instruction decode unit 5006 and dispatched (with other instructions in some embodiments) to instruction execution unit or units 5008. Typically several execution units 5008 are employed, for example an arithmetic execution unit, a floating point execution unit and a branch instruction execution unit. The instruction is executed by the execution unit, accessing operands from instruction specified registers or memory as needed. If an operand is to be accessed (loaded or stored) from memory 5002, a load/store unit 5005 typically handles the access under control of the instruction being executed. Instructions may be executed in hardware circuits or in internal microcode (firmware) or by a combination of both.

As noted, a computer system includes information in local (or main) storage, as well as addressing, protection, and reference and change recording. Some aspects of addressing include the format of addresses, the concept of address spaces, the various types of addresses, and the manner in which one type of address is translated to another type of address. Some of main storage includes permanently assigned storage locations. Main storage provides the system with directly addressable fast-access storage of data. Both data and programs are to be loaded into main storage (from input devices) before they can be processed.

Main storage may include one or more smaller, faster-access buffer storages, sometimes called caches. A cache is typically physically associated with a CPU or an I/O processor. The effects, except on performance, of the physical construction and use of distinct storage media are generally not observable by the program.

Separate caches may be maintained for instructions and for data operands. Information within a cache is maintained in contiguous bytes on an integral boundary called a cache block or cache line (or line, for short). A model may provide an EXTRACT CACHE ATTRIBUTE instruction which returns the size of a cache line in bytes. A model may also provide PREFETCH DATA and PREFETCH DATA RELATIVE LONG instructions which effects the prefetching of storage into the data or instruction cache or the releasing of data from the cache.

Storage is viewed as a long horizontal string of bits. For most operations, accesses to storage proceed in a left-to-right sequence. The string of bits is subdivided into units of eight bits. An eight-bit unit is called a byte, which is the basic building block of all information formats. Each byte location in storage is identified by a unique nonnegative integer, which is the address of that byte location or, simply, the byte address. Adjacent byte locations have consecutive addresses, starting with 0 on the left and proceeding in a left-to-right sequence. Addresses are unsigned binary integers and are 24, 31, or 64 bits.

Information is transmitted between storage and a CPU or a channel subsystem one byte, or a group of bytes, at a time. Unless otherwise specified, in, for instance, the z/Architecture®, a group of bytes in storage is addressed by the leftmost byte of the group. The number of bytes in the group is either implied or explicitly specified by the operation to be performed. When used in a CPU operation, a group of bytes is called a field. Within each group of bytes, in, for instance, the z/Architecture®, bits are numbered in a left-to-right sequence. In the z/Architecture®, the leftmost bits are sometimes referred to as the “high-order” bits and the rightmost bits as the “low-order” bits. Bit numbers are not storage addresses, however. Only bytes can be addressed. To operate on individual bits of a byte in storage, the entire byte is accessed. The bits in a byte are numbered 0 through 7, from left to right (in, e.g., the z/Architecture®). The bits in an address may be numbered 8-31 or 40-63 for 24-bit addresses, or 1-31 or 33-63 for 31-bit addresses; they are numbered 0-63 for 64-bit addresses. Within any other fixed-length format of multiple bytes, the bits making up the format are consecutively numbered starting from 0. For purposes of error detection, and in preferably for correction, one or more check bits may be transmitted with each byte or with a group of bytes. Such check bits are generated automatically by the machine and cannot be

directly controlled by the program. Storage capacities are expressed in number of bytes. When the length of a storage-operand field is implied by the operation code of an instruction, the field is said to have a fixed length, which can be one, two, four, eight, or sixteen bytes. Larger fields may be implied for some instructions. When the length of a storage-operand field is not implied but is stated explicitly, the field is said to have a variable length. Variable-length operands can vary in length by increments of one byte (or with some instructions, in multiples of two bytes or other multiples). When information is placed in storage, the contents of only those byte locations are replaced that are included in the designated field, even though the width of the physical path to storage may be greater than the length of the field being stored.

Certain units of information are to be on an integral boundary in storage. A boundary is called integral for a unit of information when its storage address is a multiple of the length of the unit in bytes. Special names are given to fields of 2, 4, 8, and 16 bytes on an integral boundary. A halfword is a group of two consecutive bytes on a two-byte boundary and is the basic building block of instructions. A word is a group of four consecutive bytes on a four-byte boundary. A doubleword is a group of eight consecutive bytes on an eight-byte boundary. A quadword is a group of 16 consecutive bytes on a 16-byte boundary. When storage addresses designate halfwords, words, doublewords, and quadwords, the binary representation of the address contains one, two, three, or four rightmost zero bits, respectively. Instructions are to be on two-byte integral boundaries. The storage operands of most instructions do not have boundary-alignment requirements.

On devices that implement separate caches for instructions and data operands, a significant delay may be experienced if the program stores into a cache line from which instructions are subsequently fetched, regardless of whether the store alters the instructions that are subsequently fetched.

In one embodiment, the invention may be practiced by software (sometimes referred to licensed internal code, firmware, micro-code, milli-code, pico-code and the like, any of which would be consistent with the present invention). Referring to FIG. 8, software program code which embodies the present invention is typically accessed by processor 5001

of the host system 5000 from long-term storage media devices 5011, such as a CD-ROM drive, tape drive or hard drive. The software program code may be embodied on any of a variety of known media for use with a data processing system, such as a diskette, hard drive, or CD-ROM. The code may be distributed on such media, or may be distributed to users
5 from computer memory 5002 or storage of one computer system over a network 5010 to other computer systems for use by users of such other systems.

The software program code includes an operating system which controls the function and interaction of the various computer components and one or more application programs.

10 Program code is normally paged from storage media device 5011 to the relatively higher-speed computer storage 5002 where it is available for processing by processor 5001. The techniques and methods for embodying software program code in memory, on physical media, and/or distributing software code via networks are well known and will not be further discussed herein. Program code, when created and stored on a tangible medium (including
15 but not limited to electronic memory modules (RAM), flash memory, Compact Discs (CDs), DVDs, Magnetic Tape and the like is often referred to as a "computer program product". The computer program product medium is typically readable by a processing circuit preferably in a computer system for execution by the processing circuit.

20 FIG. 12 illustrates a representative workstation or server hardware system in which the present invention may be practiced. The system 5020 of FIG. 12 comprises a representative base computer system 5021, such as a personal computer, a workstation or a server, including optional peripheral devices. The base computer system 5021 includes one or more
25 processors 5026 and a bus employed to connect and enable communication between the processor(s) 5026 and the other components of the system 5021 in accordance with known techniques. The bus connects the processor 5026 to memory 5025 and long-term storage 5027 which can include a hard drive (including any of magnetic media, CD, DVD and Flash Memory for example) or a tape drive for example. The system 5021 might also include a
30 user interface adapter, which connects the microprocessor 5026 via the bus to one or more interface devices, such as a keyboard 5024, a mouse 5023, a printer/scanner 5030 and/or other interface devices, which can be any user interface device, such as a touch sensitive

screen, digitized entry pad, etc. The bus also connects a display device 5022, such as an LCD screen or monitor, to the microprocessor 5026 via a display adapter.

The system 5021 may communicate with other computers or networks of computers by way of a network adapter capable of communicating 5028 with a network 5029. Example network adapters are communications channels, token ring, Ethernet or modems. Alternatively, the system 5021 may communicate using a wireless interface, such as a CDPD (cellular digital packet data) card. The system 5021 may be associated with such other computers in a Local Area Network (LAN) or a Wide Area Network (WAN), or the system 5021 can be a client in a client/server arrangement with another computer, etc. All of these configurations, as well as the appropriate communications hardware and software, are known in the art.

FIG. 13 illustrates a data processing network 5040 in which the present invention may be practiced. The data processing network 5040 may include a plurality of individual networks, such as a wireless network and a wired network, each of which may include a plurality of individual workstations 5041, 5042, 5043, 5044. Additionally, as those skilled in the art will appreciate, one or more LANs may be included, where a LAN may comprise a plurality of intelligent workstations coupled to a host processor.

Still referring to FIG. 13, the networks may also include mainframe computers or servers, such as a gateway computer (client server 5046) or application server (remote server 5048 which may access a data repository and may also be accessed directly from a workstation 5045). A gateway computer 5046 serves as a point of entry into each individual network. A gateway is needed when connecting one networking protocol to another. The gateway 5046 may be preferably coupled to another network (the Internet 5047 for example) by means of a communications link. The gateway 5046 may also be directly coupled to one or more workstations 5041, 5042, 5043, 5044 using a communications link. The gateway computer may be implemented utilizing an IBM cServer™ System z® server available from International Business Machines Corporation.

Referring concurrently to FIG. 12 and FIG. 13, software programming code which may embody the present invention may be accessed by the processor 5026 of the system 5020 from long-term storage media 5027, such as a CD-ROM drive or hard drive. The software programming code may be embodied on any of a variety of known media for use with a data processing system, such as a diskette, hard drive, or CD-ROM. The code may be distributed on such media, or may be distributed to users 5050, 5051 from the memory or storage of one computer system over a network to other computer systems for use by users of such other systems.

Alternatively, the programming code may be embodied in the memory 5025, and accessed by the processor 5026 using the processor bus. Such programming code includes an operating system which controls the function and interaction of the various computer components and one or more application programs 5032. Program code is normally paged from storage media 5027 to high-speed memory 5025 where it is available for processing by the processor 5026. The techniques and methods for embodying software programming code in memory, on physical media, and/or distributing software code via networks are well known and will not be further discussed herein. Program code, when created and stored on a tangible medium (including but not limited to electronic memory modules (RAM), flash memory, Compact Discs (CDs), DVDs, Magnetic Tape and the like is often referred to as a “computer program product”. The computer program product medium is typically readable by a processing circuit preferably in a computer system for execution by the processing circuit.

The cache that is most readily available to the processor (normally faster and smaller than other caches of the processor) is the lowest (L1 or level one) cache and main store (main memory) is the highest level cache (L3 if there are 3 levels). The lowest level cache is often divided into an instruction cache (I-Cache) holding machine instructions to be executed and a data cache (D-Cache) holding data operands.

Referring to FIG. 14, an exemplary processor embodiment is depicted for processor 5026. Typically one or more levels of cache 5053 are employed to buffer memory blocks in order to improve processor performance. The cache 5053 is a high speed buffer holding cache

lines of memory data that are likely to be used. Typical cache lines are 64, 128 or 256 bytes of memory data. Separate caches are often employed for caching instructions than for caching data. Cache coherence (synchronization of copies of lines in memory and the caches) is often provided by various “snoop” algorithms well known in the art. Main
5 memory storage 5025 of a processor system is often referred to as a cache. In a processor system having 4 levels of cache 5053, main storage 5025 is sometimes referred to as the level 5 (L5) cache since it is typically faster and only holds a portion of the non-volatile storage (DASD, tape etc) that is available to a computer system. Main storage 5025 “caches” pages of data paged in and out of the main storage 5025 by the operating system.

10 A program counter (instruction counter) 5061 keeps track of the address of the current instruction to be executed. A program counter in a z/Architecture® processor is 64 bits and can be truncated to 31 or 24 bits to support prior addressing limits. A program counter is typically embodied in a PSW (program status word) of a computer such that it persists
15 during context switching. Thus, a program in progress, having a program counter value, may be interrupted by, for example, the operating system (context switch from the program environment to the operating system environment). The PSW of the program maintains the program counter value while the program is not active, and the program counter (in the PSW) of the operating system is used while the operating system is executing. Typically,
20 the program counter is incremented by an amount equal to the number of bytes of the current instruction. RISC (Reduced Instruction Set Computing) instructions are typically fixed length while CISC (Complex Instruction Set Computing) instructions are typically variable length. Instructions of the IBM z/Architecture® are CISC instructions having a length of 2, 4 or 6 bytes. The Program counter 5061 is modified by either a context switch operation or a
25 branch taken operation of a branch instruction for example. In a context switch operation, the current program counter value is saved in the program status word along with other state information about the program being executed (such as condition codes), and a new program counter value is loaded pointing to an instruction of a new program module to be executed. A branch taken operation is performed in order to permit the program to make decisions or
30 loop within the program by loading the result of the branch instruction into the program counter 5061.

Typically an instruction fetch unit 5055 is employed to fetch instructions on behalf of the processor 5026. The fetch unit either fetches “next sequential instructions”, target instructions of branch taken instructions, or first instructions of a program following a context switch. Modern Instruction fetch units often employ prefetch techniques to speculatively prefetch instructions based on the likelihood that the prefetched instructions might be used. For example, a fetch unit may fetch 16 bytes of instruction that includes the next sequential instruction and additional bytes of further sequential instructions.

The fetched instructions are then executed by the processor 5026. In an embodiment, the fetched instruction(s) are passed to a dispatch unit 5056 of the fetch unit. The dispatch unit decodes the instruction(s) and forwards information about the decoded instruction(s) to appropriate units 5057, 5058, 5060. An execution unit 5057 will typically receive information about decoded arithmetic instructions from the instruction fetch unit 5055 and will perform arithmetic operations on operands according to the opcode of the instruction. Operands are provided to the execution unit 5057 preferably either from memory 5025, architected registers 5059 or from an immediate field of the instruction being executed. Results of the execution, when stored, are stored either in memory 5025, registers 5059 or in other machine hardware (such as control registers, PSW registers and the like).

A processor 5026 typically has one or more units 5057, 5058, 5060 for executing the function of the instruction. Referring to FIG. 15A, an execution unit 5057 may communicate with architected general registers 5059, a decode/dispatch unit 5056, a load store unit 5060, and other 5065 processor units by way of interfacing logic 5071. An execution unit 5057 may employ several register circuits 5067, 5068, 5069 to hold information that the arithmetic logic unit (ALU) 5066 will operate on. The ALU performs arithmetic operations such as add, subtract, multiply and divide as well as logical function such as and, or and exclusive-or (XOR), rotate and shift. Preferably the ALU supports specialized operations that are design dependent. Other circuits may provide other architected facilities 5072 including condition codes and recovery support logic for example. Typically the result of an ALU operation is held in an output register circuit 5070 which can forward the result to a variety of other processing functions. There are many arrangements

of processor units, the present description is only intended to provide a representative understanding of one embodiment.

5 An ADD instruction for example would be executed in an execution unit 5057 having arithmetic and logical functionality while a floating point instruction for example would be executed in a floating point execution having specialized floating point capability. Preferably, an execution unit operates on operands identified by an instruction by performing an opcode defined function on the operands. For example, an ADD instruction may be executed by an execution unit 5057 on operands found in two registers 5059 identified by
10 register fields of the instruction.

The execution unit 5057 performs the arithmetic addition on two operands and stores the result in a third operand where the third operand may be a third register or one of the two source registers. The execution unit preferably utilizes an Arithmetic Logic Unit (ALU)
15 5066 that is capable of performing a variety of logical functions such as Shift, Rotate, And, Or and XOR as well as a variety of algebraic functions including any of add, subtract, multiply, divide. Some ALUs 5066 are designed for scalar operations and some for floating point. Data may be Big Endian (where the least significant byte is at the highest byte address) or Little Endian (where the least significant byte is at the lowest byte address)
20 depending on architecture. The IBM z/Architecture® is Big Endian. Signed fields may be sign and magnitude, 1's complement or 2's complement depending on architecture. A 2's complement number is advantageous in that the ALU does not need to design a subtract capability since either a negative value or a positive value in 2's complement requires only an addition within the ALU. Numbers are commonly described in shorthand, where a 12 bit
25 field defines an address of a 4,096 byte block and is commonly described as a 4 Kbyte (Kilo-byte) block, for example.

Referring to FIG. 15B, branch instruction information for executing a branch instruction is typically sent to a branch unit 5058 which often employs a branch prediction algorithm such
30 as a branch history table 5082 to predict the outcome of the branch before other conditional operations are complete. The target of the current branch instruction will be fetched and speculatively executed before the conditional operations are complete. When the conditional

operations are completed the speculatively executed branch instructions are either completed or discarded based on the conditions of the conditional operation and the speculated outcome. A typical branch instruction may test condition codes and branch to a target address if the condition codes meet the branch requirement of the branch instruction, a target address may be calculated based on several numbers including ones found in register fields or an immediate field of the instruction for example. The branch unit 5058 may employ an ALU 5074 having a plurality of input register circuits 5075, 5076, 5077 and an output register circuit 5080. The branch unit 5058 may communicate with general registers 5059, decode dispatch unit 5056 or other circuits 5073, for example.

The execution of a group of instructions can be interrupted for a variety of reasons including a context switch initiated by an operating system, a program exception or error causing a context switch, an I/O interruption signal causing a context switch or multi-threading activity of a plurality of programs (in a multi-threaded environment), for example. Preferably a context switch action saves state information about a currently executing program and then loads state information about another program being invoked. State information may be saved in hardware registers or in memory for example. State information preferably comprises a program counter value pointing to a next instruction to be executed, condition codes, memory translation information and architected register content. A context switch activity can be exercised by hardware circuits, application programs, operating system programs or firmware code (microcode, pico-code or licensed internal code (LIC)) alone or in combination.

A processor accesses operands according to instruction defined methods. The instruction may provide an immediate operand using the value of a portion of the instruction, may provide one or more register fields explicitly pointing to either general purpose registers or special purpose registers (floating point registers for example). The instruction may utilize implied registers identified by an opcode field as operands. The instruction may utilize memory locations for operands. A memory location of an operand may be provided by a register, an immediate field, or a combination of registers and immediate field as exemplified by the z/Architecture® long displacement facility wherein the instruction defines a base register, an index register and an immediate field (displacement field) that are

added together to provide the address of the operand in memory for example. Location herein typically implies a location in main memory (main storage) unless otherwise indicated.

5 Referring to FIG. 15C, a processor accesses storage using a load/store unit 5060. The load/store unit 5060 may perform a load operation by obtaining the address of the target operand in memory 5053 and loading the operand in a register 5059 or another memory 5053 location, or may perform a store operation by obtaining the address of the target operand in memory 5053 and storing data obtained from a register 5059 or another memory
10 5053 location in the target operand location in memory 5053. The load/store unit 5060 may be speculative and may access memory in a sequence that is out-of-order relative to instruction sequence, however the load/store unit 5060 is to maintain the appearance to programs that instructions were executed in order. A load/store unit 5060 may communicate with general registers 5059, decode/dispatch unit 5056, cache/memory interface 5053 or
15 other elements 5083 and comprises various register circuits, ALUs 5085 and control logic 5090 to calculate storage addresses and to provide pipeline sequencing to keep operations in-order. Some operations may be out of order but the load/store unit provides functionality to make the out of order operations to appear to the program as having been performed in order, as is well known in the art.

20 Preferably addresses that an application program “sees” are often referred to as virtual addresses. Virtual addresses are sometimes referred to as “logical addresses” and “effective addresses”. These virtual addresses are virtual in that they are redirected to physical memory location by one of a variety of dynamic address translation (DAT) technologies
25 including, but not limited to, simply prefixing a virtual address with an offset value, translating the virtual address via one or more translation tables, the translation tables preferably comprising at least a segment table and a page table alone or in combination, preferably, the segment table having an entry pointing to the page table. In the
z/Architecture®, a hierarchy of translation is provided including a region first table, a region
30 second table, a region third table, a segment table and an optional page table. The performance of the address translation is often improved by utilizing a translation lookaside buffer (TLB) which comprises entries mapping a virtual address to an associated physical

memory location. The entries are created when the DAT translates a virtual address using the translation tables. Subsequent use of the virtual address can then utilize the entry of the fast TLB rather than the slow sequential translation table accesses. TLB content may be managed by a variety of replacement algorithms including LRU (Least Recently used).

5

In the case where the processor is a processor of a multi-processor system, each processor has responsibility to keep shared resources, such as I/O, caches, TLBs and memory, interlocked for coherency. Typically, “snoop” technologies will be utilized in maintaining cache coherency. In a snoop environment, each cache line may be marked as being in any one of a shared state, an exclusive state, a changed state, an invalid state and the like in order to facilitate sharing.

10

I/O units 5054 (FIG. 14) provide the processor with means for attaching to peripheral devices including tape, disc, printers, displays, and networks for example. I/O units are often presented to the computer program by software drivers. In mainframes, such as the System z® from IBM®, channel adapters and open system adapters are I/O units of the mainframe that provide the communications between the operating system and peripheral devices.

15

Further, other types of computing environments can benefit from one or more aspects of the present invention. As an example, an environment may include an emulator (e.g., software or other emulation mechanisms), in which a particular architecture (including, for instance, instruction execution, architected functions, such as address translation, and architected registers) or a subset thereof is emulated (e.g., on a native computer system having a processor and memory). In such an environment, one or more emulation functions of the emulator can implement one or more aspects of the present invention, even though a computer executing the emulator may have a different architecture than the capabilities being emulated. As one example, in emulation mode, the specific instruction or operation being emulated is decoded, and an appropriate emulation function is built to implement the individual instruction or operation.

20

25

30

In an emulation environment, a host computer includes, for instance, a memory to store instructions and data; an instruction fetch unit to fetch instructions from memory and to

optionally, provide local buffering for the fetched instruction; an instruction decode unit to receive the fetched instructions and to determine the type of instructions that have been fetched; and an instruction execution unit to execute the instructions. Execution may include loading data into a register from memory; storing data back to memory from a register; or performing some type of arithmetic or logical operation, as determined by the decode unit. In one example, each unit is implemented in software. For instance, the operations being performed by the units are implemented as one or more subroutines within emulator software.

More particularly, in a mainframe, architected machine instructions are used by programmers, usually today “C” programmers, often by way of a compiler application. These instructions stored in the storage medium may be executed natively in a z/Architecture® IBM® Server, or alternatively in machines executing other architectures. They can be emulated in the existing and in future IBM® mainframe servers and on other machines of IBM® (e.g., Power Systems servers and System x® Servers). They can be executed in machines running Linux on a wide variety of machines using hardware manufactured by IBM®, Intel®, AMD™, and others. Besides execution on that hardware under a z/Architecture®, Linux can be used as well as machines which use emulation by Hercules or FSI (Fundamental Software, Inc), where generally execution is in an emulation mode. In emulation mode, emulation software is executed by a native processor to emulate the architecture of an emulated processor. Information about the above-referenced emulator products is available on the World Wide Web at, respectively, www.hercules-390.org and www.funsoft.com.

The native processor typically executes emulation software comprising either firmware or a native operating system to perform emulation of the emulated processor. The emulation software is responsible for fetching and executing instructions of the emulated processor architecture. The emulation software maintains an emulated program counter to keep track of instruction boundaries. The emulation software may fetch one or more emulated machine instructions at a time and convert the one or more emulated machine instructions to a corresponding group of native machine instructions for execution by the native processor. These converted instructions may be cached such that a faster conversion can be

accomplished. Notwithstanding, the emulation software is to maintain the architecture rules of the emulated processor architecture so as to assure operating systems and applications written for the emulated processor operate correctly. Furthermore, the emulation software is to provide resources identified by the emulated processor architecture including, but not limited to, control registers, general purpose registers, floating point registers, dynamic address translation function including segment tables and page tables for example, interrupt mechanisms, context switch mechanisms, Time of Day (TOD) clocks and architected interfaces to I/O subsystems such that an operating system or an application program designed to run on the emulated processor, can be run on the native processor having the emulation software.

A specific instruction being emulated is decoded, and a subroutine is called to perform the function of the individual instruction. An emulation software function emulating a function of an emulated processor is implemented, for example, in a "C" subroutine or driver, or some other method of providing a driver for the specific hardware as will be within the skill of those in the art after understanding the description of the preferred embodiment. Various software and hardware emulation patents including, but not limited to U.S. Letters Patent No. 5,551,013, entitled "Multiprocessor for Hardware Emulation", by Beausoleil et al.; and U.S. Letters Patent No. 6,009,261, entitled "Preprocessing of Stored Target Routines for Emulating Incompatible Instructions on a Target Processor", by Scalzi et al; and U.S. Letters Patent No. 5,574,873, entitled "Decoding Guest Instruction to Directly Access Emulation Routines that Emulate the Guest Instructions", by Davidian et al; and U.S. Letters Patent No. 6,308,255, entitled "Symmetrical Multiprocessing Bus and Chipset Used for Coprocessor Support Allowing Non-Native Code to Run in a System", by Gorishek et al; and U.S. Letters Patent No. 6,463,582, entitled "Dynamic Optimizing Object Code Translator for Architecture Emulation and Dynamic Optimizing Object Code Translation Method", by Lethin et al; and U.S. Letters Patent No. 5,790,825, entitled "Method for Emulating Guest Instructions on a Host Computer Through Dynamic Recompilation of Host Instructions", by Eric Traut; and many others, illustrate a variety of known ways to achieve emulation of an instruction format architected for a different machine for a target machine available to those skilled in the art.

In FIG. 16, an example of an emulated host computer system 5092 is provided that emulates a host computer system 5000' of a host architecture. In the emulated host computer system 5092, the host processor (CPU) 5091 is an emulated host processor (or virtual host processor) and comprises an emulation processor 5093 having a different native instruction set architecture than that of the processor 5091 of the host computer 5000'. The emulated host computer system 5092 has memory 5094 accessible to the emulation processor 5093. In the example embodiment, the memory 5094 is partitioned into a host computer memory 5096 portion and an emulation routines 5097 portion. The host computer memory 5096 is available to programs of the emulated host computer 5092 according to host computer architecture. The emulation processor 5093 executes native instructions of an architected instruction set of an architecture other than that of the emulated processor 5091, the native instructions obtained from emulation routines memory 5097, and may access a host instruction for execution from a program in host computer memory 5096 by employing one or more instruction(s) obtained in a sequence & access/decode routine which may decode the host instruction(s) accessed to determine a native instruction execution routine for emulating the function of the host instruction accessed. Other facilities that are defined for the host computer system 5000' architecture may be emulated by architected facilities routines, including such facilities as general purpose registers, control registers, dynamic address translation and I/O subsystem support and processor cache, for example. The emulation routines may also take advantage of functions available in the emulation processor 5093 (such as general registers and dynamic translation of virtual addresses) to improve performance of the emulation routines. Special hardware and off-load engines may also be provided to assist the processor 5093 in emulating the function of the host computer 5000'.

The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the invention. As used herein, the singular forms "a", "an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms "comprises" and/or "comprising", when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components and/or groups thereof.

The corresponding structures, materials, acts, and equivalents of all means or step plus function elements in the claims below, if any, are intended to include any structure, material, or act for performing the function in combination with other claimed elements as specifically claimed. The description of the present invention has been presented for purposes of
5 illustration and description, but is not intended to be exhaustive or limited to the invention in the form disclosed. Many modifications and variations will be apparent to those of ordinary skill in the art without departing from the scope of the invention. The embodiment was chosen and described in order to best explain the principles of the invention and the practical application, and to enable others of ordinary skill in the art to understand the invention for
10 various embodiments with various modifications as are suited to the particular use contemplated.

CLAIMS

1. A method for executing a store instruction for storing data in an adapter, comprising the steps of:

5 a computer readable storage medium readable by a processing circuit and storing instructions for execution by the processing circuit for performing a method comprising:
obtaining a machine instruction for execution, the machine instruction being defined for computer execution according to a computer architecture, the machine instruction comprising:

10 an opcode field identifying a store to adapter instruction;
a first field identifying a first location that includes data to be stored in an adapter;
a second field identifying a second location, the contents of which include a function handle identifying the adapter, a designation of an address space within the adapter in which data is to be stored, and an offset within the address space; and

15 executing the machine instruction, the executing comprising:
using the function handle to obtain a function table entry associated with the adapter;
obtaining a data address of the adapter using at least one of information in the function table entry and the offset; and

20 storing data from the first location in a specific location in the address space identified by the designation of the address space, the specific location identified by the data address of the adapter.

2. The method of claim 1, wherein the address space to be accessed is one of a memory space or an I/O space, and wherein the obtaining the data address comprises using one or
25 more parameters of the function table entry to obtain the data address.

3. The method of claim 2, wherein the using one or more parameters comprises adding a value of a base address register of the function table entry to the offset to obtain the data address.

4. The method of claim 1, wherein the address space to be accessed is a configuration space, and wherein the data address is the offset provided by the instruction, the offset identifying a register number in the configuration space.

5. The method of claim 1, wherein the contents of the second location includes an amount of data to be stored.

6. The method of claim 1, wherein the instruction is implemented based on the architecture of the adapter.

10

7. A computer program comprising computer program code stored on a computer readable medium to, when loaded into a computer system and executed thereon, cause said computer system to perform all the steps of a method according to any of claims 1 to 6.

15 8. A computer system for executing a store instruction for storing data in an adapter, said computer system comprising:

a memory;

a processor in communications with the memory;

the processor obtaining a machine instruction for execution, the machine instruction

20 being defined for computer execution according to a computer architecture, the machine instruction comprising:

an opcode field identifying a store to adapter instruction;

a first field identifying a first location that includes data to be stored in an adapter;

a second field identifying a second location, the contents of which include a function

25 handle identifying the adapter, a designation of an address space within the adapter in which data is to be stored, and an offset within the address space; and

the processor executing the machine instruction, the executing comprising:

using the function handle to obtain a function table entry associated with the adapter;

obtaining a data address of the adapter using at least one of information in the

30 function table entry and the offset; and

storing data from the first location in a specific location in the address space identified by the designation of the address space, the specific location identified by the data address of the adapter.

5 9. A computer system for executing a store instruction for storing data in an adapter, said computer system comprising:

a memory;

a processor in communications with the memory, wherein the computer system is configured to perform a method, said method comprising:

10 obtaining a machine instruction for execution, the machine instruction being defined for computer execution according to a computer architecture, the machine instruction comprising:

an opcode field identifying a store to adapter instruction;

a first field identifying a first location that includes data to be stored in an adapter;

15 a second field identifying a second location, the contents of which include a function handle identifying the adapter, a designation of an address space within the adapter in which data is to be stored, and an offset within the address space; and

executing the machine instruction, the executing comprising:

using the function handle to obtain a function table entry associated with the adapter;

20 obtaining a data address of the adapter using at least one of information in the function table entry and the offset; and

storing data from the first location in a specific location in the address space identified by the designation of the address space, the specific location identified by the data address of the adapter.

25 10. The computer system of claim 9, wherein the address space to be accessed is one of a memory space or an I/O space, and wherein the obtaining the data address comprises using one or more parameters of the function table entry to obtain the data address.

30 11. The computer system of claim 10, wherein the using one or more parameters comprises adding a value of a base address register of the function table entry to the offset to obtain the data address.

12. The computer system of claim 9, wherein the address space to be accessed is a configuration space, and wherein the data address is the offset provided by the instruction, the offset identifying a register number in the configuration space.

5 13. The computer system of claim 9, wherein the contents of the second location includes an amount of data to be stored.

14. The computer system of claim 9, wherein the instruction is implemented based on the architecture of the adapter.

10

15. A computer program product for executing a store instruction for storing data in an adapter, said computer program product comprising:

a computer readable storage medium readable by a processing circuit and storing instructions for execution by the processing circuit for performing all the steps of the method
15 of any of claims 1 to 6.

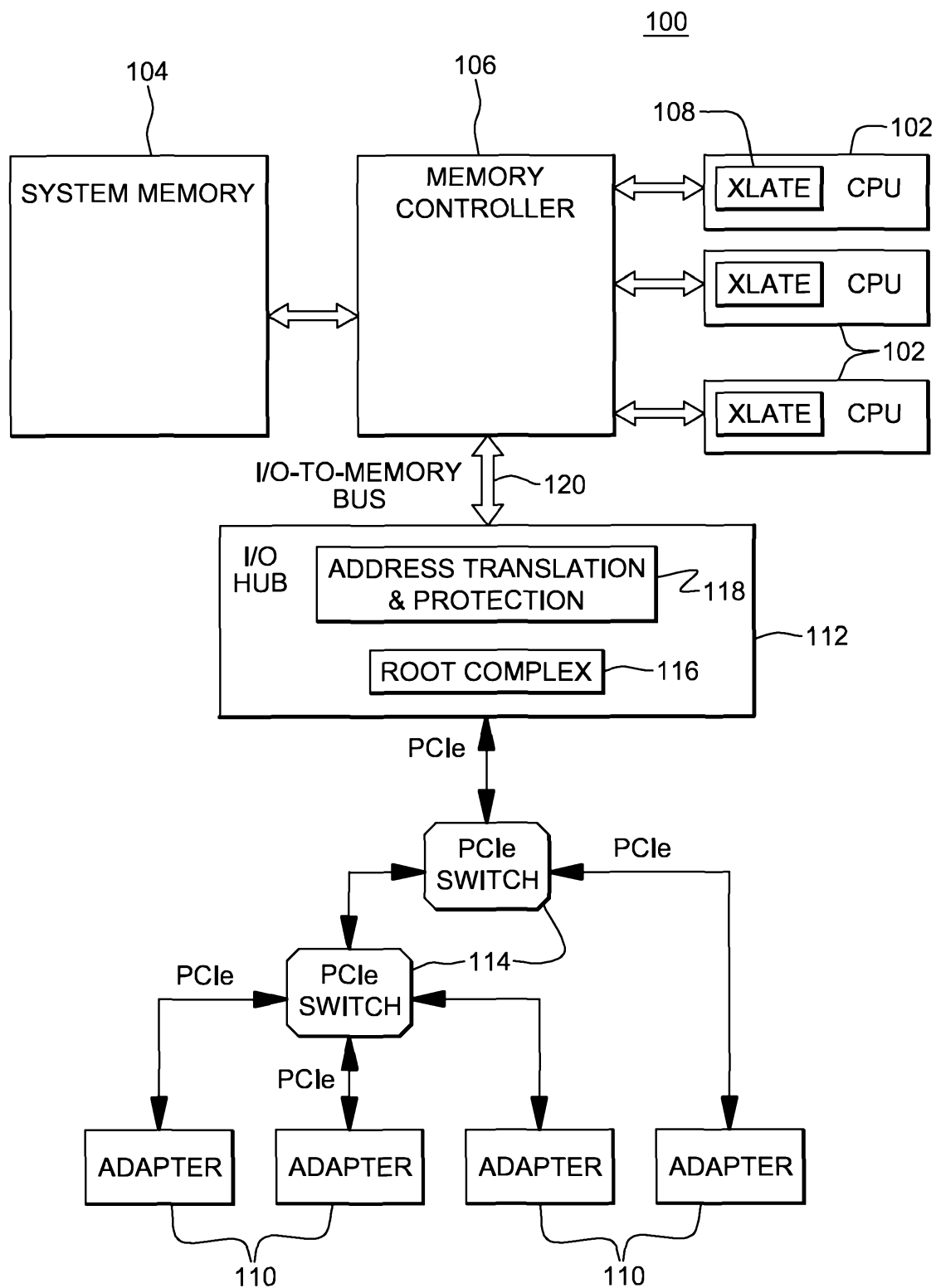


FIG. 1A

2/22

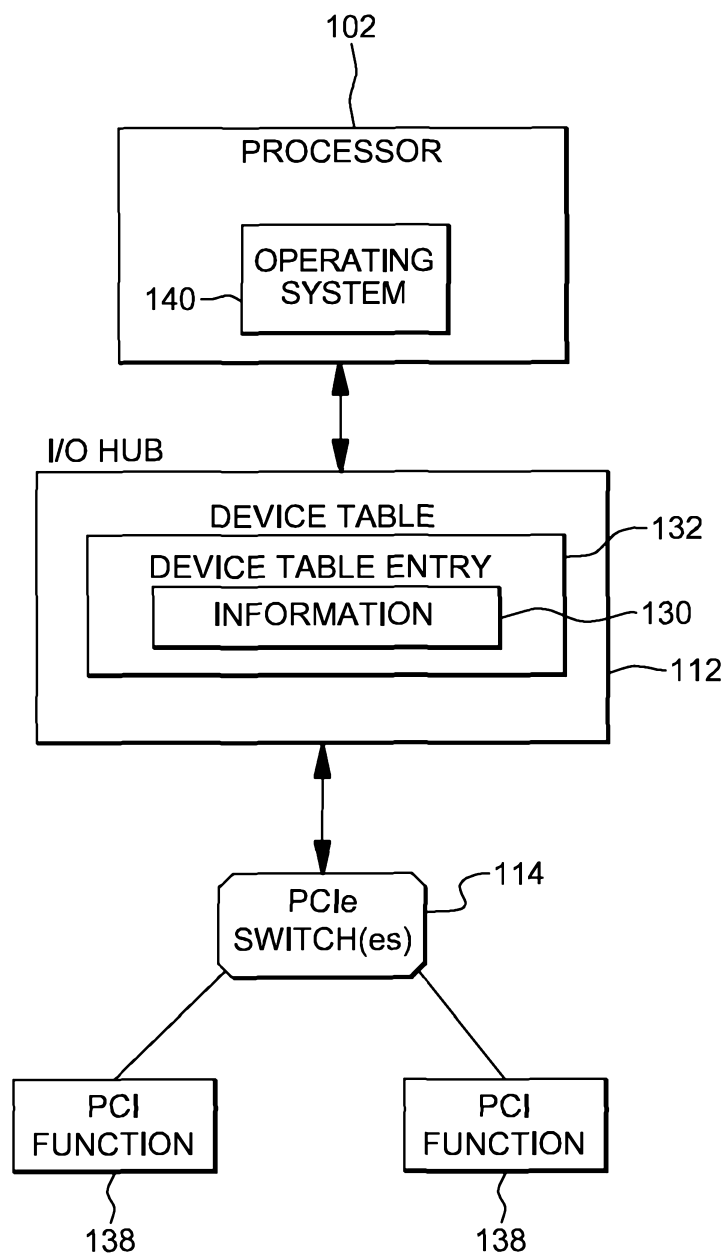


FIG. 1B

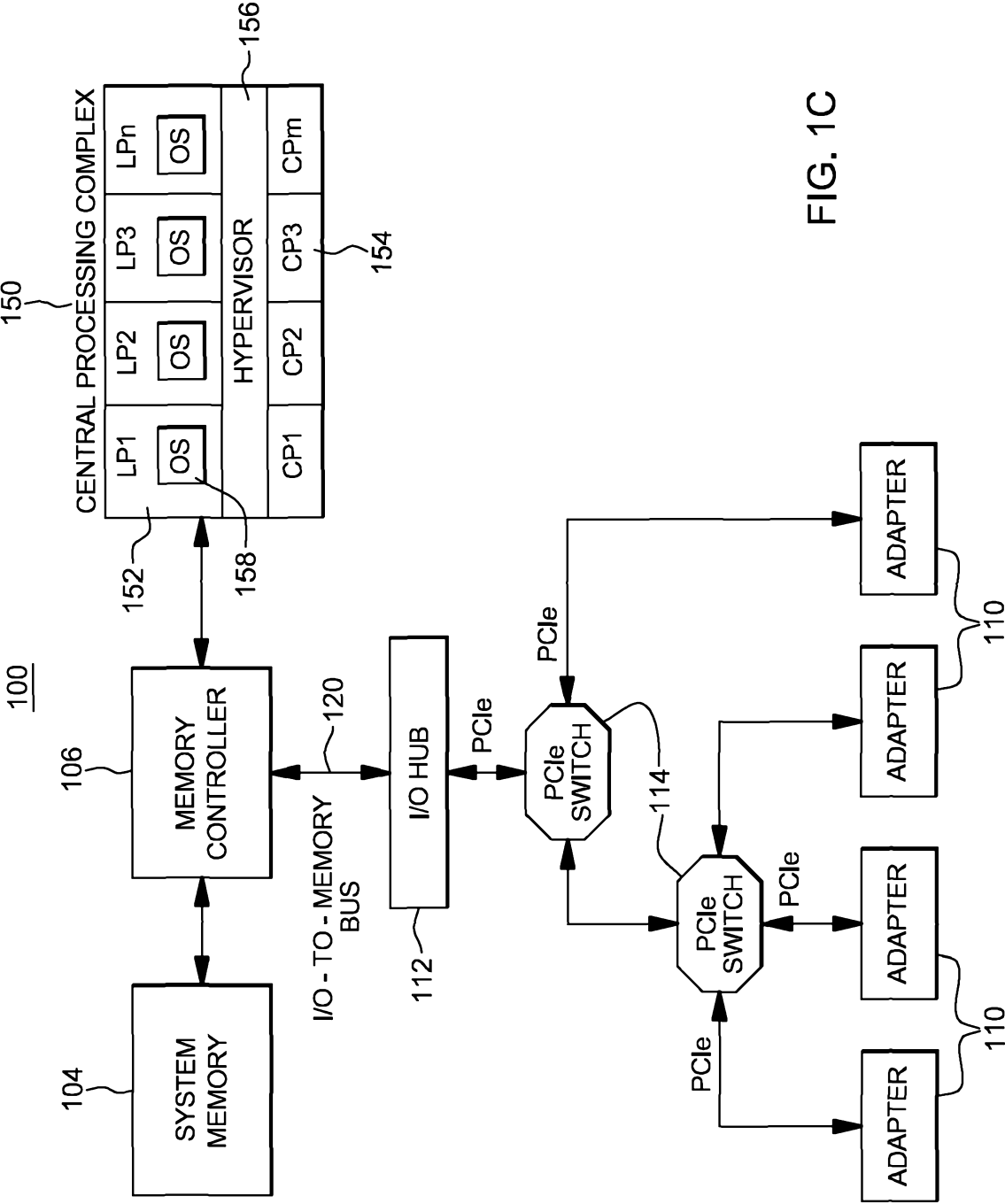


FIG. 1C

4/22

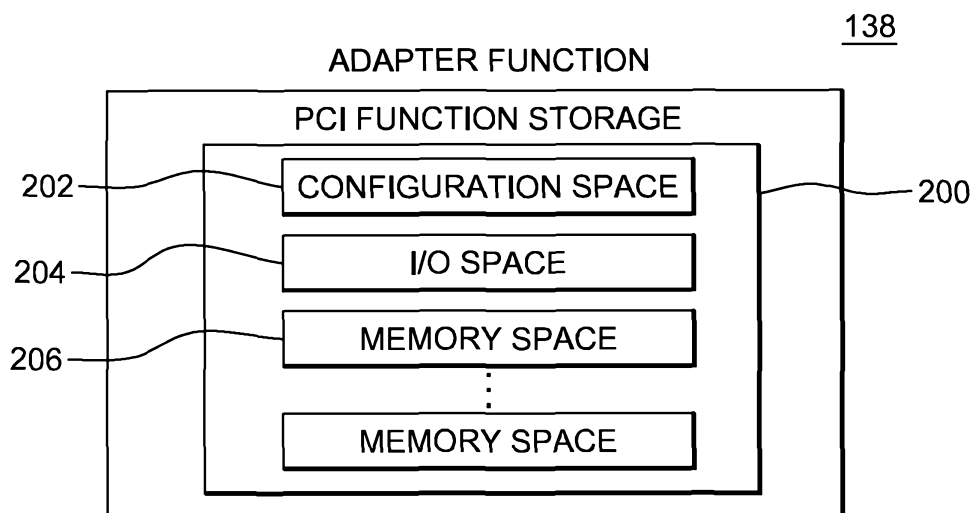


FIG. 2

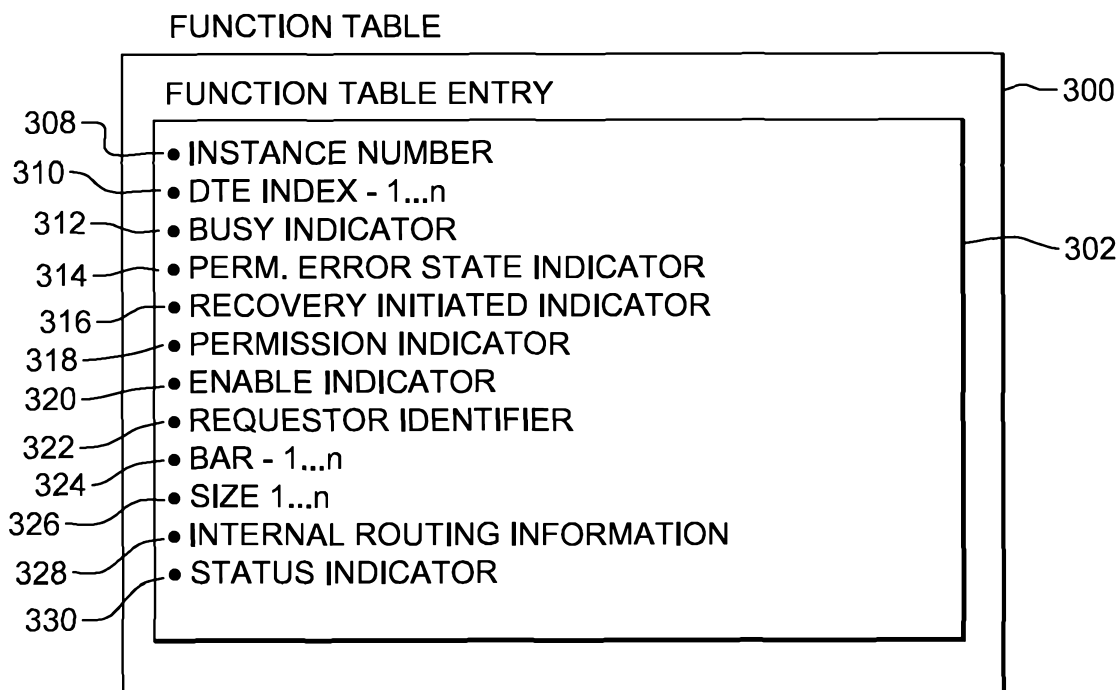


FIG. 3A

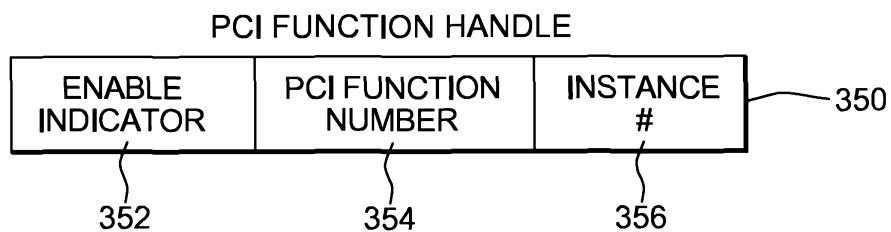


FIG. 3B

5/22

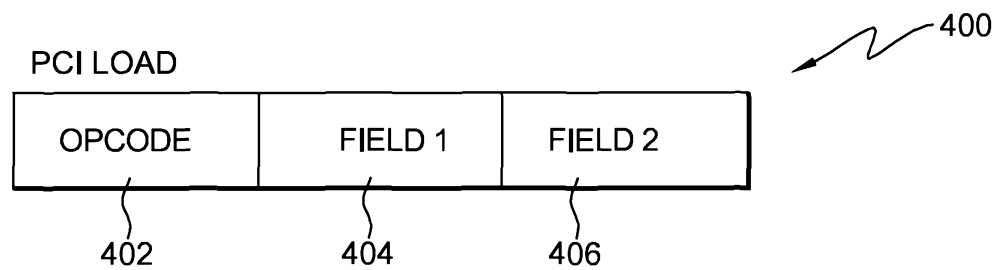


FIG. 4A

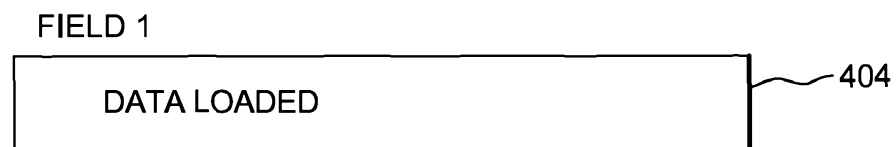


FIG. 4B

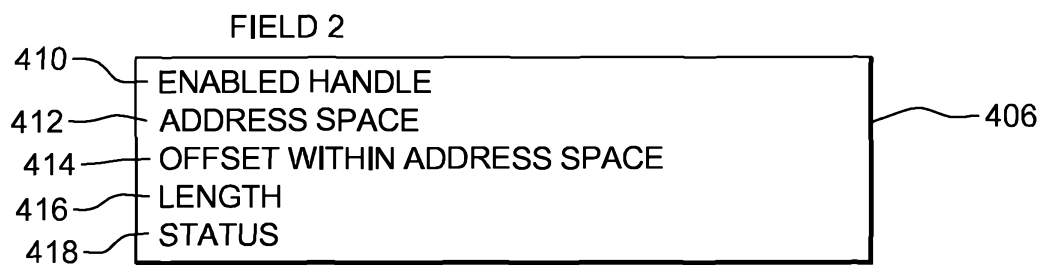
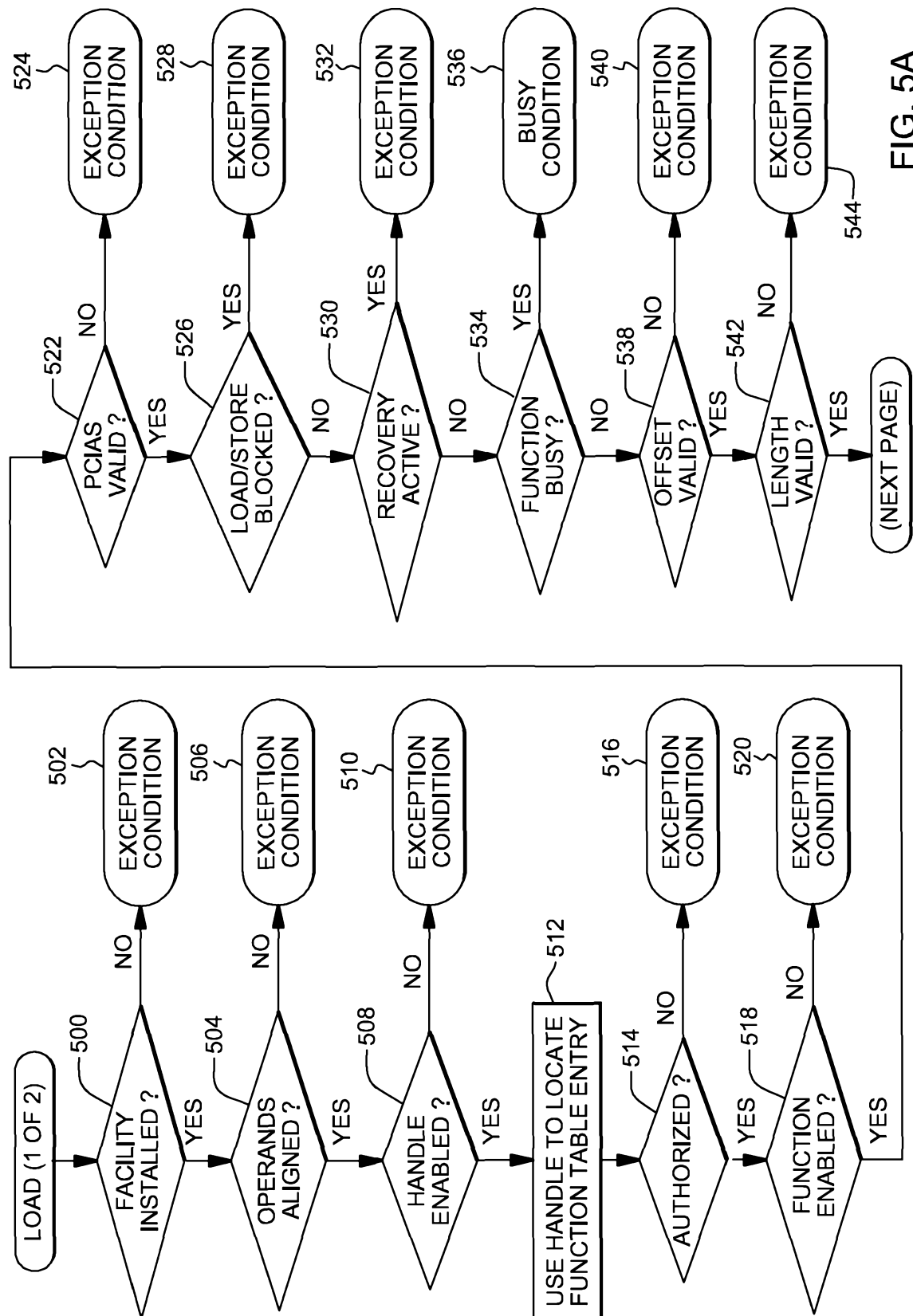


FIG. 4C

6/22



7/22

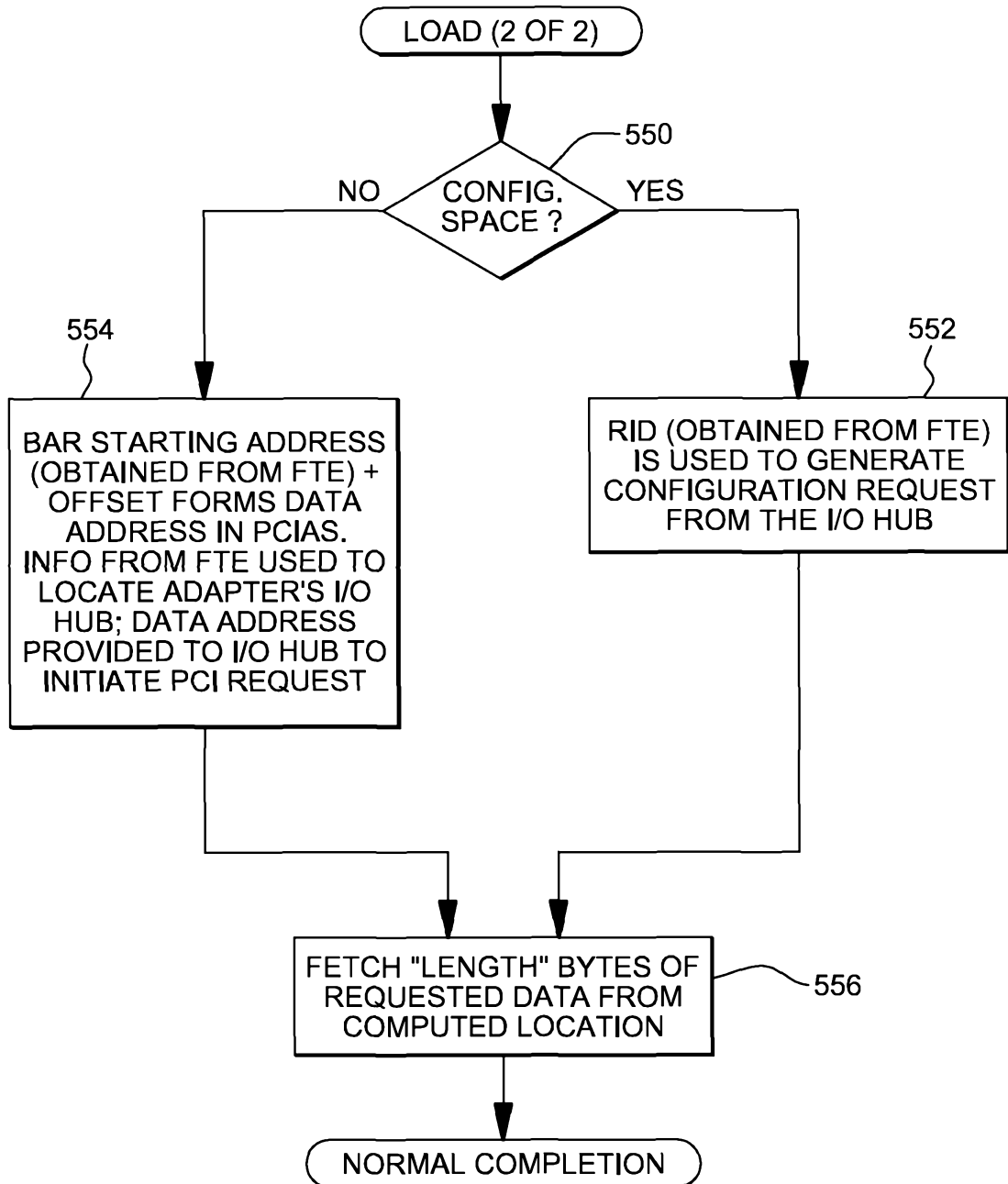


FIG. 5B

8/22

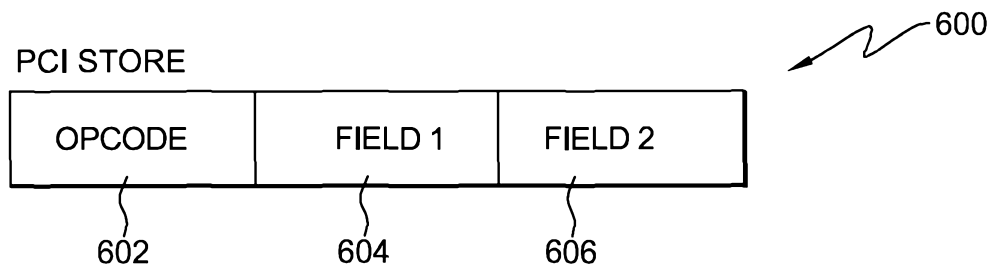


FIG. 6A

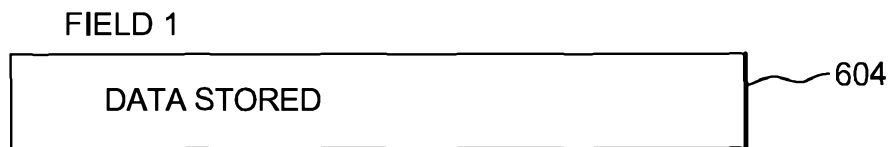
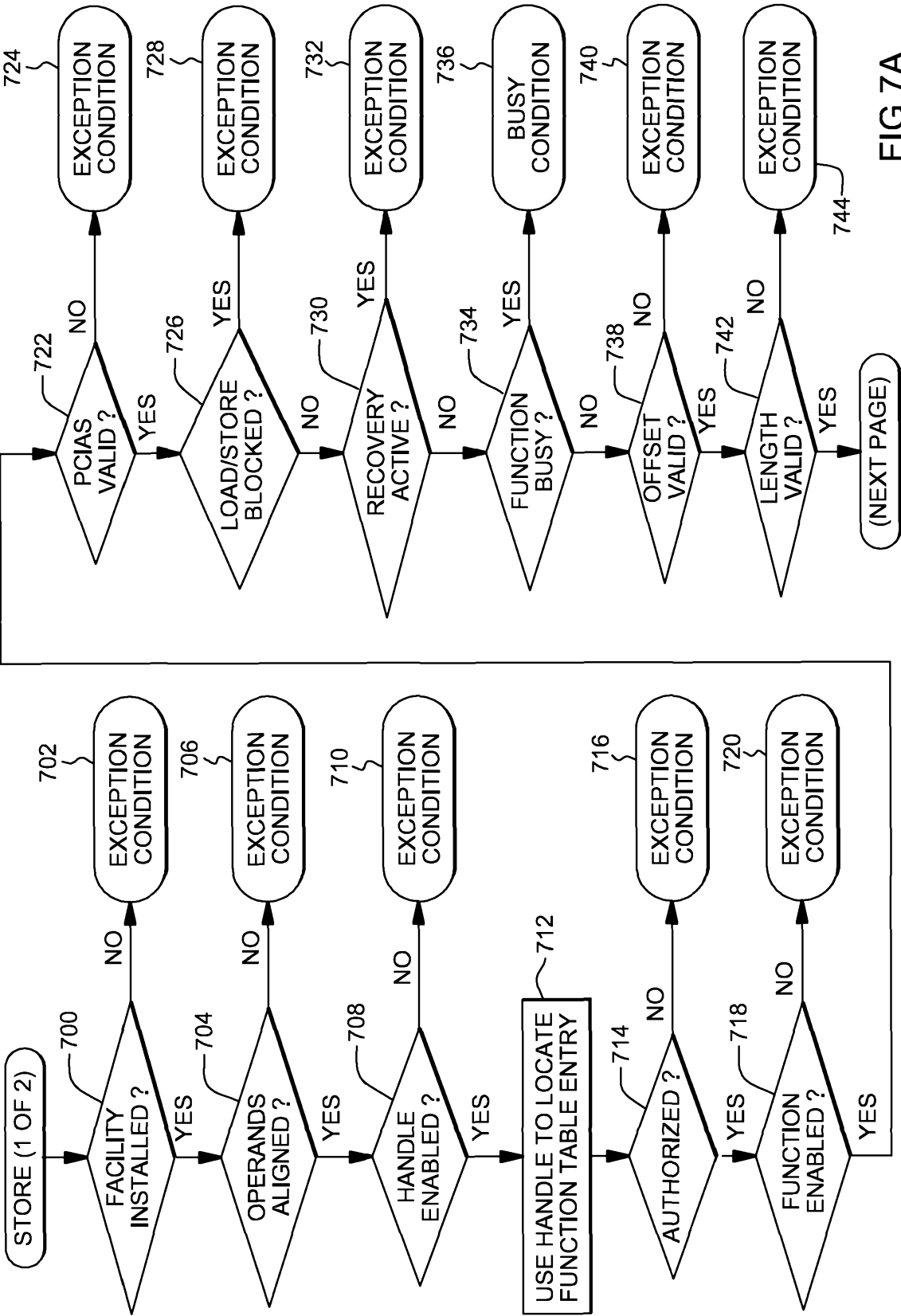


FIG. 6B



FIG. 6C



10/22

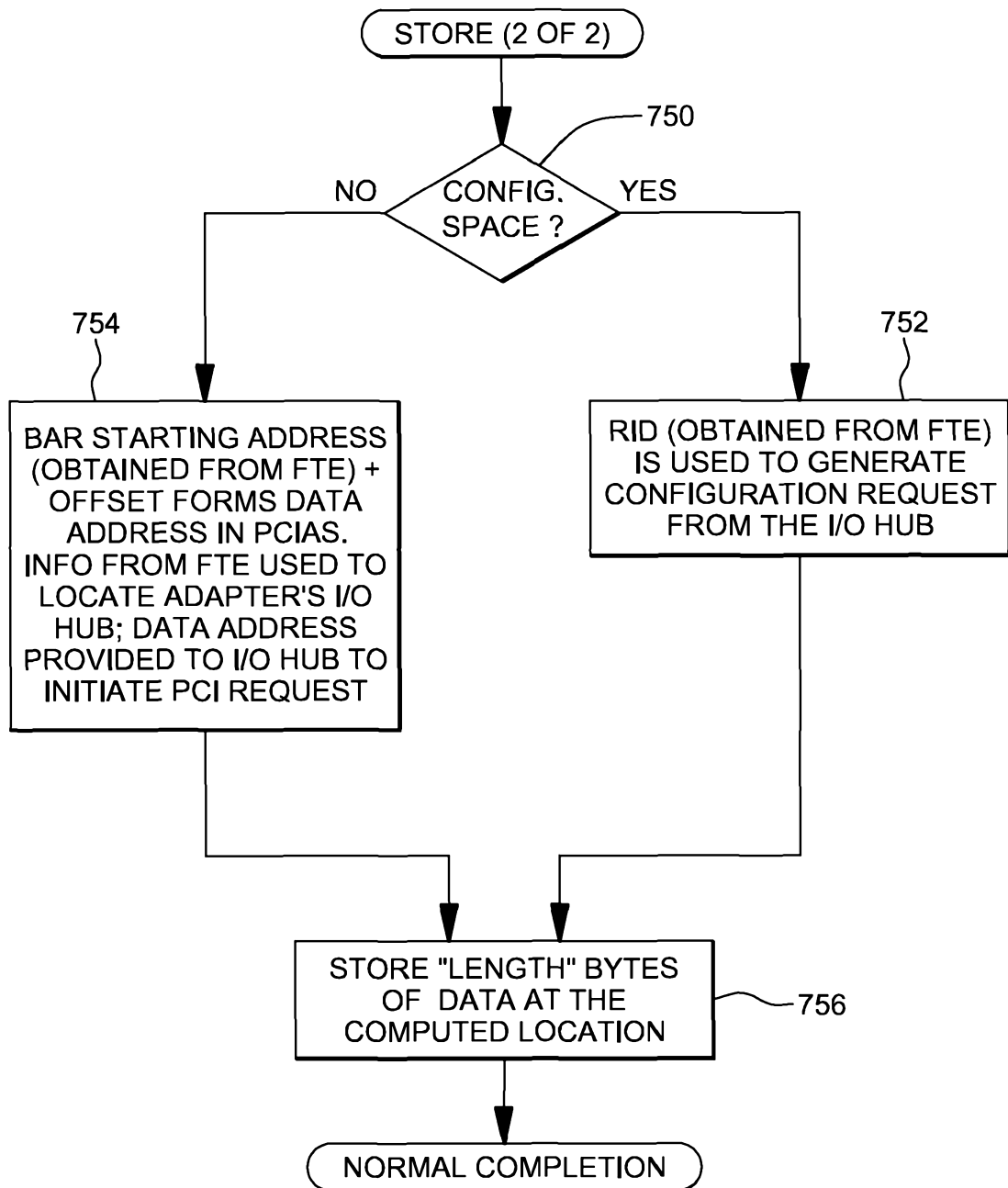


FIG. 7B

11/22

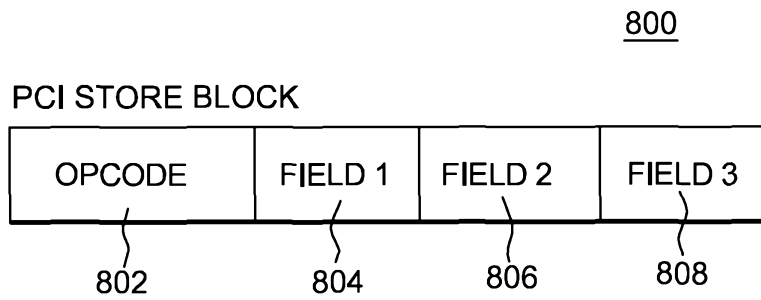


FIG. 8A

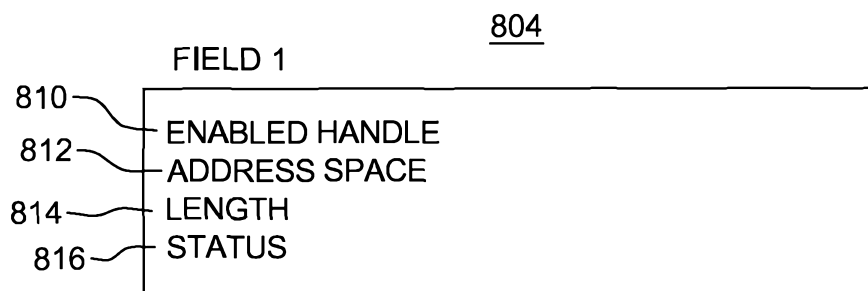


FIG. 8B



FIG. 8C

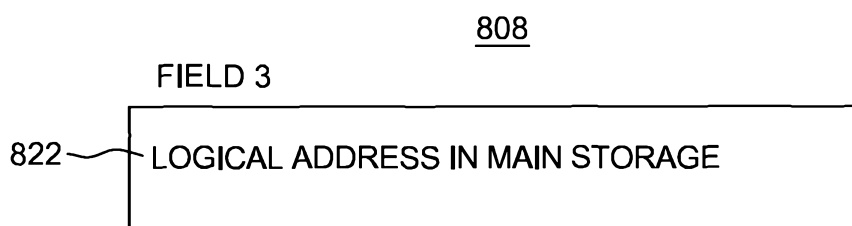
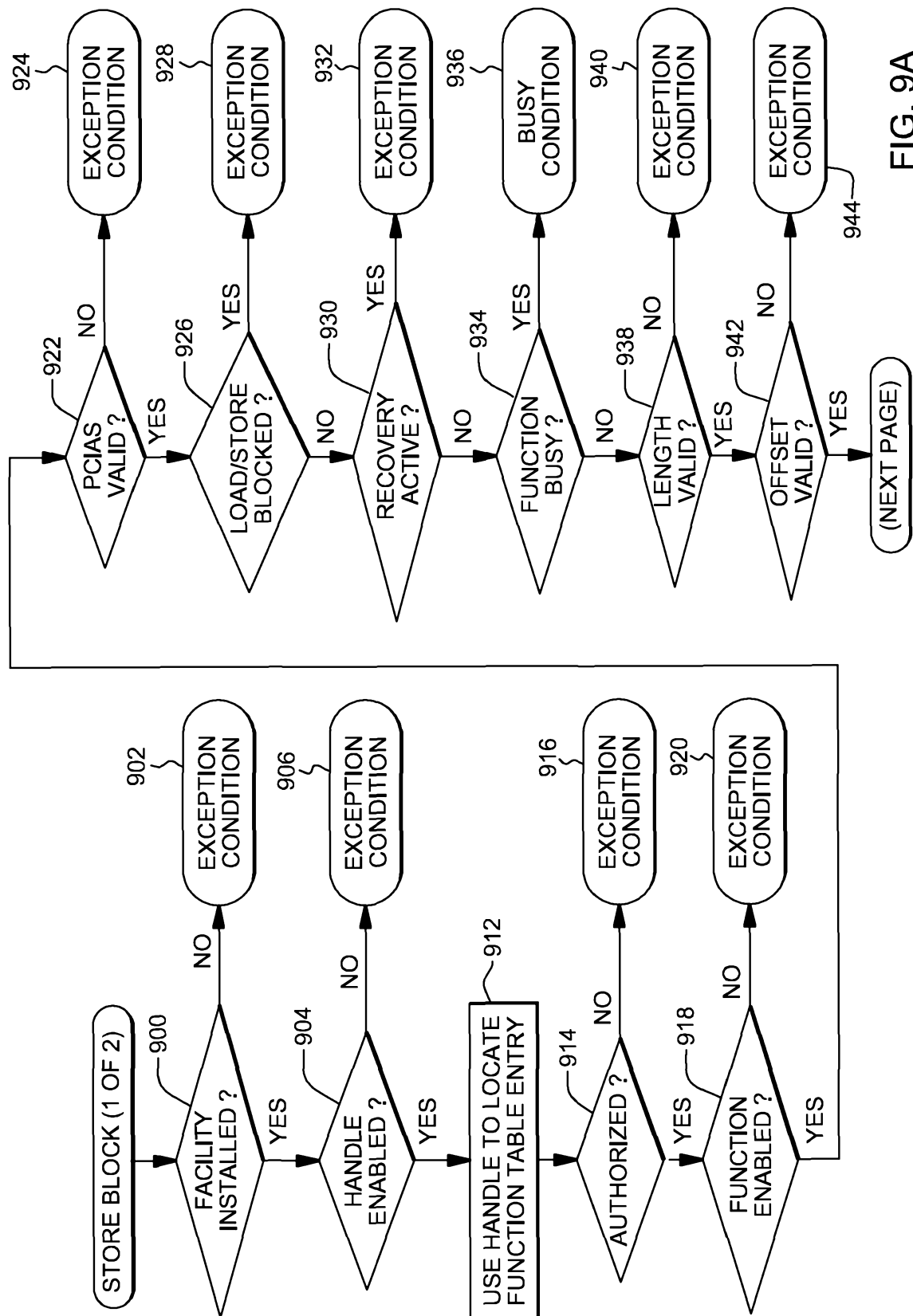


FIG. 8D

12/22



13/22

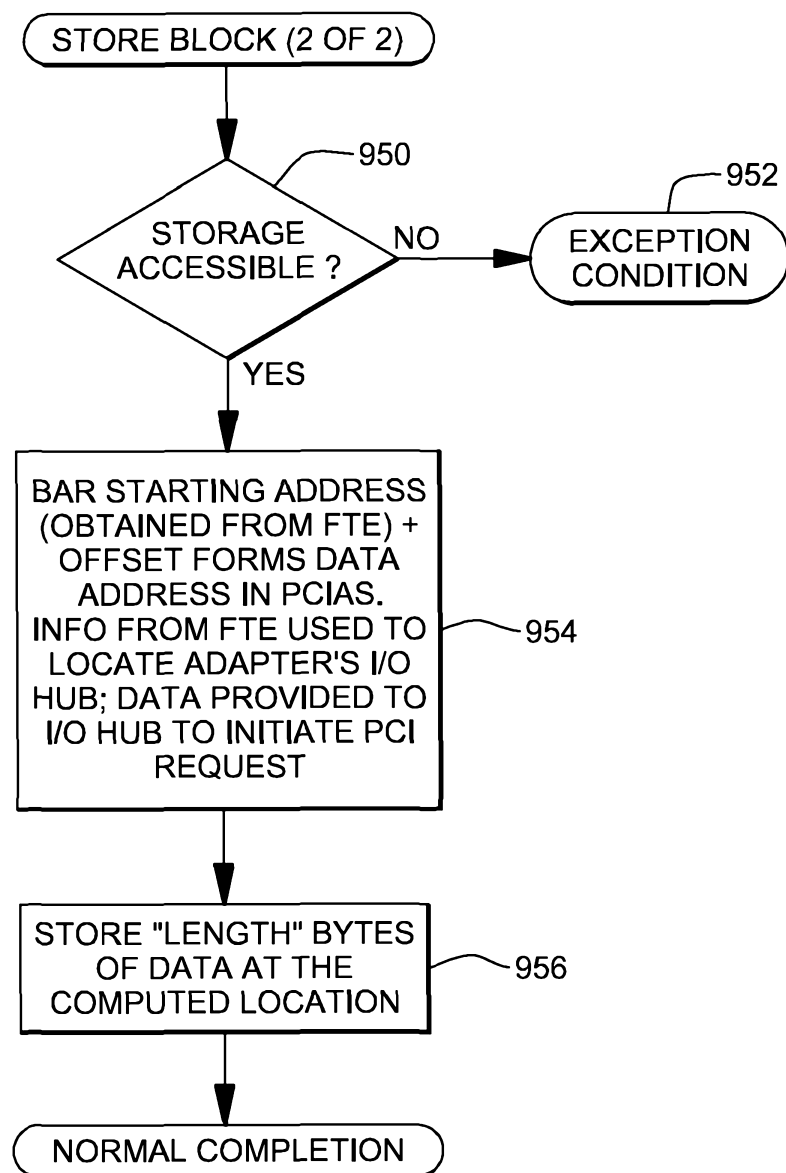


FIG. 9B

COMPUTER
PROGRAM
PRODUCT
1000

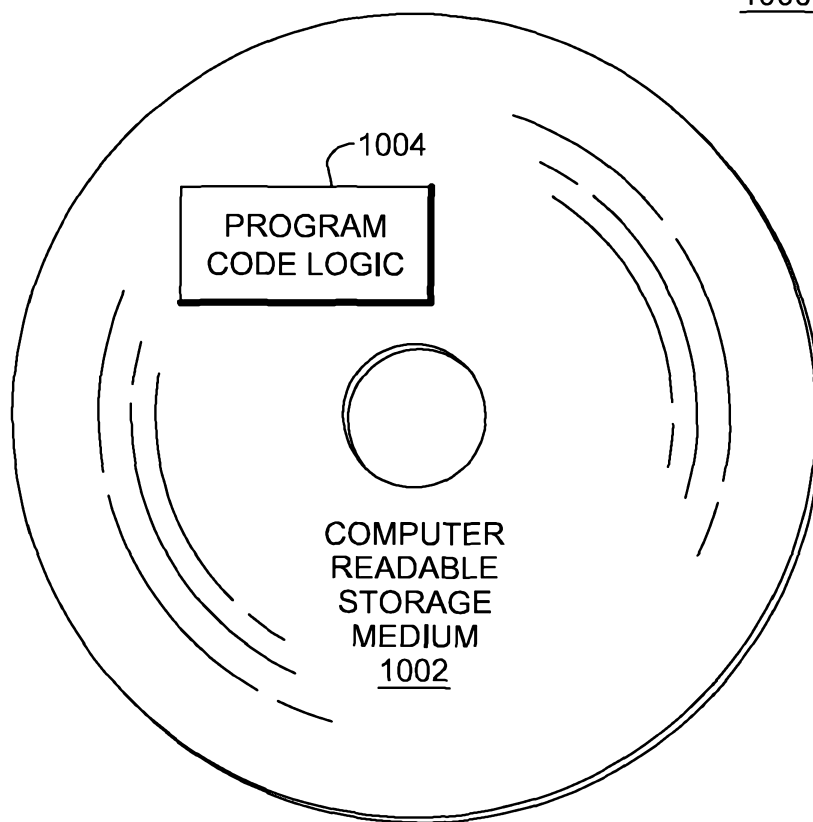


FIG. 10

15/22

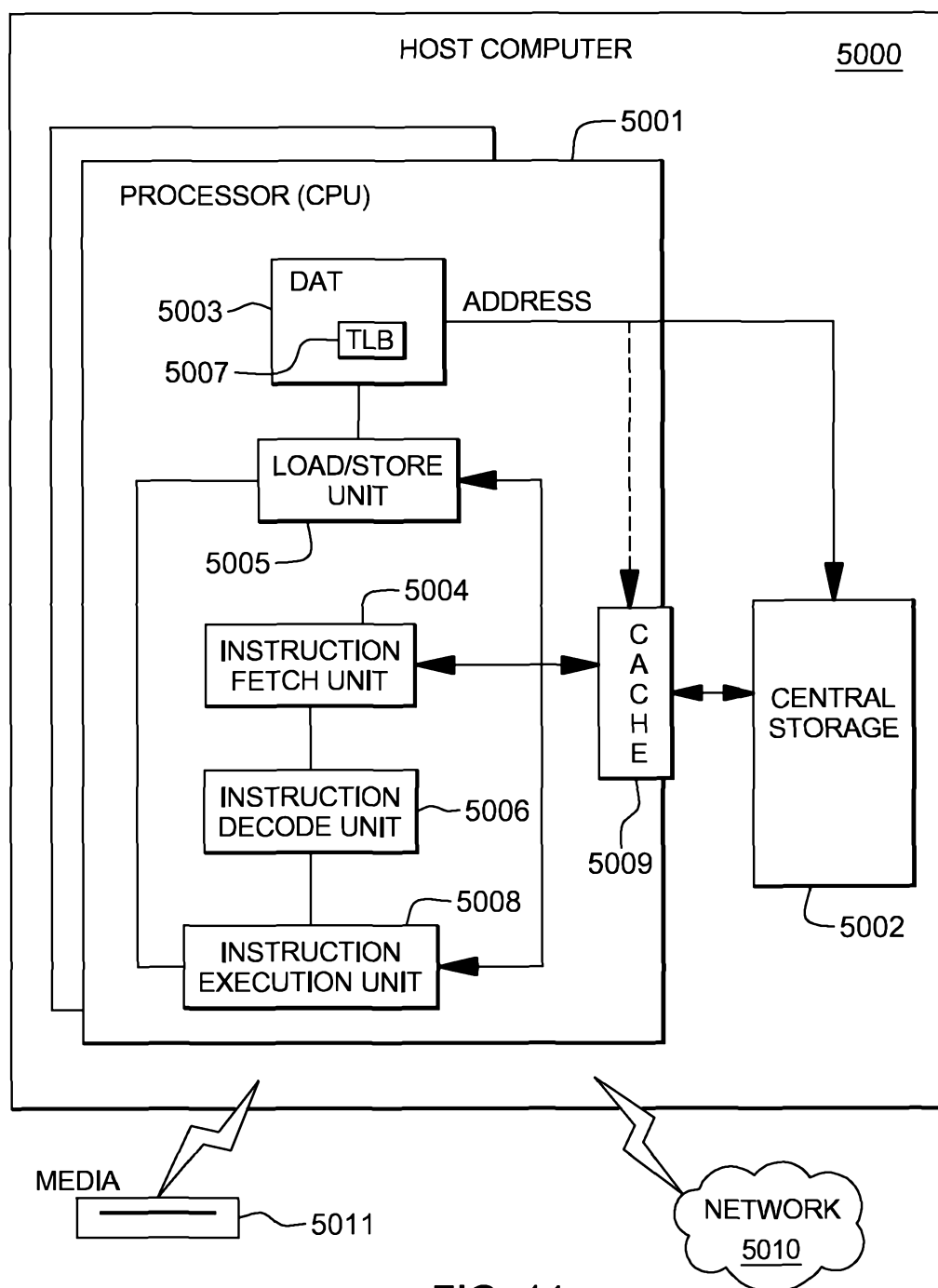


FIG. 11

16/22

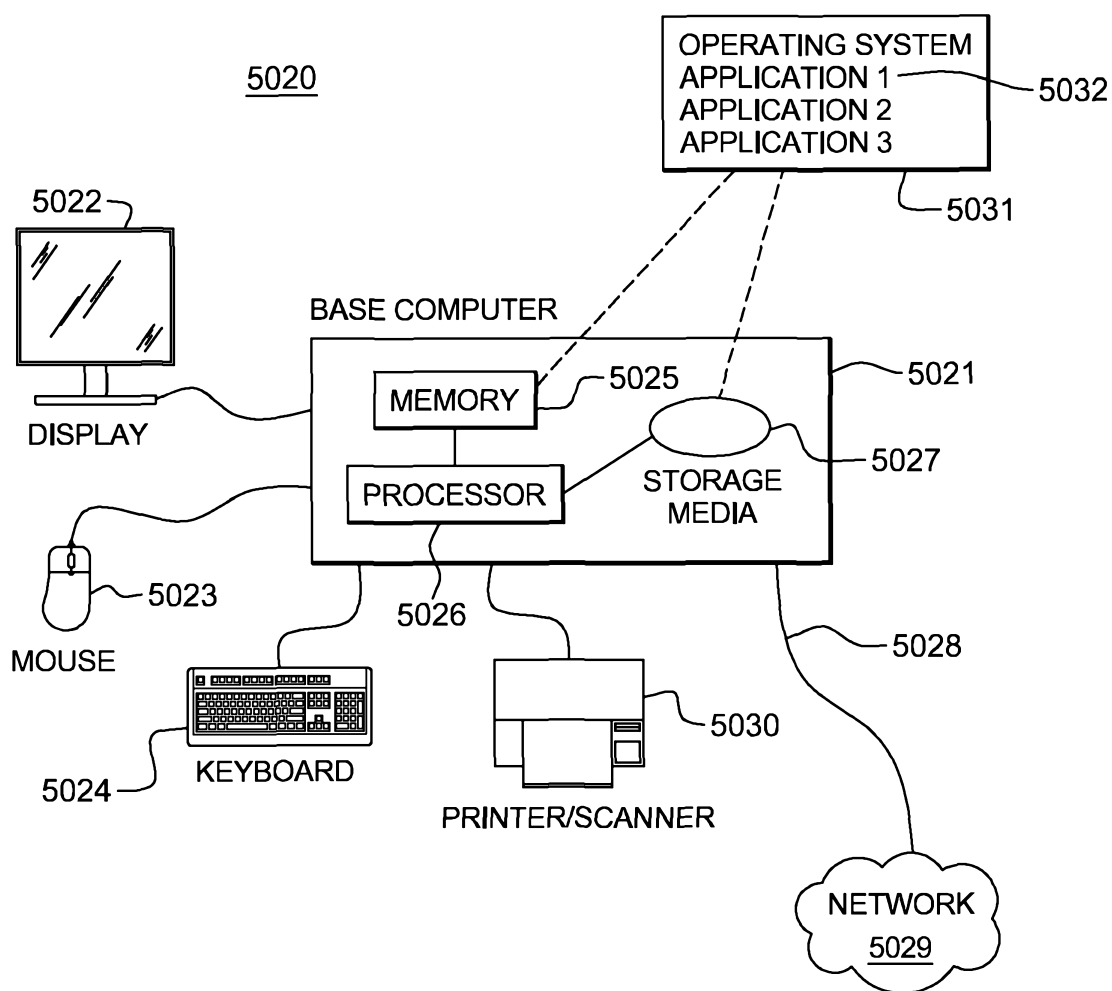


FIG. 12

17/22

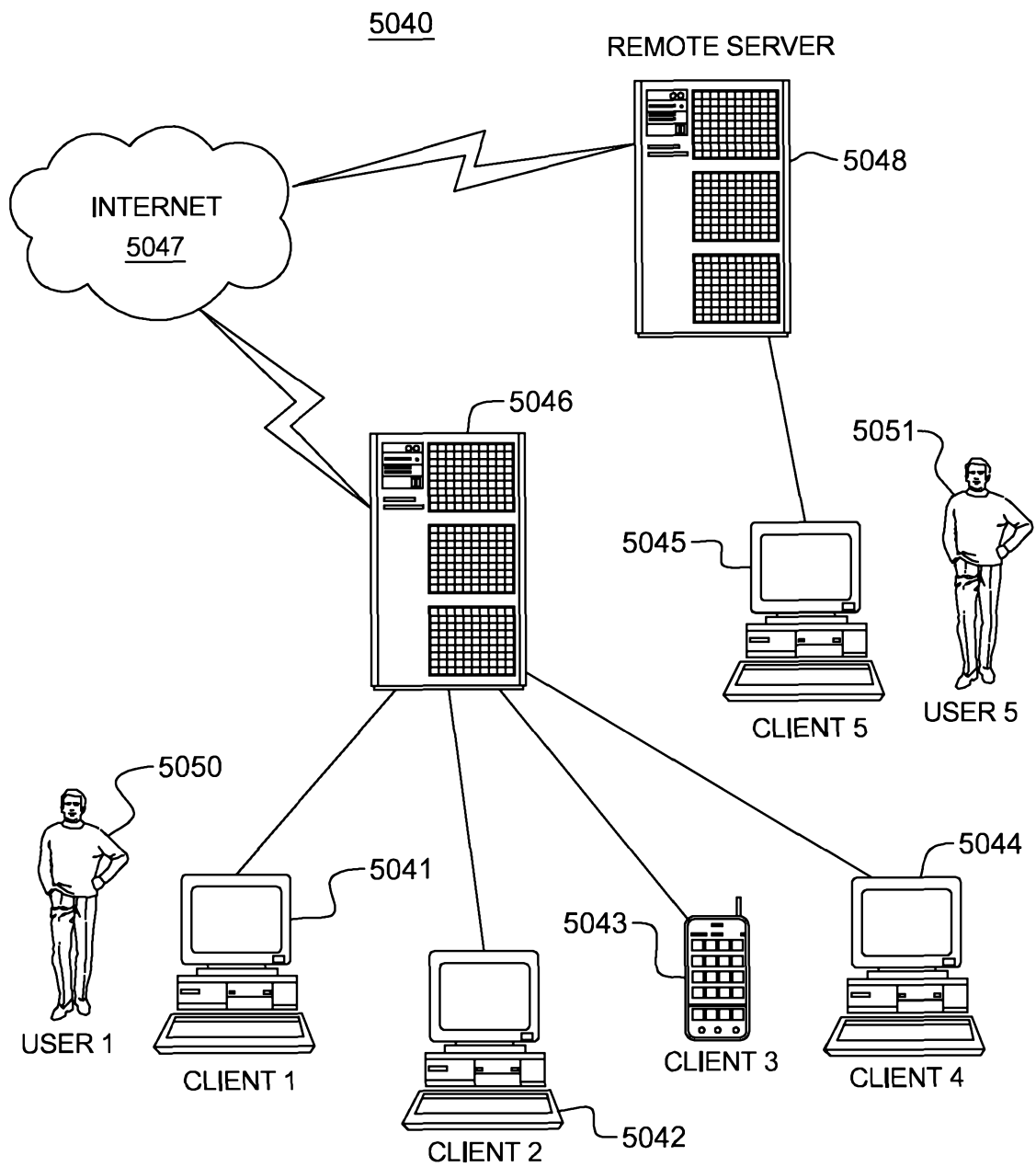


FIG. 13

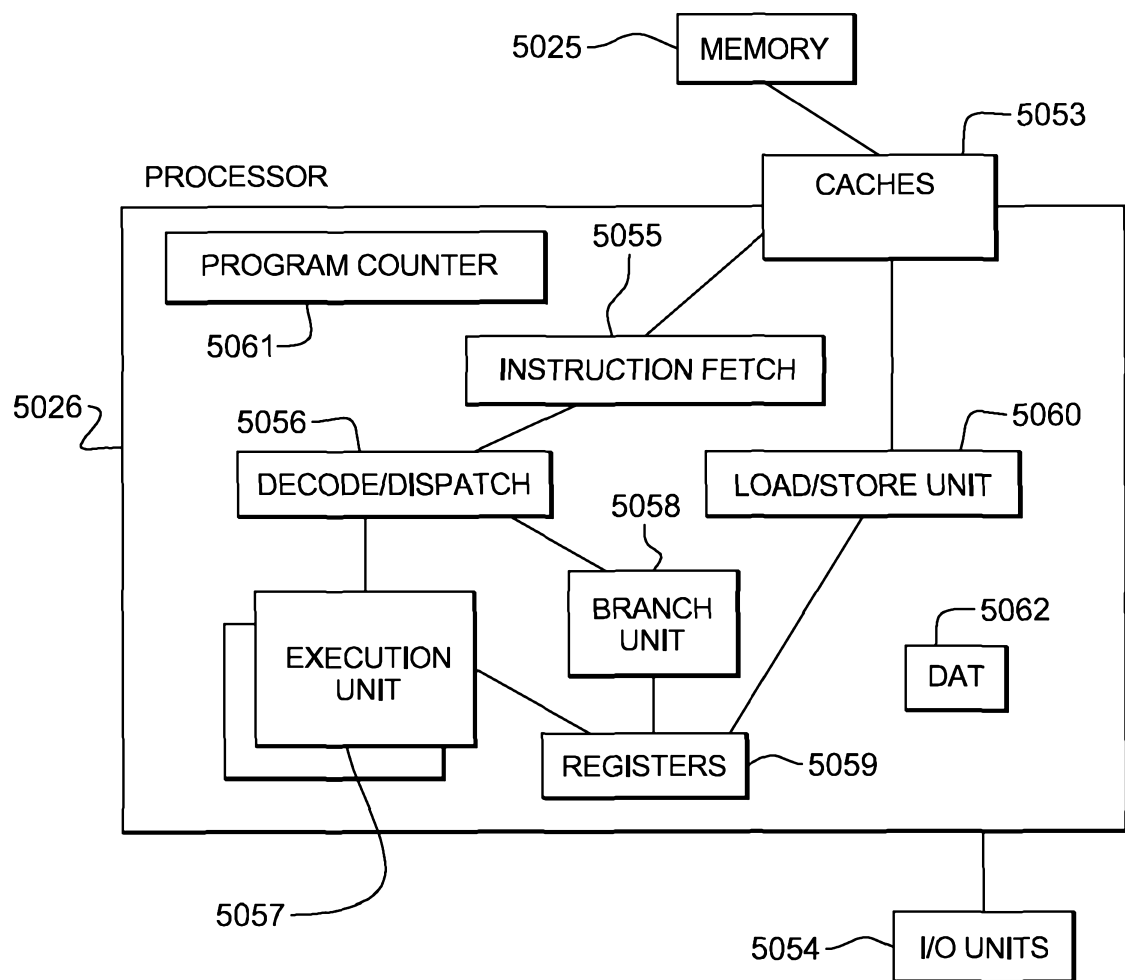


FIG. 14

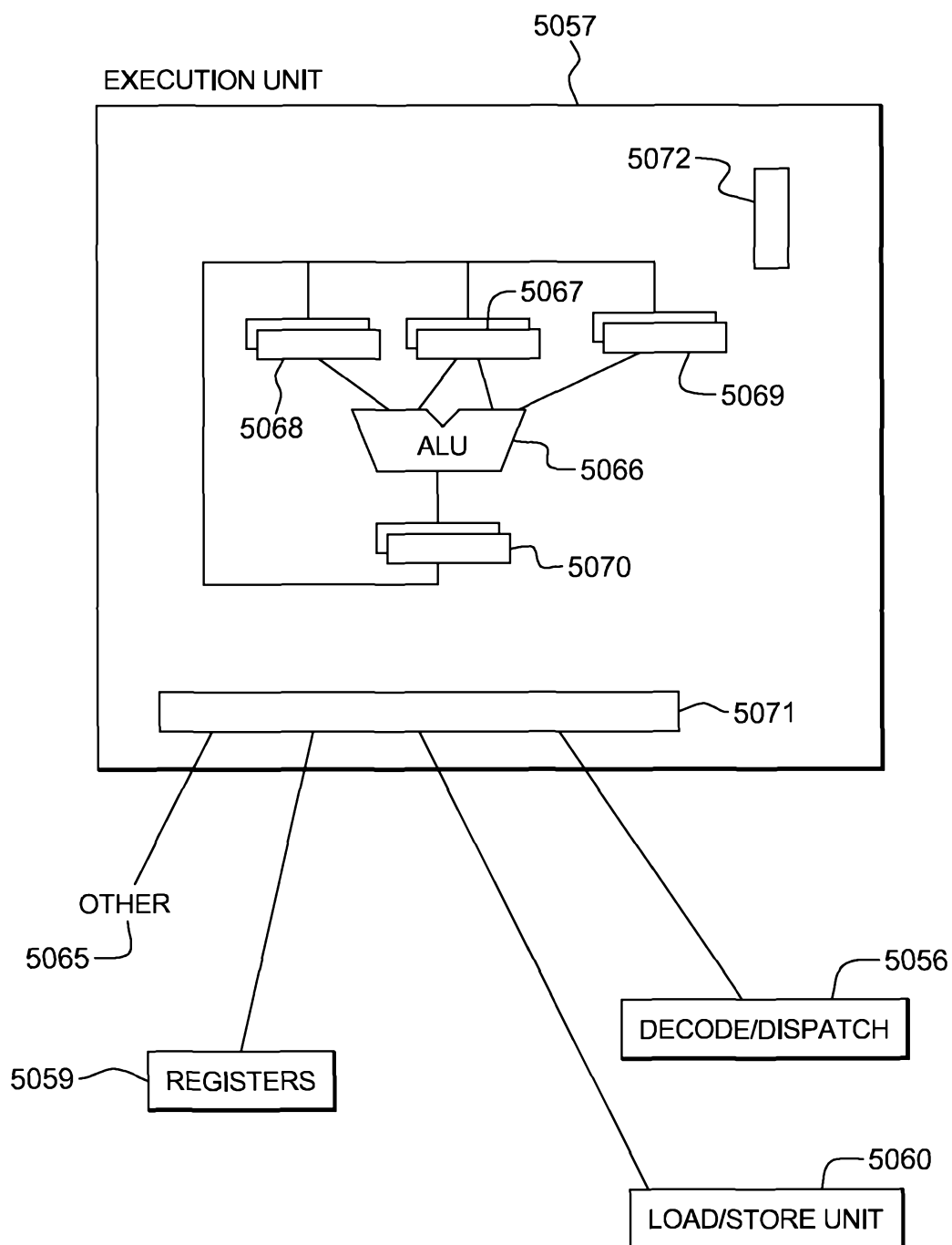


FIG. 15A

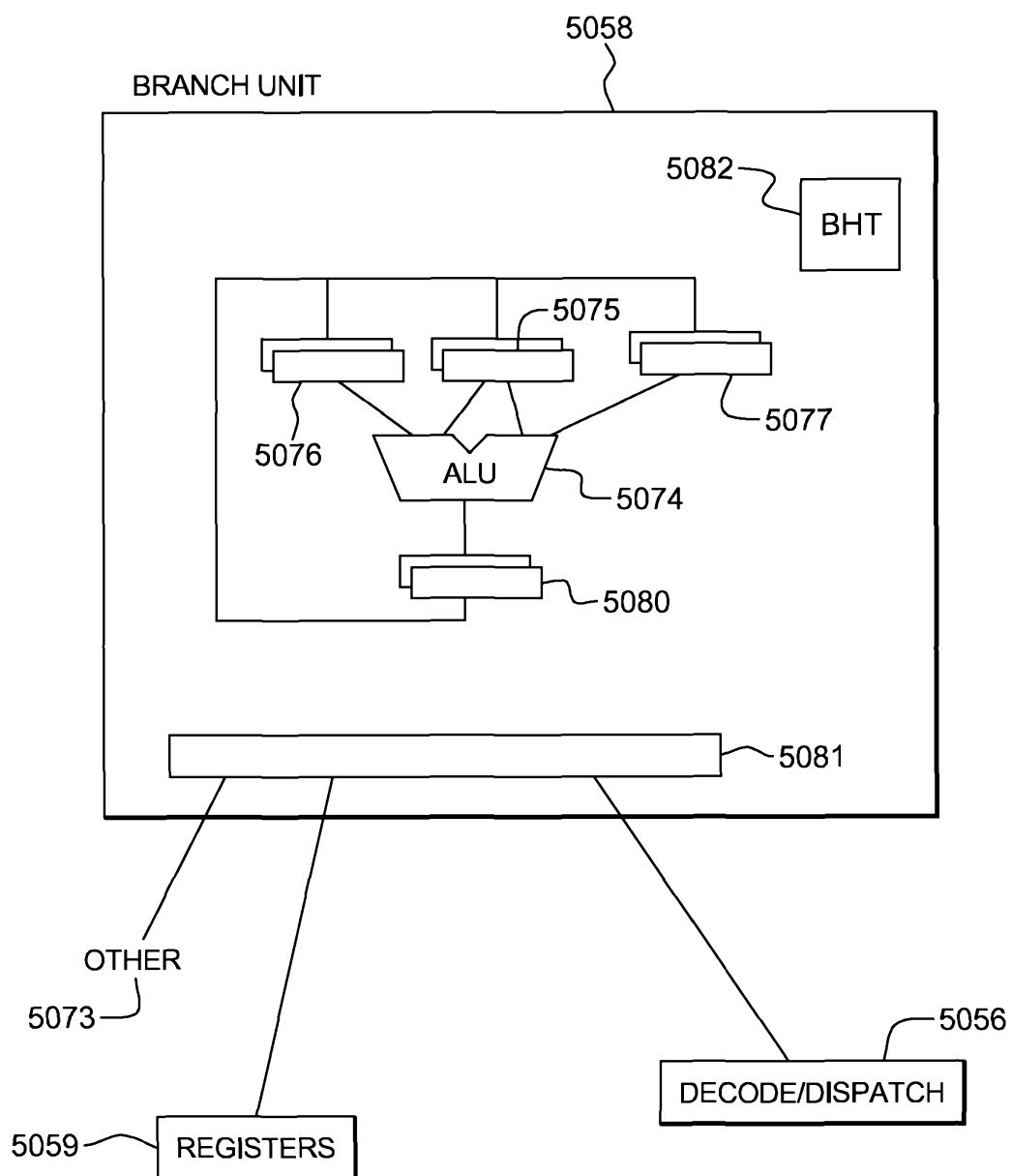


FIG. 15B

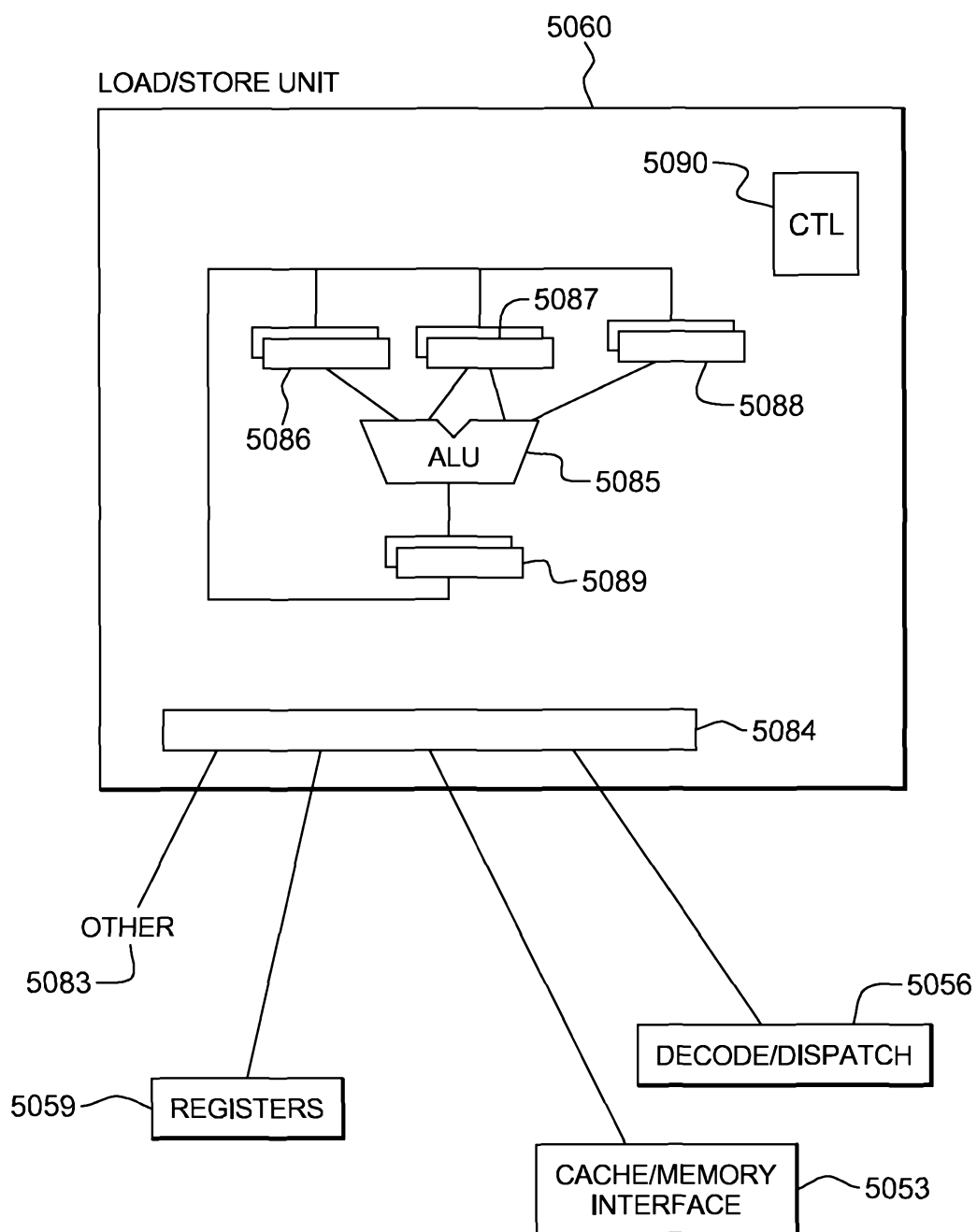


FIG. 15C

22/22

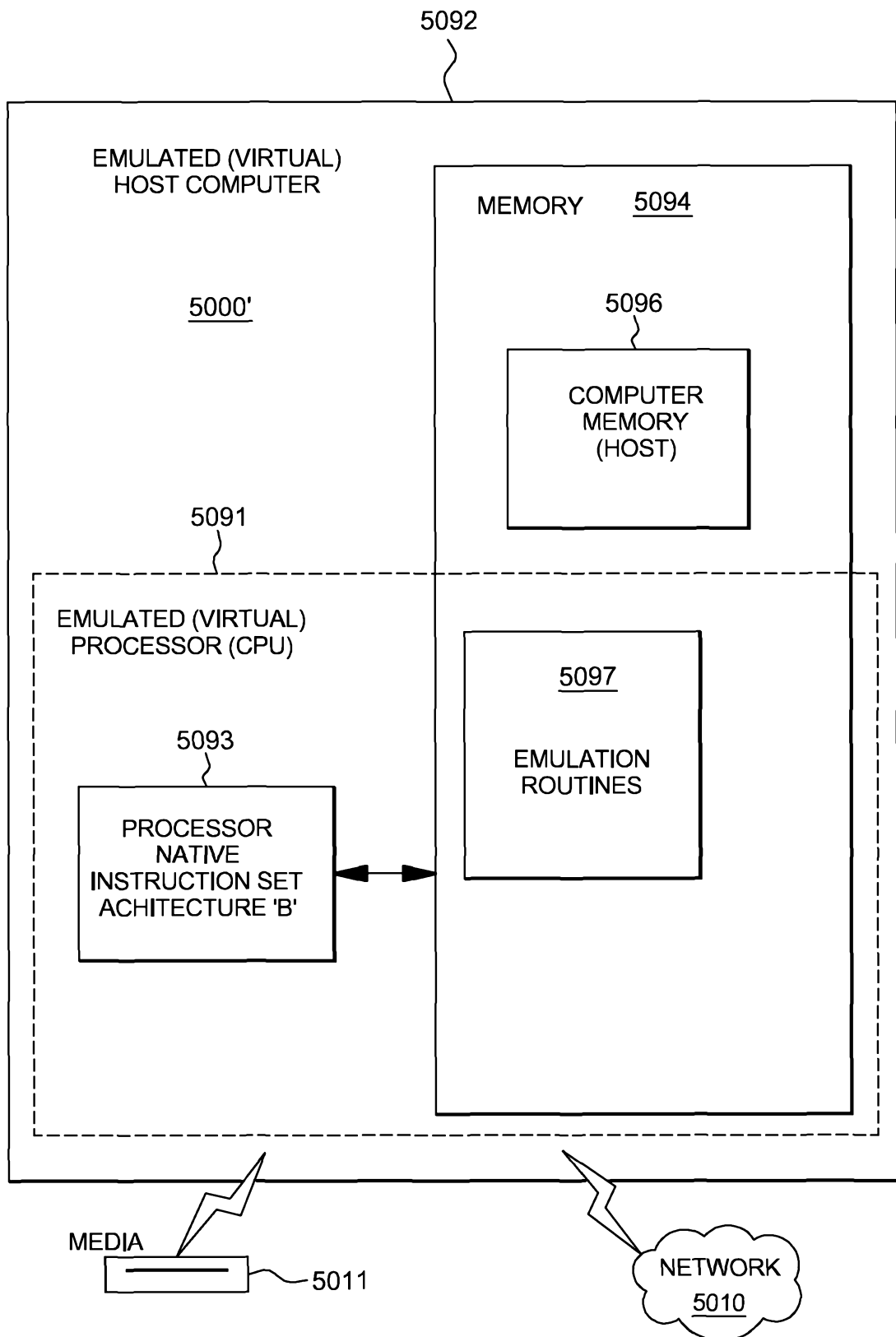


FIG. 16