(19) Deutsches Patent- und Markenamt

(10) **DE 10 2017 113 576 B4** 2019.01.24

(12) # Patentschrift

(21) Aktenzeichen: **10 2017 113 576.6**
(22) Anmeldetag: **20.06.2017**
(43) Offenlegungstag: **25.01.2018**
(45) Veröffentlichungstag der Patenterteilung: **24.01.2019**

(51) Int Cl.: ***G06F 13/40*** (2006.01)
G06F 13/12 (2006.01)
G06F 13/38 (2006.01)
H04L 12/931 (2013.01)

Innerhalb von neun Monaten nach Veröffentlichung der Patenterteilung kann nach § 59 Patentgesetz gegen das Patent Einspruch erhoben werden. Der Einspruch ist schriftlich zu erklären und zu begründen. Innerhalb der Einspruchsfrist ist eine Einspruchsgebühr in Höhe von 200 Euro zu entrichten (§ 6 Patentkostengesetz in Verbindung mit der Anlage zu § 2 Abs. 1 Patentkostengesetz).

(30) Unionspriorität:
15/215,304 **20.07.2016** **US**

(73) Patentinhaber:
Western Digital Technologies, Inc., San Jose, Calif., US

(74) Vertreter:
Dehns Germany, 80333 München, DE

(72) Erfinder:
Herman, Pinchas, San Jose, Calif., US;
Karamcheti, Vijay, San Jose, Calif., US;
Mullendore, Rodney M., San Jose, Calif., US;
Radke, William H., San Jose, Calif., US

(56) Ermittelter Stand der Technik:

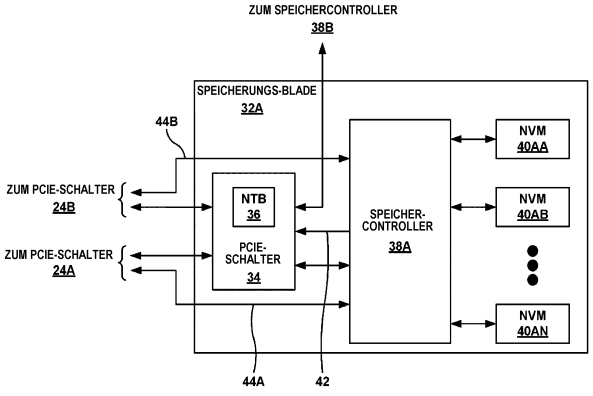
US **2012 / 0 166 699** **A1**

HOU, Rui [et al.]: Cost effective data center servers. In: High Performance Computer Architecture (HPCA2013), 2013 IEEE 19th International Symposium on. IEEE, 2013. S. 179-187.

Non-transparent Bridging with IDT 89HPES32 NT24G2 PCI Express® NTB Switch. Application Note AN-724. Integrated Device Technology, Inc, 2009. URL: https://www.idt.com/document/apn/724-non-transparent-bridging-idt-pes32nt24g2-pcie-switch [abgerufen am 27.02.2018].

(54) Bezeichnung: **Auf PCI Express basierender Speichereinsatz mit Zweifachport, der Einzelport-Speichercontroller enthält**

(57) Zusammenfassung: Ein Speichereinsatz kann einen Speichercontroller umfassen, der einen einzigen PCIe-Port und einen PCIe-Schalter umfasst. Der PCIe-Schalter kann einen kommunikativ mit einem ersten PCIe-Koppelfeld gekoppelten ersten PCIe-Port, einen kommunikativ mit einem zweiten, anderen PCIe-Koppelfeld gekoppelten zweiten PCIe-Port und einen kommunikativ mit dem einzigen PCIe-Port des Speichercontrollers gekoppelten dritten PCIe-Port umfassen. Der erste PCIe-Port und der zweite PCIe-Port können dafür ausgelegt sein, selektiv kommunikativ mit einer nicht transparenten Brücke (NTB) des PCIe-Schalters gekoppelt zu werden.



Beschreibung

TECHNISCHES GEBIET

[0001] Die vorliegende Offenbarung betrifft allgemein netzwerkangeschlossene Speicherungssysteme.

STAND DER TECHNIK

[0002] Netzwerkangeschlossene Speicherungs- bzw. NAS-Systeme, die in Unternehmensanwendungen verwendet werden, können als hoch verfügbare Systeme ausgelegt sein, die Redundanz enthalten, um Auswirkungen von Ausfall von Komponenten im System zu mindern. Zum Beispiel kann eine hoch verfügbare NAS zwei Mengen von Netzwerkports in die NAS, zwei Steuer-CPUs und zwei Schaltnetzwerke zum Anschluss an Speichereinsätze umfassen. Jeder Speichereinsatz kann zwei Ports umfassen, wobei jeder jeweilige Port zum Anschluss des Speichereinsatzes an ein jeweiliges der zwei Schaltnetzwerke verwendet wird. HOU, Rui[et al.]: Cost effective data center servers. In: High Performance Computer Architecture(HPCA2013), 2013 IEEE 19th International Symposium on. IEEE, 2013. S. 179-187 offenbart eine Architektur mit verteilten Speichern, einen Speichercontroller mit einem einzigen PCIe-Port, einen PCIe-Schalter, umfassend: einen kommunikativ mit einem ersten PCIe-Koppelfeld gekoppelten ersten PCIe-Port; einen kommunikativ mit einem zweiten, anderen PCIe-Koppelfeld gekoppelten zweiten PCIe-Port, einen kommunikativ mit dem einzigen PCIe-Port des Speichercontrolelrs gekoppelten dritten PCIe-Port, wobei der erste PCIe-Port und der zweite PCIe-Port dafür ausgelegt sind, selektiv kommunikativ mit einer nicht transparenten Brücke des PCIe-Schalters gekoppelt zu werden, weiterhin eine Archtektur, umfassend: eine erste Steuer-CPU, eine zweite Steuer-CPU, einen kommunikativ mit der ersten Steuer-CPU gekoppelten ersten PCIe-Schalter, einen kommunikativ mit der zweiten Steuer-CPU gekoppelten zweiten PCIe-Schalter und weiterhin Verfahren mit den folgenden Schritten: ein Speichercontroller empfängt einen Zuweisungsbefehl einer nicht transparenten Brücke von einer Steuer-CPU, wobei der Speichercontroller einen einzigen PCIe-Port umfasst und einen PCIe-Schalter umfasst und wobei der PCIe-Schalter einen kommunikativ mit einem ersten PCIe-Koppelfeld gekoppelten ersten PCIe-Port, einen kommunikativ mit einem zweiten anderen PCIe-Koppelfeld gekoppelten zweiten PCIe-Port und einen kommunikativ mit dem einzigen PCIe-Port des Speichercontrollers gekoppelten dritten PCIe-Port umfasst; und der Speichercontrolelr übermittelt auf der Basis des NTB-Zuweisungsbefehls einen Befehl zu dem PCe-Schalter, um zu bewirken, dass sich die NTB kommunikativ entweder mit dem ersten PCIe-Port des PCIe-Schalters oder dem zweiten PCIe-Port des PCIe-Schalters koppelt.

[0003] US 2012 0 166 699 A1 offenbart einen hoch verfügbaren PCIe-Speichereinsatz.

[0004] Non-transparent Bridging with IDT 89HPES 32NT24G2 PCI Express NTB Switch. Application Note AN-724. Integrated Device Technology, Inc, 2009. URL:https://www.idt.com/document/apn/724-non-transparent-bridging-idt-pes32nt24g2-pcie-switch [abgerufen am 27.02.2018] offenbart technischen Hintergrund zu der Erfindung.

KURZFASSUNG

[0005] In einigen Beispielen beschreibt die Offenbarung einen Speichereinsatz mit einem Speichercontroller, der einen einzigen PCIe-Port und einen PCIe-Schalter umfasst. Der PCIe-Schalter kann einen kommunikativ mit einem ersten PCIe-Koppelfeld gekoppelten ersten PCIe-Port, einen kommunikativ mit einem zweiten, anderen PCIe-Koppelfeld gekoppelten zweiten PCIe-Port und einen kommunikativ mit dem einzigen PCIe-Port des Speichercontrollers gekoppelten dritten PCIe-Port umfassen. Der erste PCIe-Port und der zweite PCIe-Port können dafür ausgelegt sein, selektiv kommunikativ mit einer nicht transparenten Brücke (NTB) des PCIe-Schalters gekoppelt zu werden.

[0006] In einigen Beispielen beschreibt die Offenbarung ein netzwerkangeschlossenes Speicherungssystem mit einer ersten Steuer-CPU, einer zweiten Steuer-CPU, einem kommunikativ mit der ersten Steuer-CPU gekoppelten ersten PCIe-Schalter, einem kommunikativ mit der zweiten Steuer-CPU gekoppelten zweiten PCIe-Schalter und einem Speichereinsatz. Der Speichereinsatz kann einen Speichercontroller umfassen, der einen einzigen PCIe-Port und einen dritten PCIe-Schalter umfasst. Der dritte PCIe-Schalter kann einen kommunikativ mit dem ersten PCIe-Schalter gekoppelten ersten PCIe-Port, einen kommunikativ mit dem zweiten PCIe-Schalter gekoppelten zweiten PCIe-Port und einen kommunikativ mit dem einzigen PCIe-Port des Speichercontrollers gekoppelten dritten PCIe-Port umfassen. Der erste PCIe-Port und der zweite PCIe-Port können dafür ausgelegt sein, selektiv kommunikativ mit einer nicht transparenten Brücke (NTB) des dritten PCIe-Schalters gekoppelt zu werden.

[0007] In einigen Beispielen beschreibt die Offenbarung ein Verfahren, das umfasst, dass ein Speichercontroller eines Speichereinsatzes einen Zuweisungsbefehl einer nicht transparenten Brücke (NTB) von einer Steuer-CPU empfängt. Der Speichercontroller kann einen einzigen PCIe-Port umfassen. Der Speichereinsatz kann den Speichercontroller und einen PCIe-Schalter umfassen. Der PCIe-Schalter kann einen kommunikativ mit einem ersten PCIe-Koppelfeld gekoppelten ersten PCIe-Port, einen kommunikativ mit einem zweiten, anderen PCIe-

Koppelfeld gekoppelten zweiten PCIe-Port und einen kommunikativ mit dem einzigen PCIe-Port des Speichercontrollers gekoppelten dritten PCIe-Port umfassen. Das Verfahren kann außerdem umfassen, dass der Speichercontroller auf der Basis des NTB-Zuweisungsbefehls einen Befehl zu dem PCIe-Schalter übermittelt, um zu bewirken, dass sich die NTB kommunikativ entweder mit dem ersten PCIe-Port des PCIe-Schalters oder dem zweiten PCIe-Port des PCIe-Schalters koppelt.

[0008] Die Einzelheiten eines oder mehrerer Beispiele werden in den beigefügten Zeichnungen und der nachfolgenden Beschreibung dargelegt. Andere Merkmale, Aufgaben und Vorteile werden aus der Beschreibung und den beigefügten Zeichnungen und aus den Ansprüchen ersichtlich.

Figurenliste

Fig. 1 ist eine konzeptuelle und schematische Blockdarstellung einer beispielhaften Speicherungsumgebung, in der ein netzwerkangeschlossenes Speicher- bzw. NAS-System als Speicherungsvorrichtung für mehrere Hostvorrichtungen fungieren kann, gemäß einem oder mehreren Aspekten der vorliegenden Offenbarung.

Fig. 2 ist eine konzeptuelle und schematische Blockdarstellung eines beispielhaften Speichereinsatzes gemäß einem oder mehreren Aspekten der vorliegenden Offenbarung.

Fig. 3 ist eine konzeptuelle und schematische Blockdarstellung eines beispielhaften Speicherungs-Blade gemäß einem oder mehreren Aspekten der vorliegenden Offenbarung.

Fig. 4 ist ein Flussdiagramm einer beispielhaften Technik, um eine nicht transparente Brücke einem PCIe-Port zuzuweisen, gemäß einem oder mehreren Aspekten der vorliegenden Offenbarung.

AUSFÜHRLICHE BESCHREIBUNG

[0009] Die vorliegende Offenbarung beschreibt einen Speichereinsatz mit mindestens einem Speichercontroller, der einen einzigen PCIe-Port (Peripheral Component Interconnect Express) und einen PCIe-Schalter umfasst, der zwei PCIe-Ports umfasst, ausgelegt zum Verbinden des Speichereinsatzes mit externen Vorrichtungen, wie etwa jeweiligen PCIe-Koppelfeldern. Der PCIe-Schalter erlaubt die Verwendung des Einzelport-Speichercontrollers in einem Zweifachport-System, wodurch Einzelport-Speichercontroller in hoch verfügbaren netzwerkangeschlossenen Speicher- bzw. NAS-Systemen verwendet werden können. Dadurch können im Handel erhältliche auf PCIe basierende Speichercontroller in hoch verfügbaren NAS-Systemen verwendet werden, statt die Entwicklung neuer auf PCIe basierender Speichercontroller mit Zweifachport für die Verwendung in hoch verfügbaren NAS-Systemen zu erfordern.

[0010] Damit der Einzelport-Speichercontroller in einem Zweifachport-System verwendet werden kann, umfasst der PCIe-Schalter eine nicht transparente Brücke (NTB). Eine NTB erlaubt die Verbindung zweier PCIe-Koppelfelder mit einem einzigen Schalter und kann verhindern, dass Vorrichtungen auf dem ersten PCIe-Koppelfeld voll in das zweite PCIe-Koppelfeld sehen und umgekehrt. Stattdessen kann die NTB ein Fenster für mit dem ersten PCIe-Koppelfeld verbundene Vorrichtungen zum Sehen in das zweite PCIe-Koppelfeld und umgekehrt bereitstellen. Da jedes PCIe-Koppelfeld ein jeweiliges Adressenschema verwendet, kann die NTB Adressenübersetzung zwischen den von den jeweiligen PCIe-Koppelfeldern verwendeten Adressen bereitstellen, wodurch Vorrichtungen aus dem zweiten PCIe-Koppelfeld auf Vorrichtungen aus dem ersten PCIe-Koppelfeld zugreifen können und umgekehrt. Auf diese Weise erlaubt die Bereitstellung eines PCIe-Schalters, der eine NTB auf einem der PCIe-Ports des PCIe-Schalters umfasst, die Verwendung eines Einzelport-Speichercontrollers in einem Zweifachport-NAS-System.

[0011] **Fig. 1** ist eine konzeptuelle und schematische Blockdarstellung einer beispielhaften Speicherungsumgebung **10**, in der ein NAS-System **16** als Datenspeicherungsvorrichtung für Hostvorrichtungen **12** fungieren kann, gemäß einer oder mehreren Techniken der vorliegenden Offenbarung. Zum Beispiel können die Hostvorrichtungen **12** nichtflüchtige Speichervorrichtungen, die in dem NAS-System **16** enthalten sind, zum Speichern und Abrufen von Daten verwenden.

[0012] Die Speicherungsumgebung **10** kann mehrere Hostvorrichtungen **12** umfassen, die Daten in einer oder mehreren Speicherungsvorrichtungen, wie etwa dem NAS-System **16** speichern und/oder daraus abrufen können. Wie in **Fig. 1** dargestellt, können die Hostvorrichtungen **12** über ein erstes Schaltnetzwerk **14a** und ein zweites Schaltnetzwerk **14b** (zusammen „Schaltnetzwerke **14**“) mit dem NAS-System **16** kommunizieren. Die Hostvorrichtungen **12** wären zum Beispiel beliebige von vielfältigen Vorrichtungen, darunter Computerserver, Datenverarbeitungscluster auf Cloud-Basis, Desktop-Computer, Notebook-(d.h. Laptop-)Computer, Tablet-Computer, Set-Top-Boxen, Telefonhandapparate, wie etwa sogenannte „Smartphones“, sogenannte „Smartpads“, Fernseher, Kameras, Anzeigevorrichtungen, digitale Medienplayer, Videospielkonsolen, Video-Streamingvorrichtungen und dergleichen.

[0013] Die Schaltnetzwerke **14** können einen Datenbus zum Austausch von Daten mit den Hostvorrich-

tungen **12** und/oder einen Steuerbus zum Austausch von Befehlen mit den Hostvorrichtungen **12** umfassen. In einigen Beispielen kann jedes der Schaltnetzwerke **14** ein geschaltetes Koppelfeld umfassen, in dem alle Hostvorrichtungen **12** über Schalter mit jedem der Netzwerkports **20A** und **20B** verbunden sind. Die Schaltnetzwerke **14** können eine beliebige geeignete Netzwerktransporttechnologie benutzen. Zum Beispiel können die Schaltnetzwerke **14** Ethernet und/oder InfiniBand und/oder Fibre Channel und/oder dergleichen benutzen. Durch Bereitstellung von zwei Schaltnetzwerken **14A** und **14B** ist die Speicherungsumgebung **10** von den Hostvorrichtungen **12** zu den Speichereinsätzen **30A**-**30N** (zusammen „Speichereinsätze **30**“) voll zweifach-portig. Wenn eines des ersten Schaltnetzwerks **14A** oder des zweiten Schaltnetzwerks **14B** ausfällt, können sich die Hostvorrichtungen **12** auf diese Weise immer noch unter Verwendung des anderen des Schaltnetzwerks **14A** oder des zweiten Schaltnetzwerks **14B** mit dem NAS-System **16** verbinden.

[0014] Jedes der Schaltnetzwerke **14** ist kommunikativ mit dem NAS-Speicherungssystem **16** gekoppelt. Zum Beispiel ist in **Fig. 1** das erste Schaltnetzwerk **14A** kommunikativ mit den ersten Netzwerkports **20A** der ersten Steuerplatine **18A** des NAS-Systems **16** gekoppelt. Ähnlich ist in **Fig. 1** das zweite Schaltnetzwerk **14B** kommunikativ mit zweiten Netzwerkports **20B** der zweiten Steuerplatine **18B** des NAS-Systems **16** gekoppelt.

[0015] Das NAS-System **16** umfasst zwei Steuerplatinen **18A** und **18B** (zusammen „Steuerplatinen 18“), wodurch wieder Redundanz für ein hoch verfügbares System bereitgestellt wird. Die Steuerplatinen **18** enthalten Komponenten zum Steuern des NAS-Systems **16**, darunter Steuer-CPUs **22A** und **22B** (zusammen „Steuer-CPUs **22**“), DRAM **26A** und **26B** (zusammen „DRAM **26**“) und PCIe-Schalter **24A** und **24B** (zusammen „PCIe-Schalter **24**“).

[0016] Die Steuerplatine **18A** umfasst einen oder mehrere erste Netzwerkports **20A**, die kommunikativ mit dem ersten Schaltnetzwerk **14A** gekoppelt sind und Kommunikation zwischen der ersten Steuerplatine **18A** (z.B. der ersten Steuer-CPU **22A**) und den Hostvorrichtungen **12** erlauben. Die ersten Netzwerkports **20A** können ein beliebiges Protokoll, eine beliebige Netzwerktransporttechnologie und einen beliebigen Verbinderformfaktor implementieren, und sie können jeweils auf der von den Schaltnetzwerken **14** verwendeten Technologie basieren.

[0017] Ein oder mehrere erste Netzwerkports **20A** sind kommunikativ mit der ersten Steuer-CPU **22A** gekoppelt. Zum Beispiel können ein oder mehrere erste Netzwerkports **20A** und die erste Steuer-CPU **22A** mit einer gemeinsamen Leiterplatte (PCB) verbunden und unter Verwendung einer oder mehrerer elektrischer Leiterbahnen auf oder in der PCB kommunikativ gekoppelt sein.

[0018] Die erste Steuer-CPU **22A** steuert den Betrieb des NAS-Systems **16** alleine oder in Kombination mit der zweiten CPU **22B** auf der zweiten Steuerplatine **18B**. Die erste Steuer-CPU **22A** kann auch als Controller des NAS-Systems **16** bezeichnet werden. Die erste Steuer-CPU **22A** kann eine beliebige Art von Prozessor sein, darunter zum Beispiel ein Mikroprozessor, ein digitaler Signalprozessor (DSB), eine anwendungsspezifische integrierte Schaltung (ASIC), ein am Einsatzort programmierbares Gatearray (FPGA) oder beliebige andere äquivalente integrierte oder diskrete Logikschaltkreise, sowie beliebige Kombinationen solcher Komponenten. In einigen Beispielen kann die erste Steuer-CPU **22A** einen Prozessor auf x86-Basis umfassen, wie etwa den von Intel® oder AMD® erhältlichen Prozessor auf x86-Basis.

[0019] Über die Schaltnetzwerke **14** empfängt die erste Steuer-CPU **22A** Befehle von den Hostvorrichtungen **12** und tauscht Daten mit diesen aus. Die erste Steuer-CPU **22A** bewirkt, dass das NAS-System **16** die Befehle zum Speichern oder Abrufen von Daten aus den Speichereinsätzen **30** ausführt. Die von den Hostvorrichtungen **12** empfangenen Befehle können Lesebefehle und Schreibbefehle umfassen. Die erste Steuer-CPU **22A** führt auch andere Funktionen aus, wie etwa eine Flash-Übersetzungsschicht (Abbildung oder Umleitung von logisch zu physisch), Befehlswarteschlangen, Schreibaggregation, Lese-Cache-Speicherung, Verschlüsselung und Entschlüsselung von Daten, Komprimierung und Dekomprimierung von Daten, Fehlerkorrekturcode zur Ermöglichung der Wiederherstellung von fehlerhaften Daten, RAID-Befehle und Hintergrundsystemaufgaben wie Abnutzungsnivellierung, Müllabfuhr, Verfolgung des Systemstatus oder dergleichen.

[0020] In einigen Beispielen kann die erste Steuerplatine **18A** ein FPGA **28A** umfassen, das als Offload-Prozessor fungiert. Das FPGA **28A** kann dafür ausgelegt sein, eine oder mehrere Operationen anstelle der Steuer-CPU **22A** auszuführen, um eine Arbeitslast der CPU **22A** zu verringern. Zum Beispiel kann das FPGA **28A** die Steuer-CPU **22A** von einem oder mehreren von Schreibwarteschlangenverwaltung, Komprimierung, Verschlüsselung, RAID-Berechnung, Dekomprimierung, Entduplikation, Entschlüsselung, Lese-Cache-Speicherung oder dergleichen entlasten. In einigen Beispielen kann die erste Steuerplatine **18A** das FPGA **28A** weglassen, die zweite Steuerplatine **18B** das FPGA **28B** weglassen oder beides.

[0021] Die erste Steuerplatine **18A** umfasst auch ersten DRAM **26A**. Der erste DRAM **26A** ist Arbeitsspeicher für die erste Steuer-CPU **22A**, und das ers-

te FPGA **28A** und Speichercontroller der Speichereinsätze **30A** können über den ersten PCIe-Schalter **24A** auf den ersten DRAM **26A** zugreifen. Der erste DRAM **26A** kann Daten in Bezug auf den Betrieb der ersten Steuer-CPU **22A**, des FPGA **28A** und der Speichercontroller der Speichereinsätze **30A** speichern, darunter zum Beispiel Task-Warteschlangen wie etwa Lesewarteschlangen, Schreibwarteschlangen oder dergleichen; Lesepuffer; Tabellen der Übersetzung von logischen in physische Adressen oder dergleichen.

[0022] Die erste Steuerplatine **18A** umfasst auch einen ersten PCIe-Schalter **24A**. Der erste PCIe-Schalter **24A** verbindet verschiedene Vorrichtungen oder Komponenten mit einem PCIe-Koppelfeld, darunter die CPU **22A**, das FPGA **28A** und die Speichereinsätze **30**. In einigen Beispielen stellt wie in **Fig. 1** gezeigt der erste PCIe-Schalter **24A** auch eine Kommunikationsverbindung mit dem zweiten PCIe-Schalter **24B** der zweiten Steuerplatine **18B** bereit. In einigen Beispielen kann, statt einen einzigen ersten PCIe-Schalter **24A** bereitzustellen, die erste Steuerplatine **18A** mehrere erste PCIe-Schalter **24A** umfassen, wie etwa einen primären PCIe-Schalter und einen Hilfs-PCIe-Schalter. In einigen Beispielen, bei denen die erste Steuerplatine **18A** einen primären PCIe-Schalter und einen Hilfs-PCIe-Schalter umfasst, kann der primäre PCIe-Schalter Zugang zu Speichereinsätzen **30** bereitstellen, die bezüglich einer Schicht der ersten Steuerplatine **18A** (z.B. einer Teilmenge von Speichereinsätzen **30**) lokal sind, und der Hilfs-PCIe-Schalter kann Zugang zu Speichereinsätzen **30** bereitstellen, die bezüglich einer Schicht der zweiten Steuerplatine **18B** (z.B. einer zweiten Teilmenge von Speichereinsätzen **30**) lokal sind. Auf diese Weise kann der erste PCIe-Schalter **24A**, gleichgültig, ob es sich um einen einzigen Schalter oder mehrere Schalter handelt, der ersten Steuer-CPU **22A** und dem ersten FPGA **28A** Zugang zu allen Speichereinsätzen **30** bereitstellen.

[0023] Ähnlich umfasst die zweite Steuerplatine **18B** zweite Netzwerkports **20B**, eine zweite Steuer-CPU **22B**, zweiten DRAM **26B**, einen zweiten PCIe-Schalter **24B** und ein zweites FPGA **28B**. Die zweiten Netzwerkports **20B**, die zweite Steuer-CPU **22B**, der zweite DRAM **26B**, der zweite PCIe-Schalter **24B** und das zweite FPGA **28B** können jeweils den ersten Netzwerkports **20A**, der ersten Steuer-CPU **22A**, dem ersten DRAM **26A**, dem ersten PCIe-Schalter **24A** und dem ersten FPGA **28A** ähnlich oder im Wesentlichen gleich sein.

[0024] In einigen Beispielen können die erste Steuer-CPU **22A** und die zweite Steuer-CPU **22B** beide gleichzeitig aktiv sein, was hier als Aktiv-Aktiv-Konfiguration bezeichnet wird. In anderen Beispielen kann von der ersten Steuer-CPU **22A** und der zweiten Steuer-CPU **22B** eine aktiv sein, und die andere kann passiv oder im Leerlauf sein, was als Aktiv-Passiv-Konfiguration bezeichnet wird. Die passive oder leerlaufende Steuer-CPU ist im Fall eines Ausfalls der aktiven Steuer-CPU anwesend.

[0025] In Beispielen, in denen die erste Steuer-CPU **22A** und die zweite Steuer-CPU **22B** in einer Aktiv-Aktiv-Konfiguration oder einer Aktiv-Passiv-Konfiguration sind, können die erste Steuer-CPU **22A** und die zweite Steuer-CPU **22B** Daten austauschen, um Zustandsinformationen zu unterhalten, wie etwa Koordination von Schreibaktualisierungen an Speichereinsätzen **30**. Wie in **Fig. 1** gezeigt, können in einigen Beispielen die erste Steuer-CPU **22A** und die zweite Steuer-CPU **22B** kommunikativ durch eine dedizierte Verbindung gekoppelt sein.

[0026] In einigen Beispielen kann das NAS-System **16** zusätzliche Komponenten umfassen, die in **Fig. 1** nicht gezeigt sind. Zum Beispiel kann das NAS-System **16** eine oder mehrere Stromversorgungen umfassen und kann eine Midplane umfassen, die Signale zwischen Speichereinsätzen **30** und Steuerplatinen **18** routet. Ferner kann in einigen Beispielen die erste Steuerplatine **18A** in eine Steuerplatine mit ersten Netzwerkports **20A**, der ersten Steuer-CPU **22A** und erstem DRAM **26A**; und eine Routerplatine mit dem ersten PCIe-Schalter **24A** und dem ersten FPGA **28A** aufgetrennt sein. Ähnlich kann die zweite Steuerplatine **18B** in eine Steuerplatine mit zweiten Netzwerkports **20B**, der zweiten Steuer-CPU **22B** und zweiten DRAM **26B**; und eine Routerplatine mit dem zweiten PCIe-Schalter **24B** und dem zweiten FPGA **28B** aufgetrennt sein.

[0027] Die erste Steuer-CPU **22A** und die zweite Steuer-CPU **22B** sind jeweils ein Wurzelport eines PCIe-Koppelfelds, das um den ersten PCIe-Schalter **24A** bzw. den zweiten PCIe-Schalter **24B** herum zentriert ist. Jedes PCIe-Koppelfeld benutzt sein eigenes Adressenschema, das während der Erhebung zugewiesen wird, die durch den PCIe-Treiber durchgeführt wird, der durch die jeweilige Steuer-CPU ausgeführt wird. Es kann verhindert werden, dass Vorrichtungen aus einem PCIe-Koppelfeld voll in das andere PCIe-Koppelfeld sehen können, um Wettbewerb zwischen Vorrichtungen und Adressenschemata zu verhindern. Dementsprechend können die PCIe-Schalter **24A** und **24B** durch einen Port mit einer nicht transparenten Brücke (NTB) verbunden werden. Eine NTB erlaubt Verbindung der zwei PCIe-Koppelfelder über PCIe-Schalter **24** und verhindert, dass Vorrichtungen auf dem ersten PCIe-Koppelfeld voll in das zweite PCIe-Koppelfeld sehen und umgekehrt. Stattdessen kann die NTB ein Fenster für eine mit dem ersten PCIe-Koppelfeld verbundene Vorrichtung zum Sehen in das zweite PCIe-Koppelfeld und umgekehrt bereitstellen. Ferner kann die NTB Adressenübersetzung zwischen den durch die jeweiligen PCIe-Koppelfelder verwendeten Adressen bereitstel-

len, wodurch Vorrichtungen aus dem zweiten PCIe-Koppelfeld auf Vorrichtungen aus dem ersten PCIe-Koppelfeld zugreifen können und umgekehrt. In einigen Beispielen ist von der ersten Steuer-CPU **22A** und der zweiten Steuer-CPU **22B** eine als Master-Steuer-CPU und die andere als Slave-Steuer-CPU festgelegt.

[0028] Da das NAS-System **16** ein hoch verfügbares Zweifachport-System ist, umfasst jeder der Speichereinsätze **30** zwei Ports - eine Verbindung mit jedem des ersten PCIe-Schalters **24A** und des zweiten PCIe-Schalters **24B**. Dementsprechend ist jeder Speichereinsatz der Speichereinsätze **30** mit dem ersten PCIe-Koppelfeld und dem zweiten PCIe-Koppelfeld verbunden.

[0029] Gemäß Beispielen der vorliegenden Offenbarung umfasst mindestens ein Speichereinsatz der Speichereinsätze **30** einen Einzelport-Speichercontroller. Zur Ermöglichung der Verwendung eines Einzelport-Controllers in einem Zweifachport-Speichereinsatz kann der Speichereinsatz auch einen PCIe-Schalter umfassen, der zwei externe Ports umfasst (die den Speichereinsatz mit den PCIe-Schaltern **24A** und **24B** verbinden). Die externen Ports können selektiv kommunikativ mit einer NTB gekoppelt werden. Die NTB erlaubt die Verbindung der zwei PCIe-Koppelfelder über den PCIe-Schalter in dem Speichereinsatz und verhindert, dass Vorrichtungen auf dem ersten PCIe-Koppelfeld voll in das zweite PCIe-Koppelfeld sehen und umgekehrt. Stattdessen kann die NTB ein Fenster für eine mit dem ersten PCIe-Koppelfeld verbundene Vorrichtung zum Sehen in das zweite PCIe-Koppelfeld und umgekehrt bereitstellen. Ferner kann die NTB Adressenübersetzung zwischen den durch die jeweiligen PCIe-Koppelfelder verwendeten Adressen bereitstellen, wodurch Vorrichtungen aus dem zweiten PCIe-Koppelfeld auf Vorrichtungen aus dem ersten PCIe-Koppelfeld zugreifen können und umgekehrt. In einigen Beispielen zählt die Master-Steuer-CPU die Speichereinsätze **30** auf und die NTB ist kommunikativ mit dem Port gekoppelt, der mit dem PCIe-Koppelfeld der Slave-Steuer-CPU gekoppelt ist. Die NTB führt dann Adressenübersetzung zwischen dem ersten PCIe-Koppelfeld (von dem die Speichereinsätze **30** Teil sind) und dem zweiten PCIe-Koppelfeld durch. Auf diese Weise können beide Steuer-CPU **22** auf die Speichereinsätze zugreifen, worin ein Einzelport-Controller enthalten ist, obwohl der Controller nur einen einzigen Port umfasst.

[0030] **Fig. 2** ist eine konzeptuelle und schematische Blockdarstellung eines beispielhaften Speichereinsatzes **30A** gemäß einem oder mehreren Aspekten der vorliegenden Offenbarung. Wie in **Fig. 2** gezeigt, umfasst der Speichereinsatz **30A** ein erstes Speicherungs-Blade **32A** und ein zweites Speicherungs-Blade **32B**. Das erste Speicherungs-Blade **32A** kann als Mother-Blade bezeichnet werden und umfasst einen PCIe-Schalter **34**. Das zweite Speicherungs-Blade **32B** kann als Daughter-Blade bezeichnet werden und ist optional.

[0031] Das erste Speicherungs-Blade **32A** umfasst den PCIe-Schalter **34** mit einer NTB **36**, einen ersten Einzelport-Speichercontroller **38A** und eine erste Vielzahl von nichtflüchtigen Speicherungsvorrichtungen 40AA-40AN (zusammen „NVM-Vorrichtungen **40A**“). Die NVM-Vorrichtungen **40A** können eine beliebige Art von nichtflüchtigen Speicherungsvorrichtungen umfassen. Einige Beispiele für NVM-Vorrichtungen **40A** wären, aber ohne Beschränkung darauf, Flash-Speicher-Vorrichtungen, Phasenänderungs-Speicher- bzw. PCM-Vorrichtungen, resistive Direktzugriffsspeicher- bzw. ReRAM-Vorrichtungen, magnetoresistive Direktzugriffs-Speicher- bzw. MRAM-Vorrichtungen, ferroelektrischer Direktzugriffsspeicher (F-RAM), holografische Speicher-Vorrichtungen und eine beliebige andere Art von nichtflüchtigen Speicherungsvorrichtungen. Jede der NVM-Vorrichtungen **40A** ist z.B. durch einen dedizierten Kanal mit einem ersten Einzelport-Speichercontroller **38A** verbunden.

[0032] Der erste Einzelport-Speichercontroller **38A** steuert den Betrieb des Speicherungs-Blade z.B. auf der Basis von Befehlen, die von den Steuer-CPUs **22** oder FPGAs **28** (**Fig. 1**) empfangen werden. In einigen Beispielen kann der erste Einzelport-Speichercontroller **38A** ein voll ausgestatteter Speichercontroller sein und kann Funktionalität für eines oder mehrere von Müllabfuhr, Abnutzungsnivellierung, Verschlüsselung, Entschlüsselung, Komprimierung, Dekomprimierung, Fehlerkorrekturcode, Befehlswarteschlangen, Lesepufferung, Schreibaggregation oder dergleichen bereitstellen oder ausführen. In anderen Beispielen kann der erste Einzelport-Speichercontroller **38A** eine verringerte Menge von Funktionalität implementieren, da die Steuer-CPUs **22** und FPGAs **28** einen Teil der Funktionalität eines besser ausgestatteten Speichercontrollers implementieren können. Zum Beispiel kann der erste Einzelport-Speichercontroller **38A** Programmier-, Lösch- und Leseoperationen ausführen, und übrige Funktionalität kann durch die Steuer-CPUs **22** und FPGAs **28** ausgeführt werden. Der erste Einzelport-Speichercontroller **38A** ist mit dem PCIe-Schalter 34 durch einen einzigen PCIe-Port kommunikativ gekoppelt (z.B. elektrisch gekoppelt). In einigen Beispielen, in denen mehrere NVM-Vorrichtungen **40A** Flash-Vorrichtungen sind, kann der erste Einzelport-Speichercontroller **38A** als Flash-Controller bezeichnet werden.

[0033] Das zweite Speicherungs-Blade **32B** ist dem ersten Speicherungs-Blade **32A** ähnlich, umfasst aber nicht den PCIe-Schalter 34. Stattdessen umfasst das zweite Speicherungs-Blade **32B** einen

zweiten Einzelport-Speichercontroller **38B** und eine zweite Vielzahl von NVM-Vorrichtungen 40BA-40BN (zusammen „zweite Vielzahl von NVM-Vorrichtungen 40B“). Der zweite Einzelport-Speichercontroller **38B** und die zweite Vielzahl von NVM-Vorrichtungen **40B**-**40B** können dem ersten Einzelport-Speichercontroller **38A** und der ersten Vielzahl von NVM-Vorrichtungen **40A**-**40A** jeweils ähnlich oder im Wesentlichen gleich sein.

[0034] In dem Beispiel von **Fig. 2** ist der PCIe-Schalter **34** ein 2-zu-2-Schalter. In einigen Beispielen, wie etwa wenn der Speichereinsatz **30A** das zweite Speicherungs-Blade **32B** weglässt, kann der PCIe-Schalter **34** ein 2-zu-1-Schalter sein. Wie in **Fig. 2** gezeigt, umfasst der PCIe-Schalter **34** einen kommunikativ (z.B. elektrisch) mit dem ersten PCIe-Schalter **24A** gekoppelten ersten PCIe-Port, einen kommunikativ (z.B. elektrisch) mit dem zweiten PCIe-Schalter **24B** gekoppelten zweiten PCIe-Port, einen kommunikativ (z.B. elektrisch) mit dem ersten Einzelport-Speichercontroller **38A** gekoppelten dritten PCIe-Port und einen kommunikativ (z.B. elektrisch) mit dem zweiten Einzelport-Speichercontroller **38B** gekoppelten vierten PCIe-Port. Auf diese Weise koppelt der PCIe-Schalter 34 kommunikativ jeden der Speichercontroller **38** mit jedem der PCIe-Schalter 24.

[0035] Der PCIe-Schalter **34** umfasst auch die NTB **36**. In dem Beispiel von **Fig. 2** ist die NTB **36** mit dem zweiten PCIe-Port gekoppelt, der kommunikativ mit dem zweiten PCIe-Schalter **24B** gekoppelt ist. Daraus folgt, dass die zweite Steuer-CPU **22B** die Slave-Steuer-CPU ist und die erste Steuer-CPU **22A** die Master-Steuer-CPU ist. In einigen Beispielen kann der PCIe-Schalter **34** Logik oder Schalter umfassen, die es der NTB erlauben, selektiv einem beliebigen Port des PCIe-Schalters **34** zugewiesen zu werden. In einigen Beispielen kann der PCIe-Schalter **34** (z.B. durch den ersten Einzelport-Speichercontroller **38A** oder eine der Steuer-CPUs **22**) so gesteuert werden, dass die NTB **36** dem PCIe-Schalter zugewiesen wird, der der Slave-Steuer-CPU zugeordnet ist.

[0036] Wie oben beschrieben erlaubt die NTB **36** Verbindung der zwei PCIe-Koppelfelder (eines ist dem ersten PCIe-Schalter **24A** zugeordnet und das andere dem zweiten PCIe-Schalter **24B**) über den PCIe-Schalter 34, verhindert, dass Vorrichtungen auf dem ersten PCIe-Koppelfeld voll in das zweite PCIe-Koppelfeld sehen und umgekehrt. Stattdessen kann die NTB **36** ein Fenster für mit dem ersten PCIe-Koppelfeld verbundene Vorrichtungen zum Sehen in das zweite PCIe-Koppelfeld und umgekehrt bereitstellen. Da der Speichereinsatz **30A** durch die erste Steuer-CPU **22A** aufgezählt werden und Teil des ersten PCIe-Koppelfelds sein kann, kann die NTB **36** der zweiten Steuer-CPU **22B** und dem zweiten FPGA **28B** erlauben, den Speichereinsatz **30A** zu sehen. Ferner kann die NTB **36** bidirektionale Adressenübersetzung zwischen den durch das erste PCIe-Koppelfeld (mit der ersten Steuer-CPU **22A**, dem ersten FPGA **28A** und dem Speichereinsatz **30A**) verwendeten Adressen und durch das zweite PCIe-Koppelfeld (mit der zweiten Steuer-CPU **22B** und dem zweiten FPGA **28B**) verwendeten Adressen bereitstellen, wodurch die zweite Steuer-CPU **22B** und das zweite FPGA **28B** aus dem zweiten PCIe-Koppelfeld auf den Speichereinsatz **30A** und die erste und zweite Vielzahl von NVM-Vorrichtungen **40A** und **40B** zugreifen können.

[0037] **Fig. 3** ist eine konzeptuelle und schematische Blockdarstellung eines beispielhaften Speicherungs-Blade **32A** gemäß einem oder mehreren Aspekten der vorliegenden Offenbarung. Wie in **Fig. 3** gezeigt und ähnlich wie das Speicherungs-Blade **32A** von **Fig. 2** umfasst das Speicherungs-Blade **32A** einen PCIe-Schalter **34**, einen ersten Einzelport-Speichercontroller **38A** und eine Vielzahl von NVM-Vorrichtungen **40A**. In dem Beispiel von **Fig. 3** umfasst der PCIe-Schalter 34 vier PCIe-Ports: einen kommunikativ (z.B. elektrisch) mit dem ersten PCIe-Schalter **24A** gekoppelten ersten PCIe-Port, einen kommunikativ (z.B. elektrisch) mit dem zweiten PCIe-Schalter **24B** gekoppelten zweiten PCIe-Port, einen kommunikativ (z.B. elektrisch) mit dem ersten Einzelport-Speichercontroller **38A** gekoppelten dritten PCIe-Port und einen kommunikativ (z.B. elektrisch) mit dem zweiten Einzelport-Speichercontroller **38B** gekoppelten vierten PCIe-Port. Wie oben beschrieben kann in anderen Beispielen der PCIe-Schalter 34 drei Ports umfassen und kann eine Verbindung mit dem zweiten Einzelport-Speichercontroller **38B** weglassen, z.B. in Beispielen, in denen der Speicherungsbehälter, der das Speicherungs-Blade **32A** umfasst, kein zweites Speicherungs-Blade umfasst.

[0038] Im Gegensatz zu dem in **Fig. 2** gezeigten Beispiel wurde in dem Beispiel von **Fig. 3** die NTB **36** nicht einem PCIe-Port zugewiesen. Ferner umfasst in dem Beispiel von **Fig. 3** das Speicherungs-Blade **32A** zusätzliche Merkmale, darunter eine Schaltsteuerschnittstelle **42**, eine erste Seitenband-Steuerschnittstelle **44A** und eine zweite Seitenband-Steuerschnittstelle **44B**.

[0039] Die Schaltsteuerschnittstelle **42** koppelt kommunikativ (z.B. elektrisch) den ersten Einzelport-Speichercontroller **38A** und den PCIe-Schalter 34. Zum Beispiel kann die Schaltsteuerschnittstelle **42** zwei oder mehr elektrische Bahnen umfassen, von denen eine oder mehrere dem Empfang und eine oder mehrere dem Senden dediziert sind. Als weiteres Beispiel kann die Schaltsteuerschnittstelle **42** nur unidirektionale Kommunikation von dem ersten Einzelport-Speichercontroller **38A** zu dem PCIe-Schalter 34 ermöglichen und kann somit eine oder mehrere elektrische Bahnen umfassen, die dem Senden von

dem ersten Einzelport-Speichercontroller **38A** dediziert sind.

[0040] Der erste Einzelport-Speichercontroller **38A** kann dafür ausgelegt sein, über die Schaltsteuerschnittstelle **42** einen Befehl zu dem PCIe-Schalter 34 zu übermitteln, um zu bewirken, dass die NTB **36** entweder dem ersten PCIe-Port des PCIe-Schalters 34 oder dem zweiten PCIe-Port des PCIe-Schalters 34 zugewiesen wird. Anders ausgedrückt, kann der erste Einzelport-Speichercontroller **38A** dafür ausgelegt sein, über die Schaltsteuerschnittstelle **42** einen Befehl zu dem PCIe-Schalter 34 zu übermitteln, um zu bewirken, dass die NTB **36** kommunikativ (z.B. elektrisch) entweder mit dem ersten PCIe-Port des PCIe-Schalters 34 oder dem zweiten PCIe-Port des PCIe-Schalters 34 gekoppelt wird. Auf diese Weise kann der erste Einzelport-Speichercontroller **38A** bewirken, dass die NTB **36** einem Port zugewiesen wird, der einem PCIe-Schalter (z.B. dem ersten PCIe-Schalter **24A** oder dem zweiten PCIe-Schalter **24B**) zugeordnet ist, der einer Slave-Steuer-CPU (entweder der ersten Steuer-CPU **22A** oder der zweiten Steuer-CPU **22B)** zugeordnet ist.

[0041] In einigen Beispielen kann der erste Einzelport-Speichercontroller **38A** während des PCIe-Aufzählungsprozesses beim Herauffahren des NAS-Systems **16** (**Fig. 1**) oder bei Failover von der ersten Steuer-CPU **22A** als Master zu der zweiten Steuer-CPU **22B** als Master oder umgekehrt einen NTB-Zuweisungsbefehl von der ersten Steuer-CPU **22A** oder der zweiten Steuer-CPU **22B** empfangen. Die erste Seitenband-Steuerschnittstelle **44A** koppelt kommunikativ (z.B. elektrisch) den ersten Einzelport-Speichercontroller **38A** mit dem ersten PCIe-Schalter, FPGA **28A** oder der CPU **22A** (z.B. über eine Midplane, einen E/A-Port oder eine Standardschnittstelle, wie etwa einen UART (Universal Asynchronous Receiver/Transmitter), einen I^2C-Controller (Inter-Integrated Circuit) oder dergleichen) und ist dafür ausgelegt, es dem ersten Einzelport-Speichercontroller **38A** zu erlauben, mit dem ersten PCIe-Schalter **24A** (einer ersten externen Vorrichtung) zu kommunizieren, während der PCIe-Schalter 34 umgangen wird. Ähnlich koppelt die zweite Seitenband-Steuerschnittstelle **44B** kommunikativ (z.B. elektrisch) den ersten Einzelport-Speichercontroller **38A** mit dem zweiten PCIe-Schalter **24B** und ist dafür ausgelegt, es dem ersten Einzelport-Speichercontroller **38A** zu erlauben, mit dem zweiten PCIe-Schalter **24B** (einer zweiten externen Vorrichtung) zu kommunizieren, während der PCIe-Schalter 34 umgangen wird. In einigen Beispielen umfasst jede der ersten Seitenband-Steuerschnittstelle **44A** und der zweiten Seitenband-Steuerschnittstelle **44B** (zusammen „Seitenband-Steuerschnittstellen 44“) eine erste Kommunikationsverbindung (z.B. elektrische Bahn), die dem Senden dediziert ist, und eine zweite Kommunikationsverbindung (z.B. elektrische Bahn), die zum Empfangen dediziert ist.

[0042] Auf diese Weise erlauben die Seitenband-Steuerschnittstellen **44** dem ersten Einzelport-Speichercontroller **38A**, einen NTB-Zuweisungsbefehl von einer Master-Steuer-CPU (z.B. entweder der ersten Steuer-CPU **22A** oder der zweiten Steuer-CPU **22B**) während des durch den PCIe-Treiber ausgeführten Vorrichtungsaufzählungsprozesses zu empfangen. Als Reaktion auf Empfang des NTB-Zuweisungsbefehls kann der erste Einzelport-Speichercontroller **38A** ausgelegt sein zum Übermitteln eines Befehls auf der Basis des NTB-Zuweisungsbefehls zu dem PCIe-Schalter **34** über die Schaltsteuerschnittstelle **42**, um zu bewirken, dass die NTB **36** kommunikativ (z.B. elektrisch) entweder mit dem ersten PCIe-Port des PCIe-Schalters **34** oder dem zweiten PCIe-Port des PCIe-Schalters **34** gekoppelt wird. Auf diese Weise kann die Master-Steuer-CPU, ohne sich auf den PCIe-Schalter zu verlassen (der mit Bezug auf die NTB **36** nicht korrekt konfiguriert sein kann), bewirken, dass die NTB **36** dem PCIe-Port des PCIe-Schalters zugewiesen wird, der der Slave-Steuer-CPU zugeordnet ist.

[0043] In einigen Beispielen benutzen die Einzelport-Speichercontroller **38** DRAM **26** (**Fig. 1**) für Task-Warteschlangen, wie etwa Schreibwarteschlangen, Lesewarteschlangen oder dergleichen. Zum Beispiel können Steuer-CPUs **22** Tasks zu jeweiligen Warteschlangen hinzufügen, die jeweiligem DRAM **26** zugeordnet sind (z.B. kann die erste Steuer-CPU **22A** Tasks zu jeweiligen Task-Warteschlangen hinzufügen, die im ersten DRAM **26A** gespeichert sind, und die zweite Steuer-CPU **22B** kann Tasks zu jeweiligen Task-Warteschlangen hinzufügen, die im zweiten DRAM **26B** gespeichert sind), und die Einzelport-Speichercontroller **38** können auf die jeweiligen Task-Warteschlangen zugreifen, um Tasks zur Ausführung zu erhalten. Ein Teil der Steuerlogik für die Task-Warteschlangen kann durch die Steuer-CPUs **22** ausgeführt werden, und ein Teil der Steuerlogik für die Task-Warteschlangen kann durch die Einzelport-Speichercontroller **38** ausgeführt werden. Da die Einzelport-Speichercontroller **38** einen einzigen Port aufweisen, können die Einzelport-Speichercontroller **38** jedoch nicht ohne weiteres dafür ausgelegt werden, auf Task-Warteschlangen zuzugreifen, die beiden Steuer-CPUs **22** zugeordnet sind.

[0044] Gemäß einigen Beispielen der vorliegenden Offenbarung kann jeder Einzelport-Speichercontroller **38** eine definierte Anzahl von Task-Warteschlangen unterstützen. Für jeden Einzelport-Speichercontroller **38** kann die Gesamtzahl von Task-Warteschlangen durch die Anzahl der Steuer-CPUs **22** dividiert werden, und eine jeweilige Anzahl von Task-Warteschlangen kann jeder jeweiligen Steuer-CPU der Steuer-CPUs **22** dediziert werden. In einigen Bei-

spielen können die Task-Warteschlangen nummeriert oder anderweitig identifiziert werden; somit können die Einzelport-Speichercontroller **38** und Steuer-CPUs **22** wissen, welche Speicherungswarteschlangen sich in welchem DRAM **26** befinden.

[0045] Zum Beispiel kann jeder Einzelport-Speichercontroller **38 1024** Task-Warteschlangen unterstützen. Da die NAS **16** zwei Steuer-CPUs **22A** und **22B** umfasst, können die Task-Warteschlangen halbiert werden, und für jeden Einzelport-Speichercontroller **38** können 512 Task-Warteschlangen der ersten Steuer-CPU **22A** zugeordnet oder zugewiesen werden und 512 Task-Warteschlangen können der zweiten Steuer-CPU **22B** zugeordnet oder zugewiesen werden. In einigen Beispielen können die Einzelport-Speichercontroller **38** eine Angabe der zugewiesenen Ports zu der ersten Steuer-CPU **22A** und zweiten Steuer-CPU **22B** z.B. durch ein Signal über den Midplane-Verbinder übermitteln.

[0046] **Fig. 4** ist ein Flussdiagramm einer beispielhaften Technik, um eine nicht transparente Brücke einem PCIe-Port zuzuweisen, gemäß einem oder mehreren Aspekten der vorliegenden Offenbarung. Die Technik von **Fig. 3** wird mit Bezug auf die Speicherungsumgebung **10** von **Fig. 1**, den Speichereinsatz **30A** von **Fig. 2** und das Speicherungs-Blade **32A** von **Fig. 3** beschrieben, obwohl andere Vorrichtungen und Systeme die Technik von **Fig. 4** implementieren können und die Speicherungsumgebung **10** von **Fig. 1**, der Speichereinsatz **30A** von **Fig. 2** und das Speicherungs-Blade **32A** von **Fig. 3** andere Techniken ausführen können.

[0047] Die Technik von **Fig. 4** umfasst, dass ein Speichercontroller (z.B. erster Einzelport-Speichercontroller **38A**) eines Speichereinsatzes **30A** einen NTB-Zuweisungsbefehl von einer Steuer-CPU (z.B. der ersten Steuer-CPU **22A** oder der zweiten Steuer-CPU **22B**) empfängt (52). Wie mit Bezug auf **Fig. 2** und **Fig. 3** beschrieben umfasst der Speichercontroller (z.B. der erste Einzelport-Speichercontroller **38A**) einen einzigen PCIe-Port. Der Speichereinsatz **30A** umfasst den Speichercontroller (z.B. den ersten Einzelport-Speichercontroller **38A**) und einen PCIe-Schalter **34**. Der PCIe-Schalter **34** umfasst einen kommunikativ mit einem ersten PCIe-Koppelfeld gekoppelten ersten PCIe-Port, einen kommunikativ mit einem zweiten, anderen PCIe-Koppelfeld gekoppelten zweiten PCIe-Port und einen kommunikativ mit dem einzigen PCIe-Port des Speichercontrollers (z.B. des ersten Einzelport-Speichercontrollers **38A)** gekoppelten dritten PCIe-Port. Der NTB-Zuweisungsbefehl kann eine Angabe umfassen, welche Steuer-CPU die Master-Steuer-CPU ist (z.B. die erste Steuer-CPU **22A** oder die zweite Steuer-CPU **22B**), welcher PCIe-Port des PCIe-Schalters **34** der NTB **36** zuzuweisen ist, welcher PCIe-Port des PCIe-Schalters **34** mit einem PCIe-Schalter (**24A** oder **24B**) verbunden ist, der der Master-Steuer-CPU (z.B. der ersten Steuer-CPU **22A** oder der zweiten Steuer-CPU **22B**) zugeordnet ist, welcher PCIe-Port des PCIe-Schalters 34 mit einem PCIe-Schalter (**24A** oder **24B**) verbunden ist, der der Slave-Steuer-CPU (z.B. der ersten Steuer-CPU **22A** oder der zweiten Steuer-CPU **22B**) zugeordnet ist oder dergleichen.

[0048] Die Technik von **Fig. 4** umfasst außerdem, dass der Speichercontroller (z.B. der erste Einzelport-Speichercontroller **38A**) auf der Basis des NTB-Zuweisungsbefehls einen Befehl zu dem PCIe-Schalter 34 übermittelt, um zu bewirken, dass sich die NTB **36** entweder mit dem ersten PCIe-Port des PCIe-Schalters 34 oder dem zweiten PCIe-Port des PCIe-Schalters 34 kommunikativ koppelt (z.B. elektrisch koppelt oder diesem zugewiesen wird) (54). Wie oben beschrieben kann in einigen Beispielen die NTB **36** mit dem PCIe-Port des PCIe-Schalters 34, der mit der Slave-Steuer-CPU (z.B. der ersten Steuer-CPU **22A** oder der zweiten Steuer-CPU **22B**) verbunden ist, über den jeweiligen PCIe-Schalter **24a** oder **24B** kommunikativ gekoppelt (z.B. elektrisch gekoppelt oder zugewiesen) werden. Als Ergebnis wird der andere PCIe-Port (der nicht der NTB **36** zugeordnet ist) der Master-Steuer-CPU (z.B. der ersten Steuer-CPU **22A** oder der zweiten Steuer-CPU **22B**) zugeordnet und stellt eine transparente Verbindung mit dem PCIe-Koppelfeld bereit, das der Master-Steuer-CPU zugeordnet ist (z.B. ist keine Adressenübersetzung für Vorrichtungen erforderlich, die dem PCIe-Koppelfeld zugeordnet sind, das der Master-Steuer-CPU zugeordnet ist, um auf den Speichereinsatz **30A** zuzugreifen.

[0049] Obwohl die obigen Beispiele mit Bezug auf einen Controller einer Speicherungsvorrichtung beschrieben wurden, können die hier beschriebenen Beispiele in anderen Szenarien durch einen anderen Prozessor, wie etwa einen Vielzweckprozessor, implementiert werden, und die Tabelle der Übersetzung von logischen in physische Datenadressen kann zum Beispiel ein Übersetzungs-Lookaside-Puffer sein.

[0050] Die in der vorliegenden Offenbarung beschriebenen Techniken können mindestens teilweise in Hardware, Software, Firmware oder einer beliebigen Kombination davon implementiert werden. Zum Beispiel können verschiedene Aspekte der beschriebenen Techniken in einem oder mehreren Prozessoren implementiert werden, darunter ein oder mehrere Mikroprozessoren, digitale Signalprozessoren (DSP), anwendungsspezifische integrierte Schaltungen (ASIC), am Einsatzort programmierbare Gatearrays (FPGA) oder beliebige andere äquivalente integrierte oder diskrete Logikschaltkreise, sowie beliebige Kombinationen solcher Komponenten. Der Ausdruck „Prozessor“ oder „Verarbeitungsschaltkreise“ kann sich im Allgemeinen auf beliebige der obigen Logikschaltkreise alleine oder in Kombination mit an-

deren Logikschaltkreisen oder beliebige andere äquivalente Schaltkreise beziehen. Eine Steuereinheit, die Hardware umfasst, kann auch eine oder mehrere der Techniken der vorliegenden Offenbarung ausführen.

[0051] Solche Hardware, Software und Firmware kann in derselben Vorrichtung oder in getrennten Vorrichtungen implementiert werden, um die verschiedenen in der vorliegenden Offenbarung beschriebenen Techniken zu unterstützen. Zusätzlich können beliebige der beschriebenen Einheiten, Module oder Komponenten zusammen oder getrennt als diskrete, aber interoperable Logikvorrichtungen implementiert werden. Die Abbildung verschiedener Merkmale als Module oder Einheiten soll verschiedene Funktionsaspekte hervorheben und bedeutet nicht unbedingt, dass solche Module oder Einheiten durch getrennte Hardware-, Firmware- oder Softwarekomponenten realisiert werden müssen. Funktionalität, die einem oder mehreren Modulen oder einer oder mehreren Einheiten zugeordnet ist, kann stattdessen durch getrennte Hardware-, Firmware- oder Softwarekomponenten ausgeführt oder in gemeinsame oder getrennte Hardware-, Firmware- oder Softwarekomponenten integriert werden.

[0052] Die in der vorliegenden Offenbarung beschriebenen Techniken können auch in einem Herstellungsartikel realisiert oder codiert werden, der ein mit Anweisungen codiertes computerlesbares Speichermedium umfasst. In einem Herstellungsartikel, einschließlich eines codierten computerlesbaren Speichermediums, eingebettete oder codierte Anweisungen können bewirken, dass ein oder mehrere programmierbare Prozessoren oder andere Prozessoren eine oder mehrere der hier beschriebenen Techniken implementieren, wie etwa, wenn in dem computerlesbaren Speichermedium enthaltene oder codierte Anweisungen durch den einen oder die mehreren Prozessoren ausgeführt werden. Computerlesbare Speichermedien wären zum Beispiel RAM (Random Access Memory), ROM (Read Only Memory), PROM (Programmable Read Only Memory), EPROM (Erasable Programmable Read Only Memory), EEPROM (Electronically Erasable Programmable Read Only Memory), Flash-Speicher, eine Festplatte, CD-ROM (Compact Disc ROM), eine Diskette, eine Kassette, magnetische Medien, optische Medien oder andere computerlesbare Medien. In einigen Beispielen kann ein Herstellungsartikel ein oder mehrere computerlesbare Speichermedien umfassen.

[0053] In einigen Beispielen kann ein computerlesbares Speichermedium ein nicht transitorisches Medium umfassen. Der Ausdruck „nicht transitorisch“ kann angeben, dass das Speichermedium nicht in einer Trägerwelle oder einem ausgebreiteten Signal realisiert ist. In bestimmten Beispielen kann ein nicht transitorisches Speichermedium Daten speichern, die sich mit der Zeit (z.B. in RAM oder Cache) ändern.

Patentansprüche

1. Speichereinsatz, umfassend:
einen Speichercontroller mit einem einzigen PCIe-Port;
einen PCIe-Schalter, umfassend:
einen kommunikativ mit einem ersten PCIe-Koppelfeld gekoppelten ersten PCIe-Port;
einen kommunikativ mit einem zweiten, anderen PCIe-Koppelfeld gekoppelten zweiten PCIe-Port;
einen kommunikativ mit dem einzigen PCIe-Port des Speichercontrollers gekoppelten dritten PCIe-Port, wobei der erste PCIe-Port und der zweite PCIe-Port dafür ausgelegt sind, selektiv kommunikativ mit einer nicht transparenten Brücke (NTB) des PCIe-Schalters gekoppelt zu werden, ferner umfassend:
eine erste Seitenband-Steuerschnittstelle, die kommunikativ mit dem Speichercontroller gekoppelt und dafür ausgelegt ist, dem Speichercontroller zu erlauben, mit einer ersten externen Vorrichtung zu kommunizieren, während der PCIe-Schalter umgangen wird;
eine zweite Seitenband-Steuerschnittstelle, die kommunikativ mit dem Speichercontroller gekoppelt und dafür ausgelegt ist, dem Speichercontroller zu erlauben, mit einer zweiten externen Vorrichtung zu kommunizieren, während der PCIe-Schalter umgangen wird, wobei der Speichercontroller einen NTB-Zuweisungsbefehl von der ersten externen Vorrichtung oder der zweiten externen Vorrichtung empfängt, wobei der Speichereinsatz ferner Folgendes umfasst:
eine Schaltsteuerschnittstelle, die den Speichercontroller kommunikativ mit dem PCIe-Schalter koppelt, wobei der Speichercontroller dafür ausgelegt ist, auf der Basis des NTB-Zuweisungsbefehls über die Schaltsteuerschnittstelle einen Befehl zu dem PCIe-Schalter zu übermitteln, um zu bewirken, dass sich die NTB kommunikativ entweder mit dem ersten PCIe-Port des PCIe-Schalters oder mit dem zweiten PCIe-Port des PCIe-Schalters koppelt.

2. Speichereinsatz nach Anspruch 1, ferner umfassend:
eine Schaltsteuerschnittstelle, die den Speichercontroller kommunikativ mit dem PCIe-Schalter koppelt, wobei der Speichercontroller dafür ausgelegt ist, einen NTB-Zuweisungsbefehl über die Schaltsteuerschnittstelle zu dem PCIe-Schalter zu übermitteln, um zu bewirken, dass sich die NTB kommunikativ entweder mit dem ersten PCIe-Port des Schalters oder dem zweiten PCIe-Port des PCIe-Schalters koppelt.

3. Speichereinsatz nach Anspruch 1, wobei
der Speichercontroller einen ersten Speichercontroller umfasst;

der Speichereinsatz ferner einen zweiten Speichercontroller umfasst, der einen einzigen PCIe-Port umfasst; und
der PCIe-Schalter ferner einen kommunikativ mit dem einzigen PCIe-Port des zweiten Speichercontrollers gekoppelten vierten PCIe-Port umfasst.

4. Speichereinsatz nach Anspruch 3, der ferner ein erstes Speicherungs-Blade und ein zweites Speicherungs-Blade umfasst, wobei das erste Speicherungs-Blade den PCIe-Schalter, den ersten Speichercontroller und eine erste Vielzahl von nichtflüchtigen Speichervorrichtungen umfasst und wobei das zweite Speicherungs-Blade den zweiten Speichercontroller und eine zweite Vielzahl von nichtflüchtigen Speicherungsvorrichtungen umfasst.

5. Netzwerkangeschlossenes Speicherungs- bzw. NAS-System, umfassend:
eine erste Steuer-CPU;
eine zweite Steuer-CPU;
einen kommunikativ mit der ersten Steuer-CPU gekoppelten ersten PCIe-Schalter;
einen kommunikativ mit der zweiten Steuer-CPU gekoppelten zweiten PCIe-Schalter; und
einen Speichereinsatz, wobei der Speichereinsatz Folgendes umfasst:
einen Speichercontroller, der einen einzigen PCIe-Port umfasst; und
einen dritten PCIe-Schalter, umfassend:
einen kommunikativ mit dem ersten PCIe-Schalter gekoppelten ersten PCIe-Port;
einen kommunikativ mit dem zweiten PCIe-Schalter gekoppelten zweiten PCIe-Port;
einen kommunikativ mit dem einzigen PCIe-Port des Speichercontrollers gekoppelten dritten PCIe-Port, wobei der erste PCIe-Port und der zweite PCIe-Port dafür ausgelegt sind, selektiv kommunikativ mit einer nicht transparenten Brücke (NTB) des dritten PCIe-Schalters gekoppelt zu werden, wobei der Speichereinsatz ferner Folgendes umfasst:
eine erste Seitenband-Steuerschnittstelle, die kommunikativ mit dem Speichercontroller gekoppelt und dafür ausgelegt ist, dem Speichercontroller zu erlauben, mit der ersten Steuer-CPU zu kommunizieren, während der dritte PCIe-Schalter umgangen wird;
eine zweite Seitenband-Steuerschnittstelle, die kommunikativ mit dem Speichercontroller gekoppelt und dafür ausgelegt ist, dem Speichercontroller zu erlauben, mit der zweiten Steuer-CPU zu kommunizieren, während der dritte PCIe-Schalter umgangen wird, wobei der Speichercontroller einen NTB-Zuweisungsbefehl von der ersten Steuer-CPU oder der zweiten Steuer-CPU empfängt, wobei der Speichereinsatz ferner Folgendes umfasst:
eine Schaltsteuerschnittstelle, die den Speichercontroller kommunikativ mit dem dritten PCIe-Schalter koppelt, wobei der Speichercontroller dafür ausgelegt ist, auf der Basis des NTB-Zuweisungsbefehls über die Schaltsteuerschnittstelle einen Befehl zu dem dritten PCIe-Schalter zu übermitteln, um zu bewirken, dass sich die NTB kommunikativ entweder mit dem ersten PCIe-Port des dritten PCIe-Schalters oder mit dem zweiten PCIe-Port des dritten PCIe-Schalters koppelt.

6. NAS-System nach Anspruch 5, wobei der Speichereinsatz ferner Folgendes umfasst:
eine Schaltsteuerschnittstelle, die den Speichercontroller kommunikativ mit dem dritten PCIe-Schalter koppelt, wobei der Speichercontroller dafür ausgelegt ist, einen NTB-Zuweisungsbefehl über die Schaltsteuerschnittstelle zu dem dritten PCIe-Schalter zu übermitteln, um zu bewirken, dass sich die NTB kommunikativ entweder mit dem ersten PCIe-Port des PCIe-Schalters oder mit dem zweiten PCIe-Port des PCIe-Schalters koppelt.

7. NAS-System nach Anspruch 5, wobei
der Speichercontroller einen ersten Speichercontroller umfasst;
der Speichereinsatz ferner einen zweiten Speichercontroller umfasst, der einen einzigen PCIe-Port umfasst; und
der PCIe-Schalter ferner einen kommunikativ mit dem einzigen PCIe-Port des zweiten Speichercontrollers gekoppelten vierten PCIe-Port umfasst.

8. NAS-System nach Anspruch 7, wobei der Speichereinsatz ferner ein erstes Speicherungs-Blade und ein zweites Speicherungs-Blade umfasst, wobei das erste Speicherungs-Blade den dritten PCIe-Schalter, den ersten Speichercontroller und eine erste Vielzahl von nichtflüchtigen Speichervorrichtungen umfasst und wobei das zweite Speicherungs-Blade den zweiten Speichercontroller und eine zweite Vielzahl von nichtflüchtigen Speicherungsvorrichtungen umfasst.

9. NAS-System nach Anspruch 5, wobei der Speichereinsatz eine Vielzahl von Speichereinsätzen umfasst, wobei jeder jeweilige Speichereinsatz der Vielzahl von Speichereinsätzen Folgendes umfasst:
einen jeweiligen Speichercontroller, der einen einzigen PCIe-Port umfasst;
einen jeweiligen dritten PCIe-Schalter, umfassend:
ein jeweiliger erster PCIe-Port ist kommunikativ mit dem ersten PCIe-Schalter gekoppelt;
ein jeweiliger zweiter PCIe-Port ist kommunikativ mit dem zweiten PCIe-Schalter gekoppelt;
ein jeweiliger dritter PCIe-Port ist kommunikativ mit dem einzigen PCIe-Port des jeweiligen Speichercontrollers gekoppelt, und wobei der jeweilige erste PCIe-Port und der jeweilige zweite PCIe-Port dafür ausgelegt sind, selektiv kommunikativ mit einer jeweiligen nicht transparenten Brücke (NTB) des jeweiligen dritten PCIe-Schalters gekoppelt zu werden.

10. Verfahren mit den folgenden Schritten:

ein Speichercontroller eines Speichereinsatzes empfängt einen Zuweisungsbefehl einer nicht transparenten Brücke (NTB) von einer Steuer-CPU, wobei der Speichercontroller einen einzigen PCIe-Port umfasst, wobei der Speichereinsatz den Speichercontroller und einen PCIe-Schalter umfasst und wobei der PCIe-Schalter einen kommunikativ mit einem ersten PCIe-Koppelfeld gekoppelten ersten PCIe-Port, einen kommunikativ mit einem zweiten, anderen PCIe-Koppelfeld gekoppelten zweiten PCIe-Port und einen kommunikativ mit dem einzigen PCIe-Port des Speichercontrollers gekoppelten dritten PCIe-Port umfasst; und
der Speichercontroller übermittelt auf der Basis des NTB-Zuweisungsbefehls einen Befehl zu dem PCIe-Schalter, um zu bewirken, dass sich die NTB kommunikativ entweder mit dem ersten PCIe-Port des PCIe-Schalters oder dem zweiten PCIe-Port des PCIe-Schalters koppelt, wobei
der Speichereinsatz ferner Folgendes umfasst:
eine erste Seitenband-Steuerschnittstelle, die kommunikativ mit dem Speichercontroller gekoppelt und dafür ausgelegt ist, dem Speichercontroller zu erlauben, mit der Steuer-CPU zu kommunizieren, während der PCIe-Schalter umgangen wird;
eine zweite Seitenband-Steuerschnittstelle, die kommunikativ mit dem Speichercontroller gekoppelt und dafür ausgelegt ist, dem Speichercontroller zu erlauben, mit einer zweiten externen Vorrichtung zu kommunizieren, während der PCIe-Schalter umgangen wird; und
der Speichercontroller den NTB-Zuweisungsbefehl über die erste Seitenband-Steuerschnittstelle von der Steuer-CPU empfängt, wobei
der Speichereinsatz ferner eine Schaltsteuerschnittstelle umfasst, die den Speichercontroller kommunikativ mit dem PCIe-Schalter koppelt; und
der Speichercontroller den Befehl auf der Basis des NTB-Zuweisungsbefehls über die Schaltsteuerschnittstelle zu dem PCIe-Schalter übermittelt.

11. Verfahren nach Anspruch 10, wobei
der Speichereinsatz ferner eine Schaltsteuerschnittstelle umfasst, die den Speichercontroller kommunikativ mit dem PCIe-Schalter koppelt; und
der Speichercontroller den Befehl über die Schaltsteuerschnittstelle zu dem PCIe-Schalter übermittelt.

12. Verfahren nach Anspruch 10, wobei
der Speichercontroller einen ersten Speichercontroller umfasst;
der Speichereinsatz ferner einen zweiten Speichercontroller umfasst, der einen einzigen PCIe-Port umfasst; und
der PCIe-Schalter ferner einen kommunikativ mit dem einzigen PCIe-Port des zweiten Speichercontrollers gekoppelten vierten PCIe-Port umfasst.

Es folgen 4 Seiten Zeichnungen

Anhängende Zeichnungen

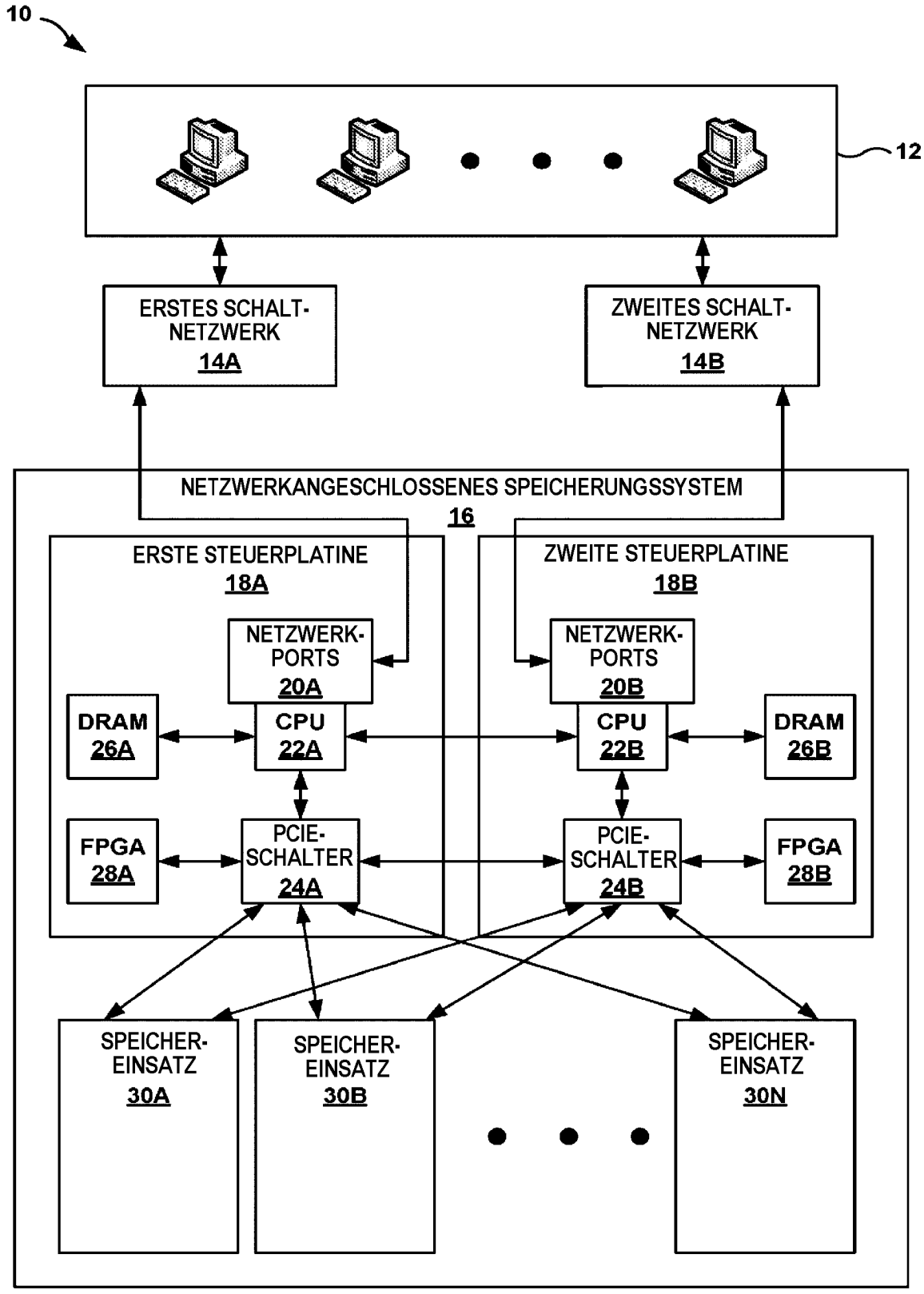


FIG. 1

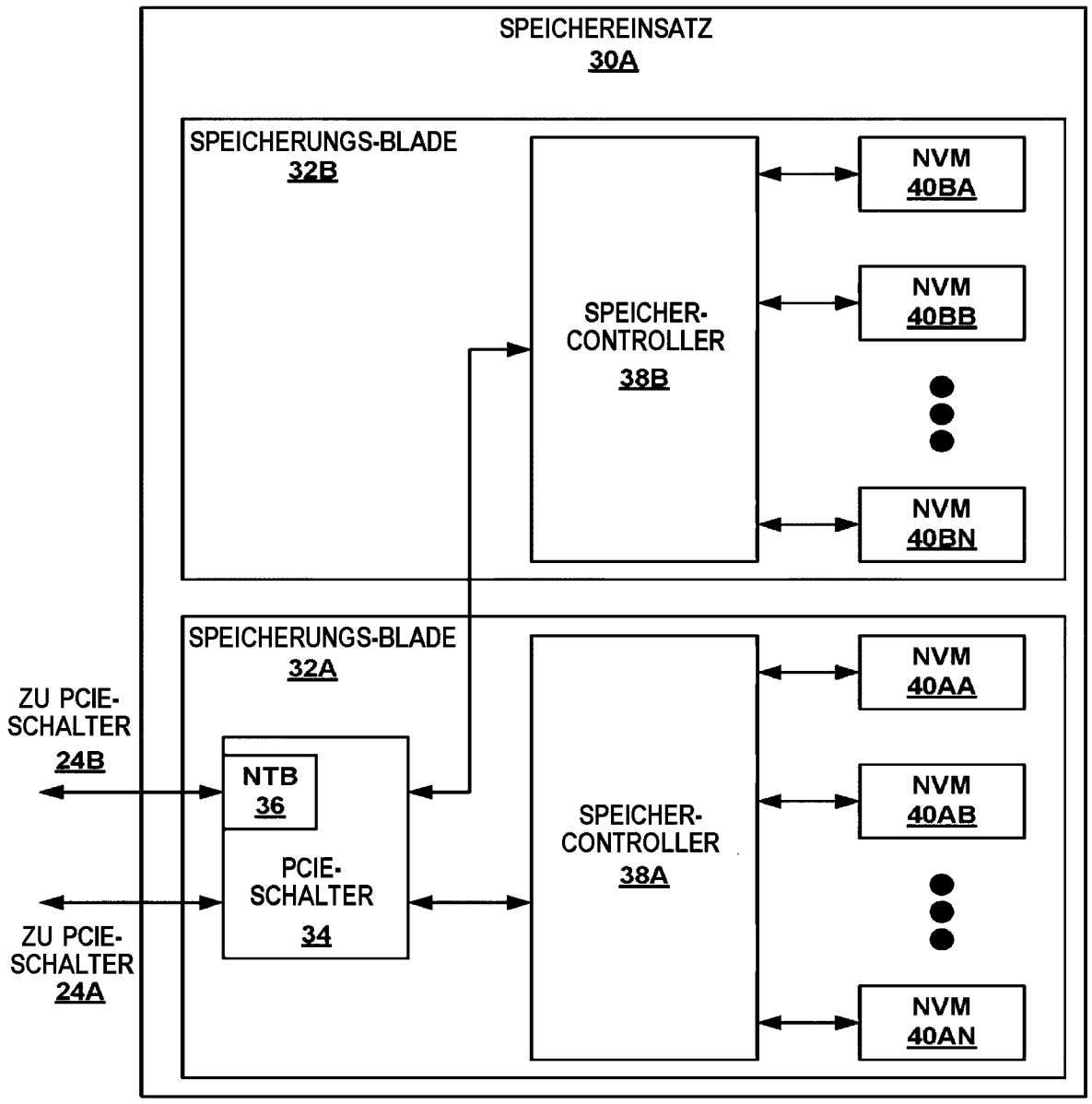

SPEICHEREINSATZ
30A
SPEICHERUNGS-BLADE
32B
SPEICHER-
CONTROLLER
38B
NVM
40BA
NVM
40BB
NVM
40BN
SPEICHERUNGS-BLADE
32A
ZU PCIE-
SCHALTER
24B
NTB
36
PCIE-
SCHALTER
34
ZU PCIE-
SCHALTER
24A
SPEICHER-
CONTROLLER
38A
NVM
40AA
NVM
40AB
NVM
40AN

FIG. 2

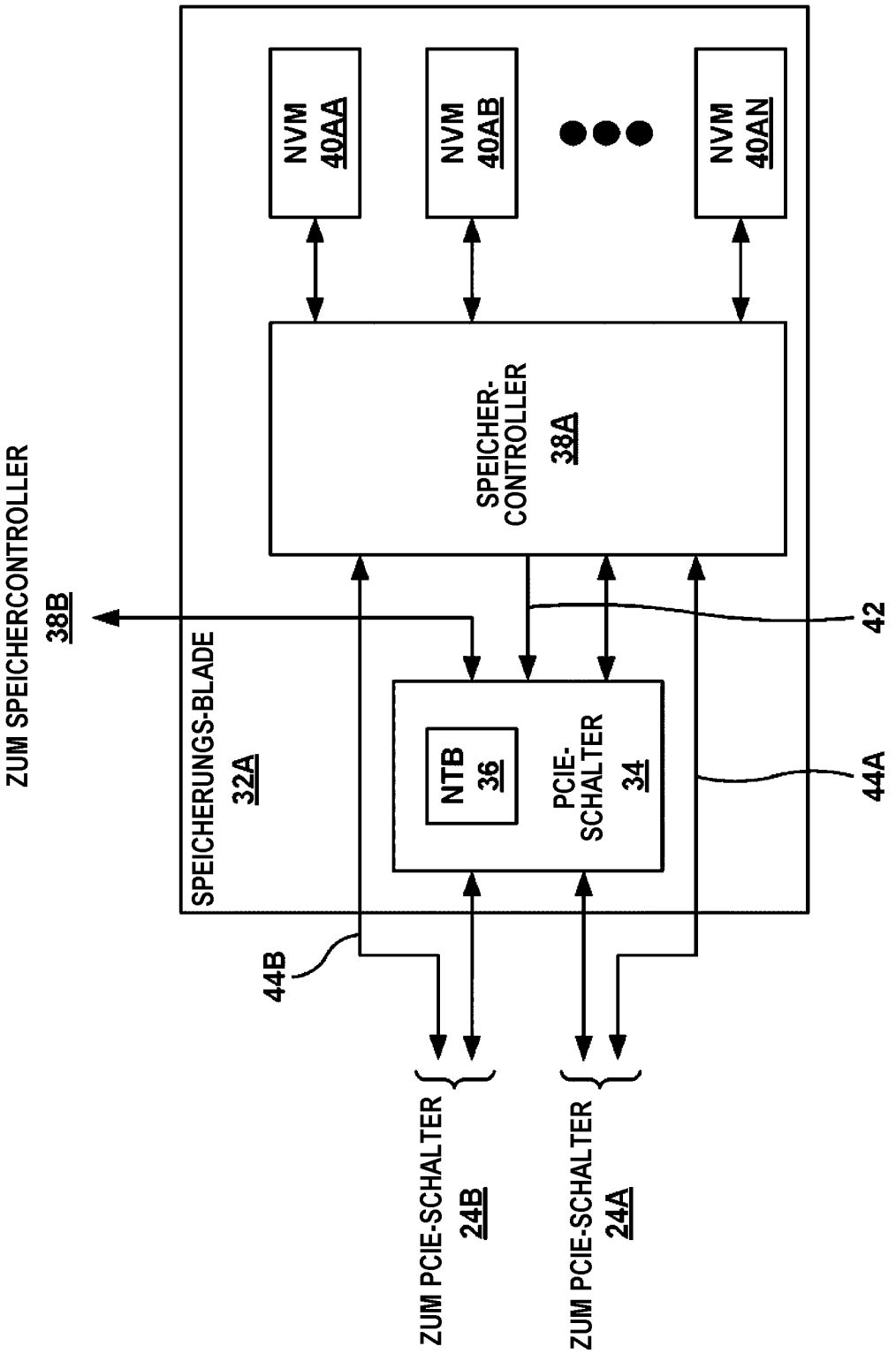

NVM 40AA
NVM 40AB
NVM 40AN
SPEICHER-CONTROLLER 38A
ZUM SPEICHERCONTROLLER 38B
SPEICHERUNGS-BLADE 32A
NTB 36
PCIE-SCHALTER 34
42
44A
44B
ZUM PCIE-SCHALTER 24B
ZUM PCIE-SCHALTER 24A

FIG. 3

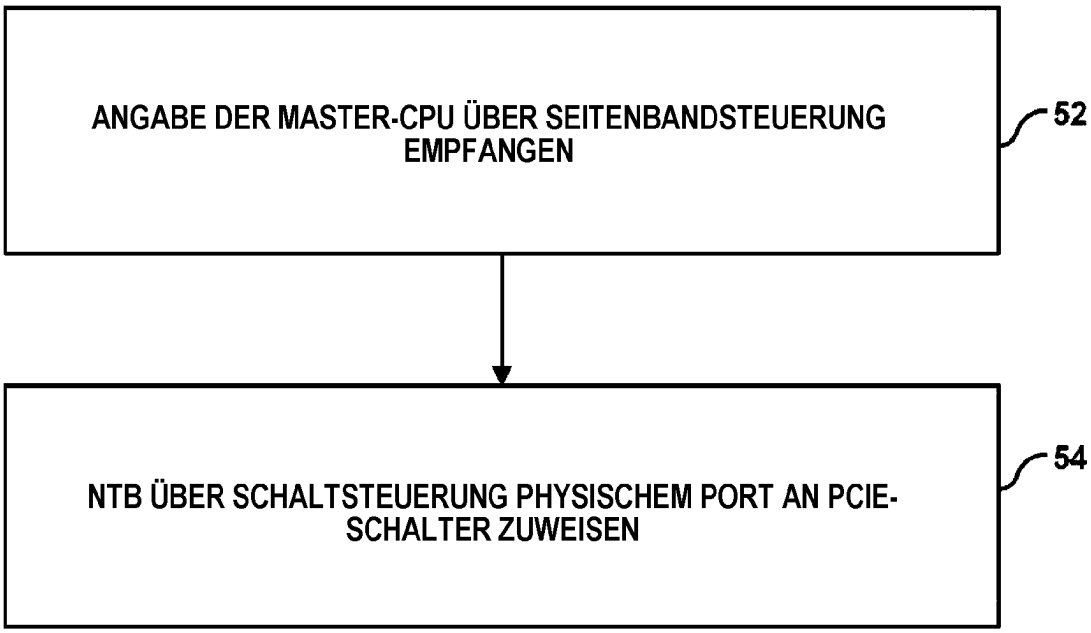

ANGABE DER MASTER-CPU ÜBER SEITENBANDSTEUERUNG EMPFANGEN
52
NTB ÜBER SCHALTSTEUERUNG PHYSISCHEM PORT AN PCIE-SCHALTER ZUWEISEN
54

FIG. 4