(19) **Bundesrepublik Deutschland**
Deutsches Patent- und Markenamt

(10) **DE 699 19 584 T2** 2005.08.11

(12) # Übersetzung der europäischen Patentschrift

(97) **EP 1 145 131 B1**
(21) Deutsches Aktenzeichen: **699 19 584.5**
(86) PCT-Aktenzeichen: **PCT/US99/12605**
(96) Europäisches Aktenzeichen: **99 928 406.0**
(87) PCT-Veröffentlichungs-Nr.: **WO 99/066416**
(86) PCT-Anmeldetag: **04.06.1999**
(87) Veröffentlichungstag der PCT-Anmeldung: **23.12.1999**
(97) Erstveröffentlichung durch das EPA: **17.10.2001**
(97) Veröffentlichungstag der Patenterteilung beim EPA: **18.08.2004**
(47) Veröffentlichungstag im Patentblatt: **11.08.2005**

(51) Int Cl.[7]: **G06F 13/40**

(30) Unionspriorität:
94847 **15.06.1998** **US**

(84) Benannte Vertragsstaaten:
DE, GB, IE

(73) Patentinhaber:
Sun Microsystems, Inc., Santa Clara, Calif., US

(74) Vertreter:
Dr. Weber, Dipl.-Phys. Seiffert, Dr. Lieke, 65183 Wiesbaden

(72) Erfinder:
ROWLINSON, Stephen, Reading, GB; OYELAKIN, A., Femi, Hayes, GB; WILLIAMS, J., Emrys, Milton Keynes, GB; GARNETT, J., Paul, Merseyside WA12 9PW, GB

(54) Bezeichnung: **BETRIEBSMITTELSTEUERUNG IN EINER DATENVERARBEITUNGSANLAGE**

Beschreibung

HINTERGRUND DER ERFINDUNG

[0001] Die vorliegende Erfindung bezieht sich auf eine Ressourcensteuerung in einem Computersystem. Die Erfindung kann beispielsweise in einem Multi-Prozessorsystem angewendet werden, in welchem erste und zweite Verarbeitungssätze (die jeweils einen oder mehrere Prozessoren aufweisen können) in Kommunikationsverbindung mit einem I/O-Gerätebus stehen. Die Erfindung kann insbesondere, jedoch nicht ausschließlich, auf fehlertolerante Computersysteme angewendet werden, in welchen zwei oder mehr Verarbeitungssätze in einem schrittweise eng gekoppelten (Lockstep-) Betrieb mit einem I/O-Gerätebus kommunizieren müssen.

[0002] In derartigen fehlertoleranten Computersystemen besteht ein Ziel darin, nicht nur in der Lage zu sein, Fehler zu identifizieren, sondern auch eine Struktur bereitzustellen, die in der Lage ist, ein hohes Maß an Systemverfügbarkeit und Systemwiderstandsfähigkeit gegenüber internen oder externen Störungen bereitzustellen. Um hohe Niveaus der Systemwiderstandsfähigkeit gegenüber internen Störungen bereitzustellen, wie z.B. gegenüber dem Ausfall eines I/O-Gerätes, wäre es wünschenswert, wenn solche Systeme automatisch den Zugriff auf und von jeder Einrichtung automatisch kontrollieren, die anscheinend Probleme verursacht bzw. Probleme verursachen könnte.

[0003] Automatische Zugangssteuerung bringt beträchtliche technische Herausforderungen mit sich, insofern, als das System nicht nur die in Frage kommenden Geräte überwachen muß, um Fehler zu erfassen, sondern auch eine Umgebung bereitstellen muß, in welcher das System als Ganzes trotz eines Ausfalls einer oder mehrerer der Systemkomponenten weiterhin arbeiten kann.

[0004] Dementsprechend besteht ein Ziel der vorliegenden Erfindung darin, diese technischen Probleme anzugehen.

[0005] Aspekte der Erfindung führen außerdem zu beträchtlichen Vorteilen, wenn sie in nicht fehlertoleranten Computersystemen mit mehreren Prozessoren verwendet werden, in welchen die Verarbeitungssätze unabhängig arbeiten. In dieser Situation können jedem unabhängig arbeitenden Verarbeitungssatz Systemeinrichtungen zugeordnet werden, die an einen gemeinsamen Bus angeschlossen sind. Diese Anordnung ermöglicht es, daß die Architektur eines Mehrprozessor-Computersystems vereinfacht wird, und sie wird ermöglicht durch Bereitstellung einer Brücke, die zwischen Verarbeitungssätzen für die Benutzung der Systemeinrichtungen vermittelt, während sie außerdem für die Verarbeitungssätze Informationen bezüglich der Geräte, sofern vorhanden, die für den Gebrauch verfügbar sind.

ZUSAMMENFASSUNG DER ERFINDUNG

[0006] Besondere und bevorzugte Aspekte der Erfindung sind in den beigefügten unabhängigen und abhängigen Ansprüchen dargestellt.

[0007] Gemäß einem Aspekt der Erfindung wird eine Brücke für ein Computersystem bereitgestellt. Die Brücke weist zumindest einen ersten Verarbeitungssatz und einen zweiten Verarbeitungssatz auf, die jeweils über einen I/O-Bus mit der Brücke verbunden sind. Ein Ressourcensteuermechanismus in der Brücke weist eine Schnittstelle für den Austausch von Signalen mit einem oder mehreren Ressourcen-Steckplätzen eines Gerätebusses auf, der mit der Brücke verbindbar ist, wobei jeder der Ressourcensteckplätze in der Lage ist, mit einer Systemressource zu kommunizieren, wobei die Brücke weiterhin ein Register aufweist, welches jeder der Systemressourcen zugeordnet ist, wobei das Register umschaltbare Kennzeichen hat, die einen Betriebszustand der zugehörigen Systemressource kennzeichnen, wobei der Steuermechanismus derart betrieben werden kann, daß er Signale zu und/oder von entsprechenden Systemressourcen des Computersystems leitet. Zumindest eine der Ressourcen kann eine I/O-Einrichtung sein. Das Register kann Speicher zum Speichern der Kennzeichen aufweisen. Das Register kann eine 4-Bit-Speichereinheit aufweisen.

[0008] Das Computersystem kann zwei Verarbeitungssätze aufweisen, wobei jeder der Verarbeitungssätze einen oder mehrere Prozessoren aufweist. Zumindest einige der umschaltbaren Kennzeichnen können verwendet werden, um anzuzeigen, ob die zugehörige Ressource einem der Verarbeitungssätze zugeordnet ist. Die zumindest einigen der umschaltbaren Kennzeichen können zusätzlich anzeigen, welchem der Verarbeitungssätze die dem Register zugehörige Ressource zugeordnet worden ist.

[0009] Das Register kann eine 4-Bit-Speichereinheit aufweisen und das Computersystem kann einen ersten Verarbeitungssatz und einen zweiten Verarbeitungssatz aufweisen, wobei ein zweites Bit und ein drittes Bit der Einheit umschaltbar sind, um anzuzeigen, ob die zugehörige Ressource im Besitz des ersten Verarbeitungssatzes, des zweiten Verarbeitungssatzes oder weder im Besitz des ersten Verarbeitungssatzes noch im Besitz des zweiten Verarbeitungssatzes ist.

[0010] Zumindest eines der umschaltbaren Kennzeichen kann anzeigen, ob die zugehörige Ressource Zugriff auf einen der Verarbeitungssätze hat oder nicht. Zumindest eines der umschaltbaren Kennzeichen kann wahlweise so betreibbar sein, daß es unbestimmte Daten erzeugt, wenn ein Lesen der dem Register zugehörigen Ressource versucht wird.

[0011] Ein weiterer Aspekt der Erfindung stellt eine Brücke für ein Computersystem bereit. Das Computersystem weist zumindest zwei Verarbeitungssätze und eine Routingmatrix auf. Ein Ressourcensteuermechanismus in der Brücke weist ein Register auf, das eine Mehrzahl umschaltbarer Kennzeichen hat, die jeweils einen Betriebszustand einer zugehörigen Ressource des Computersystems anzeigen. Die Routingmatrix ist in der Lage, Befehle und/oder Daten zu und von einer Ressource zu leiten, welche durch zumindest eines der Identifikationskennzeichen als im Besitz eines der Verarbeitungssätze befindlich identifiziert wird.

[0012] Ein weiterer Aspekt der Erfindung stellt eine Brücke für ein Computersystem bereit. Das Computersystem weist einen ersten Verarbeitungssatz und einen zweiten Verarbeitungssatz auf. Ein Ressourcensteuermechanismus in der Brücke weist ein 4-Bit-Register auf, wobei 2 der 4 Bits umschaltbar sind zu oder von: einem ersten Zustand, welcher anzeigt, daß die zu dem Register gehörige Ressource weder dem ersten Verarbeitungssatz noch dem zweiten Verarbeitungssatz zugeordnet ist, einen zweiten Zustand, welcher anzeigt, daß die zu dem Register gehörige Ressource dem ersten Verarbeitungssatz zugeordnet ist, und einen dritten Zustand, welcher anzeigt, daß die dem Register zugehörige Ressource dem zweiten Verarbeitungssatz zugeordnet ist.

[0013] Ein erstes Bit der verbleibenden zwei Bits in dem 4-Bit-Register kann umschaltbar sein zu oder von: einem ersten Zustand, welcher anzeigt, daß die zugehörige Ressource Zugriff auf einen der ersten und zweiten Verarbeitungssätze erhalten hat, und einem zweiten Zustand, welcher anzeigt, daß die zugehörige Ressource weder auf die ersten noch auf die zweiten Verarbeitungssätze Zugriff erhalten, wobei der erste Zustand nur ausgelöst wird, wenn die 2 der 4 Bits des Registers anzeigen, daß die zugehörige Ressource im Besitz entweder des ersten oder des zweiten Verarbeitungssatzes ist.

[0014] Ein zweites Bit der verbleibenden 2 Bits in dem 4-Bit-Register kann umschaltbar sein zu und/oder von: einem ersten Zustand, welcher anzeigt, daß Schreibevorgänge in die zugehörige Ressource erlaubt sind, und einem zweiten Zustand, welcher anzeigt, daß Schreibvorgänge in die zugehörige Ressource ignoriert werden sollen und daß unbestimmte Daten in Reaktion auf irgendeinen Schreibversuch in die zugehörige Ressource erzeugt werden.

[0015] Ein weiterer Aspekt der Erfindung stellt ein Verfahren zum Steuern der Ressourcen eines Computersystems bereit. Das Verfahren weist die Schritte auf, daß ein Register in einer Brücke des Computersystems bereitgestellt wird, wobei das Register eine Mehrzahl von Indizes hat, die so umschaltbar sind, daß sie die Betriebszustände einer Ressource, welche zu einem Steckplatz eines Busses gehört, der mit der Brücke verbunden ist, anzeigt, und daß die Ressource entsprechend den Betriebszuständen gesteuert wird, welche durch die Kennzeichen des Registers angezeigt werden. Das Register kann in einem Speicher mit wahlfreiem Zugriff (RAM) implementiert sein. Das Register kann ein 4-Bit-Register aufweisen. Das Verfahren kann den Schritt aufweisen, daß das Register aktualisiert wird, um Veränderungen im Betriebszustand der Ressource wiederzugeben.

[0016] Ein weiterer Aspekt der Erfindung stellt ein Verfahren für das Leiten (Routing) von Signalen von einem ersten Verarbeitungssatz oder einem zweiten Verarbeitungssatz zu zumindest einer Ressource bereit, die an einem Ressourcenbus vorgesehen ist, welcher wahlweise mit dem ersten oder dem zweiten Verarbeitungssatz verbindbar ist. Das Verfahren weist die Schritte auf: Aussenden von Signalen von den ersten und zweiten Verarbeitungssätzen, wobei die Signale für eine Ressource des Computersystems bestimmt sind, Abfragen eines Registers, um festzustellen, ob der eine der ersten und zweiten Verarbeitungssätze Zugriffsrechte auf die Ressource hat, und Lenken der Signale zu der Ressource, wenn das Register anzeigt, daß Zugang zu der Ressource für den einen der ersten und zweiten Verarbeitungssätze gewährt ist.

[0017] Ein weiterer Aspekt der Erfindung stellt ein Computersystem bereit, welches aufweist: eine Mehrzahl

von Verarbeitungssätzen, die jeweils einen oder mehrere Prozessoren haben und die jeweils mit einem Prozessorbus verbunden sind, eine Mehrzahl von Geräten, denen jeweils ein Steckplatz eines I/O-Gerätebusses zugeordnet ist, und eine Brücke, die mit der Mehrzahl von Prozessorbussen und mit dem I/O-Gerätebus verbunden ist, wobei die Brücke einen Gerätesteuermechanismus aufweist, der eine Schnittstelle für den Austausch von Signalen mit einem oder mehreren der Steckplätze und den zugehörigen Geräten bzw. Einrichtungen hat, und ein jedem Gerät bzw. jeder Einrichtung zugeordnetes Register, wobei das Register umschaltbare Kennzeichen hat, die einen Betriebszustand der zugehörigen Einrichtung bzw. des zugehörigen Gerätes anzeigen, wobei der Steuermechanismus so betreibbar ist, daß er im Gebrauch Signale zu und/oder von entsprechenden Systemressourcen des Computersystems lenkt.

[0018] Ein weiterer Aspekt der Erfindung stellt eine Brücke für ein Mehrprozessor-Computersystem bereit. Die Brücke weist auf: einen Ressourcensteuermechanismus, der eine Schnittstelle für den Austausch von Signalen mit einem oder mehreren Ressourcensteckplätzen hat, wobei jeder der Ressourcensteckplätze in der Lage ist, mit einer Systemressource zu kommunizieren, und ein Register, das jeder Systemressource zugeordnet ist, wobei das Register umschaltbare Kennzeichen hat, die einen Betriebszustand der zugehörigen Systemressource anzeigen, wobei der Steuermechanismus derart betreibbar ist, daß er im Gebrauch Signale zu und/oder von entsprechenden Systemressourcen des Computersystems leitet.

KURZE BESCHREIBUNG DER FIGUREN

[0019] Beispielhafte Ausführungsformen der vorliegenden Erfindung werden im folgenden nur beispielhaft beschrieben, wobei auf die beiliegenden Zeichnungen Bezug genommen wird, in welchen gleiche Bezugszeichen sich auf gleiche Elemente beziehen und in welchen:

[0020] Fig. 1 eine schematische Übersicht eines fehlertoleranten Computersystems ist, welches eine Ausführungsform der Erfindung beinhaltet,

[0021] Fig. 2 eine schematische Übersicht einer speziellen Implementierung eines Systems ist, welches auf dem von Fig. 1 beruht,

[0022] Fig. 3 eine schematische Wiedergabe einer Implementierung eines Verarbeitungssatzes ist,

[0023] Fig. 4 eine schematische Wiedergabe eines anderen Beispiels eines Verarbeitungssatzes ist,

[0024] Fig. 5 eine schematische Wiedergabe eines weiteren Verarbeitungssatzes ist,

[0025] Fig. 6 ein schematisches Blockdiagramm einer Ausführungsform einer Brücke für das System nach Fig. 1 ist,

[0026] Fig. 7 ein schematisches Blockdiagramm eines Speichers für die Brücke nach Fig. 6 ist,

[0027] Fig. 8 ein schematisches Blockdiagramm einer Steuerlogik für die Brücke nach Fig. 6 ist,

[0028] Fig. 9 eine schematische Wiedergabe einer Routingmatrix der Brücke nach Fig. 6 ist,

[0029] Fig. 10 eine beispielhafte Implementierung der Brücke nach Fig. 6 ist,

[0030] Fig. 11 ein Zustandsdiagramm ist, welches Betriebszustände der Brücke nach Fig. 6 veranschaulicht,

[0031] Fig. 12 ein Flußdiagramm ist, welches Stufen in dem Betrieb der Brücke nach Fig. 6 veranschaulicht,

[0032] Fig. 13 eine Einzelheit einer Betriebsstufe aus Fig. 12 ist,

[0033] Fig. 14 das Anordnen bzw. Vorbringen von I/O-Zyklen in dem System nach Fig. 1 veranschaulicht,

[0034] Fig. 15 die in einem Puffer für anstehende Schreibvorgänge gespeicherten Daten veranschaulicht,

[0035] Fig. 16 eine schematische Wiedergabe eines Steckplatzantwortregisters ist,

[0036] Fig. 17 eine Schreibstufe für verschiedenartige Daten veranschaulicht,

[0037] Fig. 18 eine Modifikation von Fig. 17 veranschaulicht,

[0038] Fig. 19 eine Lesestufe für verschiedenartige Daten veranschaulicht,

[0039] Fig. 20 eine alternative Lesestufe für verschiedenartige Daten veranschaulicht,

[0040] Fig. 21 ein Flußdiagramm ist, welches die Betriebsweise des Schreibmechanismus für verschiedenartige Daten zusammenfaßt,

[0041] Fig. 22 ein schematisches Blockdiagramm ist, welches eine Vermittlung innerhalb des Systems nach Fig. 1 erläutert,

[0042] Fig. 23 ein Zustandsdiagramm ist, welches die Betriebsweise eines Gerätebusvermittlers veranschaulicht,

[0043] Fig. 24 ein Zustandsdiagramm ist, welches die Betriebsweise eines Brückenvermittlers veranschaulicht,

[0044] Fig. 25 ein Zeitablaufdiagramm für PCI-Signale ist,

[0045] Fig. 26 ein schematisches Diagramm ist, welches die Betriebsweise der Brücke nach Fig. 6 für einen direkten Speicherzugriff veranschaulicht,

[0046] Fig. 27 ein Flußdiagramm ist, welches ein Verfahren zum direkten Speicherzugriff in der Brücke nach Fig. 6 zeigt, und

[0047] Fig. 28 ein Flußdiagramm eines Reintegrationsprozesses ist, welcher die Überwachung eines "verunreinigten" RAMs umfaßt.

BESCHREIBUNG DER BEVORZUGTEN AUSFÜHRUNGSFORMEN

[0048] Fig. 1 ist eine schematische Übersicht eines fehlertoleranten Computersystems **10**, welches eine Mehrzahl von CPU-Sätzen (Verarbeitungssätzen) **14** und **16** sowie eine Brücke **12** aufweist. Wie in Fig. 1 dargestellt ist, gibt es zwei Verarbeitungssätze **14** und **16**, auch wenn in anderen Ausführungsformen möglicherweise drei oder mehr Verarbeitungssätze sein mögen. Die Brücke **12** bildet eine Schnittstelle zwischen den Verarbeitungssätzen und den I/O-Geräten, wie z.B. den Geräten **28**, **29**, **30**, **31** und **32**. In dem vorliegenden Dokument wird der Begriff "Verarbeitungssatz" verwendet, um eine Gruppe von einem oder mehreren Prozessoren zu bezeichnen, möglicherweise einschließlich Speicher, mit Ausgeben und Empfangen von gemeinsamen Ausgangsgrößen und Eingangsgrößen. Es versteht sich, daß der alternative Begriff, der oben erwähnt wurde, "CPU-Satz", stattdessen verwendet werden könnte, und daß diese Begriffe in der gesamten vorliegenden Druckschrift austauschbar verwendet werden könnten. Es versteht sich auch, daß der Term "Brücke" verwendet wird, um irgendein Gerät, eine Vorrichtung oder Anordnung zu bezeichnen, die für das Verbinden von zwei oder mehr Bussen desselben oder unterschiedlicher Typen geeignet ist.

[0049] Der erste Verarbeitungssatz **14** ist mit der Brücke **12** über einen ersten Verarbeitungssatz-I/O-Bus (PA-Bus) **24** verbunden, im vorliegenden Fall mit einem peripheren Component-Interconnect-Bus (PCI-Bus). Der zweite Verarbeitungssatz **16** ist mit der Brücke über einen I/O-Bus des zweiten Verarbeitungssatzes (PB-Bus) **26** desselben Typs wie im Falle des PA-Busses **24** verbunden (das heißt hier einem PCI-Bus). Die I/O-Geräte sind mit der Brücke **12** über einen Geräte-I/O-Bus (D-Bus) **22** verbunden, im vorliegenden Fall also einem PCI-Bus.

[0050] Auch wenn in dem speziell beschriebenen Beispiel die Busse **22**, **24** und **26** alle PCI-Busse sind, gilt dies nur für dieses Beispiel und in anderen Ausführungsformen können andere Busprotokolle verwendet werden, und der D-Bus **22** kann ein gegenüber dem PA-Bus und dem PB-Bus (P-Busse) **24** und **26** unterschiedliches Protokoll haben.

[0051] Die Verarbeitungssätze **14** und **16** sowie die Brücke **12** sind synchron unter der Steuerung eines gemeinsamen Taktes **20** zu betreiben, welcher mit Taktsignalleitungen **21** verbunden ist.

[0052] Einige der Geräte, die eine Ethernet- (E-Net-) Schnittstelle **28** und eine Kleincomputersystemschnitt-

stelle (SCSI-Schnittstelle) umfassen, sind dauerhaft mit dem Gerätebus **22** verbunden, andere I/O-Einrichtungen, wie z.B. die I/O-Einrichtungen **30**, **31** und **32**, können in individuell geschaltete Steckplätze **33**, **34** und **35** „heiß" (im Betrieb) eingefügt werden. Ein Umschalten von dynamischen Feldeffekttransistoren (FET) kann für die Schlitze bzw. Steckplätze **33**, **34** und **35** bereitgestellt werden, um das Einsetzen der Geräte, wie z.B. der Geräte **30**, **31** und **32**, freizuschalten. Die Bereitstellung der FETs ermöglicht eine Längenzunahme des D-Busses **22**, da nur diejenigen Einrichtungen, die aktiv sind, eingeschaltet werden, was die effektive Gesamtbuslänge vermindert. Es versteht sich, daß die Anzahl von I/O-Einrichtungen, die mit dem D-Bus **22** verbunden werden können, und die Anzahl von dafür vorgesehenen Steckplätzen gemäß einer bestimmten Implementierung in Übereinstimmung mit den speziellen Auslegungsanforderungen eingestellt werden können.

[0053] Fig. 2 ist eine schematische Übersicht einer besonderen Implementierung von fehlertoleranten Computern, die eine Brückenstruktur des in Fig. 1 dargestellten Typs verwenden. In Fig. 2 umfaßt das fehlertolerante Computersystem eine Mehrzahl (in diesem Fall vier) von Brücken **12** auf ersten und zweiten I/O-Hauptplatinen (MB40 und MB42), um die Anzahl von I/O-Geräten, die angeschlossen werden können, zu vergrößern, und auch um die Anzahl der I/O-Geräte, die verbunden werden können, zu reduzieren und um außerdem die Zuverlässigkeit und Redundanz zu verbessern. Demnach sind in der in Fig. 2 dargestellten Ausführungsform zwei Verarbeitungssätze **14** und **16** jeweils auf einem entsprechenden Steckboard **44** und **46** vorgesehen, wobei die Verarbeitungssatzplatinen **44** und **46** die I/O-Hauptplatinen MB40 und MB42 überbrücken. Eine erste Haupttaktquelle **20A** ist auf der ersten Hauptplatine **40** montiert und eine zweite Sklaventaktquelle **20B** ist auf der zweiten Hauptplatine **42** montiert. Taktsignale werden den Verarbeitungssatzplatinen **44** und **46** über entsprechende Verbindungen zugeführt (in Fig. 2 nicht dargestellt).

[0054] Die ersten und zweiten Brücken **12.1** und **12.2** sind auf der ersten I/O-Hauptplatine **40** montiert. Die erste Brücke **12.1** ist über Busse **24.1** bzw. **26.1** mit den Verarbeitungssätzen **14** bzw. **16** verbunden. In ähnlicher Weise ist die zweite Brücke **12.2** über Busse **24.2** bzw. **26.2** mit den Verarbeitungssätzen **14** und **16** verbunden. Die Brücke **12.1** ist mit einem I/O-Datenbus (D-Bus) **22.1** verbunden, und die Brücke **12.2** ist mit einem I/O-Datenbus (D-Bus) **22.2** verbunden.

[0055] Dritte und vierte Brücken **12.3** und **12.4** sind auf der zweiten I/O-Hauptplatine **42** montiert. Die Brücke **12.3** ist über Busse **24.3** bzw. **26.3** mit den Verarbeitungssätzen **14** bzw. **16** verbunden. In ähnlicher Weise ist die Brücke **12.4** über Busse **24.4** bzw. **26.4** mit den Verarbeitungssätzen **14** bzw. **16** verbunden. Die Brücke **12.3** ist mit einem I/O-Datenbus (D-Bus) **22.3** verbunden, und die Brücke **12.4** ist mit einem I/O-Datenbus (D-Bus **22.4**) verbunden.

[0056] Man erkennt, daß die in Fig. 2 dargestellte Anordnung es ermöglichen kann, daß eine große Anzahl von I/O-Einrichtungen mit den beiden Verarbeitungssätzen **14** und **16** über Datenbusse **22.1**, **22.2**, **22.3** und **22.4** verbunden werden kann, um entweder den Bereich der verfügbaren I/O-Einrichtungen zu erhöhen oder um einen höheren Grad an Redundanz bereitzustellen oder beides.

[0057] Fig. 3 ist eine schematische Übersicht einer möglichen Ausgestaltung eines Verarbeitungssatzes, wie z.B. des Verarbeitungssatzes **14** in Fig. 1. Der Verarbeitungssatz **16** könnte dieselbe Ausgestaltung haben. In Fig. 3 ist eine Mehrzahl von Prozessoren (in diesem Fall vier) **52** über einen oder mehrere Busse **54** mit einer Bussteuerung **50** des Verarbeitungssatzes verbunden. Wie in Fig. 3 dargestellt, sind mehrere Ausgangsbusse **24** des Verarbeitungssatzes mit der Bussteuerung **50** des Verarbeitungssatzes verbunden, wobei jeder Ausgangsbus **24** des Verarbeitungssatzes mit einer entsprechenden Brücke **12** verbunden ist. Beispielsweise würde in der Anordnung nach Fig. 4 nur ein I/O-Bus (P-Bus) **24** für den Verarbeitungssatz vorgesehen sein, wohingegen in der Anordnung gemäß Fig. 2 vier derartige I/O-Busse (P-Busse) **24** des Verarbeitungssatzes vorgesehen wären. Bei dem in Fig. 3 dargestellten Verarbeitungssatz **14** arbeiten individuelle Prozessoren unter Verwendung eines gemeinsamen Speichers **56** und sie empfangen Eingaben und liefern Ausgaben an dem gemeinsamen P-Bus (den gemeinsamen P-Bussen) **24**.

[0058] Fig. 4 ist eine alternative Ausgestaltung eines Verarbeitungssatzes, wie z.B. des Verarbeitungssatzes **14** nach Fig. 1. In diesem Fall ist eine Mehrzahl von Prozessor-/Speichergruppen **61** mit einem gemeinsamen internen Bus **64** verbunden. Jede Prozessor-/Speichergruppe **61** umfaßt einen oder mehrere Prozessoren **62** und einen zugehörigen Speicher **66**, der mit einem internen Gruppenbus **63** verbunden ist. Eine Schnittstelle **65** verbindet den internen Gruppenbus **63** mit dem gemeinsamen internen Bus **64**. Dementsprechend sind in der Anordnung, die in Fig. 4 dargestellt ist, individuelle Verarbeitungsgruppen jeweils mit jedem der Prozessoren **62** und dem zugehörigen Speicher **66** über einen gemeinsamen internen Bus **64** mit einer Bussteuerung **60** des Verarbeitungssatzes verbunden. Die Schnittstellen **65** ermöglichen, daß ein Prozessor **62** einer Verarbeitungsgruppe nicht nur mit den Daten in seinem lokalen Speicher **66**, sondern auch denjenigen in dem Spei-

cher einer anderen Verarbeitungsgruppe **61** innerhalb des Verarbeitungssatzes **14** arbeitet. Die Bussteuerung **60** des Verarbeitungssatzes stellt eine gemeinsame Schnittstelle zwischen dem gemeinsamen internen Bus **64** und den Verarbeitungssatz-I/O-Bussen (P-Bussen) **24** bereit, die mit der Brücke (den Brücken) **12** verbunden sind.

[0059] Es versteht sich, daß, obwohl nur zwei Verarbeitungsgruppen **61** in Fig. 4 dargestellt sind, eine solche Struktur selbstverständlich nicht auf diese Anzahl von Verarbeitungsgruppen beschränkt ist.

[0060] Fig. 5 veranschaulicht eine alternative Ausgestaltung eines Verarbeitungssatzes, wie z.B. des Verarbeitungssatzes **14** nach Fig. 1. Hier umfaßt ein einfacher Verarbeitungssatz einen einzelnen Prozessor **72** und einen zugehörigen Speicher **76**, der über einen gemeinsamen Bus **74** mit einer Bussteuerung **70** eines Verarbeitungssatzes verbunden ist. Die Bussteuerung **70** des Verarbeitungssatzes stellt eine Schnittstelle zwischen dem internen Bus **74** und den I/O-Bussen (P-Bussen) **24** für die Verbindung mit der Brücke (den Brücken) **12** bereit.

[0061] Dementsprechend erkennt man aus den Fig. 3, Fig. 4 und Fig. 5, daß der Verarbeitungssatz viele verschiedene Formen annehmen kann und daß die spezielle Wahl einer bestimmten Verarbeitungssatzstruktur auf Basis der Verarbeitungserfordernisse einer bestimmten Anwendung und des Ausmaßes der erforderlichen Redundanz erfolgen kann. In der folgenden Beschreibung wird angenommen, daß die oben erwähnten Verarbeitungssätze **14** und **16** eine Struktur haben, wie sie in Fig. 3 dargestellt ist, auch wenn es sich versteht, daß eine andere Form des Verarbeitungssatzes vorgesehen werden könnte.

[0062] Die Brücke(n) **12** ist bzw. sind in einer Anzahl von Betriebszuständen betreibbar. Diese Betriebsarten werden später noch genauer beschrieben. Um jedoch das allgemeine Verständnis der Struktur der Brücke zu fördern, werden hier zwei Betriebsarten kurz zusammengefaßt. In einem ersten kombinierten Betriebszustand ist eine Brücke **12** so betreibbar, daß sie Adressen und Daten zwischen den Verarbeitungssätzen **14** und **16** (über die PA- bzw. PB-Busse **24** bzw. **26**) und die Geräte (über den D-Bus **22**) leitet. In diesem kombinierten Betriebszustand werden I/O-Zyklen, die durch die Verarbeitungssätze **14** und **16** erzeugt werden, miteinander verglichen, um sicherzustellen, daß beide Verarbeitungssätze korrekt arbeiten. Vergleichsfehler erzwingen, daß die Brücke **12** in einen Fehlerbegrenzungsbetrieb übergeht (E-Zustand), in welchem die I/O (Eingabe/Ausgabe) verhindert und eine Diagnoseinformation gesammelt wird. In dem zweiten, aufgespaltenen Betrieb leitet die Brücke **12** Adressen und Daten und vermittelt diese von einem der Verarbeitungssätze **14** und **16** auf den D-Bus **22** und/oder auf den jeweils anderen der Verarbeitungssätze **16** bzw. **14**. Bei dieser Betriebsart sind die Verarbeitungssätze **14** und **16** nicht synchronisiert und es werden keine I/O-Vergleiche vorgenommen. DMA-Operationen sind ebenfalls in beiden Betriebsarten zulässig. Wie oben erwähnt, werden die verschiedenen Betriebsarten, einschließlich des kombinierten und aufgespaltenen Betriebs, später noch genauer beschrieben. Nunmehr folgt jedoch eine Beschreibung der Grundstruktur eines Beispiels der Brücke **12**.

[0063] Fig. 6 ist eine schematische funktionelle Übersicht über die Brücke **12** nach Fig. 1. Erste und zweite I/O-Busschnittstellen des Verarbeitungssatzes, die PA-Busschnittstelle **84** und die PB-Busschnittstelle **86**, sind mit PA- bzw. PB-Bussen **24** bzw. **26** verbunden. Eine Geräte-I/O-Busschnittstelle, die D-Busschnittstelle **82**, ist mit dem D-Bus **22** verbunden. Es versteht sich, daß die PA-, PB- und D-Busschnittstellen nicht als separate Bauteile ausgestaltet sein müssen, sondern auch in die anderen Elemente der Brücke integriert sein könnten. Dementsprechend erfordert im Zusammenhang mit dieser Beschreibung, dort, wo auf eine Busschnittstelle Bezug genommen wird, dies nicht das Vorhandensein eines speziellen getrennten Bauteiles, sondern lediglich die Fähigkeit der Brücke, den betreffenden Bus zu schalten, beispielsweise mit Hilfe von physikalischen oder logischen Brückenverbindungen für die Leitungen der betreffenden Busse.

[0064] Die Führung (im folgenden als Routingmatrix bezeichnet) **80** ist über einen ersten internen Pfad **94** mit der PA-Busschnittstelle **84** und über einen zweiten internen Pfad **96** mit der PB-Busschnittstelle **86** verbunden. Die Routingmatrix **80** ist weiterhin über einen dritten internen Pfad **92** mit der D-Busschnittstelle **82** verbunden. Die Routingmatrix **80** ist demnach in der Lage, eine I/O-Bustransaktionsleitung in beiden Richtungen zwischen den PA- und PB-Busschnittstellen **84** und **86** bereitzustellen. Sie ist auch in der Lage, ein Leiten bzw. Routing in beiden Richtungen zwischen einem oder beiden, der PA- und der PB-Busschnittstellen, und der D-Busschnittstelle **82** bereitzustellen. Die Routingmatrix **80** ist über einen weiteren internen Pfad **100** mit der Speichersteuerlogik **90** verbunden. Die Speichersteuerlogik **90** steuert den Zugriff auf Brückenregister **110** und auf einen Speicher mit Direktzugriff (SRAM) **126**. Die Routingmatrix **80** ist daher auch so betreibbar, daß sie ein Routing in beiden Richtungen zwischen den PA-, PB- und D-Busschnittstellen **84**, **86** und **82** und der Speichersteuerlogik **90** bereitstellt. Die Routingmatrix **80** wird durch die Brückensteuerlogik **88** über Steuerpfade **98** und **99** gesteuert. Die Brückensteuerlogik **88** reagiert auf Steuersignale, Daten und Adressen auf den internen Pfa-

den **93**, **95** und **97** und auch auf Taktsignale von der Taktleitung (den Taktleitungen) **21**.

[0065] In der Ausführungsform gemäß der vorliegenden Erfindung arbeitet jeder der P-Busse (PA-Bus **24** und PB-Bus **26**) unter einem PCI-Protokoll. Die Bussteuerungen **50** des Verarbeitungssatzes (siehe Fig. 3) arbeiten ebenfalls unter PCI-Protokoll. Dementsprechend liefern die PA- und PB-Busschnittstellen **84** und **86** jeweils die gesamte Funktionalität, die für eine kompatible Schnittstelle erforderlich ist, welche sowohl einen Master- als auch einen Slave-Betrieb für zu und von dem D-Bus **22** oder internen Speichern und Registern der Brücke in dem Speicherteilsystem **90** übertragene Daten bereitstellt. Die Busschnittstellen **84** und **86** können beim Übergang der Brücke in einen Fehlerzustand (E-State) oder beim Erfassen eines I/O-Fehlers diagnostische Information an interne Brückenstatusregister in dem Speicherteilsystem **90** liefern.

[0066] Die Gerätebusschnittstelle **82** führt die gesamte Funktionalität durch, die für eine mit PCI in Einklang stehende Master- und Slaveschnittstelle für das Übertragen von Daten zu und von den PA- und PB-Bussen **84** und **86** erforderlich ist. Der D-Bus **82** ist während Übertragungen durch direkten Speicherzugriff (DMA) beim Übergang in einen E-Zustand oder bei Erfassung eines I/O-Fehlers so betreibbar, daß er diagnostische Information für interne Statusregister in dem Speicherteilsystem **90** der Brücke bereitstellt.

[0067] Fig. 7 veranschaulicht genauer die Brückenregister **110** und den SRAM **124**. Die Speichersteuerlogik **90** ist über einen Pfad (z.B. einen Bus) **112** mit einer Anzahl von Registerbauteilen **114**, **116**, **118**, **120** verbunden. Die Speichersteuerlogik ist außerdem über einen Pfad (z.B. Bus) **128** mit dem SRAM **126** verbunden, in welchem eine angezeigte Schreibpufferkomponente **122** und eine Komponente **124** eines verunreinigten Speichers zugeordnet sind. Auch wenn eine bestimmte Konfiguration der Komponenten **114**, **116**, **118**, **120**, **122** und **124** in Fig. 7 dargestellt ist, können diese Komponenten auch auf andere Art und Weise ausgestaltet werden, wobei andere Komponenten als Bereiche eines gemeinsamen Speichers definiert sind (z.B. ein Speicher mit Direktzugriff, wie z.B. der SRAM **126**, bei welchem der Pfad **112**/**128** durch interne Adressierung der Bereiche des Speichers gebildet wird). Wie in Fig. 7 dargestellt, sind der Puffer für angekündigte Schreibvorgänge **122** und der dirty RAM **124** unterschiedlichen Bereichen des SRAM-Speichers **126** zugeordnet, wohingegen die Register **114**, **116**, **118** und **120** so ausgestaltet sind, daß sie von dem SRAM-Speicher getrennt sind.

[0068] Steuer- und Statusregister (CSRs) **114** bilden interne Register, die die Steuerung verschiedener Betriebsarten der Brücke ermöglichen, das Einfangen diagnostischer Information für einen E-Zustand und für I/O-Fehler ermöglichen, und den Zugriff des Verarbeitungssatzes auf PCI-Steckplätze und -Geräte steuern, die mit dem D-Bus **22** verbunden sind. Diese Register werden durch Signale von der Routingmatrix **80** eingestellt.

[0069] Register für verschiedenartige Daten (DDRs) **116** liefern Orte für das Aufnehmen bzw. Enthalten verschiedenartiger Daten für verschiedene Verarbeitungssätze, um zu ermöglichen, daß nicht deterministische Datenereignisse gehandhabt werden können. Diese Register werden durch Signale von den PA- und PB-Bussen eingestellt.

[0070] Eine Brückendecodierlogik ermöglicht ein gemeinsames Schreiben, um einen Datenkomparator abzuschalten und ermöglicht ein Schreiben in zwei DDRs **116**, einen für jeden Verarbeitungssatz **14** und **16**.

[0071] Ein ausgewählter der DDRs kann dann synchron durch die Verarbeitungssätze **14** und **16** eingelesen werden. Die DDRs liefern damit einen Mechanismus, der es ermöglicht, daß eine Stelle von einem Verarbeitungssatz (**14**, **16**) zu einem anderen (**16**, **14**) reflektiert wird.

[0072] Steckplatzreaktionsregister (SRR) **118** bestimmen die Besitzverhältnisse von Gerätesteckplätzen auf dem D-Bus **22** und erlauben es, daß DMA zu dem passenden Verarbeitungssatz (Verarbeitungssätzen) geleitet wird. Diese Register sind mit der Adreßdecodierlogik verknüpft.

[0073] Trennregister **120** werden verwendet für die Speicherung von Datenphasen eines I/O-Zyklus, der abgebrochen wird, während sich Daten auf dem Weg zu einem anderen Bus in der Brücke befinden. Die Trenn- bzw. Abschaltregister **120** empfangen alle in einer Schlange in der Brücke befindlichen Daten, wenn ein Zielgerät eine Transaktion unterbricht, oder wenn der E-Zustand erfaßt wird. Diese Register sind mit der Routingmatrix **80** verbunden. Die Routingmatrix kann bis zu drei Datenworte und Bytefreigaben in einer Schlange halten. Vorausgesetzt, daß die anfänglichen Adressen als gleich bewertet wurden, leiten Adreßzielsteuerungen Adressen ab, die um einen Schritt heraufgesetzt werden, wenn Daten zwischen der Brücke und dem Bestimmungsort (oder Ziel) ausgetauscht werden. Wenn in einem Schreibvorgang (beispielsweise ein I/O-Schreiben durch einen Prozessor oder ein DVMA (Zugriff vom D-Bus auf P-Bus)) Daten in ein Ziel (Zieladresse) geschrieben werden, können diese Daten in der Brücke eingefangen werden, wenn ein Fehler auftritt. Dementspre-

chend werden diese Daten in den Trenn- bzw. Abschaltregistern **120** gespeichert, wenn ein Fehler auftritt. Auf diese Abschaltregister kann dann bei einer Reparatur aus einem E-Zustand zugegriffen werden, um die zu dem Schreibe- oder Lesezyklus gehörigen Daten wiederzugewinnen, die unterwegs waren, als der E-Zustand ausgelöst wurde.

[0074] Auch wenn die DDRs **116**, die SRRs **118** und die Abschaltregister getrennt dargestellt sind, können sie einen integralen Teil der CSRs **114** bilden.

[0075] Die E-Zustand- und Fehler-CSRs **114**, die für das Einfangen eines fehlerhaften Zyklus auf den P-Bussen **24** und **26** vorgesehen sind, umfassen eine Anzeige der fehlerhaften Daten. Im Anschluß an den Übergang in einen E-Zustand werden alte Schreibvorgänge, die für die P-Busse ausgelöst wurden, in dem Puffer **122** für angekündigte Schreibvorgänge protokolliert. Dies können andere Schreibvorgänge sein, als diejenigen, die in den Bussteuerungen **50** des Verarbeitungssatzes angekündigt worden waren oder die durch Software ausgelöst wurden, bevor eine E-Zustandsunterbrechung veranlaßt hat, daß die Prozessoren das Ausführen von Schreibvorgängen auf den P-Bussen **24** und **26** stoppen.

[0076] Ein Speicher für "verunreinigte" Daten **124** (dirty RAM **124**) wird verwendet, um anzuzeigen, welche Seiten des Hauptspeichers **56** der Verarbeitungssätze **14** und **16** durch direkte Speicherzugriffs- (DMA-) Transaktionen von einem oder mehreren Geräten auf dem D-Bus **22** modifiziert worden sind. Jede Seite (beispielsweise jede 8K-Seite) wird durch ein einzelnes Bit in dem dirty RAM **124** markiert, welches dann gesetzt wird, wenn ein DMA-Schreibzugriff auftritt und wieder gelöscht wird durch ein Lesen und einen Löschzyklus, der auf dem dirty RAM **124** ausgelöst wird durch einen Prozessor **52** eines Verarbeitungssatzes **14** und **16**.

[0077] Der dirty RAM **124** und der Puffer **122** für angekündigte Schreibvorgänge können beide in dem dirty RAM **124** in der Brücke **12** zugeordnet sein. Auf diesen Speicherraum kann für Testzwekke während normaler Lese- und Schreibzyklen zugegriffen werden.

[0078] Fig. 8 ist eine schematische Funktionsübersicht über die Brückensteuerlogik **88**, die in Fig. 6 dargestellt ist. Alle mit dem D-Bus **22** verbundenen Geräte sind geographisch adressiert. Dementsprechend führt die Brücke eine Decodierung aus, die erforderlich ist, um die isolierenden FETs für jeden Steckplatz (Schlitz) freizuschalten, bevor ein Zugriff auf diese Steckplätze ausgelöst wird.

[0079] Die Adreßdecodierung, die durch die Adreßdecodierlogik **136**, **138** ausgeführt wird, ermöglicht im wesentlichen vier grundlegende Zugriffstypen:

– nicht synchronisierter Zugriff (das heißt nicht in dem kombinierten Betriebszustand) durch einen Verarbeitungssatz (z.B. den Verarbeitungssatz **14** nach Fig. 1) oder den anderen Verarbeitungssatz (z.B. den Verarbeitungssatz **16** nach Fig. 1), wobei in diesem Fall der Zugriff von der PA-Busschnittstelle **84** zu der PB-Busschnittstelle **86** geleitet wird;
– Zugriff durch einen der Verarbeitungssätze **14** und **16** in dem aufgespaltenen Betriebszustand oder von beiden Verarbeitungssätzen **14** und **16** in dem kombinierten Betriebszustand auf ein I/O-Gerät auf dem D-Bus **22**, wobei in diesem Fall der Zugriff über die D-Busschnittstelle **82** geleitet wird;
– ein DMA-Zugriff durch ein Gerät auf dem D-Bus **22**, auf einen oder beide der Verarbeitungssätze **14** und **16**, welcher im kombinierten Betriebszustand an beide Verarbeitungssätze **14** und **16** gerichtet würde oder im nicht synchronisierten Betrieb zu dem relevanten Verarbeitungssatz **14** oder **16**, und, in einem aufgespaltenen Betrieb, auf einen Verarbeitungssatz **14** oder **16**, der einen Schlitz bzw. Steckplatz aufweist, an welchem das Gerät lokalisiert ist; und
– ein PCI-Konfigurationszugriff auf Geräte in I/O-Steckplätzen.

[0080] Wie oben erwähnt, wird eine geographische Adressierung verwendet. Demnach hat beispielsweise der Schlitz bzw. Steckplatz **0** auf der Hauptplatine A dieselbe Adresse, wenn darauf entweder durch den Verarbeitungssatz **14** oder den Verarbeitungssatz **16** Bezug genommen wird.

[0081] Eine geographische Adressierung wird in Kombination mit der FET-Umschaltung des PCI-Schlitzes verwendet. Während eines oben erwähnten Konfigurationszugriffs werden getrennte Geräteauswahlsignale für Geräte bereitgestellt, die nicht über einen FET isoliert sind. Ein einzelnes Geräteauswahlsignal kann für die geschalteten PCI-Schlitze bzw. -Steckplätze vorgesehen werden, während die FET-Signale verwendet werden können, um eine richtige Karte freizuschalten. Getrennte FET-Schaltleitungen sind für jeden Steckplatz vorgesehen, um die FETs für die Steckplätze getrennt zu schalten.

[0082] Die SRRs **118**, die in die CSR-Register **114** inkorporiert sein könnten, sind den Adreßdecodierfunktio-

nen zugeordnet. Die SRRs **118** dienen einer Anzahl verschiedener Rollen bzw. Funktionen, die später noch genauer beschrieben werden. Einige dieser Rollen bzw. Funktionen werden jedoch hier zusammengefaßt.

[0083] In einem kombinierten Betriebszustand kann jeder Steckplatz gesperrt werden, so daß Schreibvorgänge einfach bestätigt werden, ohne daß irgendeine Transaktion auf dem Gerätebus **22** erfolgt, wodurch die Daten verloren gehen. Lesevorgänge liefern sinnlose Daten, erneut ohne irgendeine Transaktion auf der Geräteplatine zu bewirken.

[0084] In dem gespaltenen Betriebszustand kann jeder Steckplatz sich in einem von drei Zuständen befinden. Die Zustände sind:

– nicht in Besitz;
– in Besitz durch Verarbeitungssatz A14;
– Besitz durch Verarbeitungssatz B16.

[0085] Auf einen Steckplatz, der nicht im Besitz eines Verarbeitungssatzes **14** oder **16** ist, welcher einen Zugriff macht (dies umfaßt nicht im Besitz befindliche Steckplätze) kann nicht zugegriffen werden. Dementsprechend wird ein solcher Zugriff abgewiesen.

[0086] Wenn ein Verarbeitungssatz **14** oder **16** abgeschaltet wird, bewegen sich alle Steckplätze, die in seinem Besitz waren, in den Zustand nicht im Besitz befindlich. Ein Verarbeitungssatz **14** oder **16** kann nur einen nicht im Besitz befindlichen Steckplatz beanspruchen, er kann einem anderen Verarbeitungssatz nicht den Besitz wegnehmen. Dies kann nur geschehen durch Abschalten des anderen Verarbeitungssatzes, oder indem man den anderen Verarbeitungssatz dazu bringt, den Besitz aufzugeben.

[0087] Auf die Besitzzustandsbits kann zugegriffen werden und sie können eingestellt werden, wenn man sich in dem kombinierten Betriebszustand befindet, sie haben jedoch keinen Effekt, bevor nicht ein Eintritt in den aufgespaltenen Zustand stattgefunden hat. Dies ermöglicht es, die Konfiguration eines aufgespaltenen Systems festzulegen, während man sich noch immer im kombinierten Betriebszustand befindet.

[0088] Jedem PCI-Gerät wird ein Bereich der Adressenkarte bzw. des Adressenfeldes des VErarbeitungssatzes zugeordnet. Die oberen Bits der Adresse werden durch den PCI-Steckplatz bzw. – Steckplatz bestimmt. Wenn das Gerät eine DMA ausführt, ist die Brücke in der Lage zu überprüfen, ob das Gerät die korrekte Adresse verwendet, weil ein D-Bus-Vermittler die Brücke darüber informiert, welches Gerät den Bus zu einem bestimmten Zeitpunkt benutzt. Wenn ein Gerätezugriff eine Adresse des Prozessorsatzes hat, die dafür nicht gültig ist, so wird der Gerätezugriff ignoriert. Es versteht sich, daß eine durch ein Gerät präsentierte Adresse eine virtuelle Adresse ist, die durch eine I/O-Speicherverwaltungseinheit in der Bussteuerung **50** des Verarbeitungssatzes in eine tatsächliche Speicheradresse übersetzt wird.

[0089] Die Adressen, die durch die Adreßdecoder ausgegeben werden, werden über die Initiator- und Zielsteuerungen **138** und **140** und über die Leitungen **98** unter der Steuerung einer Brückensteuerung **132** und eines Vermittlers **134** zu der Routingmatrix **80** geleitet.

[0090] Ein Vermittler **134** ist in verschiedenen unterschiedlichen Betriebsarten betreibbar, um die Benutzung der Brücke nach dem Prinzip "wer zuerst kommt, wird zuerst bedient" und unter Verwendung konventioneller PCI-Bussignale auf den P- und D-Bussen zu vermitteln.

[0091] In einem kombinierten Betriebszustand ist der Vermittler **134** so betreibbar, daß er zwischen den synchronisierten Verarbeitungssätzen **14** und **16** und irgendwelchen Auslösern auf dem Gerätebus **22** für die Benutzung der Brücke **12** vermittelt. Mögliche Szenarien sind:

– Zugriff eines Verarbeitungssatzes auf den Gerätebus **22**,
– Zugriff eines Verarbeitungssatzes auf die internen Register in der Brücke **12**,
– Gerätezugriff auf den Speicher **56** eines Verarbeitungssatzes.

[0092] In dem aufgespaltenen Betriebszustand müssen beide Verarbeitungssätze **14** und **16** die Verwendung der Brücke vermitteln und damit auch den Zugriff auf den Gerätebus **22** und die internen Brückenregister (z.B. CSR-Register **114**). Die Brücke **12** steht auch in Konkurrenz mit den Auslösern auf dem Gerätebus **22** für die Verwendung dieses Gerätebusses **22**.

[0093] Jeder Steckplatz auf dem Gerätebus hat ein ihm zugeordnetes Vermittlungsfreigabebit. Diese Vermittlungsfreigabebits werden nach einem Reset gelöscht und müssen gesetzt werden, um einem Steckplatz die

Anfrage an einen Bus zu erlauben. Wenn ein Gerät auf dem Gerätebus **22** in Verdacht steht, einen I/O-Fehler zu liefern, so wird das Vermittlungsfreigabebit von diesem Gerät durch die Brücke automatisch zurückgesetzt.

[0094] Eine PCI-Busschnittstelle in der Bussteuerung (den Bussteuerungen) **50** des Verarbeitungssatzes erwartet, die Hauptbussteuerung für den betroffenen P-Bus zu sein, das heißt sie enthält den PCI-Bus-Vermittler für den PA- oder PB-Bus, mit welchem sie verbunden ist. Die Brücke **12** kann nicht direkt den Zugriff auf die PA- und PB-Busse **24** und **26** kontrollieren. Für einen Zugriff auf den PA- oder PB-Bus steht die Brücke **12** mit dem Verarbeitungssatz auf dem betreffenden Bus in Konkurrenz und zwar unter der Steuerung der Bussteuerung **50** auf dem betreffenden Bus.

[0095] Weiterhin ist in Fig. 8 ein Komparator **130** und eine Brückensteuerung **132** dargestellt. Der Komparator **130** ist so betreibbar, daß er I/O-Zyklen von den Verarbeitungssätzen **14** und **16** vergleicht, um irgendwelche nicht synchronisierten Ereignisse festzustellen. Bei der Feststellung eines nicht synchronisierten Ereignisses ist der Komparator **130** so betreibbar, daß er die Brückensteuerung **132** veranlaßt, einen E-Zustand für die Analyse des nicht synchronen Ereignisses und für eine mögliche Reparatur desselben zu aktivieren.

[0096] Fig. 9 ist eine schematische Funktionsübersicht der Routingmatrix **80**.

[0097] Die Routingmatrix **80** weist einen Multiplexer **143** auf, der auf Auslösersteuersignale **98** von der Auslösersteuerung **138** nach Fig. 8 reagiert, um einen der folgenden Pfade, PA-Buspfad **94**, PB-Buspfad **96**, D-Buspfad **92** oder interner Buspfad **100** als den aktuellen Eingang zu der Routingmatrix auszuwählen. Getrennte Ausgangspuffer **144**, **145**, **146** und **147** sind für die Ausgabe an jeden der Pfade **94**, **96**, **92** und **100** vorgesehen, wobei diese Puffer wahlweise durch Signale **99** von der Zielsteuerung **140** nach Fig. 8 ausgewählt werden. Zwischen dem Multiplexer und den Puffern **144-147** werden Signale in einem Puffer **149** gehalten. In der vorliegenden Ausführungsform werden drei Datenzyklen für einen I/O-Zyklus in der Pipeline gehalten, die durch den Multiplexer **143**, den Puffer **149** und die Puffer **144** repräsentiert wird.

[0098] Zu den Fig. 6 bis Fig. 9 wurde eine funktionelle Beschreibung von Elementen der Brücke gegeben. Fig. 10 ist eine schematische Wiedergabe eines räumlich-körperlichen Aufbaus der Brücke, bei welchem die Brückensteuerlogik **88**, die Speichersteuerlogik **90** und die Brückenregister **110** in einem ersten feldprogrammierbaren Gatearray (FPGA) **89** implementiert sind, die Routingmatrix **80** in weiteren FPGAs **80.1** und **80.2** implementiert ist und der SRAM **126** in Form eines oder mehrerer getrennter SRAMs implementiert ist, die durch Adreßsteuerleitungen **127** adressiert werden. Die Busschnittstellen **82**, **84** und **86**, welche in Fig. 6 dargestellt sind, sind keine getrennten Elemente, sondern in den FPGAs **80.1**, **80**,**2** und **89** integriert. Die FPGAs **80.1** und **80.2** werden für die oberen 32 Bits 32-63 eines 64-Bit-PCI-Busses verwendet und die unteren 32 Bits 0-31 des 64-Bit-PCI-Busses. Es versteht sich, daß ein einzelner FPGA für die Routingmatrix **80** verwendet werden könnte, wenn die notwendige Logik innerhalb des Gerätes aufgenommen werden könnte. In der Tat könnten, wenn ein FPGA ausreichender Kapazität verfügbar ist, die Brückensteuerlogik, die Speichersteuerlogik und die Brückenregister in demselben FPGA inkorporiert werden wie die Routingmatrix. In der Tat können viele andere Konfigurationen ins Auge gefaßt werden und auch eine andere Technologie als FPGAs, beispielsweise einer oder mehrere anwendungsspezifische integrierte Schaltkreise (ASICs) können verwendet werden. Wie in Fig. 10 dargestellt, sind die FPGAs **89**, **80.1** und **80.2** sowie der SRAM **126** über interne Buspfade **85** und Pfadsteuerleitungen **87** miteinander verbunden.

[0099] Fig. 11 ist ein Übergangsdiagramm, welches die verschiedenen Betriebsarten der Brücke genauer illustriert. Der Brückenbetrieb kann aufgeteilt werden in drei grundlegende Betriebszustände, nämlich einen Fehlerzustand (EState oder E-Zustand), Betriebsart **150**, einen aufgespaltenen Betriebszustand **156** und einen kombinierten Betriebszustand **158**. Der EState-Betriebszustand **150** kann noch weiter in zwei Zustände aufgeteilt sein.

[0100] Nach dem anfänglichen Zurücksetzen beim Einschalten der Brücke oder im Anschluß an einen nicht synchronisierten Vorgang befindet sich die Brücke in dem anfänglichen E-Zustand **152**. In diesem Zustand werden alle Schreibvorgänge in dem Puffer **122** für angekündigte bzw. anstehende Schreibzugriffe gespeichert und Lesevorgänge aus den internen Brückenregistern (z.B. die CSR-Register **116**) sind zulässig, während alle anderen Lesevorgänge als Fehler behandelt werden (das heißt sie werden abgelehnt). In diesem Zustand führen die individuellen Verarbeitungssätze **14** und **16** Auswertungen für die Bestimmung einer erneuten Startzeit durch. Jeder Verarbeitungssatz **14** und **16** bestimmt den Zeittakt seines eigenen Neustart-Zeitgebers. Die Zeitgebereinstellung hängt von einem "Schuld"-Faktor für den Übergang in den E-Zustand ab. Ein Prozessorsatz, der festgestellt hat, daß er wahrscheinlich den Fehler verursacht hat, setzt für den Zeitgeber eine längere Zeit. Ein Verarbeitungssatz, der meint, daß es unwahrscheinlich sei, daß er den Fehler verursacht hat, setzt eine

kurze Zeit für den Zeitgeber. Der erste der Verarbeitungssätze **14** und **16**, dessen gesetzte Zeit abläuft, wird ein primärer Verarbeitungssatz. Dementsprechend bewegt sich die Brücke, wenn dies festgestellt worden ist, gemäß **153** in den primären E-Zustand **154**.

[0101] Wenn irgendein Verarbeitungssatz **14**/**16** der primäre Verarbeitungssatz geworden ist, so arbeitet die Brücke anschließend in dem primären E-Zustand **154**. Dieser Zustand ermöglicht es dem primären Verarbeitungssatz, daß er in die Brückenregister schreibt (insbesondere in die SRRs **118**). Andere Schreibvorgänge werden nicht mehr in dem Puffer für anstehende Schreibvorgänge gespeichert, sondern gehen einfach verloren. Lesevorgänge des Gerätebusses werden in dem primären E-Zustand **124** noch immer abgelehnt.

[0102] Sobald der Zustand E-State (E-Zustand) beseitigt ist, begibt sich die Brücke gemäß **155** in den aufgespaltenen Zustand **156**. In dem aufgespaltenen Zustand **156** wird ein Zugriff des Gerätebusses **52** durch die SRR-Register **118** gesteuert, während ein Zugriff auf den Brückenspeicher einfach vermittelt wird. Der primäre Status der Verarbeitungssätze **14** und **16** wird ignoriert. Der Übergang in einen kombinierten Betrieb wird erreicht mit Hilfe eines sync_reset (**157**) (Synchronisationsrückstellung). Nach der Ausgabe des sync_reset-Betriebes ist die Brücke dann in dem kombinierten Zustand **158** betreibbar, in welchem alle Lese- und Schreibzugriffe auf den D-Bus **22** und die PA- und PB-Busse **24** und **26** zulässig sind. Alle derartigen Zugriffe auf die PA- und PB-Busse **24** und **26** werden in dem Komparator **130** verglichen. Das Erfassen einer Fehlanpassung zwischen irgendwelchen Lese- und Schreibzyklen (mit Ausnahme von speziellen I/O-Zyklen für verschiedenartige Daten) bewirkt einen Übergang **151** in den E-Zustand **150**. Die verschiedenen beschriebenen Zustände werden durch die Brückensteuerung **132** gesteuert.

[0103] Die Rolle des Komparators **130** besteht darin, I/O-Operationen auf den PA- und PB-Bussen in dem kombinierten Zustand **158** zu überwachen und zu vergleichen und der Brückensteuerung **132** in Reaktion auf ein fehlangepaßtes Signal eine Meldung zu machen, wodurch die Brückensteuerung **132** den Übergang **151** in den Fehlerzustand **150** veranlaßt. Die I/O-Operationen können alle I/O-Operationen umfassen, die durch die Verarbeitungssätze ausgelöst werden, ebenso wie DMA-Transfers bezüglich eines DMA, der durch ein Gerät auf dem Gerätebus ausgelöst wird.

[0104] Die nachstehende Tabelle 1 faßt die verschiedenen Zugriffsvorgänge zusammen, die in jedem der Betriebszustände erlaubt sind.

TABELLE 1

	D-Bus-Lesen	D-Bus-Schreiben
E-Zustand	Master-Ablehnung	gespeichert im Puffer für angekündigte Schreibvorgänge
primärer E-Zustand	Master-Ablehnung	verloren
aufgespalten	gesteuert durch SRR-Bits und vermittelt	gesteuert durch SRR-Bits und vermittelt
kombiniert	zugelassen und verglichen	zugelassen und verglichen

[0105] Wie oben beschrieben, befindet sich nach einer anfänglichen Rückstellung das System in dem anfänglichen E-Zustand **152**. In diesem Zustand können weder der Verarbeitungssatz **14** noch der Verarbeitungssatz **16** auf den D-Bus **22** oder den P-Bus **26** oder **24** des anderen Verarbeitungssatzes **16** oder **14** zugreifen. Die internen Brückenregister **110** der Brücke sind zugänglich, jedoch nur für das Lesen.

[0106] Ein System, welches im kombinierten Betriebszustand **158** läuft, geht in den E-Zustand **150** über, wenn es einen Vergleichsfehler gibt, der in dieser Brücke erfaßt wurde oder es wird alternativ ein Vergleichsfehler erfaßt in einer anderen Brücke in einem Mehrfachbrückensystem, wie es beispielsweise in Fig. 2 dargestellt ist. Außerdem können Übergänge in einen E-Zustand **150** in anderen Situationen auftreten, beispielsweise im Falle eines softwaregesteuerten Ereignisses, welches Teil eines Selbstüberprüfungsvorganges bildet.

[0107] Beim Übergehen in den E-Zustand **150** wird allen oder einem Teilsatz der Prozessoren der Verarbeitungssätze über eine Interrruptleitung **95** ein Interrupt angezeigt. Im Anschluß daran führen alle I/O-Zyklen, die auf einem P-Bus **24** oder **26** erzeugt werden, dazu, daß Lesevorgänge zurückgegeben werden mit einer Aus-

nahme und daß Schreibvorgänge bzw. -anforderungen in dem Puffer für angekündigte bzw. anstehende Schreibvorgänge aufgezeichnet werden.

[0108] Die Betriebsweise des Komparators **130** wird nun genauer beschrieben. Der Komparator ist mit Pfaden **94**, **95**, **96** und **97** verbunden, um Adressen, Daten und ausgewählte Steuersignale von den PA- und PB-Busschnittstellen **84** und **86** zu vergleichen. Ein fehlgeschlagener Vergleich von synchronisierten Zugriffen auf Geräte des Geräte-I/O-Busses **22** bewirkt einen Übergang von dem kombinierten Zustand **158** in den E-Zustand **150**. Für I/O-Lesezyklen des Verarbeitungssatzes werden die Adresse, der Befehl, die Adreßparität, Bytefreigaben und Partitätsfehlerparameter verglichen.

[0109] Wenn der Vergleich während der Adressierphase fehlschlägt, startet die Brücke einen erneuten Versuch mit den Bussteuerungen **50** des Verarbeitungssatzes, was verhindert, daß Daten die I/O-Bussteuerungen **50** verlassen. In diesem Fall tritt keinerlei Aktivität auf dem Geräte-I/O-Bus **22** auf. Beim erneuten Versuch des Prozessors (der Prozessoren) wird kein Fehler gemeldet.

[0110] Wenn der Vergleich während einer Datenphase fehlschlägt (es werden nur Steuersignale und Bytefreischaltungen überprüft), so signalisiert die Brücke den Bussteuerungen des Verarbeitungssatzes bzw. der Verarbeitungssätze eine Zielablehnung. Den Prozessoren wird eine Fehlermeldung gegeben.

[0111] Im Falle von Schreibzyklen des Verarbeitungssatz-I/O-Busses, werden die Adreß-, Befehls-, Paritäts-, Bytefreigabe- und Datenparameter verglichen.

[0112] Wenn der Vergleich während der Adressierphase fehlschlägt, so unternimmt die Brücke einen erneuten Versuch mit den Bussteuerungen **50** des Verarbeitungssatzes, was dazu führt, daß die Bussteuerungen **50** des Verarbeitungssatzes den Zyklus erneut versuchen. Der Puffer **122** für anstehende Schreibvorgänge ist dann aktiv. Auf dem Geräte-I/O-Bus **22** tritt keine Aktivität auf.

[0113] Wenn der Vergleich während der Datenphase eines Schreibvorganges fehlschlägt, so werden keinerlei Daten an den D-Bus **22** weitergeleitet. Die fehlerhaften Daten und andere ÜBertragungsmerkmale von beiden Verarbeitungssätzen **14** und **16** werden in den Abschaltregistern **120** gespeichert und jegliche nachfolgende anstehende Schreibzyklen werden in dem Puffer **122** für angekündigte Schreibvorgänge aufgezeichnet.

[0114] Im Falle von Lesevorgängen mit einem direkten Zugriff auf einen virtuellen Speicher (DVMA) werden die Datensteuerung und die Parität für jeden Datenwert überprüft. Wenn die Daten nicht passen, so beendet die Brücke **12** die Übertragung auf dem P-Bus. Im Falle von DVMA-Schreibvorgängen werden die Steuer- und Paritätsfehlersignale auf Richtigkeit überprüft.

[0115] Andere Signale zusätzlich zu denjenigen, die oben speziell erwähnt wurden, können verglichen werden, um eine Anzeige einer Divergenz der Verarbeitungssätze zu liefern. Beispiele hierfür sind Busbewilligungen und verschiedene spezielle Signale während Transfers durch die Verarbeitungssätze und während DMA-Transfers.

[0116] Fehler fallen grob gesprochen in zwei Gruppen bzw. Typen, diejenigen, welche durch die Bussteuerung **50** des Verarbeitungssatzes für die Software sichtbar gemacht werden, und diejenigen, die durch die Bussteuerung **50** des Verarbeitungssatzes nicht sichtbar gemacht werden und demnach durch einen Interrupt von der Brücke **12** sichtbar gemacht werden müssen. Dementsprechend ist die Brücke so betreibbar, daß sie Fehler, die in Verbindung mit Lese- und Schreibzyklen der Verarbeitungssätze sowie DMA-Lesevorgängen und Schreibvorgängen berichtet wurden, einzufangen.

[0117] Die Taktsteuerung für die Brücke wird durch die Brückensteuerung **132** in Reaktion auf die Taktsignale von der Taktleitung **21** durchgeführt. Individuelle Steuerleitungen von der Steuerung **32** zu den verschiedenen Elementen der Brücke sind in den Fig. 6 bis Fig. 10 nicht dargestellt.

[0118] Fig. 12 ist ein Flußdiagramm, welches eine mögliche Folge von Betriebsstufen veranschaulicht, in welchen Verriegelungsschrittfehler während eines kombinierten Betriebszustandes erfaßt werden.

[0119] Stufe S1 entspricht der kombinierten Betriebsart, in welcher die Überprüfung auf einen Verriegelungsschrittfehler durch den in Fig. 8 dargestellten Komparator **130** durchgeführt wird.

[0120] In Stufe S2 wird angenommen, daß ein Verriegelungsschrittfehler durch den Komparator **130** erfaßt

worden ist.

[0121] In Stufe S3 wird der aktuelle Zustand in den CSR-Registern **114** gespeichert und anstehende Schreibvorgänge werden in dem Puffer **132** für angekündigte Schreibvorgänge und/oder in den Abschaltregistern **120** gesichert.

[0122] Fig. 13 veranschaulicht die Stufe S3 im Detail. Demgemäß erfaßt in der Stufe S31 die Brückensteuerung **132**, ob der durch den Komparator **130** gemeldete Verriegelungsschrittfehler während einer Datenphase aufgetreten ist, in welcher es möglich ist, Daten an den Gerätebus **22** zu leiten. In diesem Fall wird in Stufe S32 der Buszyklus beendet. Dann werden in Stufe S33 die Datenphasen in den Abschaltregistern **120** gespeichert und die Steuerung geht dann weiter in Stufe S35, wo eine Auswertung vorgenommen wird bezüglich der Frage, ob weitere I/O-Zyklen gespeichert werden müssen. Alternativ werden, wenn in der Stufe S31 festgestellt wird, daß der Verriegelungsschrittfehler nicht während einer Datenphase auftrat, die Adreß- und Datenphasen für jegliche anstehende Schreib-I/O-Zyklen in dem Puffer **122** für anstehende Schreibvorgänge gespeichert. In Stufe S34 werden, wenn es weitere anstehende und anhängige I/O-Schreibvorgänge gibt, diese ebenfalls in dem Puffer **122** für angekündigte bzw. anstehende Schreibvorgänge gespeichert.

[0123] Die Stufe S3 wird bei der Auslösung durch den anfänglichen Fehlerzustand **152** ausgeführt, der in Fig. 11 dargestellt ist. In diesem Zustand vermitteln die ersten und zweiten Verarbeitungssätze den Zugang zu der Brücke. Dementsprechend werden in der Stufe S31–S35 die anstehenden Schreibadreß- und Datenphasen für jeden der Verarbeitungssätze **14** und **16** in getrennten Bereichen des Puffers **122** für anstehende Schreibvorgänge gespeichert, und/oder in dem einzelnen Satz von Abschaltregistern, wie es oben beschrieben ist.

[0124] Fig. 14 veranschaulicht die Quelle der anstehenden Schreibe-I/O-Zyklen, die in dem Puffer **122** für anstehende Schreibvorgänge gespeichert werden müssen. Während des Normalbetriebs der Verarbeitungssätze **14** und **16** enthalten die Ausgangspuffer **162** in den individuellen Prozessoren I/O-Zyklen, die für eine Übertragung über die Bussteuerungen **50** eines Verarbeitungssatzes zu der Brücke **12** und eventuell auch zu dem Gerätebus **22** angekündigt wurden. In ähnlicher Weise enthalten auch die Puffer **160** in den Steuerungen **50** des Verarbeitungssatzes angekündigte I/O-Zyklen für die Übertragung über die Busse **24** und **26** zu der Brücke **12** und eventuell auch zu dem Gerätebus **22**.

[0125] Dementsprechend kann man sehen, daß dann, wenn ein Fehlerzustand auftritt, I/O-Schreibzyklen schon durch die Prozessoren **52** angekündigt sein können, entweder in ihren eigenen Puffern **162**, oder bereits an die Puffer **160** der Bussteuerungen **50** der Verarbeitungssätze übertragen worden sind. Es sind die I/O-Schreibzyklen in den Puffern **162** und **160**, die allmählich hindurch- und voranschreiten und die in dem Puffer **122** für angekündigte Schreibvorgänge gespeichert werden müssen.

[0126] Wie in Fig. 15 dargestellt, kann ein Schreibzyklus **164**, der dem Puffer **122** für angekündigte Schreibvorgänge angezeigt wurde, ein Adreßfeld **165** aufweisen, welches eine Adresse und einen Adreßtyp umfaßt, und zwischen einem und 16 Datenfeldern **166**, einschließlich eines Bytefreigabefeldes und der Daten selbst aufweist.

[0127] Die Daten werden in dem E-Zustand in den Puffer **122** für angekündigte Schreivorgänge geschrieben, wenn nicht der auslösende Verarbeitungssatz als ein primärer CPU-Satz gekennzeichnet worden ist. Zu diesem Zeitpunkt gehen nicht primäre Schreibvorgänge in einem E-Zustand dennoch in den Puffer für angekündigte Schreibvorgänge, selbst nachdem einer der CPU-Sätze ein primärer Verarbeitungssatz geworden ist. Ein Adreßzeiger in den CSR-Registern **114** zeigt auf die als nächste verfügbare Pufferadresse für einen angekündigten Schreibvorgang und liefert auch ein Überlaufbit, welches gesetzt wird, wenn die Brücke versucht, über den oberen Rand des Puffers für angekündigte Schreibvorgänge für irgendeinen der Verarbeitungssätze **14** und **16** zu schreiben. In der Tat werden gemäß der vorliegenden Implementierung nur die ersten 16K an Daten in jedem Puffer aufgezeichnet. Versuche, über den Rand des Puffers für angekündigte Schreibvorgänge hinauszuschreiben, werden ignoriert. Beim Reset kann der Wert des Zeigers des Puffers für angekündigte Schreibvorgänge gelöscht werden oder durch Software, die einen Schreibvorgang unter der Steuerung eines primären Verarbeitungssatzes verwendet.

[0128] Gemäß Fig. 12 versuchen die individuellen Verarbeitungssätze in Stufe S4 nach dem Sichern des Status und der angekündigten Schreibvorgänge, unabhängig den Fehlerzustand auszuwerten und zu bestimmen, ob einer der Verarbeitungssätze fehlerhaft ist. Diese Bestimmung wird durch die individuellen Prozessoren in einem Fehlerzustand vorgenommen, in welchem sie individuell den Status aus den Steuerzustands- und E-Zu-

standsregistern **114** lesen. Während dieses Fehlerbetriebes vermittelt der Vermittler (die Zugangslogik) **134** den Zugriff auf die Brücke **12**.

[0129] In der Stufe S5 erklärt sich einer der Verarbeitungssätze **14** und **16** selbst als primären Verarbeitungssatz. Dies wird dadurch bestimmt, daß jeder der Verarbeitungssätze einen Zeitfaktor auf der Basis des abgeschätzten Ausmaßes der Verantwortlichkeit für den Fehler kennzeichnet, wodurch der erste Prozessorsatz, der dies abgeschlossen hat, der primäre Verarbeitungssatz wird. In der Stufe S5 wird der Zustand für den Verarbeitungssatz wiedergewonnen und wird auf den anderen Verarbeitungssatz kopiert. Der primäre Verarbeitungssatz ist in der Lage, auf den Puffer **122** für angekündigte Schreibvorgänge und die Abschaltregister **120** zuzugreifen.

[0130] In Stufe S6 ist die Brücke in einem aufgespaltenen Betriebszustand betreibbar. Falls es möglich ist, einen äquivalenten Zustand für die ersten und zweiten Verarbeitungssätze aufrechtzuerhalten, so wird in Stufe S7 ein Reset ausgegeben, um die Verarbeitungssätze in den kombinierten Betriebszustand bei Stufe S1 zu versetzen. Es kann jedoch sein, daß es nicht möglich ist, einen äquivalenten Zustand wieder bereitzustellen, bevor ein fehlerhafter Verarbeitungssatz ausgetauscht ist. Dementsprechend bleibt das System in dem aufgespaltenen Betriebszustand gemäß Stufe S6, um den Betrieb auf der Basis eines einzigen Verarbeitungssatzes fortzusetzen. Nach dem Austausch des fehlerhaften Verarbeitungssatzes könnte das System dann einen äquivalenten bzw. ausgewogenen Zustand herstellen und über Stufe S7 in die Stufe S1 übergehen.

[0131] Wie oben beschrieben, ist der Komparator **130** in dem kombinierten Betrieb so betreibbar, daß er die I/O-Vorgänge, die durch die ersten und zweiten Verarbeitungssätze **14** und **16** ausgegeben werden, vergleicht. Dies ist in Ordnung, solange alle I/O-Vorgänge der ersten und zweiten Verarbeitungssätze **14** und **16** vollständig synchronisiert und deterministisch sind. Irgendeine Abweichung hiervon wird durch den Komparator **130** als Verlust des Verriegelungsschrittbetriebes interpretiert. Dies ist im Prinzip korrekt, da sogar eine kleinere Abweichung von identischen Ausgangswerten, falls sie durch den Komparator **130** nicht erfasst werden, dazu führen könnte, daß die Verarbeitungssätze weiter voneinander divergieren, wenn die individuellen Prozessorsätze mit den voneinander abweichenden Ausgangswerten arbeiten. Jedoch bringt eine strenge Anwendung dieses Prinzips beträchtliche Einschränkungen in der Auslegung der individuellen Verarbeitungssätze mit sich. Ein Beispiel hierfür ist, daß es nicht möglich wäre, Uhren bzw. Zeitgeber mit unterschiedlichen Tageszeiten in den individuellen Verarbeitungssätzen zu haben, die unter ihrem eigenen Takt bzw. ihrer eigenen Uhr arbeiten. Dies liegt daran, daß es unmöglich ist, zwei Kristalle zu erhalten, die im Betrieb 100-prozentig identisch sind. Selbst geringe Unterschiede in der Phase der Takte könnten im Hinblick darauf kritisch sein, ob dieselbe Abtastung gleichzeitig vorgenommen wird, beispielsweise auf jeder Seite eines Taktüberganges für die entsprechenden Verarbeitungssätze.

[0132] Dementsprechend verwendet eine Lösung diese Problems die Datenregister für verschiedenartige Daten (DDR) **116**, die zuvor bereits erwähnt wurden. Die Lösung besteht darin, Daten von den Verarbeitungssätzen in entsprechende DDRs in der Brücke zu schreiben, während der Vergleich der Datenphasen der Schreibvorgänge abgeschaltet wird, und dann einen ausgewählten der DDRs zurück in jeden Verarbeitungssatz zu lesen, wobei jeder der Verarbeitungssätze in der Lage ist, mit den selben Daten zu arbeiten.

[0133] Fig. 17 ist eine schematische Wiedergabe von Einzelheiten der Brücke nach den Fig. 6 bis Fig. 10. Wie man sieht, sind Einzelheiten der Brücke, die in Fig. 6 bis Fig. 8 nicht dargestellt sind, in Fig. 17 dargestellt, wohingegen andere Einzelheiten der Brücke, die in den Fig. 6 bis Fig. 8 dargestellt sind, aus Gründen der Klarheit in Fig. 17 nicht dargestellt sind.

[0134] Die DDRs **116** sind in den Brückenregistern **110** nach Fig. 7 vorgesehen, könnten jedoch auf in anderen Ausführungsformen auch an einem anderen Ort in der Brücke vorgesehen sein. Ein DDR **116** ist für jeden Verarbeitungssatz vorgesehen. In dem Beispiel des Mehrprozessorsystems nach Fig. 1, bei welchem zwei Verarbeitungssätze **14** und **16** vorgesehen sind, sind zwei DDRs **116A** und **116B** vorgesehen, und zwar einer für jeden der ersten und zweiten Verarbeitungssätze **14** bzw. **16**.

[0135] Fig. 17 gibt eine Schreibstufe für verschiedenartige Daten wieder. Die Adressierlogik **136** ist schematisch dargestellt mit zwei Decodierabschnitten, wobei ein Decodierabschnitt **136A** für den ersten Verarbeitungssatz und ein Decodierabschnitt **136B** für den zweiten Verarbeitungssatz **16** vorgesehen ist. Während einer Adressierphase eines I/O-Schreibvorganges für verschiedenartige Daten gibt jeder der Verarbeitungssätze **14** und **16** die selbe vorbestimmte Adresse DDR-W aus, die durch die entsprechenden ersten und zweiten Decodierabschnitte **136A** bzw. **136B** getrennt interpretiert wird, wenn die entsprechenden ersten und zweiten DDRs **116A** und **116B** adressiert werden.

[0136] Wenn die selbe Adresse durch die ersten und zweiten Verarbeitungssätze **14** und **16** ausgegeben wird, wird dies durch den Komparator **130** nicht als Verriegelungsschrittfehler interpretiert.

[0137] Der Decodierabschnitt **136A** oder der Decodierabschnitt **136B** oder beide sind dafür ausgelegt, daß sie weiterhin ein Freigabesignal **137** in Reaktion auf die vorbestimmte Schreibadresse ausgeben, die durch die ersten und zweiten Verarbeitungssätze **14** und **16** zugeführt wird. Dieses Freigabesignal wird dem Komparator **130** zugeführt und ist während der Datenphase des Schreibvorganges wirksam, um den Komparator abzuschalten. Im Ergebnis können die Daten, die durch den ersten Verarbeitungssatz ausgegeben werden, in dem ersten DDR **116A** gespeichert werden, und die Daten, die durch den zweiten Verarbeitungssatz ausgegeben werden, können in dem zweiten DDR **116B** gespeichert werden, ohne daß der Komparator in Betrieb ist, um einen Unterschied zu erfassen, selbst wenn die Daten von den ersten und zweiten Verarbeitungssätzen unterschiedlich sein sollten. Der erste Decodierabschnitt ist so betreibbar, daß er die Routingmatrix veranlaßt, die Daten von dem ersten Verarbeitungssatz **14** in dem ersten DDR **116A** zu speichern und der zweite Decodierabschnitt ist so betreibbar, daß er die Routingmatrix veranlaßt, daß sie die Daten von dem zweiten Verarbeitungssatz **16** in dem zweiten DDR **116B** speichert. Am Ende der Datenphase wird der Komparator **130** wieder eingeschaltet, um irgendwelche Unterschiede zwischen I/O-Adress- und/oder Daten-Phasen als Kennzeichen für einen Verriegelungsschrittfehler zu erfassen.

[0138] Im Anschluss an das Schreiben der verschiedenartigen Daten in die ersten und zweiten DDRs **116A** und **116B** sind die Verarbeitungssätze dann so betreibbar, daß sie die Daten von einem ausgewählten der DDRs **116A**/**116B** lesen.

[0139] Fig. 18 veranschaulicht eine alternative Anordnung, bei welcher das Freigabesignal **137** verneint wird und verwendet wird, um ein Gate **131** am Ausgang des Komparator **130** zu steuern. Wenn das Freigabesignal aktiv ist, wird der Ausgang des Komparators abgeschaltet, wohingegen dann, wenn das Sperrsignal inaktiv ist, der Ausgang des Komparators freigeschaltet wird.

[0140] Fig. 19 veranschaulicht das Lesen des ersten DDRs **116A** in einem nachfolgenden Lesezustand für verschiedenartige Daten. Wie in Fig. 19 dargestellt, gibt jeder der Verarbeitungssätze **14** und **16** die selbe vorbestimmte Adresse DDR-RA aus, die durch die entsprechenden ersten und zweiten Decodierabschnitte **136A** und **136B** als Adressierung für den selben DDR, nämlich den ersten DDR **116A**, getrennt interpretiert werden. Im Ergebnis wird der Inhalt des ersten DDR **116A** von beiden Verarbeitungssätzen **14** und **16** gelesen, und ermöglicht dadurch, daß diese Verarbeitungssätze die selben Daten empfangen. Dieses setzt die beiden Verarbeitungssätze **14** und **16** in die Lage, ein deterministisches Verhalten anzunehmen, selbst wenn die Quelle der Daten, die durch die Verarbeitungssätze **14** und **16** in die DDRs **116** geschrieben wurden, nicht deterministisch war.

[0141] Als Alternative könnten die Verarbeitungssätze jeweils die Daten von dem zweiten DDR **116B** lesen. Fig. 20 veranschaulicht das Lesen des zweiten DDR **116B** in einem Lesezustand für verschiedenartige Daten im Anschluß an die Schreibstufe für verschiedenartige Daten gemäß Fig. 15. Wie in Fig. 20 dargestellt, gibt jeder der Verarbeitungssätze **14** und **16** die selbe vorbestimmte Adresse DDR-RB aus, die durch die entsprechenden ersten und zweiten Decodierabschnitte **136A** und **136B** als Adressierung für den selben DDR, nämlich den DDR **116B**, getrennt interpretiert wird.

[0142] Im Ergebnis wird der Inhalt des zweiten DDR **116B** durch beide Verarbeitungssätze **14** und **16** gelesen, was es diesen Verarbeitungssätzen erlaubt, die selben Daten zu empfangen. Wie bei der Lesestufe für verschiedenartige Daten gemäß Fig. 19 ermöglicht dies es den beiden Verarbeitungssätzen **14** und **16**, ein deterministisches Verhalten anzunehmen, selbst wenn die Quelle der Daten, welche in die DDRs **116** durch die Verarbeitungssätze **14** und **16** geschrieben wurden, nicht deterministisch war.

[0143] Die Auswahl, welchen der ersten und zweiten DDRs **116A** und **116B** gelesen werden soll, kann auf irgendeine geeignete Art und Weise durch die Software bestimmt werden, die auf den Verarbeitungsmodulen läuft. Dies könnte auf Basis einer einfachen Auswahl eines oder des anderen DDRs geschehen, oder auf statistischer Basis oder zufällig auf irgendeine andere Art und Weise, solange die selbe Wahl eines DDR von beiden oder allen Verarbeitungssätzen getroffen wird.

[0144] Fig. 21 ist ein Flußdiagramm, welches die verschiedenen Stufen des Betriebs des oben beschriebenen DDR-Mechanismus zusammenfaßt.

[0145] In Stufe S10 wird eine Schreibadresse DDR-W empfangen und während der Adressierphase des

DDR-Schreibvorganges durch die Adreßdecodierabschnitte **136A** und **136B** decodiert.

[0146] In Stufe S11 wird der Komparator **130** gesperrt.

[0147] In Stufe S12 werden die während der Datenphase des DDR-Schreibvorganges von den Verarbeitungssätzen **14** und **16** empfangenen Daten in den ersten bzw. zweiten DDRs **116A** bzw. **116B** gespeichert, so wie sie durch die ersten bzw. zweiten Decodierabschnitte **136A** bzw. **136B** ausgewählt wurden.

[0148] In Stufe S13 wird eine DDR-Leseadresse von den ersten und zweiten Verarbeitungssätzen empfangen und durch die Decodierabschnitte **136A** bzw. **136B** decodiert.

[0149] Wenn die empfangene Adresse DDR-RA diejenige für den ersten DDR **116A** ist, so wird in Stufe S14 der Inhalt des DDR **116A** durch beide der Verarbeitungssätze **14** und **16** gelesen.

[0150] Wenn alternativ die erhaltene Adresse DDR-RB diejenige für den zweiten DDR **116B** ist, so wird in Stufe S15 der Inhalt des DDR **116B** in beide Verarbeitungssätze **14** und **16** gelesen.

[0151] Fig. 22 ist eine schematische Wiedergabe der mit den entsprechenden Bussen **22**, **24** und **26** durchgeführten Vermittlung und der Vermittlung für die Brücke selbst.

[0152] Jede der Bussteuerungen **50** des Verarbeitungssatzes in den entsprechenden Verarbeitungssätzen **14** und **16** umfaßt einen konventionellen PCI-Masterbusvermittler **180**, um eine Vermittlung für entsprechende Busse **24** und **26** vorzusehen. Jeder der Haupt- bzw. Mastervermittler **180** reagiert auf Anforderungssignale von der Bussteuerung **50** des zugehörigen Verarbeitungssatzes und der Brücke **12** auf entsprechenden Anforderungs(REQ)-Leitungen **181** und **182**. Die Hauptvermittler **180** ordnen den Zugriff auf den Bus nach dem Prinzip wer zuerst kommt, wird zuerst bedient, zu und geben ein „Gewährt"(GNT)-Signal an den Sieger auf einer entsprechenden Gewährungsleitung **183** oder **184** aus.

[0153] Eine konventionelle PCI-Busvermittlung (Zugangslogikschaltung) **185**, stellt eine Vermittlung auf dem D-Bus **22** bereit. Der D-Busvermittler **185** kann als Teil der D-Busschnittstelle **82** nach Fig. 6 ausgestaltet werden oder könnte getrennt von dieser sein. Wie im Falle des Hauptvermittlers **180** des P-Busses reagiert auch der D-Busvermittler auf Anforderungssignale von den miteinander in Konkurrenz stehenden Geräten, einschließlich der Brücke und der Geräte **30**, **31**, etc., die an den Gerätebus **22** angeschlossen sind. Entsprechende Anforderungsleitungen **186**, **187**, **188** etc. für jede der Einheiten, die im Wettbewerb um einen Zugriff auf den D-Bus **22** stehen, werden für Anforderungssignale (REQ) vorgesehen. Der D-Busvermittler **186** ordnet einen Zugriff auf den D-Bus nach dem Prinzip wer zuerst kommt, wird zuerst bedient, zu und gibt ein „Gewährt"(GNT)-Signal über entsprechende Gewährungsleitungen **189**, **190**, **192**, etc. an die obsiegende Einheit.

[0154] Fig. 23 ist ein Zustandsdiagramm, welches die Betriebsweise der D-Busvermittlung **185** zusammenfaßt. In einer besonderen Ausführungsform können bis zu sechs Anforderungssignale durch entsprechende D-Busgeräte erzeugt werden und einer durch die Brücke selbst. Bei dem Übergang in den GRANT-Zustand (Gewährungszustand) werden diese durch einen Prioritätsencoder sortiert und ein Anforderungssignal (REQ#) mit der höchsten Priorität wird als Sieger registriert und erhält ein Gewährungs(GNT#)-Signal. Jeder Gewinner bzw. Sieger, der ausgewählt wurde, modifiziert die Prioritäten in einem Prioritätsencoder um die selben REQ#-Signale beim nächsten Schritt zur Gewährung zu ergeben. Ein anderes Gerät hat die höchste Priorität, so daß jedes Gerät eine „faire" Chance des Zugriffs auf DEVs hat. Die Brücke REQ# hat ein höheres Gewicht als D-Busgeräte und erhält, bei sehr starkem Betrieb, den Bus nach jedem zweiten Gerät.

[0155] Wenn ein Gerät, das den Bus anfordert, eine Transaktion nicht innerhalb von 16 Zyklen durchführen kann, so kann es sein GNT# über den BACKOFF-Zustand verlieren. BACKOFF ist erforderlich, da gemäß PCI-Regeln ein Gerät auf den Bus zugreifen kann, einen Zyklus nachdem GNT# entfernt worden ist. Geräten kann nur dann Zugriff auf den D-Bus gewährt werden, wenn die Brücke sich nicht in dem E-Zustand befindet. Zu dem Zeitpunkt, wenn der Bus sich im Leerlauf befindet wird ein neues GNT# erzeugt.

[0156] Durch die Zustände GRANT und BUSY, werden die FETs freigeschaltet und ein zugreifendes Gerät ist bekannt und wird an die Adressdecodierlogik des D-Busses weitergeleitet, um eine Überprüfung gegenüber einer durch das Gerät gelieferten DMA-Adresse vorzunehmen.

[0157] Betrachten wir nun den Brückenvermittler **134**, welcher den Zugriff auf die Brücke für das erste Gerät erlaubt, welches das PCI FRAME-Signal vorbringt, das eine Adressierphase anzeigt. Fig. 24 ist ein Zustands-

diagramm, welches die Betriebsweise des Brückenvermittlers **134** zusammenfaßt.

[0158] Wie im Falle des D-Busvermittlers, kann ein Prioritätsencoder vorgesehen werden, um miteinander kollidierende Zugriffsversuche aufzulösen. In diesem Fall einer „Kollision" versucht der Verlierer/die Verlierer es erneut, was ihn/sie zwingt, den Bus aufzugeben. Gemäß PCI-Regeln müssen Geräte in einem neuen Versuch erneut versuchen auf die Brücke zuzugreifen und man kann erwarten, das dies geschieht.

[0159] Um zu verhindern, daß Geräte, die sehr schnell mit ihren Wiederholungsversuchen sind, die Brücke blockieren, findet eine Erinnerung der im erneuten Versuch befindlichen Schnittstellen statt, die eine höhere Priorität erhalten. Diese erinnerten Neuversuche erhalten in der selben Art wie Adressphasen eine Prioritätsordnung. Als Vorsichtsmaßnahme wird dieser Mechanismus jedoch zeitlich begrenzt, so daß man nicht mit dem Warten auf eine fehlerhafte oder tote Einrichtung festhängt. Der verwendete Algorithmus verhindert, daß ein Gerät, welches noch keinen erneuten Versuch vorgenommen hat, was jedoch bei einem erneuten Versuch eine höhere Priorität haben würde als Gerät, das derzeit darauf wartet, bereits beim ersten Versuch wiederholt wird.

[0160] Im kombinierten Betrieb wählt ein PA- oder PB-Buseingang aus, welche P-Busschnittstelle einen Brückenzugriff gewinnt. Beiden wird mitgeteilt, sie hätten gewonnen. Eine zugelassene Auswahl ermöglicht während des Normalbetriebs eine latente Fehlerüberprüfung. Der E-Zustand verhindert, daß der D-Bus gewinnt.

[0161] Der Brückenvermittler **134** reagiert auf Standard-PCI-Signale, die auf standardmäßigen PCI-Steuerleitungen **22**, **24** und **25** für die Steuerung des Zugriffs auf die Brücke **12** bereitgestellt werden.

[0162] Fig. 25 veranschaulicht Signale, die zu einem I/O-Vorgangszyklus auf dem PCI-Bus gehören. Ein PCI-Rahmensignal (FRAME#) wird zu Beginn vorgebracht. Gleichzeitig sind auf dem Datenbus (DATABUS) Adress(A)-Signale verfügbar und die geeigneten Befehls-(Schreibe/Lese)Signale (C) sind auf dem Befehlsbus (CMD-Bus) verfügbar. Kurz nachdem das Rahmensignal auf low gesetzt wurde, wird auch das „Auslösebereit"-Signal (IRDY#) auf low gesetzt. Wenn das Gerät reagiert, wird ein Signal „Gerät ausgewählt" (DEVSEL#) auf low vorgebracht. Wenn ein „Ziel bereit"-Signal auf low ausgegeben wird (TRDY#), kann eine Datenübertragung (D) auf dem Datenbus stattfinden.

[0163] Die Brücke ist so betreibbar, daß sie Zugriff auf Brückenressourcen zuordnet und dabei die Zuordnung eines Zielbusses in Reaktion auf das Setzen von FRAME# auf low für den betreffenden Auslöserbus verhandelt. Dementsprechend ist der Brückenvermittler **134** so betreibbar, daß er Zugriff auf Brückenressourcen und/oder einen Zielbus nach einem „wer zuerst kommt, wird zuerst bedient"-Prinzip in Reaktion darauf zuordnet, daß das FRAME#-Signal auf low gesetzt wird. Ebenso wie bei der einfachen „wer zuerst kommt, wird zuerst bedient"-Methode können die Vermittler zusätzlich mit einem Mechanismus für das Anhängigmachen der Vermittlungsanforderungen versehen sein und sie können auf der Basis der Anforderungs- und Zuordnungsgeschichte eine Konfliktlösung implizieren, wenn zwei Anforderungen zu identischer Zeit empfangen wurden. Alternativ kann eine einfache Priorität den verschiedenen Anforderern zugeordnet werden, wobei im Falle von Anforderungen zu identischen Zeitpunkten ein bestimmter Anforderer immer den Zuordnungsprozeß gewinnt.

[0164] Jeder der Steckplätze auf dem Gerätebus **22** hat ein Reaktionsregister für den Steckplatz (SRR) **118**, ebenso wie andere Geräte, die mit dem Bus verbunden sind, wie z.B. ein SCSI-Interface. Jeder der SRRs **118** enthält Bits, welche den Besitzzustand der Steckplätze definieren oder der Geräte, die mit den Steckplätzen auf dem direkten Speicherzugriffsbus verbunden sind. In dieser Ausführungsform und aus den oben dargelegten Gründen weist jeder SRR **118** ein vier Bit-Register auf. Es versteht sich jedoch, daß zur Bestimmung des Besitzzustandes bei mehr als zwei Verarbeitungssätzen ein größeres Register erforderlich ist. Beispielsweise ist, wenn drei Verarbeitungssätze vorgesehen sind, für jeden Steckplatz ein fünf Bit-Register erforderlich.

[0165] Fig. 16 veranschaulicht schematisch ein solches vier Bit-Register **600**. Wie in Fig. 16 dargestellt, ist ein erstes Bit **602** mit SRR[0] gekennzeichnet, ein zweites Bit **604** ist gekennzeichnet als SRR[1], ein drittes Bit **606** ist gekennzeichnet als SRR[2] und ein viertes Bit **608** ist gekennzeichnet als SRR[3].

[0166] Bit SRR[0] ist ein Bit, welches gesetzt wird, wenn Schreibvorgänge für gültige Transaktionen unterdrückt werden sollen.

[0167] Bit SRR[1] wird gesetzt, wenn der Gerätesteckplatz dem ersten Verarbeitungssatz **14** gehört. Dies definiert den Zugangspfad zwischen dem ersten Verarbeitungssatz **14** und dem Gerätesteckplatz. Wenn dieses Bit gesetzt ist, kann der erste Verarbeitungssatz **14** immer der Master eines Gerätesteckplatzes **22** sein, wäh-

rend die Fähigkeit, des Gerätesteckplatzes, Master zu sein, davon abhängt, ob Bit SRR[3] gesetzt ist.

[0168] Bit SRR[2] wird gesetzt, wenn der Gerätesteckplatz dem zweiten Verarbeitungssatz **16** gehört. Dies definiert den Zugriffspfad zwischen dem zweiten Verarbeitungssatz **16** und dem Gerätesteckplatz (Slot). Wenn dieses Bit gesetzt ist, kann der zweite Verarbeitungssatz **16** immer der Master eines Gerätesteckplatzes oder Busses **22** sein. Während die Fähigkeit des Gerätesteckplatzes, Master zu sein, davon abhängt, ob Bit SRR[3] gesetzt ist.

[0169] Bit SRR[3] ist ein Vermittlungsbit, welches dem Gerätesteckplatz die Fähigkeit gibt, Master des Gerätebusses **22** zu werden, jedoch nur dann, wenn er einem der Verarbeitungssätze **14** und **16** gehört, d.h. wenn eines der SRR[1] und SRR[2]-Bits gesetzt ist.

[0170] Wenn das Täuschbit (SRR[0]) eines SRR **118** gesetzt ist, werden Schreibvorgänge in das Gerät für diesen Steckplatz bzw. Slot ignoriert und erscheinen nicht auf dem Gerätebus **22**. Lesevorgänge liefern unbestimmte Daten zurück, ohne daß eine Transaktion auf dem Gerätebus **22** stattfindet. Im Falle eines I/O-Fehlers wird das Täuschbit SRR[0] des SRR **118**, welcher dem Gerät entspricht, das den Fehler verursacht hat, durch die Hardwarekonfiguration der Brücke gesetzt, um einen weiteren Zugriff auf den betroffenen Gerätesteckplatz zu sperren. Durch die Brücke kann auch ein Interrupt erzeugt werden, um die Software, die den Zugriff ausgelöst hat, welche zu dem I/O-Fehler führte, zu informieren, daß der Fehler aufgetreten ist. Das Täuschbit hat Wirkung, gleich ob das System in dem aufgespaltenen oder in dem kombinierten Betrieb arbeitet.

[0171] Die Besitzbits haben jedoch nur in dem aufgespaltenen Systembetrieb Wirkung. In diesem Betriebszustand kann jeder Steckplatz bzw. Slot drei Stufen haben:
nicht in Besitz,
in Besitz durch Verarbeitungssatz **14**, und
in Besitz durch Verarbeitungssatz **16**.

[0172] Dieses wird durch die beiden SRR-Bits SRR[1] und SRR[2] bestimmt, wobei SRR[1] gesetzt wird, wenn der Steckplatz dem Verarbeitungssatz **14** gehört und SRR[2] gesetzt ist, wenn der Steckplatz dem Verarbeitungssatz **16** gehört. Wenn der Steckplatz nicht in Besitz ist, so ist keines der Bits gesetzt (das Setzen beider Bits ist ein unzulässiger Zustand und wird durch die Hardware verhindert).

[0173] Auf einen Steckplatz, der nicht dem Verarbeitungssatz gehört, welcher den Zugriff macht (dies umfaßt nicht-im-Besitz-befindliche Steckplätze) kann nicht zugegriffen werden und führt zu einer Ablehnung. Ein Verarbeitungssatz kann nur einen nicht in Besitz befindlichen Steckplatz beanspruchen. Er kann nicht den Besitz von einem anderen Verarbeitungssatz fortnehmen. Dies kann nur dadurch geschehen, daß man den anderen Verarbeitungssatz abschaltet. Wenn ein Verarbeitungssatz abgeschaltet ist, gehen alle in seinem Besitz befindliche Steckplätze in den Zustand nicht in Besitz befindlich. Während es für einen Verarbeitungssatz nicht möglich ist, den Besitz von einem anderen Verarbeitungssatz zu übernehmen, ist es jedoch für einen Verarbeitungssatz möglich, den Besitz an einen anderen Verarbeitungssatz zu übergeben.

[0174] Die Besitzbits können während des kombinierten Betriebszustandes geändert werden, haben jedoch in soweit keinen Effekt, bis der Eintritt in einen aufgespaltenen Zustand erfolgt.

[0175] Die nachstehende Tabelle 2 faßt die Zugriffsrechte zusammen, wie sie durch einen SRR **118** festgelegt werden.

[0176] Aus Tabelle 2 kann man erkennen, daß dann, wenn der 4-Bit SRR für ein gegebenes Gerät beispielsweise auf 1100 gesetzt ist, der Steckplatz dem Verarbeitungssatz B gehört (d.h. SRR[2] ist logisch high) und der Verarbeitungssatz A kann aus dem Gerät nicht lesen oder in dieses schreiben (d.h. SRR[1] ist logisch low), auch wenn er von der Brücke lesen und in diese schreiben kann. „FAKE_AT" ist logisch niedrig gesetzt (d.h. SRR[0] ist logisch niedrig bzw. low), was anzeigt, daß Zugriff auf den Gerätebus erlaubt ist, solange es auf dem Bus keine Fehler gibt. Wenn „ARB_EN" auf logisch high gesetzt ist (d.h. SRR[3] ist logisch high), so kann das Gerät, zu dem das Register gehört, Master des D-Busses werden. Dieses Beispiel veranschaulicht die Betriebsweise des Registers, wenn der Bus und die zugehörigen Geräte korrekt arbeiten.

TABELLE 2

SRR [3][2][1][0]	PA-Bus	PB-Bus	Geräteschnittstelle
0000 x00x	Lesen/Schreiben Brücke SRR	Lesen/Schreiben Brücke SRR	Zugriff verweigert
0010	Lesen/Schreiben Brücke D-Slot in Besitz	Lesen/Schreiben Brücke kein Zugriff auf D-Slot	Zugriff verweigert, weil Vermittlungsbit aus ist
0100	Lesen/Schreiben Brücke kein Zugriff auf D-Slot	Lesen/Schreiben Brücke Zugriff auf D-Slot	Zugriff verweigert, weil Vermittlungsbit aus ist
1010	Lesen/Schreiben Brücke D-Slot in Besitz	Lesen/Schreiben Brücke kein Zugriff auf D-Slot	Zugriff auf CPU B verweigert Zugriff auf CPU A OK
1100	Lesen/Schreiben Brücke kein Zugriff auf D-Slot	Lesen/Schreiben Brücke Zugriff auf D-Slot	Zugriff auf CPU A verweigert Zugriff auf CPU B OK
0011	Lesen/Schreiben Brücke Brücke sondert Schreibvorgänge aus	Lesen/Schreiben Brücke kein Zugriff auf D-Slot	Zugriff verweigert, weil Vermittlungsbit aus ist
0101	Lesen/Schreiben Brücke kein Zugriff auf D-Slot	Lesen/Schreiben Brücke Brücke sondert Schreibvorgänge aus	Zugriff verweigert, weil Vermittlungsbit aus ist
1011	Lesen/Schreiben Brücke Brücke sondert Schreibvorgänge aus	Lesen/Schreiben Brücke kein Zugriff auf D-Slot	Zugriff auf CPU B verweigert Zugriff auf CPU A OK
1101	Lesen/Schreiben Brücke kein Zugriff auf D-Slot	Lesen/Schreiben Brücke Brücke sondert Schreibvorgänge aus	Zugriff auf CPU B verweigert Zugriff auf CPU A OK

[0177] In einem alternativen Beispiel zeigt, wenn der SRR für das Gerät auf 0101 gesetzt ist, das Setzen von SRR[2] auf logisch high an, daß das Gerät dem Verarbeitungssatz B gehört. Wenn jedoch das Gerät fehlerhaft arbeitet, wird SRR[3] auf logisch low gesetzt und das Gerät hat keinen Zugriff auf den Verarbeitungssatz. SRR[0] ist auf high gesetzt, so daß jegliche Schreibvorgänge auf das Gerät ignoriert werden und ein Lesen von diesem unbestimmte Daten liefert. Auf diese Weise wird das fehlerhaft funktionierende Gerät effektiv von dem Verarbeitungssatz isoliert und liefert unbestimmte Daten, um beispielsweise irgendwelche Gerätetreiber zufrieden zu stellen, die auf eine Reaktion von dem Gerät warten.

[0178] Fig. 26 veranschaulicht die Betriebsweise der Brücke **12** für einen direkten Speicherzugriff durch ein Gerät, wie z.B. eines der Geräte **28**, **29**, **30**, **31** und **32** auf den Speicher **56** der Verarbeitungssätze **14** und **16**. Wenn der D-Busvermittler **185** eine Anforderung **193** für eine direkten Speicherzugriff (DMA) von einem Gerät (z.B. dem Gerät **30** im Steckplatz **33**) auf dem Gerätebus empfängt, so bestimmt der D-Busvermittler, ob der Bus diesem Steckplatz zugeordnet wird. Als Ergebnis dieses Gewährungsvorganges kennt der D-Busvermittler den Steckplatz, der die DMA-Anforderung **193** vorgebracht hat. Die DMA-Anforderung wird dem Adreßdecodierer **142** in der Brücke zugeführt, wo die zu der Anforderung gehörenden Adressen decodiert werden. Der Adreßdecodierer reagiert auf das Gewährungssignal **124** des D-Busses für den betreffenden Steckplatz, um den Steckplatz zu identifizieren, welchem für die DMA-Anforderung Zugriff auf den D-Bus gewährt wurde.

[0179] Die Adreßdecodierlogik **142** hält einen geographischen Adressenplan **196** oder hat Zugriff auf einen solchen, der die Beziehung zwischen dem Prozessoradreßraum und den Steckplätzen als Ergebnis der verwendeten geographischen Adresse identifiziert bzw. kennzeichnet. Der geographische Adreßplan **196** könnte

als Tabelle in dem Brückenspeicher **126** gehalten werden, zusammen mit dem Puffer **122** für angekündigte Schreibvorgänge und dem dirty RAM **124**. Alternativ könnte er als eine Tabelle in einem getrennten Speicherelement gehalten werden, welches möglicherweise Teil des Adreßdecodierers **142** selbst bildet. Der Plan **182** könnte auch in einer anderen Form als einer Tabelle ausgestaltet sein.

[0180] Die Adreßdecodierlogik **142** ist so ausgestaltet, daß sie die Richtigkeit der durch das Gerät **30** zugeführten DMA-Adressen verifiziert. In einer Ausführungsform der Erfindung wird dies erreicht durch Vergleichen von vier signifikanten Adreßbits der durch das Gerät zugeführten Adresse mit entsprechenden vier Adreßbits der Adresse, die für den durch das D-Busgewährungssignal für die DMA-Anforderung gekennzeichneten Steckplatz in dem geographischen Adreßplan **196** gehalten wird. In diesem Beispiel sind vier Adreßbits ausreichend, um festzulegen, ob die zugeführte Adresse innerhalb des korrekten Adressbereiches liegt. Bei diesem speziellen Beispiel werden 32-Bit-PCI-Busadressen verwendet, wobei die Bits 31 und 30 immer auf 1 gesetzt sind, Bit 29 so zugeordnet wird, daß es kennzeichnet, welche von zwei Brücken einer Hauptplatine adressiert wird (siehe Fig. 2) und die Bits 28 bis 26 ein PCI-Gerät kennzeichnen. Die Bits 25-0 definieren einen Offset bzw. eine Verschiebung gegenüber der Basisadresse für den Adreßbereich für jeden Steckplatz. Dementsprechend ist es durch Vergleichen der Bits 29-26 möglich, festzustellen, ob die zugeführte(n) Adresse(n) in den passenden Adreßbereich für den betroffenen Steckplatz fällt (fallen). Es versteht sich, daß in anderen Ausgangsformen eine andere Zahl von Bits verglichen werden muß, um in Abhängigkeit von der Zuordnung der Adressen diese Feststellung zu treffen.

[0181] Die Adreßdecodierlogik **142** könnte so ausgelegt sein, daß sie das Busfreigabesignal **184** (Busgewährungssignal) für den in Rede stehenden Steckplatz verwendet, um einen Tabelleneintrag für den betreffenden Steckplatz zu identifizieren und dann die Adresse in diesen Eintrag mit der Adresse (den Adressen) zu vergleichen, die mit der DMA-Anforderung empfangen wurde(n), wie oben beschrieben. Alternativ könnte die Adreßdecodierlogik **142** dafür ausgelegt sein, daß die Adresse(n), die mit der DMA-Anforderung empfangen wurde(n), für das Adressieren eines dazu in Beziehung stehenden geographischen Adreßplanes verwendet, und um daraus eine Steckplatzzahl zu bestimmen, die mit dem Steckplatz verglichen werden könnte, für welchen das Busfreigabesignal **194 194** bestimmt ist, und dadurch zu bestimmen, ob die Adressen in den für den betreffenden Steckplatz passenden Bereich fallen.

[0182] Wie auch immer, so ist die Adreßdecodierlogik **142** dafür ausgelegt, das Fortschreiten eines DMA zuzulassen, wenn die DMA-Adressen für den betreffenden Steckplatz in den erwarteten Adreßraum fallen. Ansonsten ist der Adreßdecodierer dafür ausgelegt, die Steckplätze und die physikalischen Adressen zu ignorieren.

[0183] Die Adreßdecodierlogik **142** ist weiterhin so betreibbar, daß sie den Pfad (das Routing) der DMA-Anforderung zu dem passenden Verarbeitungssatz (-sätzen) **14**/**16** kontrolliert. Wenn die Brücke sich in dem kombinierten Betriebszustand befindet, wird der DMA-Zugriff automatisch allen der miteinander synchronisierten Verarbeitungssätze **14**/**16** zugeordnet. Die Adreßdecodierlogik **142** weiß, daß die Brücke sich im kombinierten Betriebszustand befindet, da sie unter der Steuerung bzw. Kontrolle der Brückensteuerung **132** läuft (siehe Fig. 8). Wenn jedoch die Brücke sich in dem aufgespaltenen Betriebszustand befindet, muß eine Entscheidung fallen, an welchen der Verarbeitungssätze, wenn überhaupt an einen, die DMA-Anforderung geschickt werden soll.

[0184] Wenn das System sich im aufgespaltenen Betriebszustand befindet, wird der Zugriff auf denjenigen Verarbeitungssatz **14** oder **16** geleitet, dem der betreffende Steckplatz gehört. Wenn der Steckplatz nicht im Besitz befindlich ist, so reagiert die Brücke nicht auf die DMA-Anforderung. In dem aufgespaltenen Betriebszustand ist die Adreßdecodierlogik **142** so betreibbar, daß sie den Besitzzustand des Gerätes bestimmt, von welchem die DMA-Anforderung stammt, in dem sie auf den SRR **118** für den betreffenden Steckplatz zugreift. Der passende Steckplatz kann durch das D-Busfreigabesignal identifiziert werden. Die Adreßdecodierlogik **142** ist so betreibbar, daß sie die Zielsteuerung **140** (siehe Fig. 8) steuert bzw. kontrolliert, um die DMA-Anforderung an den geeigneten bzw. richtigen Verarbeitungssatz (-sätze) **14**/**16** zu leiten, und zwar auf der Basis der Besitzbits SRR[1] und SRR[2]. Wenn das Bit SRR[1] gesetzt ist, ist der erste Verarbeitungssatz **14** der Besitzer und die DMA-Anforderung wird an den ersten Verarbeitungssatz geleitet. Wenn das Bit SRR[2] gesetzt ist, so ist der zweite Verarbeitungssatz **16** der Besitzer und die DMA-Anforderung wird an den zweiten Verarbeitungssatz geleitet. Wenn weder das Bit SRR[1] noch SRR[2] gesetzt ist, so wird die DMA-Anforderung von dem Adreßdecodierer ignoriert und wird an keinen der Verarbeitungssätze **14** und **16** geleitet.

[0185] Fig. 27 ist ein Flußdiagramm, welches den DMA-Verifizierungsprozeß zusammenfaßt, wie er in Bezug auf Fig. 24 veranschaulicht wurde.

[0186] In der Stufe S20 vermittelt der D-Busvermittler **160** einen Zugriff auf den D-Bus **22**.

[0187] In Stufe S21 verifiziert der Adreßdecodierer **142** die mit der DMA-Anforderung zugeführten Adressen, in dem er auf den geographischen Adreßplan zugreift.

[0188] In Stufe S22 ignoriert der Adreßdecodierer den DMA-Zugriff, wenn die Adresse außerhalb des für den betreffenden Steckplatz erwarteten Bereichs fällt.

[0189] Alternativ hängen, wie es durch die Stufe S23 wiedergegeben wird, die Aktionen des Adreßdecodierers davon ab, ob die Brücke sich in dem kombinierten oder in dem aufgespaltenen Betriebszustand befindet.

[0190] Wenn die Brücke sich in dem kombinierten Betriebszustand befindet, so steuert in Stufe S24 der Adreßdecodierer die Zielsteuerung **140** (siehe Fig. 8), um die Routingmatrix **800** zu veranlassen (siehe Fig. 6), die DMA-Anforderung an beide Verarbeitungssätze **14** und **16** zu leiten.

[0191] Wenn die Brücke sich in dem aufgespaltenen Betriebszustand befindet, ist der Adreßdecodierer so betreibbar, daß er den Besitz des betreffenden Steckplatzes durch Bezugnahme auf den SRR **118** für diesen Steckplatz in Stufe S25 verifiziert.

[0192] Wenn der Steckplatz dem ersten Verarbeitungssatz **14** zugeordnet ist (d.h. daß SRR[1]-Bit ist gesetzt), dann steuert in Stufe S26 der Adreßdecodierer **142** die Zielsteuerung **140** so (siehe Fig. 8), daß die Routingmatrix (**80**) veranlaßt wird (siehe Fig. 6), die DMA-Anforderung an den ersten Verarbeitungssatz **14** zu leiten.

[0193] Wenn der Steckplatz dem zweiten Verarbeitungssatz **16** zugeordnet ist (d.h. daß SRR[2]-Bit ist gesetzt), so steuert in Stufe S27 der Adreßdecodierer **142** die Zielsteuerung **140** (siehe Fig. 8) so, daß sie die Routingmatrix (**80**) veranlaßt (siehe Fig. 6), die DMA-Anforderung an den zweiten Verarbeitungssatz **16** zu leiten.

[0194] Wenn der Steckplatz nicht zugeordnet ist (d.h. weder das SRR[1]-Bit noch das SRR[2]-Bit ist gesetzt), so ignoriert in Schritt S18 der Adreßdecodierer **142** die DMA-Anforderung oder sondert diese aus, und die DMA-Anforderung wird den Verarbeitungssätzen **14** und **16** nicht zugeleitet.

[0195] Eine DMA-Anforderung oder eine direkte Vektorspeicherzugriffsanforderung (DVMA), die an einen oder mehrere der Verarbeitungssätze übermittelt wurde, bewirkt, daß die erforderlichen Speichervorgänge (lesen oder schreiben, je nachdem was angezeigt ist), auf dem Verarbeitungssatzspeicher bewirkt werden.

[0196] Es folgt nun eine Beschreibung eines Beispiels eines Mechanismus, um eine automatische Reparatur aus einem E-Zustand zu ermöglichen (siehe Fig. 11).

[0197] Der automatische Reparaturvorgang umfaßt die Reintegration des Zustandes der Verarbeitungssätze in einen gemeinsamen Zustand, um einen erneuten Start im Verriegelungsschritt zu versuchen. Um dieses zu erreichen, kopiert der Verarbeitungssatz, der sich selbst als primären Verarbeitungssatz meldet, wie es oben beschrieben wurde, seinen gesamten Zustand auf den anderen Verarbeitungssatz. Dies umfaßt das Sicherstellen, daß der Inhalt des Speichers beider Prozessoren der selbe ist, bevor man einen erneuten Start im Verriegelungsschrittbetrieb versucht.

[0198] Jedoch besteht ein Problem beim Kopieren des Inhaltes des Speichers von einem Verarbeitungssatz in den anderen darin, daß während dieses Kopiervorganges ein mit dem D-Bus **22** verbundenes Gerät eine direkte Speicherzugriffs(DMA)-Anforderung für einen Zugriff auf den Speicher des primären Verarbeitungssatzes versuchen könnte. Wenn DMA freigeschaltet ist, so würde ein Schreibvorgang in einen Speicherbereich, der bereits kopiert worden ist, zu dem Ergebnis führen, daß der Speicherzustand der beiden Prozessoren am Ende des Kopiervorganges nicht der selbe wäre. Im Prinzip wäre es möglich, ein DMA für den gesamten Speichervorgang zu verhindern. Dies wäre jedoch nicht wünschenswert, wenn man bedenkt, daß es wünschenswert ist, die Zeit, während welcher die Systemressourcen nicht verfügbar sind, minimal zu machen. Als Alternative wäre es möglich, den gesamten Kopiervorgang zu wiederholen, wenn ein DMA-Vorgang während der Kopierdauer stattgefunden hat. Es ist jedoch wahrscheinlich, daß während des erneuten Kopierversuchs weitere DMA-Vorgänge durchgeführt werden würden, und dementsprechend ist auch dies keine gute Option. Dementsprechend ist in dem vorliegenden System ein dirty RAM **124** in der Brücke vorgesehen. Wie zuvor bereits beschrieben, ist der dirty RAM **124** als Teil des SRAM-Speichers **126** der Brücke ausgestaltet.

[0199] Der dirty RAM **124** weist einen Bitplan auf, der einen Verunreinigungsanzeiger, beispielsweise ein verunreinigtes Bit für jeden Block oder jede Seite des Speichers hat. Das Bit für eine Speicherseite wird gesetzt, wenn ein Schreibzugriff auf den betroffenen Speicherbereich stattfindet. In einer Ausführungsform der Erfindung wird ein Bit für jeweils eine 8K-Seite des Hauptspeichers des Verarbeitungssatzes bereitgestellt. Das Bit für eine Seite des Verarbeitungssatzspeichers wird automatisch durch den Adreßdecodierer **142** gesetzt, wenn dieser eine DMA-Anforderung für diese Speicherseite entweder für den Verarbeitungssatz **14** oder den Verarbeitungssatz **16** von einem mit dem D-Bus **22** verbundenen Gerät decodiert. Der dirty RAM kann zurückgesetzt oder gelöscht werden, wenn er von einem Verarbeitungssatz gelesen worden ist, beispielsweise mit Hilfe von Lese- und Löschanweisungen zu Beginn eines Kopierdurchganges, so daß er mit dem Aufzeichnen von Seiten beginnen kann, die seit einem gegebenen Zeitpunkt „verunreinigt" bzw. verändert worden sind.

[0200] Der dirty RAM **124** kann Wort für Wort gelesen werden. Wenn für das Lesen des dirty RAM **124** eine erhebliche Wortgröße aus gewählt wird, so optimiert dies das Lesen und Zurücksetzen des dirty RAM **124**.

[0201] Dementsprechend zeigen am Ende des Kopierdurchlaufs die Bits in dem dirty RAM **124** diejenigen Seiten des Verarbeitungssatzspeichers an, die durch DMA-Schreibvorgänge während der Kopierdauer verändert oder „verunreinigt" wurden. Ein weiterer Kopierdurchlauf kann dann durchgeführt werden und zwar nur für diejenigen Speicherseiten, die „verunreinigt" worden waren. Dies erfordertet weniger Zeit als ein vollständiges Kopieren des Speichers. Dementsprechend sind am Ende des nächsten Kopierdurchlaufes typischerweise weniger Seiten als verunreinigt markiert und im Ergebnis werden die Kopierdurchgänge immer kürzer. Zu einem gewissen Zeitpunkt ist es erforderlich zu entscheiden, daß DMA-Schreibvorgänge für eine kurze Zeitdauer für einen endgültigen, kurzen Kopierdurchlauf verhindert werden, so daß am Ende des selben die Speicher der beiden Verarbeitungssätze die selben sind und der primäre Verarbeitungssatz einen Resetvorgang ausgeben bzw. beginnen kann, um den kombinierten Betriebszustand erneut zu starten.

[0202] Der dirty RAM **124** wird sowohl in dem kombinierten als auch in dem aufgespaltenen Betriebszustand gesetzt und gelöscht. Dies bedeutet, daß im aufgespaltenen Betriebszustand der dirty RAM **124** durch irgendeinen der Verarbeitungssätze gelöscht werden kann.

[0203] Die Adresse des dirty RAM **124** wird aus den Bits 13 bis 28 der durch das D-Busgerät gelieferten PCI-Adresse decodiert. Fehlerhafte Zugriffe mit vorliegenden unzulässigen Kombinationen der Adressbits 29 bis 31 werden in dem dirty RAM **124** (planmäßig) zugeordnet und bei einem Schreiben wird ein Bit „verunreinigt", auch wenn die Brücke diese Transaktion nicht an die Verarbeitungssätze weiterleitet.

[0204] Wenn der dirty RAM **124** gelesen wird, definiert die Brücke den gesamten Bereich von 0x00008000 bis 0x0000ffff als dirty RAM und löscht den Inhalt jedes Platzes in diesem Bereich nach einem Lesen.

[0205] Als Alternative für das Bereitstellen eines einzelnen dirty RAM **124**, der beim Lesen gelöscht wird, bestünde eine weitere Alternative darin, zwei dirty RAMs bereitzustellen, die in einem Knebel- bzw. Verknüpfungsbetrieb verwendet würden, wobei in einen geschrieben wird, während der andere gelesen wird.

[0206] Fig. 28 ist ein Flußdiagramm, welches die Betriebsweise des dirty RAM **124** zusammenfaßt.

[0207] In der Stufe S41 liest der primäre Verarbeitungssatz den dirty RAM **124**, was den Effekt hat, daß der dirty RAM **124** zurückgesetzt wird (ein Reset erfährt).

[0208] In der Stufe S42 kopiert der primäre Prozessor (z.B. der Verarbeitungssatz **14**) den Gesamtinhalt seines Speichers **56** in den Speicher **56** des anderen Verarbeitungssatzes (z.B, des Verarbeitungssatzes **16**).

[0209] In Stufe S43 liest der primäre Verarbeitungssatz den dirty RAM **124**, was den Effekt hat, daß der dirty RAM **124** zurückgesetzt wird.

[0210] In Stufe S44 bestimmt der primäre Prozessor, ob weniger als eine vorbestimmte Anzahl von Bits in den dirty RAM **124** geschrieben worden sind.

[0211] Wenn mehr als eine vorbestimmte Anzahl von Bits gesetzt worden sind, so kopiert der Prozessor in Stufe S45 Kopien dieser Seiten seines Speichers **56**, welche verunreinigt worden sind, wie es durch die Verunreinigungsbits angezeigt wird, die von dem dirty RAM **124** in Stufe S43 gelesen wurden, in den Speicher **56** des anderen Verarbeitungssatzes. Die Steuerung geht dann zurück auf die Stufe S43.

[0212] Wenn in Stufe S44 festgestellt wurde, daß weniger als eine vorbestimmte Anzahl von Bits in dem dirty RAM **124** geschrieben worden sind, so bewirkt in Stufe S45 der Primärprozessor, daß die Brücke DMA-Anforderungen von den mit dem D-Bus **22** verbundenen Geräten verhindert werden. Diese könnte beispielsweise dadurch erreicht werden, daß das Vermittlungsfreigabebit für jeden der Gerätesteckplätze gelöscht wird, wodurch der Zugriff der DMA-Geräte auf den D-Bus **22** verweigert wird. Alternativ könnte der Adreßdecodierer **142** so ausgestaltet werden, daß er DMA-Anforderungen unter Anweisungen von den Primärprozessor ignoriert. Während der Zeitdauer, in welcher DMA-Zugriffe verhindert werden, macht der Primärprozessor dann einen endgültigen Kopierdurchlauf von seinem Speicher in den Speicher **56** des anderen Prozessors für diese Speicherseiten, welchen den in dem dirty RAM **124** gesetzten Bits entsprechen.

[0213] In Stufe S47 kann der Primärprozessor einen Resetvorgang ausgeben bzw. ausführen, um den kombinierten Betriebszustand zu iniziieren.

[0214] In Stufe S48 werden DMA-Zugriffe wieder erlaubt.

[0215] Es versteht sich, daß, auch wenn besondere Ausführungsformen der Erfindung hier beschrieben worden sind, dennoch viele Modifikationen/Ergänzungen und/oder Ersetzungen innerhalb des Schutzumfangs der vorliegenden Erfindung vorgenommen werden können. Beispielsweise versteht es sich, daß, auch wenn in der speziellen Beschreibung zwei Verarbeitungssätze vorgesehen sind, die speziell beschriebenen Merkmale für drei oder mehr Verarbeitungssätze modifiziert werden können.

Patentansprüche

1. Brücke für ein Computersystem mit zumindest einem ersten Verarbeitungssatz (**14**) und einem zweiten Verarbeitungssatz (**16**), die jeweils mit der Brücke (**12**) über einen I/O-Bus (**24**, **26**) verbunden sind, und mit einem Steuermechanismus in der Brücke für Ressourcen bzw. Hilfseinrichtungen, mit:
einer Schnittstelle (**82**) zum Austauschen von Signalen mit einem oder mehreren Hilfseinrichtungssteckplätzen bzw. -slots eines Gerätebusses, der in der Lage ist, mit der Brücke verbunden zu werden, wobei jeder der Hilfseinrichtungssteckplätze in der Lage ist, mit einer Hilfseinrichtung bzw. einer Ressource des Systems zu kommunizieren, und
einem Register (**600**), welches jede Systemhilfseinrichtung zugeordnet ist, wobei das Register umschaltbare Indizes (**602**–**608**) hat, welche einen Betriebszustand der zugehörigen Systemhilfseinrichtung anzeigen, wobei der Steuermechanismus so betreibbar ist, daß er Signale zu und von den jeweiligen Systemhilfseinrichtungen des Computersystems leitet.

2. Brücke nach Anspruch 1, wobei zumindest eine der Hilfseinrichtungen eine I/O-Einrichtung ist.

3. Brücke nach Anspruch 1 oder 2, wobei das Register einen Speicher zum Speichern der Indizes aufweist.

4. Brücke nach Anspruch 3, wobei das Register eine 4-Bit-Speichereinheit aufweist.

5. Brücke nach einem der vorstehenden Ansprüche, wobei das Computersystem zwei Verarbeitungssätze bzw. Prozessorsätze aufweist, wobei jeder der Verarbeitungssätze einen oder mehrere Prozessoren (**52**; **62**; **72**) aufweist.

6. Brücke nach Anspruch 5, wobei zumindest einige der umschaltbaren Indizes verwendet werden, um anzuzeigen, ob die zugehörige Hilfseinrichtung einem der Prozessorsätze zugeordnet worden ist.

7. Brücke nach Anspruch 6, wobei zumindest einige der umschaltbaren Indizes zusätzlich anzeigen, welchen der Prozessorsätze die zu dem Register gehörige Hilfseinrichtung zugewiesen worden ist.

8. Brücke nach Anspruch 5 oder 6, wobei das Register eine 4-Bit-Speichereinheit aufweist und das Computersystem einen ersten Verarbeitungssatz und einen zweiten Verarbeitungssatz aufweist, wobei ein zweites Bit und ein drittes Bit der Einheit umschaltbar ist, um anzuzeigen, ob die zugehörige Hilfseinrichtung im Besitz des ersten Verarbeitungssatzes ist, im Besitz des zweiten Verarbeitungssatzes ist oder weder im Besitz des ersten Verarbeitungssatzes noch im Besitz des zweiten Verarbeitungssatzes ist.

9. Brücke nach einem der Ansprüche 5 bis 8, wobei zumindest einer der umschaltbaren Indizes anzeigt, ob der zugehörigen Hilfseinrichtung Zugriff auf einen der Verarbeitungssätze gewährt worden ist.

10. Brücke nach einem der Ansprüche 5 bis 9, wobei zumindest einer der umschaltbaren Indizes wahlweise so betreibbar ist, daß er unbestimmte Daten erzeugt, wenn ein Lesen der dem Register zugeordneten Hilfseinrichtung versucht wird.

11. Brücke nach einem der vorstehenden Ansprüche mit einer Routingmatrix, wobei die Routingmatrix so betreibbar ist, daß sie Befehle und/oder Daten zu oder von einer Hilfseinrichtung leitet, die durch zumindest einen der Indizes als im Besitz eines der Prozessorsätze befindlich gekennzeichnet ist.

12. Brücke nach einem der vorstehenden Ansprüche mit einem 4-Bit-Register, wobei zwei der vier Bits umschaltbar sind zu einem oder von einem
ersten Zustand, welcher anzeigt, daß eine dem Register zugeordnete Hilfseinrichtung keinem der ersten oder zweiten Prozessorsätze zugewiesen ist,
einen zweiten Zustand, welcher anzeigt, daß die Hilfseinrichtung, welche zu dem Register gehört, dem ersten Prozessorsatz zugewiesen ist, und
einen dritten Zustand, welcher anzeigt, daß die Hilfseinrichtung, welche zu dem Register gehört, dem zweiten Prozessorsatz zugewiesen ist.

13. Brücke nach Anspruch 12, wobei ein erstes Bit der verbleibenden zwei Bits des 4-Bit-Registers umschaltbar ist zu oder von:
einem ersten Zustand, welcher anzeigt, daß der zugehörigen Hilfseinrichtung ein Zugriff auf einen der ersten und zweiten Prozessorsätze gewährt worden ist, und
einem zweiten Zustand, welcher anzeigt, daß die zugehörige Hilfseinrichtung keinen Zugriff auf die ersten oder zweiten Prozessorsätze erhalten hat, wobei der erste Zustand nur ausgelöst wird, wenn die beiden der vier Bits des Registers anzeigen, daß die zugehörige Hilfseinrichtung sich im Besitz eines der ersten und zweiten Prozessorsätze befindet.

14. Brücke nach Anspruch 13, wobei ein zweites Bit der verbleibenden zwei Bits des 4-Bit-Registers umschaltbar ist zu oder von
einem ersten Zustand, welcher anzeigt, daß Schreibvorgänge in die zugehörige Hilfseinrichtung zugelassen sind, und
einem zweiten Zustand, welcher anzeigt, daß Schreibvorgänge in die zugehörige Hilfseinrichtung zu ignorieren sind und daß in Reaktion auf irgendwelche versuchte Schreibvorgänge in die zugehörige Hilfseinrichtung unbestimmte Daten erzeugt werden.

15. Computersystem mit:
einer Mehrzahl von Verarbeitungssätzen (**14**, **16**), die jeweils einen oder mehrere Prozessoren (**52**; **62**; **72**) haben und die jeweils mit einem Prozessorbus (**24**, **26**) verbunden sind,
einer Mehrzahl von Einrichtungen (**30**–**32**), die jeweils einem Steckplatz (**33**–**35**) eines I/O-Gerätebusses (**22**) zugeordnet sind, und
einer Brücke gemäß einem der vorstehenden Ansprüche, welche mit der Mehrzahl von Prozessorbussen und dem I/O-Gerätebus verbunden ist, wobei die Brücke einen Gerätesteuermechanismus aufweist, der eine Schnittstelle für den Austausch von Signalen mit einem oder mehreren der Steckplätze und der zugehörigen Geräte hat, und ein Register, welches jedem Gerät zugeordnet ist, wobei das Register umschaltbare Indizes hat, welche einen Betriebszustand des zugehörigen Gerätes anzeigen, wobei der Steuermechanismus im Gebrauch so betreibbar ist, daß er Signale zu und/oder von entsprechenden Systemhilfseinrichtungen des Computersystems leitet.

16. Verfahren zum Betreiben einer Brücke in einem Computersystem, wobei das Computersystem zumindest einen ersten Verarbeitungssatz (**14**) und einen zweiten Verarbeitungssatz (**16**) aufweist, die jeweils mit der Brücke (**12**) über einen I/O-Bus (**24**, **26**) verbunden sind, wobei das Verfahren aufweist:
Austauschen von Signalen einer Schnittstelle in der Brücke mit einem oder mehreren Steckplätzen eines Gerätebusses für Hilfseinrichtungen, der mit der Brücke verbunden ist, wobei jeder Steckplätze für Hilfseinrichtungen mit einer Hilfseinrichtung des Systems in Kommunikationsverbindung steht,
Speichern umschaltbarer Indizes in einem Register der Brücke, welches jeder Systemhilfseinrichtung zugeordnet ist, wobei die umschaltbaren Indizes einen Betriebszustand der zugehörigen Systemhilfseinrichtung anzeigen, und
Leiten von Signalen zu und/oder von entsprechenden Systemhilfseinrichtungen des Computersystems in Abhängigkeit von den umschaltbaren Indizes.

17. Verfahren nach Anspruch 16, wobei das Register in einem Speicher mit wahlfreiem Zugriff (RAM) im-

plementiert ist.

18. Verfahren nach Anspruch 16, wobei das Register ein 4-Bit-Register aufweist.

19. Verfahren nach Anspruch 16, mit Aktualisieren des Registers, um Veränderungen in dem Betriebszustand der Hilfseinrichtung wiederzugeben.

20. Verfahren nach einem der Ansprüche 16 bis 19, welches weiterhin aufweist:
Weiterleiten von Signalen von einem der ersten und zweiten Prozessorsätze, wobei die Signale für eine Hilfseinrichtung des Computersystems bestimmt sind,
Abfragen eines Registers, um festzustellen, ob der eine der ersten und zweiten Prozessorsätze für einen Zugriff auf die Hilfseinrichtung zugelassen ist, und
Leiten der Signale zu der Hilfseinrichtung, wenn das Register anzeigt, daß der Zugriff auf die Hilfseinrichtung für den einen der ersten und zweiten Prozessorsätze zulässig ist.

Es folgen 23 Blatt Zeichnungen

Anhängende Zeichnungen

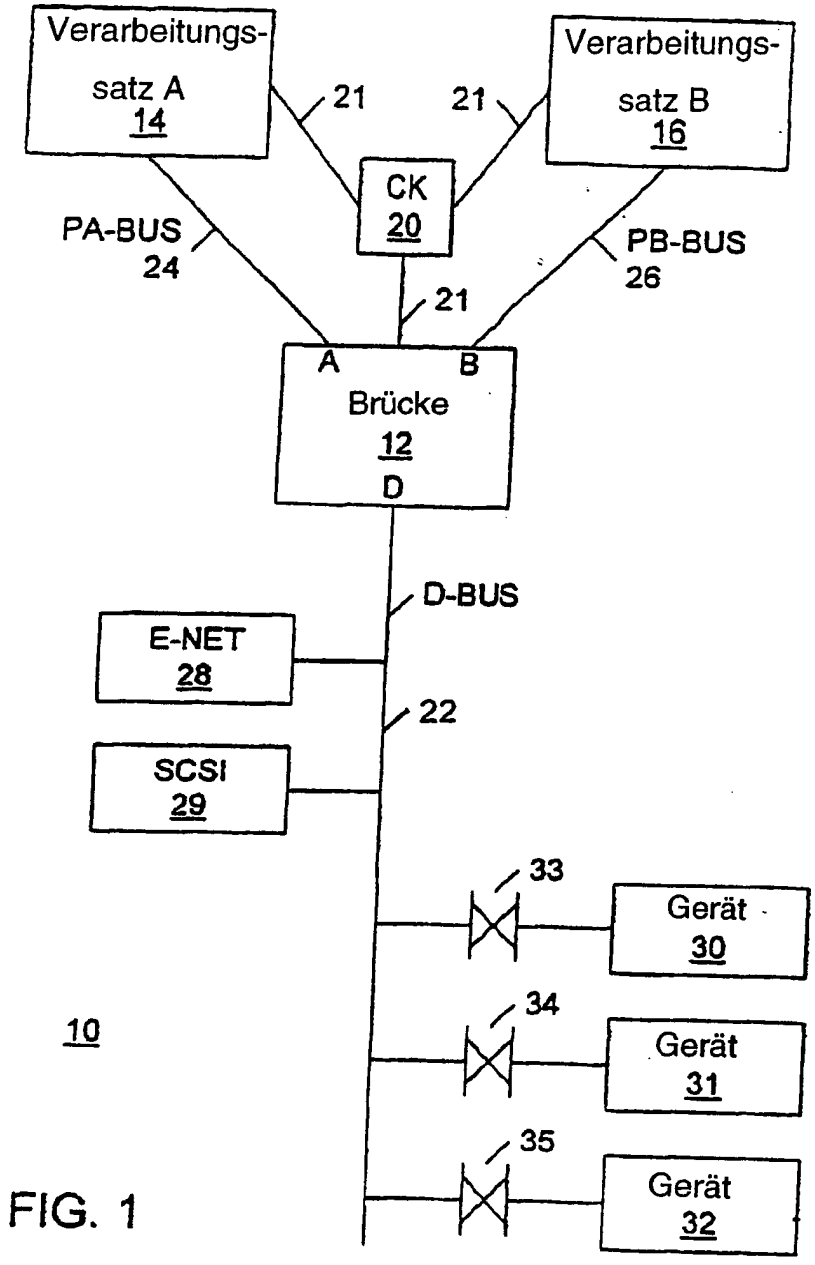

FIG. 1

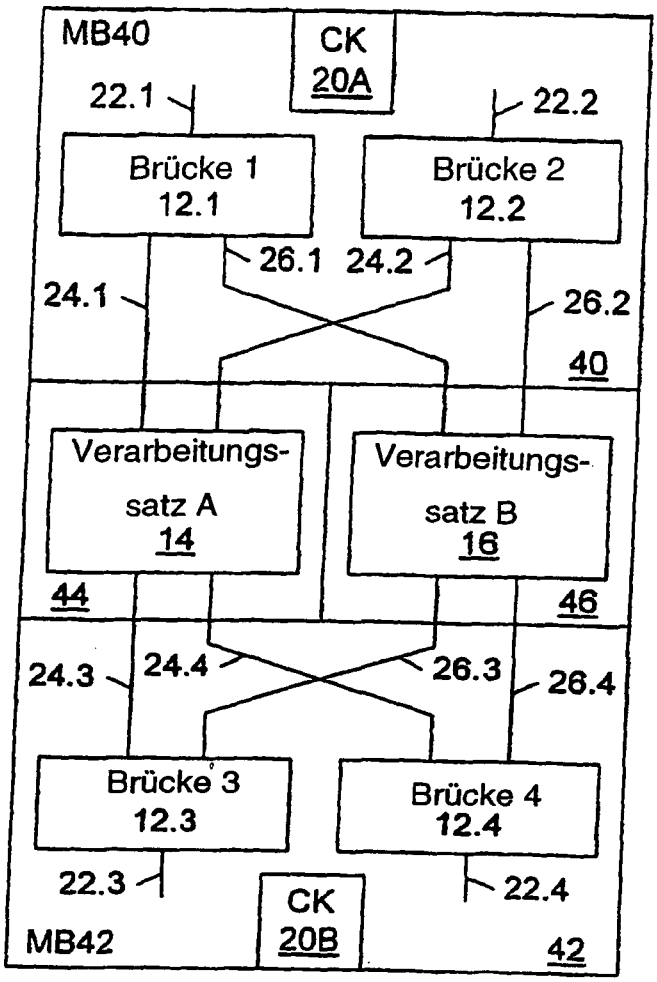
MB40
CK
20A
22.1
22.2
Brücke 1
12.1
Brücke 2
12.2
26.1
24.2
24.1
26.2
40
Verarbeitungs-
satz A
14
Verarbeitungs-
satz B
16
44
46
24.3
24.4
26.3
26.4
Brücke 3
12.3
Brücke 4
12.4
22.3
22.4
CK
20B
MB42
42

FIG. 2

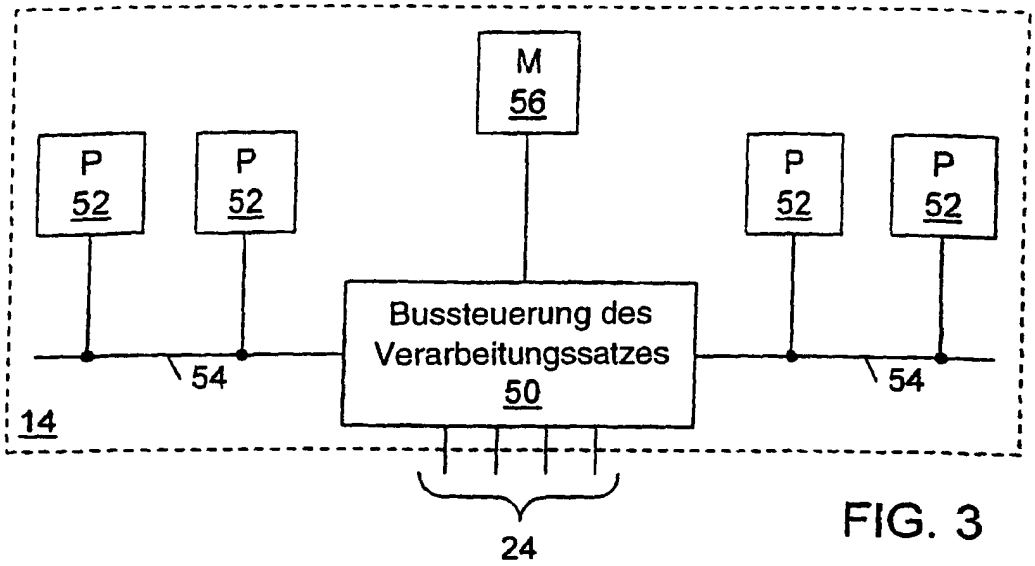

FIG. 3

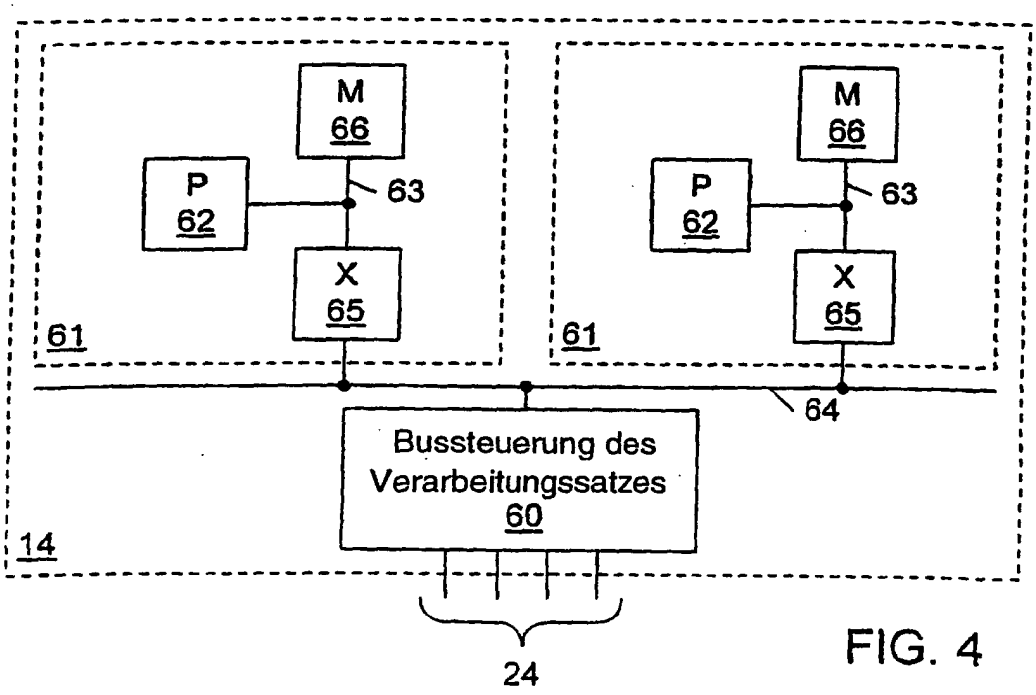

FIG. 4

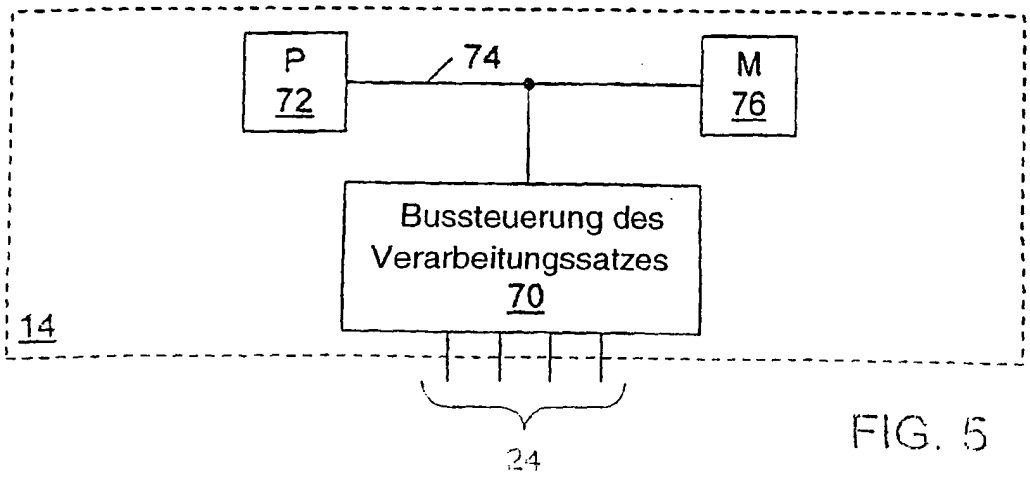

FIG. 5

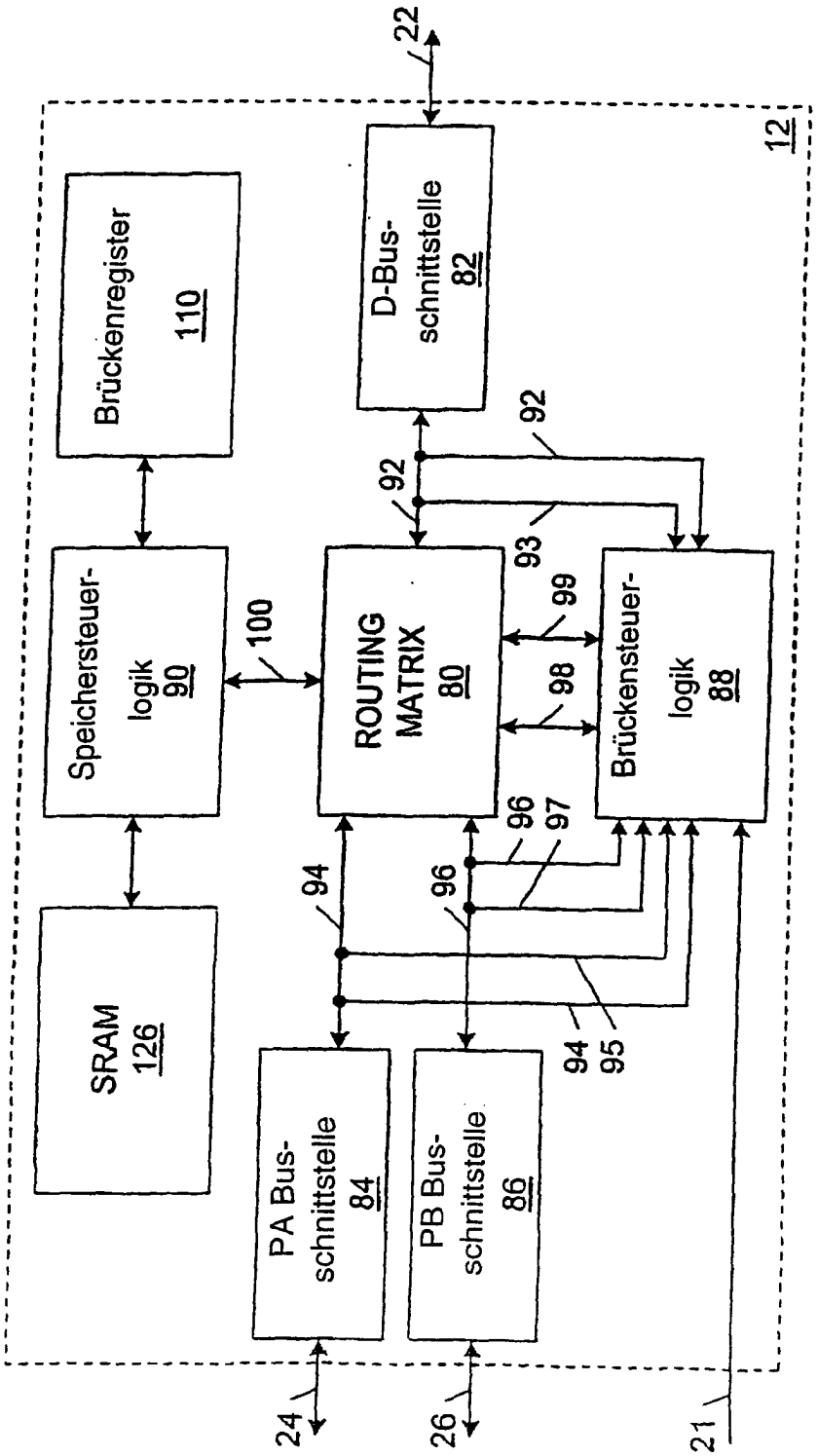
22
12
Brückenregister 110
D-Bus-schnittstelle 82
92
92
93
Speichersteuer-logik 90
100
ROUTING MATRIX 80
99
98
Brückensteuer-logik 88
96
97
94
96
SRAM 126
94
95
PA Bus-schnittstelle 84
PB Bus-schnittstelle 86
24
26
21

FIG. 6

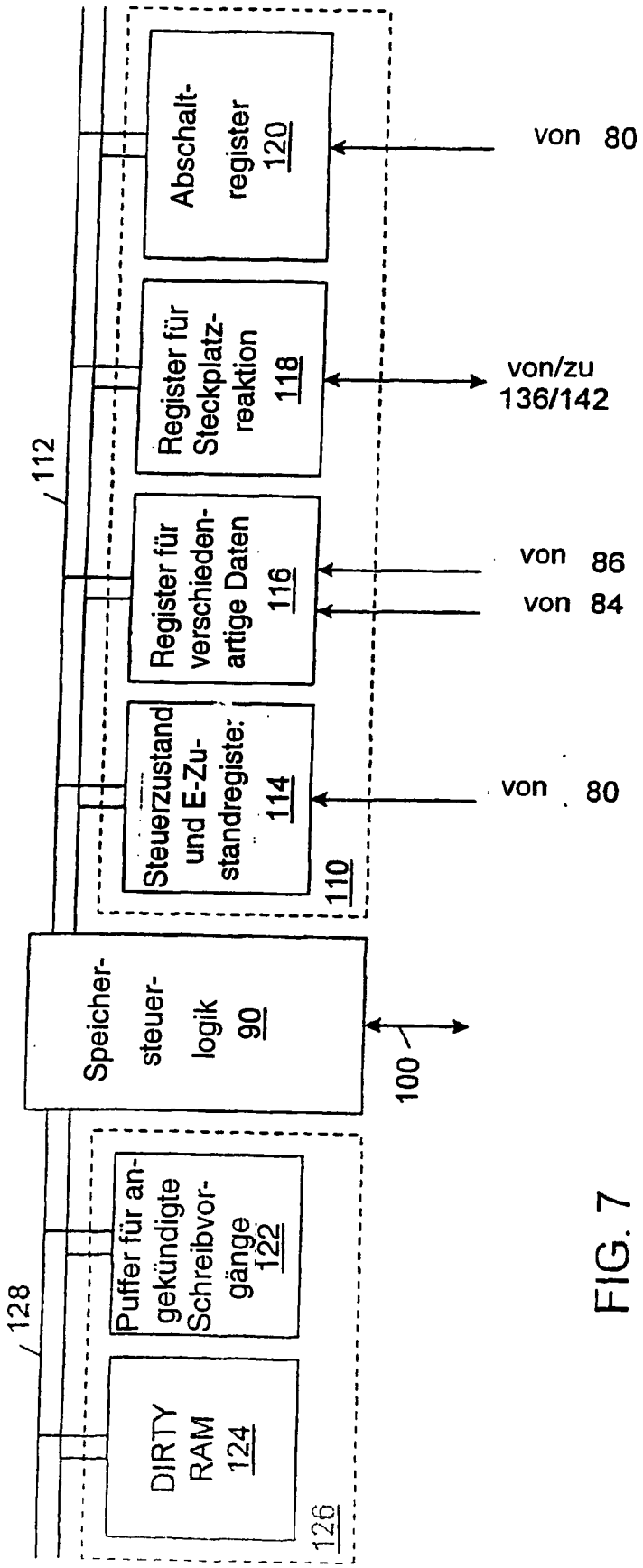
Abschalt-register 120
von 80
Register für Steckplatz-reaktion 118
von/zu 136/142
Register für verschieden-artige Daten 116
von 86
von 84
Steuerzustand und E-Zu-standregister 114
von 80
110
112
Speicher-steuer-logik 90
100
Puffer für an-gekündigte Schreibvor-gänge 122
DIRTY RAM 124
126
128

FIG. 7

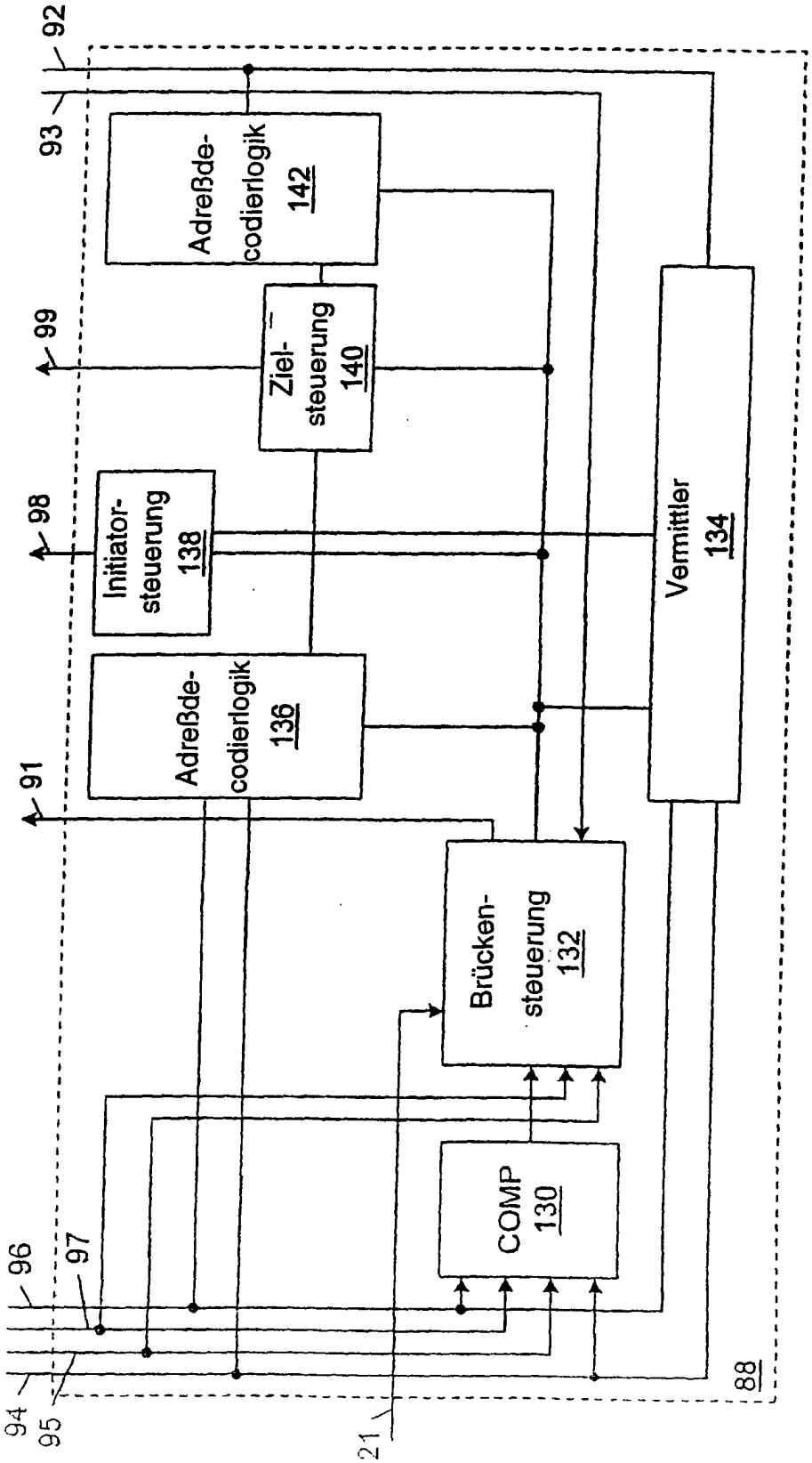
92
93
99
98
91
96
97
94
95
21
88
Adreßde-codierlogik 142
Ziel-steuerung 140
Initiator-steuerung 138
Adreßde-codierlogik 136
Brücken-steuerung 132
COMP 130
Vermittler 134

FIG. 8

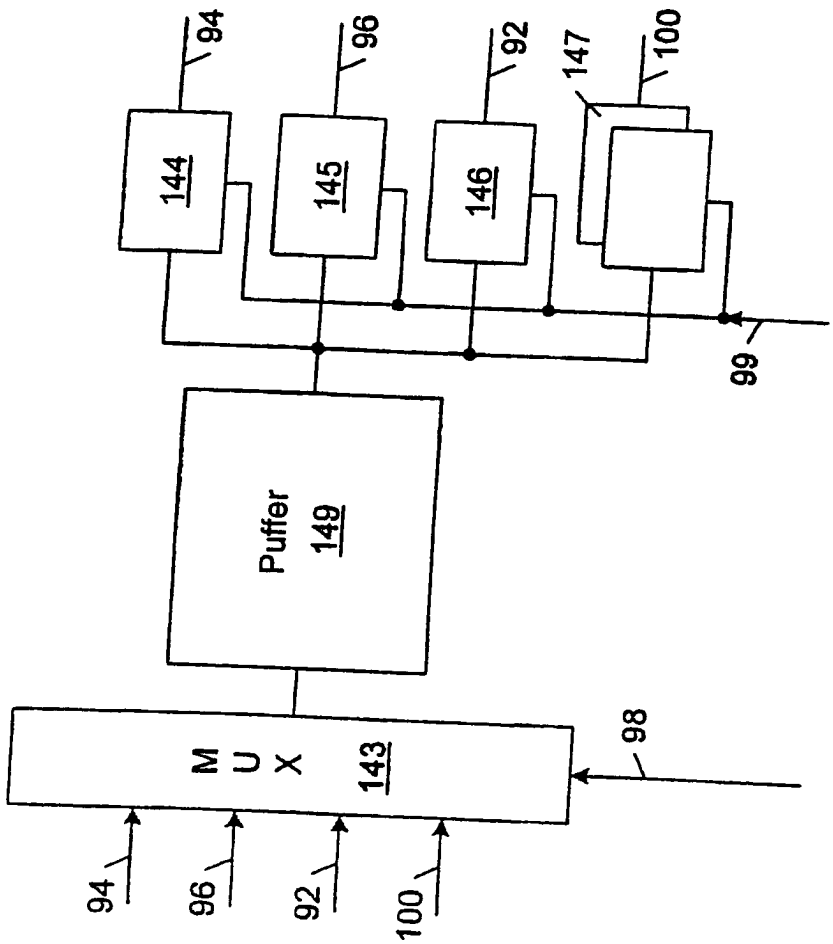

FIG. 9

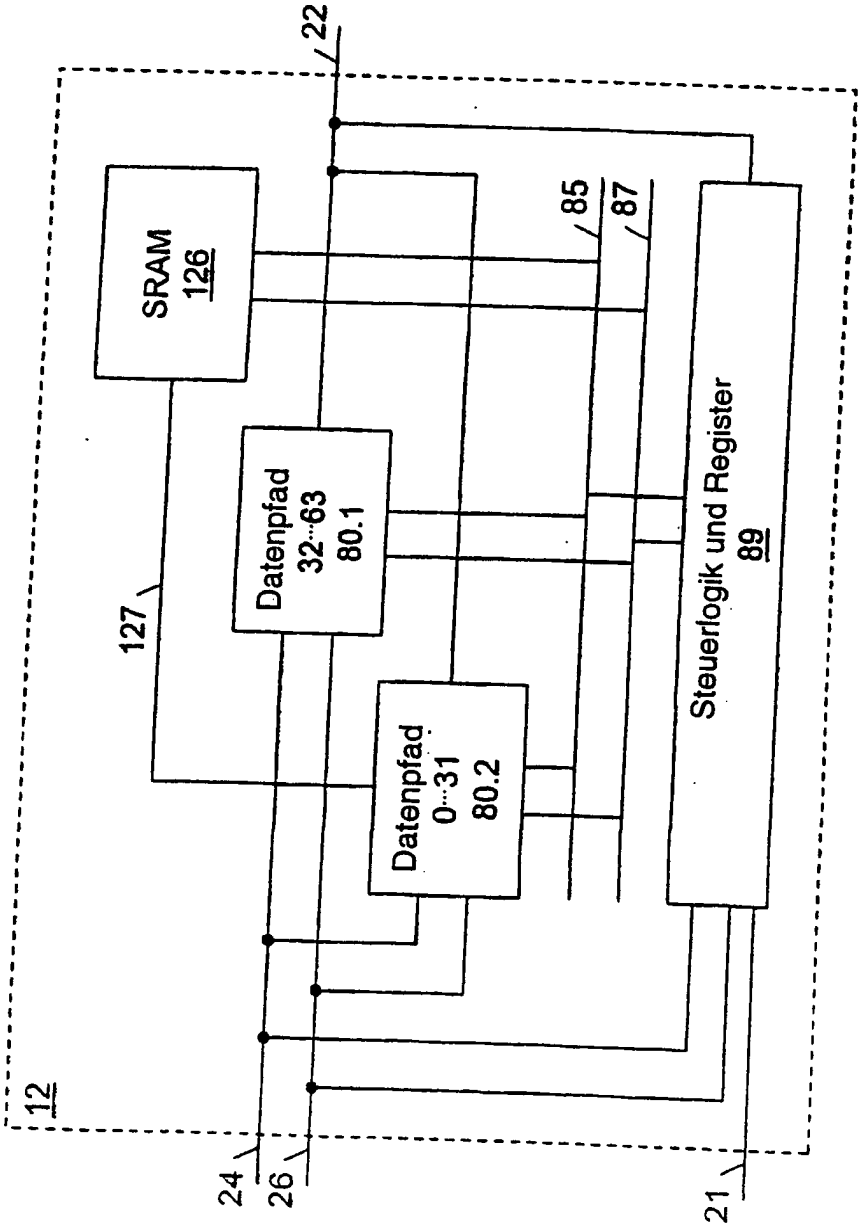
22
SRAM
126
85
87
Datenpfad
32...63
80.1
127
Datenpfad
0...31
80.2
Steuerlogik und Register
89
12
24
26
21

FIG. 10

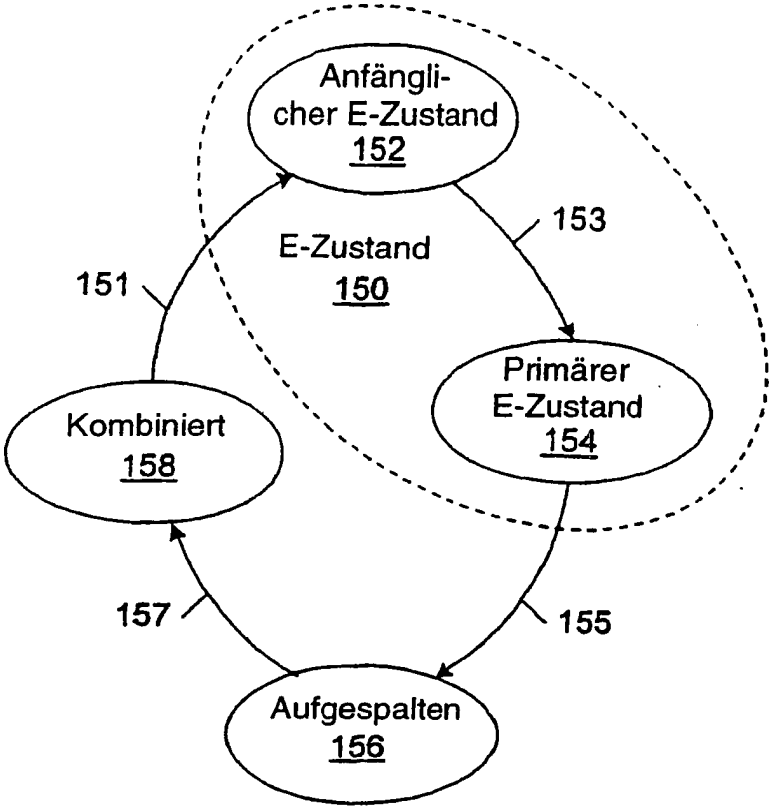
Anfängli-
cher E-Zustand
152
153
E-Zustand
150
151
Primärer
E-Zustand
154
Kombiniert
158
157
155
Aufgespalten
156

FIG. 11

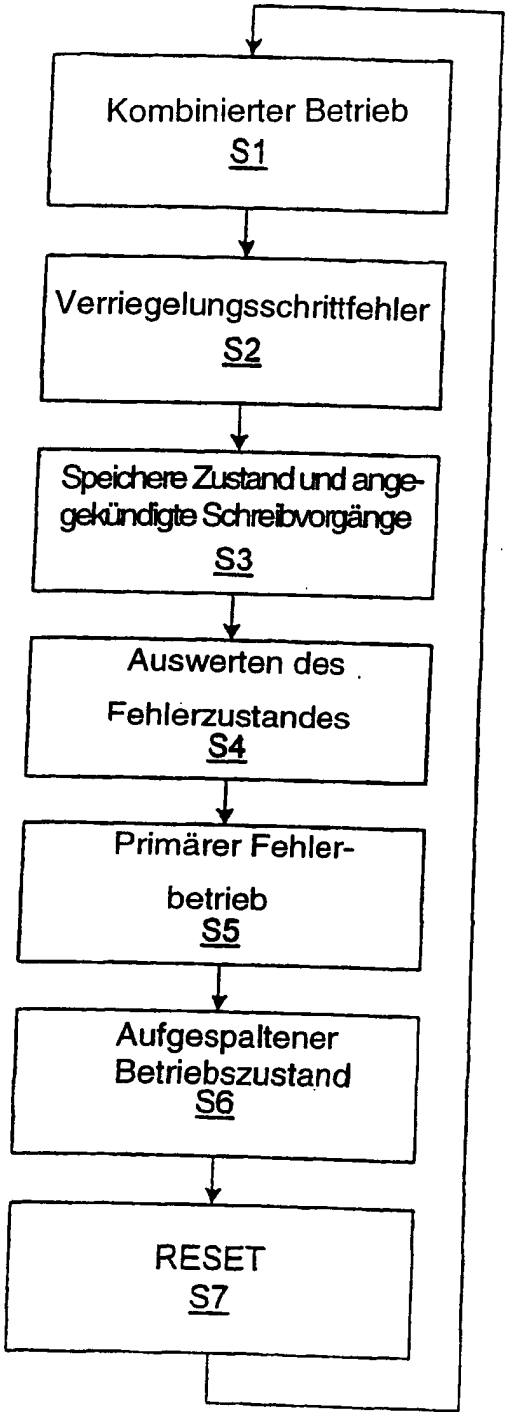
Kombinierter Betrieb
S1
Verriegelungsschrittfehler
S2
Speichere Zustand und angekündigte Schreibvorgänge
S3
Auswerten des Fehlerzustandes
S4
Primärer Fehlerbetrieb
S5
Aufgespaltener Betriebszustand
S6
RESET
S7

FIG. 12

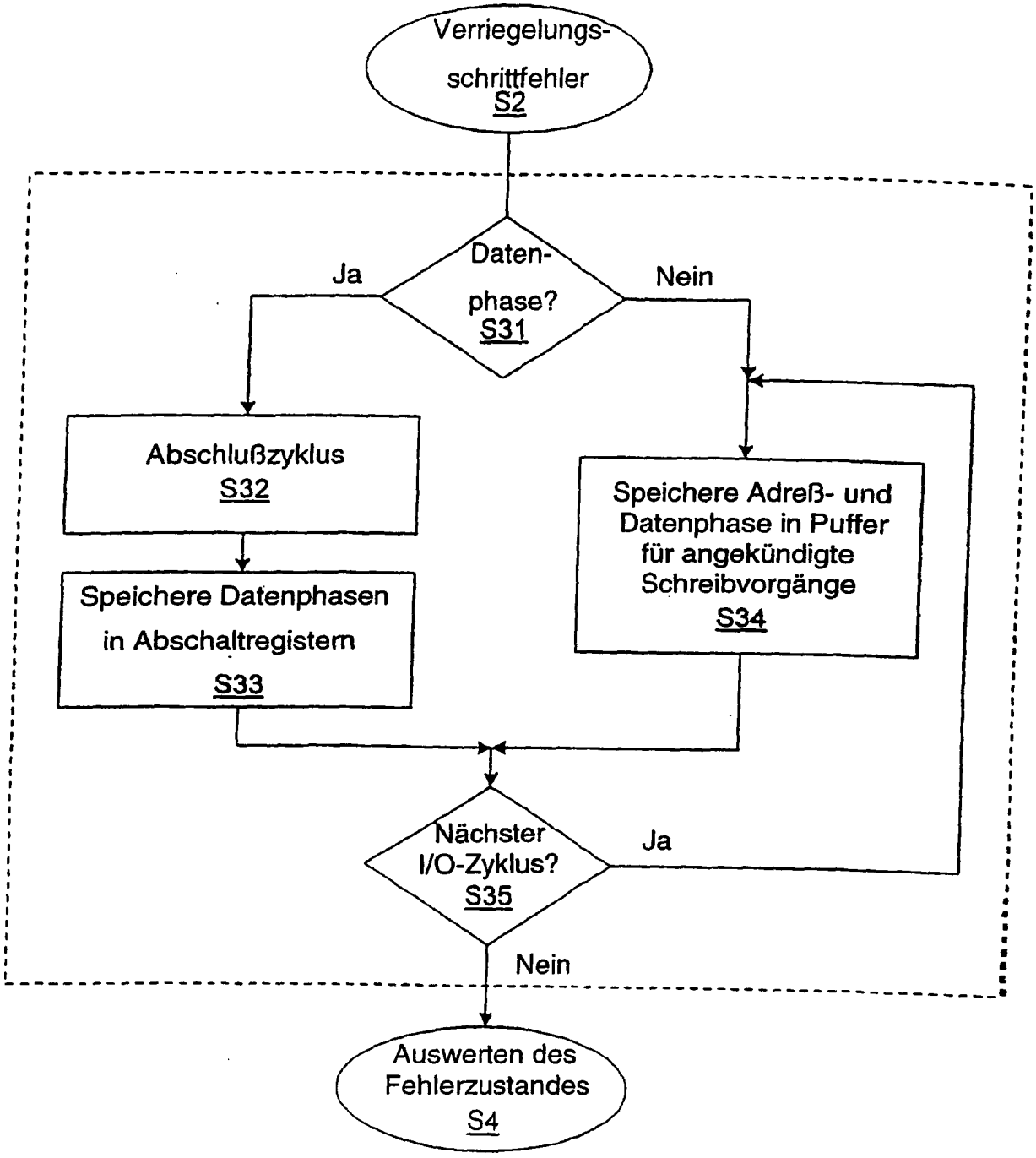
Verriegelungs-
schrittfehler
S2
Daten-
phase?
S31
Ja
Nein
Abschlußzyklus
S32
Speichere Datenphasen
in Abschaltregistern
S33
Speichere Adreß- und
Datenphase in Puffer
für angekündigte
Schreibvorgänge
S34
Nächster
I/O-Zyklus?
S35
Ja
Nein
Auswerten des
Fehlerzustandes
S4

FIG. 13

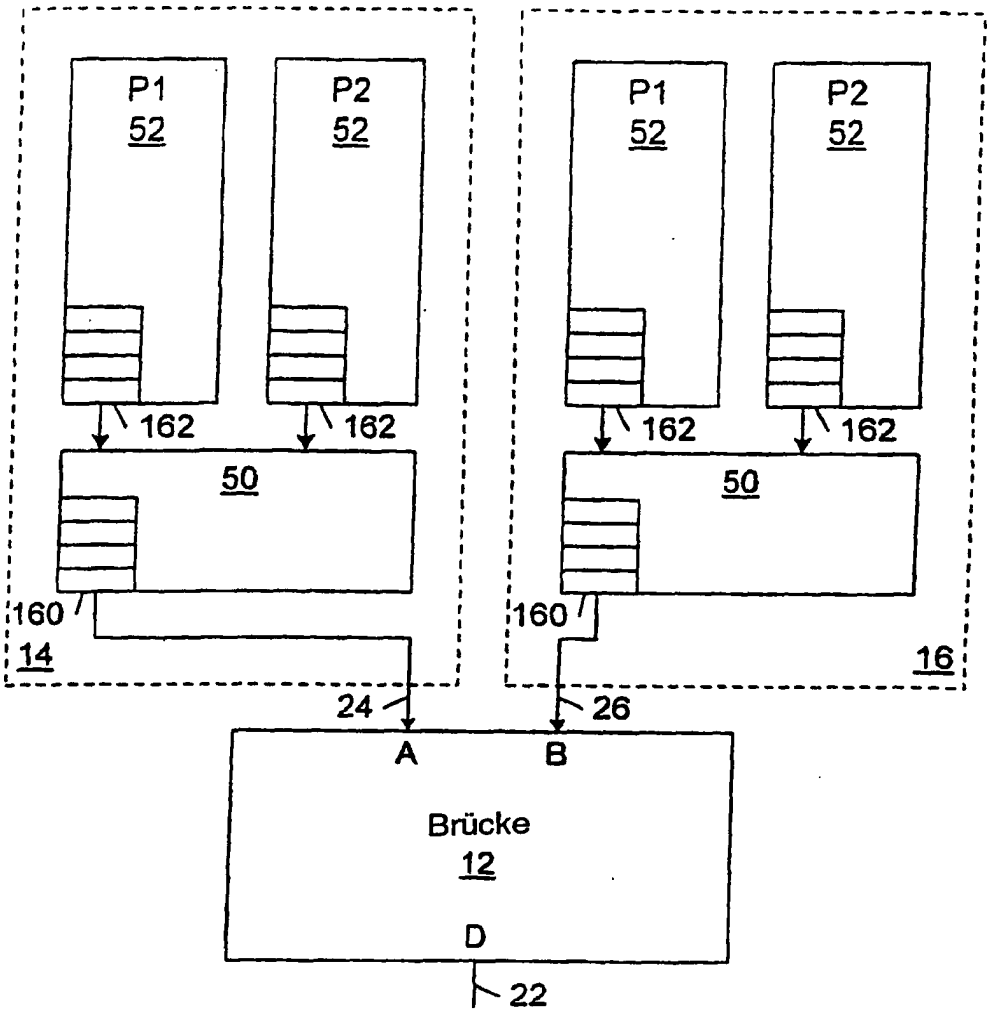
P1
52
P2
52
162
162
50
160
14
P1
52
P2
52
162
162
50
160
16
24
26
A
B
Brücke
12
D
22

FIG. 14

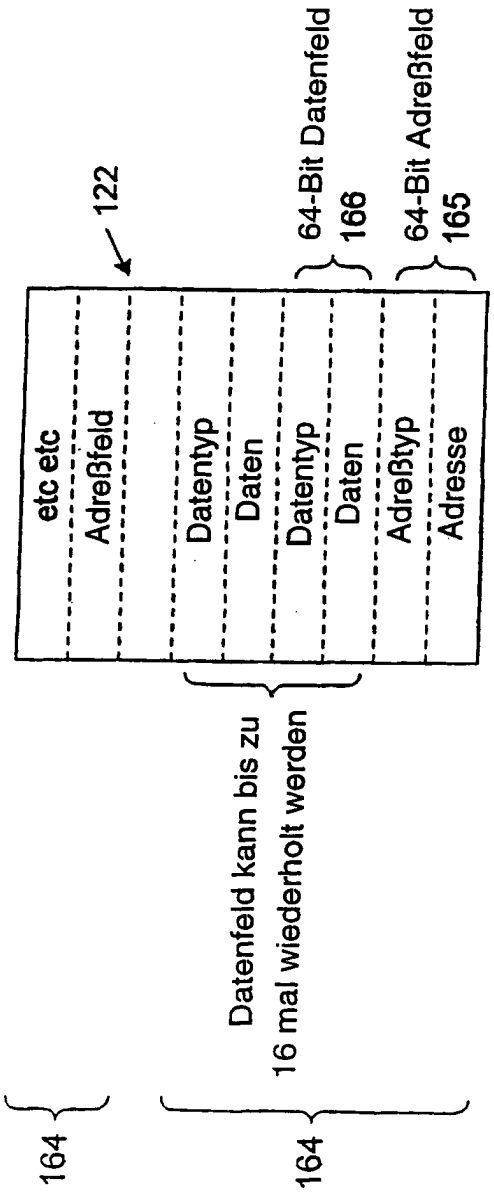
122
64-Bit Datenfeld
166
64-Bit Adreßfeld
165
etc etc
Adreßfeld
Datentyp
Daten
Datentyp
Daten
Adreßtyp
Adresse
Datenfeld kann bis zu
16 mal wiederholt werden
164
164

FIG. 15

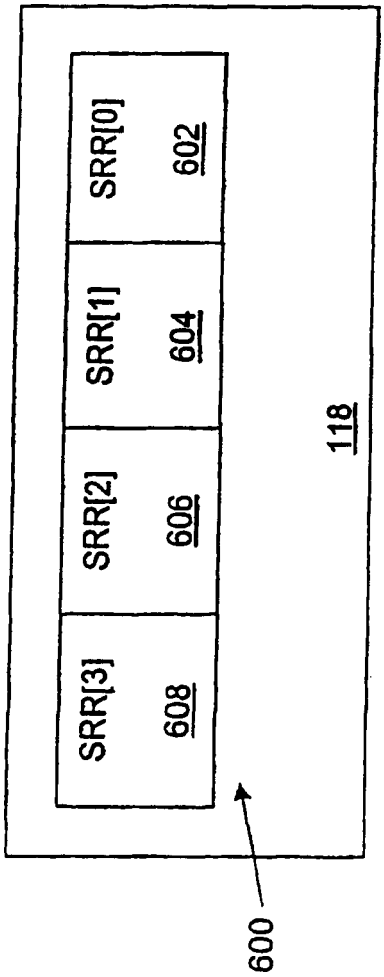

FIG. 16

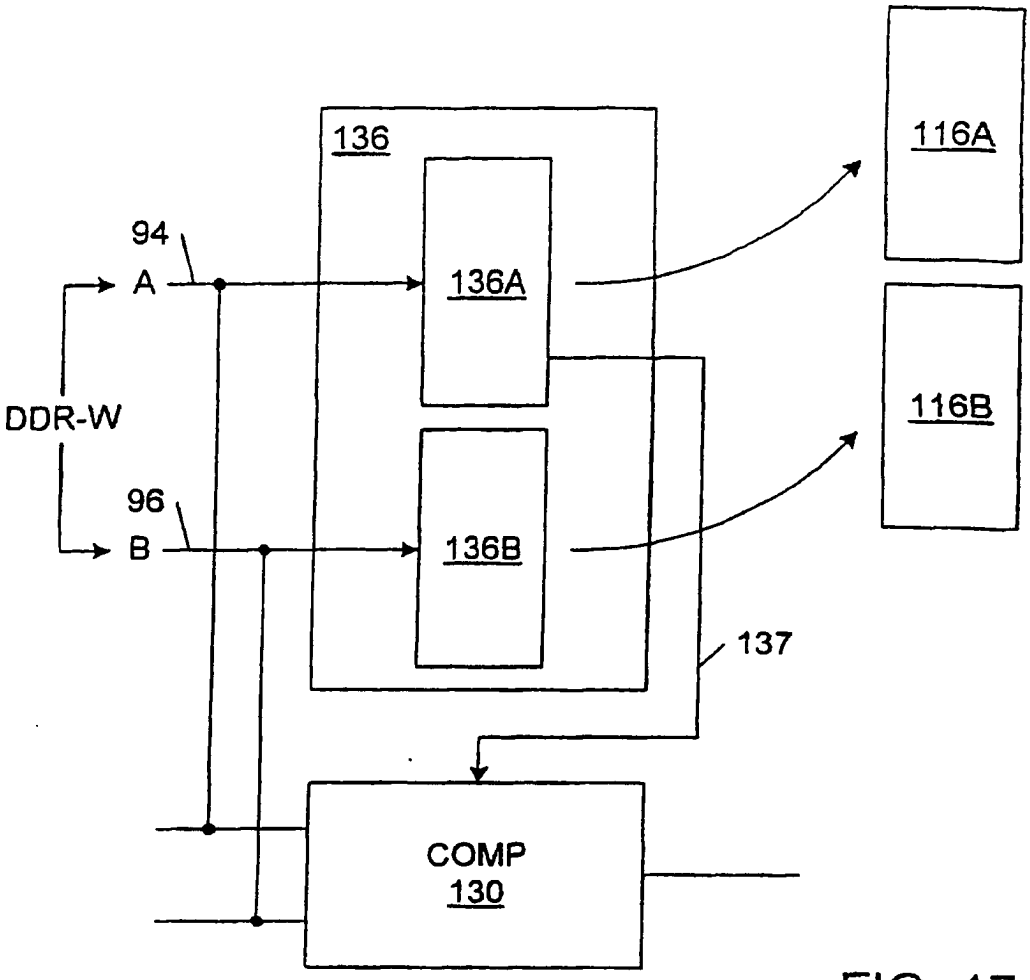
136
116A
94
A
136A
116B
DDR-W
96
B
136B
137
COMP
130

FIG. 17

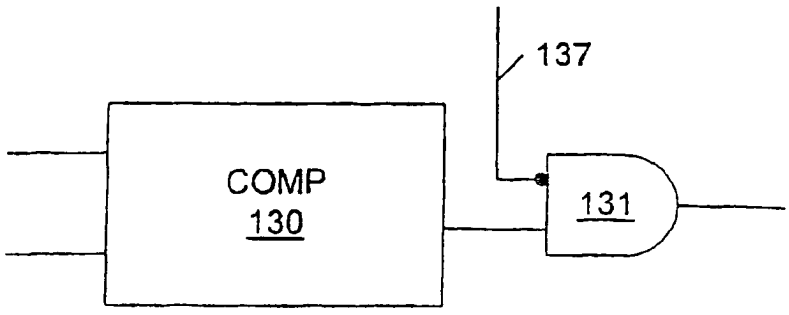
137
COMP
130
131

FIG. 18

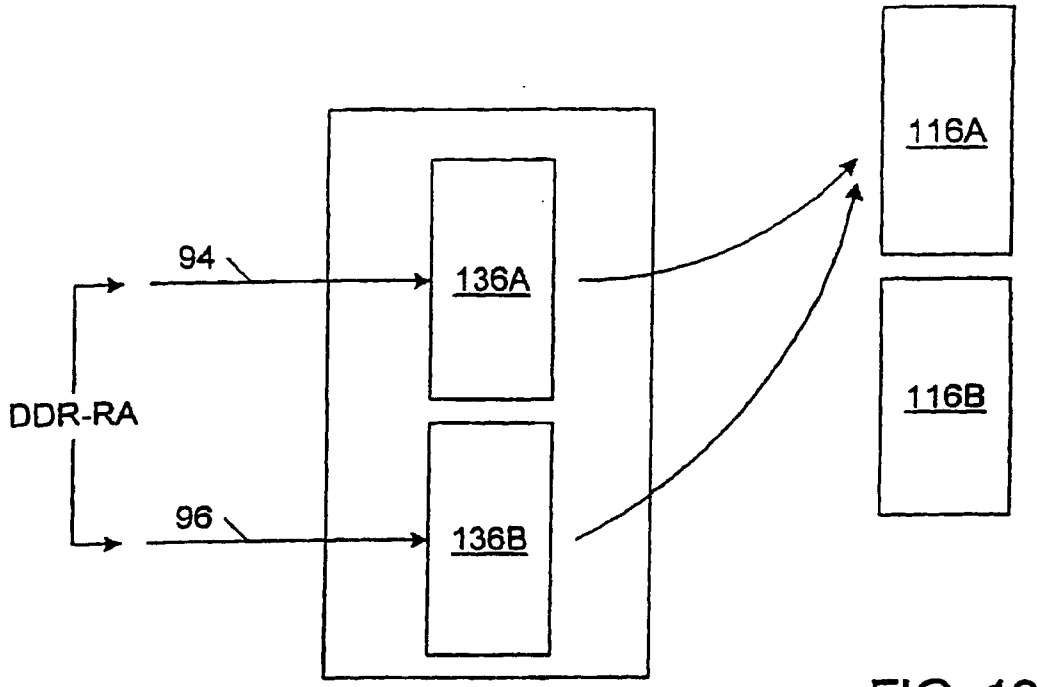
116A
94
136A
DDR-RA
116B
96
136B

FIG. 19

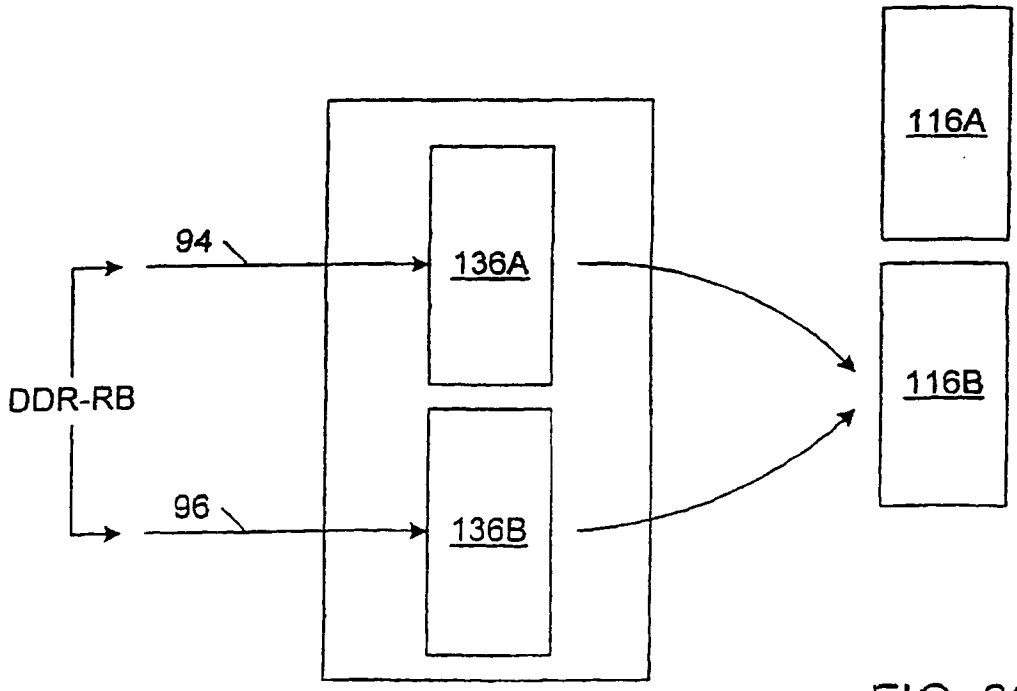
116A
94
136A
DDR-RB
116B
96
136B

FIG. 20

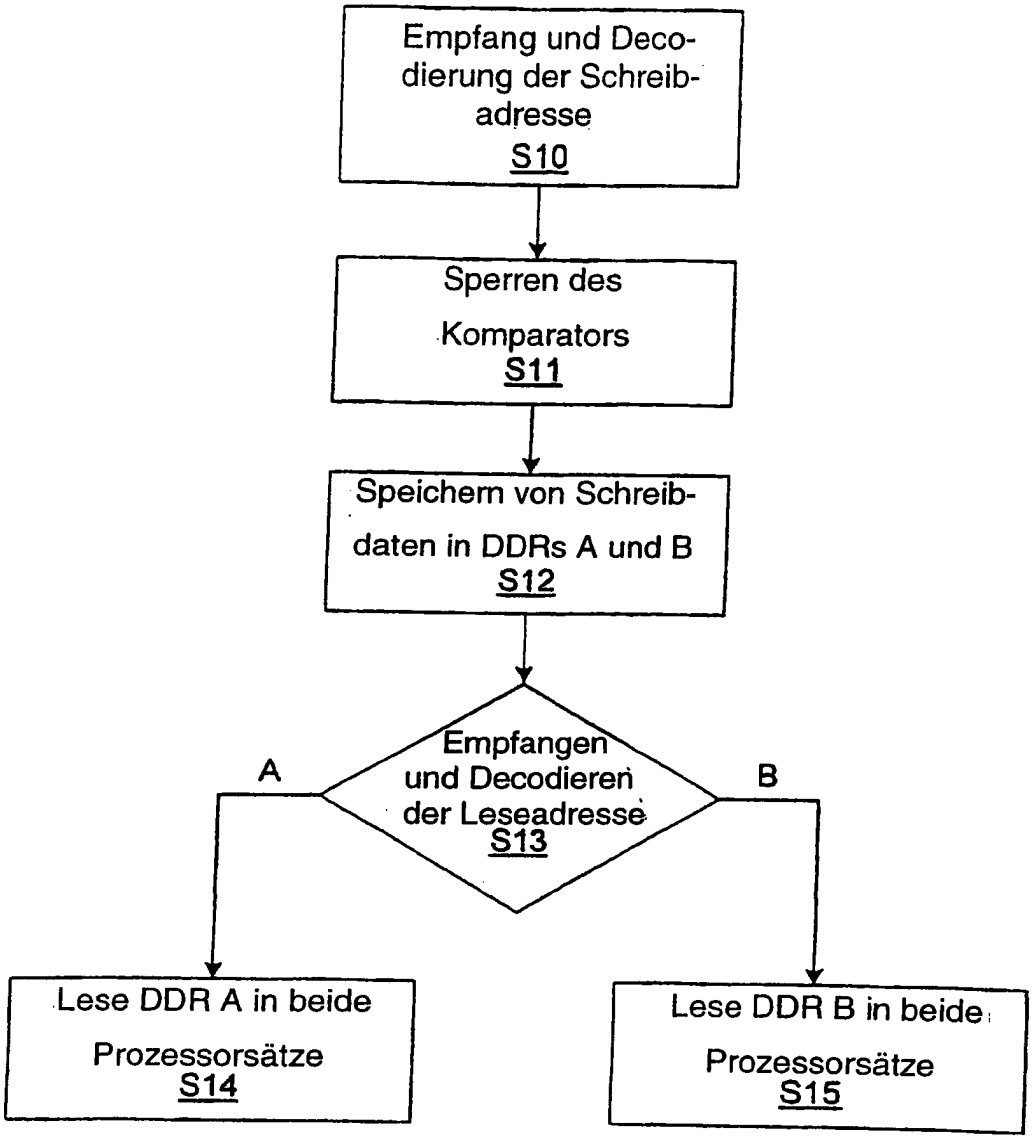
Empfang und Decodierung der Schreibadresse
S10
Sperren des Komparators
S11
Speichern von Schreibdaten in DDRs A und B
S12
Empfangen und Decodieren der Leseadresse
S13
A
B
Lese DDR A in beide Prozessorsätze
S14
Lese DDR B in beide Prozessorsätze
S15

FIG. 21

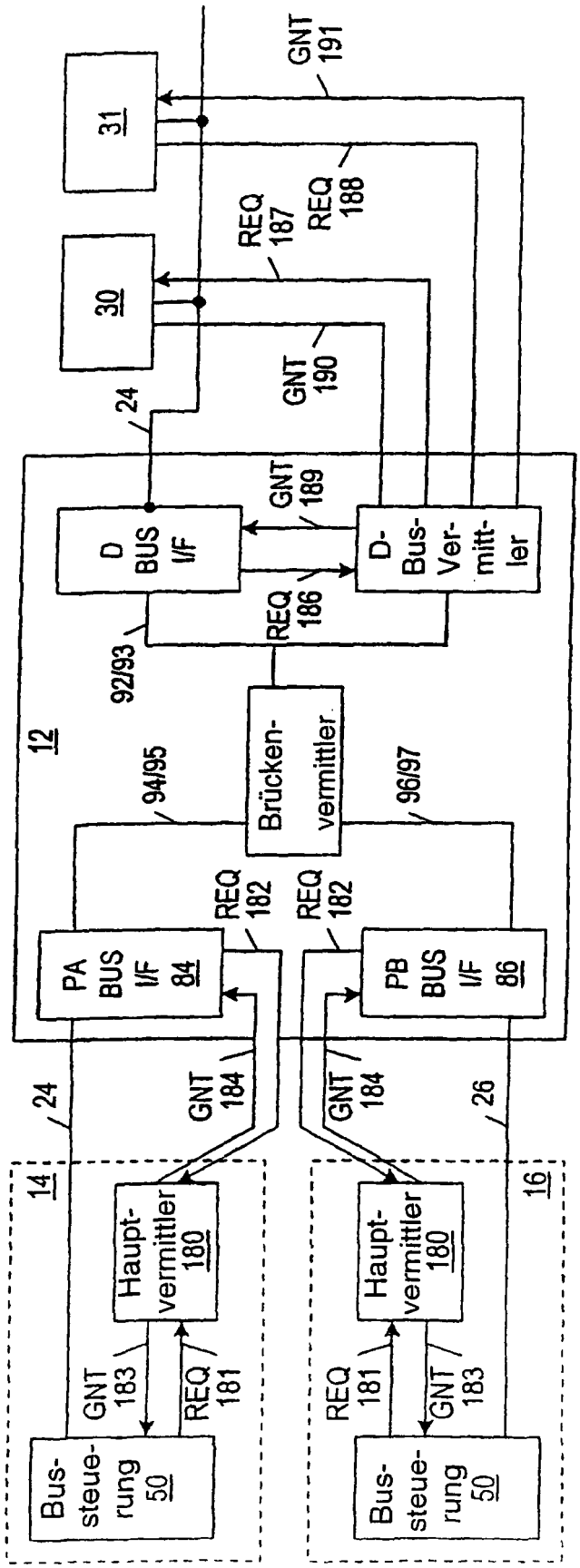
31
30
GNT 191
REQ 188
REQ 187
GNT 190
24
GNT 189
D BUS I/F
D-Bus-Ver-mitt-ler
REQ 186
92/93
12
94/95
Brücken-vermittler
96/97
REQ 182
REQ 182
PA BUS I/F 84
PB BUS I/F 86
GNT 184
GNT 184
24
26
14
16
Haupt-vermittler 180
Haupt-vermittler 180
GNT 183
REQ 181
REQ 181
GNT 183
Bus-steue-rung 50
Bus-steue-rung 50

FIG. 22

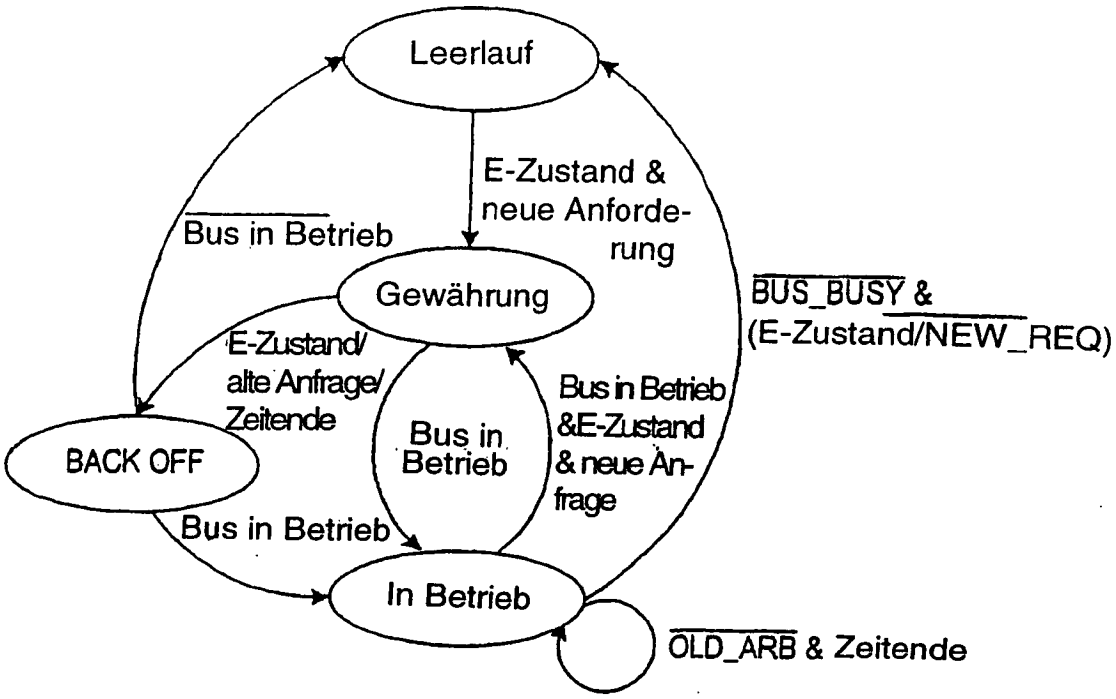
Leerlauf
E-Zustand &
neue Anforde-
rung
Bus in Betrieb
Gewährung
BUS_BUSY &
(E-Zustand/NEW_REQ)
E-Zustand/
alte Anfrage/
Zeitende
Bus in
Betrieb
Bus in Betrieb
&E-Zustand
& neue An-
frage
BACK OFF
Bus in Betrieb
In Betrieb
OLD_ARB & Zeitende

FIG. 23

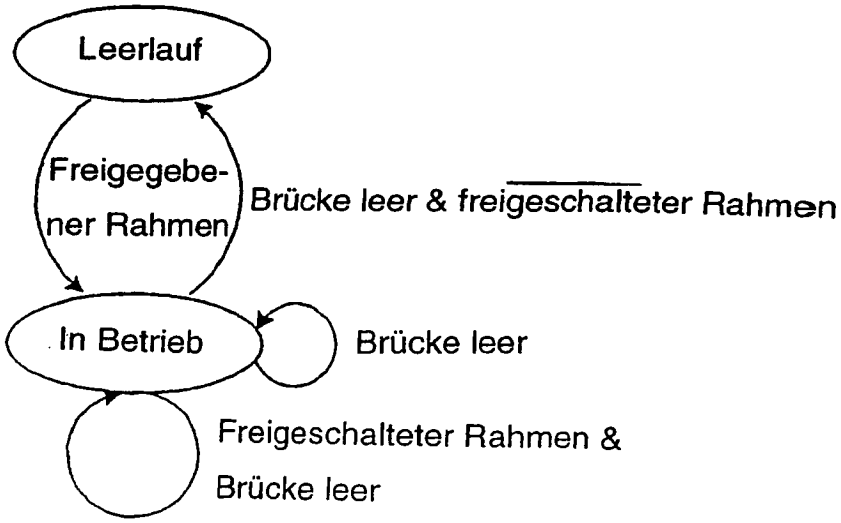
Leerlauf
Freigegebe-
ner Rahmen
Brücke leer & freigeschalteter Rahmen
In Betrieb
Brücke leer
Freigeschalteter Rahmen &
Brücke leer

FIG. 24

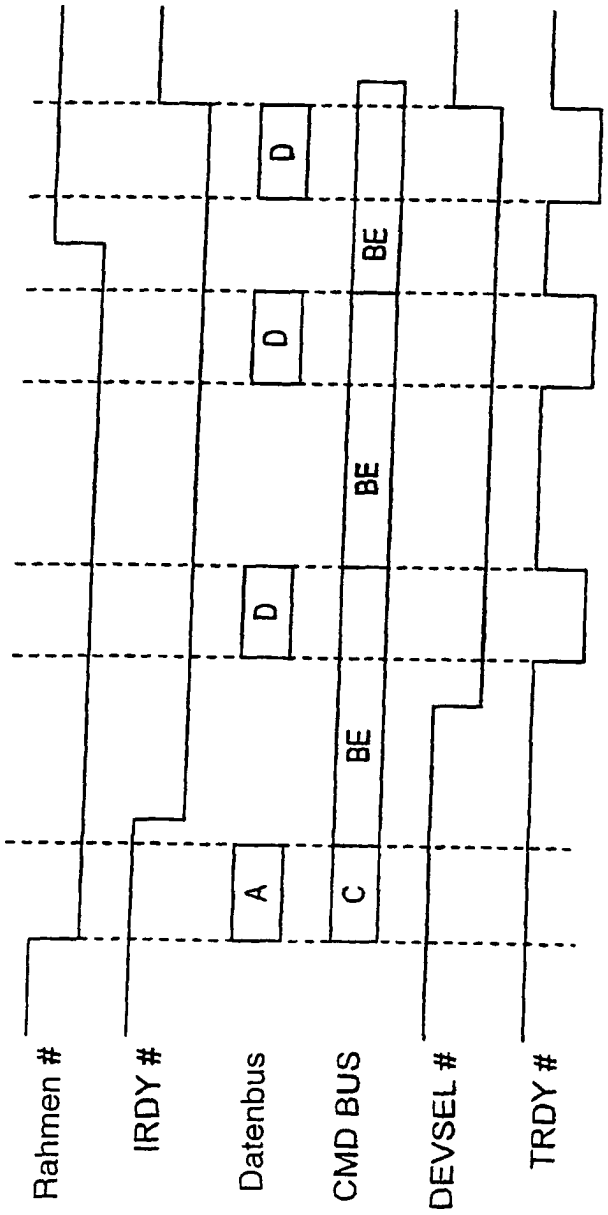

FIG. 25

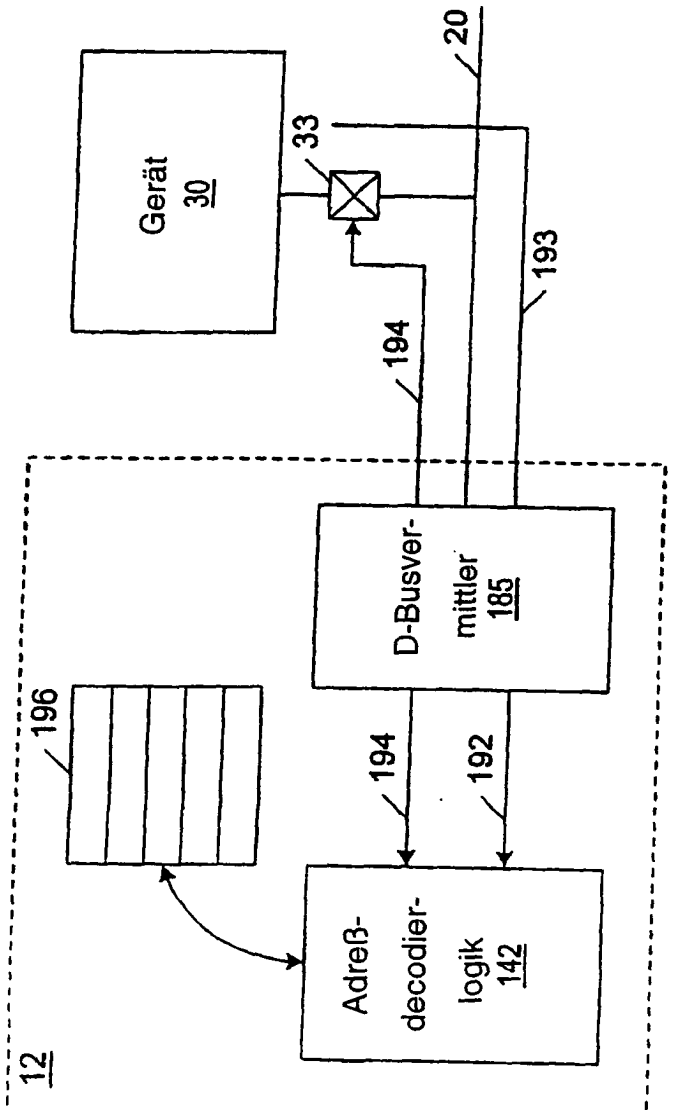
Gerät
30
33
20
193
194
D-Busver-
mittler
185
196
194
192
Adreß-
decodier-
logik
142
12

FIG. 26

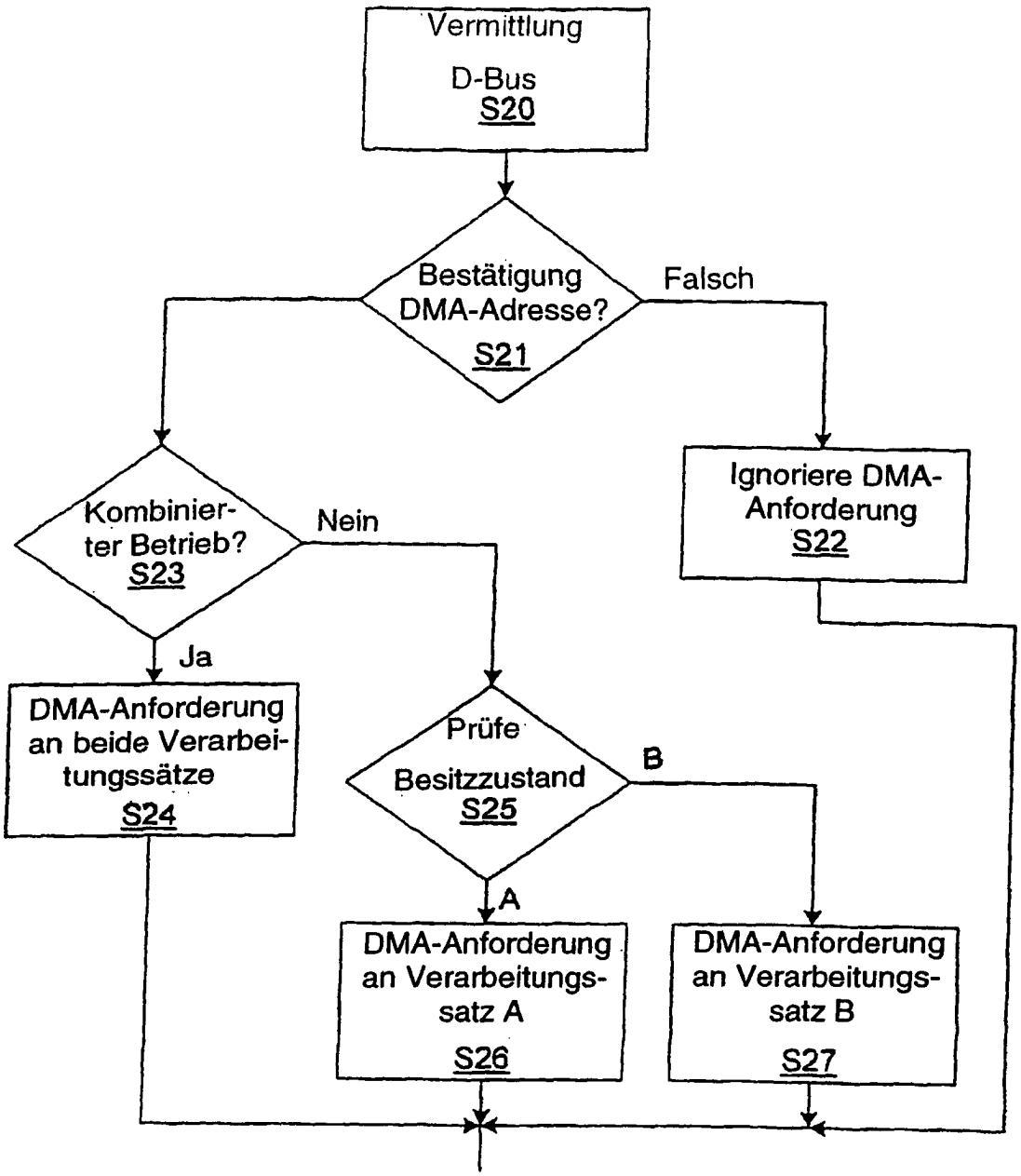
Vermittlung
D-Bus
S20
Bestätigung DMA-Adresse?
S21
Falsch
Ignoriere DMA-Anforderung
S22
Kombinierter Betrieb?
S23
Nein
Ja
DMA-Anforderung an beide Verarbeitungssätze
S24
Prüfe Besitzzustand
S25
B
A
DMA-Anforderung an Verarbeitungssatz A
S26
DMA-Anforderung an Verarbeitungssatz B
S27

FIG. 27

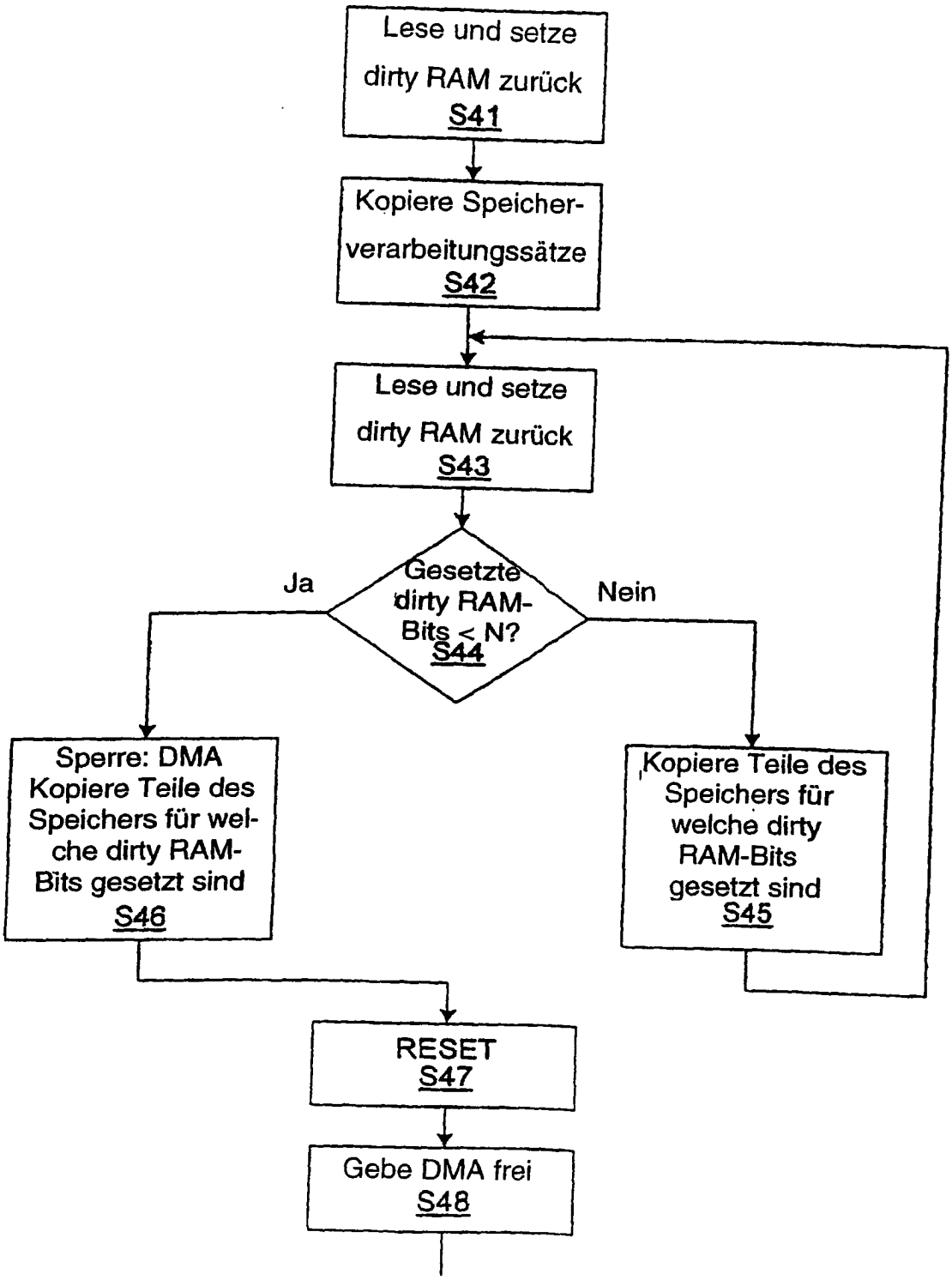
Lese und setze dirty RAM zurück S41
Kopiere Speicher-verarbeitungssätze S42
Lese und setze dirty RAM zurück S43
Gesetzte dirty RAM-Bits < N? S44
Ja
Nein
Sperre: DMA Kopiere Teile des Speichers für welche dirty RAM-Bits gesetzt sind S46
Kopiere Teile des Speichers für welche dirty RAM-Bits gesetzt sind S45
RESET S47
Gebe DMA frei S48

FIG. 28