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

(10) **DE 695 34 363 T2** 2006.07.13

(12) # Übersetzung der europäischen Patentschrift

(97) **EP 0 706 113 B1**
(21) Deutsches Aktenzeichen: **695 34 363.7**
(96) Europäisches Aktenzeichen: **95 108 146.2**
(96) Europäischer Anmeldetag: **26.05.1995**
(97) Erstveröffentlichung durch das EPA: **10.04.1996**
(97) Veröffentlichungstag der Patenterteilung beim EPA: **10.08.2005**
(47) Veröffentlichungstag im Patentblatt: **13.07.2006**

(51) Int Cl.[8]: ***G06F 3/06*** (2006.01)

(30) Unionspriorität:
319385 **05.10.1994** **US**

(84) Benannte Vertragsstaaten:
DE, FR, GB

(73) Patentinhaber:
Hewlett-Packard Development Co., L.P., Houston, Tex., US

(74) Vertreter:
Schoppe, Zimmermann, Stöckeler & Zinkler, 82049 Pullach

(72) Erfinder:
Jacobson, Michael Brent, Boise, Idaho 83714, US; Voigt, Douglas L., Boise, Idaho 83702, US; Burkes, Theresa A., Meridian, Idaho 83642, US; Diamond, Bryan, Boise, Idaho 83704, US

(54) Bezeichnung: **Verfahren um Festplatten zu einer Festplattenanordnung hinzuzufügen und gleichzeitig die Datenerreichbarkeit zu gewährleisten**

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

Beschreibung

GEBIET DER ERFINDUNG

[0001] Diese Erfindung bezieht sich auf Datenspeichersysteme, z.B. hierarchische Plattenarraydatenspeichersysteme, und Verfahren zum Hinzufügen von Platten zu solchen Systemen.

HINTERGRUND DER ERFINDUNG

[0002] Computersysteme werden in Bezug auf Geschwindigkeit, Zuverlässigkeit und Verarbeitungsfähigkeit ständig verbessert. Folglich sind Computer in der Lage, komplexere und anspruchsvollere Anwendungen zu handhaben. Während jedoch Computer immer besser werden, nehmen auch die Leistungsanforderungen zu, die an Massenspeicher- und Eingabe-/Ausgabevorrichtungen (I-/O-Vorrichtungen) gestellt werden. Es besteht weiterhin ein Bedarf, Massenspeichersysteme zu entwerfen, die in Bezug auf die Leistungsfähigkeit mit sich weiter entwickelnden Computersystemen Schritt halten.

[0003] Diese Erfindung betrifft insbesondere die Massenspeichersysteme des Plattenarray-Typs. Plattenarraydatenspeichersysteme weisen Mehrfach-Speicherplattenlaufwerksvorrichtungen auf, die angeordnet und koordiniert sind, um ein einziges Massenspeichersystem zu bilden. Für derartige Speichersysteme gibt es drei Hauptentwurfskriterien: Kosten, Leistungsfähigkeit und Verfügbarkeit. Am wünschenswertesten ist es, Speichervorrichtungen zu erzeugen, die geringe Kosten pro Megabyte, eine hohe Eingangs-/Ausgangsleistungsfähigkeit und eine hohe Datenverfügbarkeit aufweisen. „Verfügbarkeit" ist die Fähigkeit, auf Daten, die in dem Speichersystem gespeichert sind, zuzugreifen, und die Fähigkeit, im Fall eines Ausfalls einen fortgesetzten Betrieb zu gewährleisten. In der Regel wird eine Datenverfügbarkeit durch die Verwendung einer Redundanz geliefert, wobei Daten, oder Beziehungen zwischen Daten, an mehreren Plätzen gespeichert sind. Für den Fall, dass eine Speicherplatte in dem Plattearray teilweise oder vollständig ausfällt, können die Benutzerdaten über die auf den verbleibenden Platten gespeicherten redundanten Daten rekonstruiert werden.

[0004] Es gibt zwei übliche Verfahren zum Speichern von redundanten Daten. Gemäß dem ersten bzw. „Spiegel"-Verfahren werden Daten dupliziert und in zwei getrennten Bereichen des Speichersystems gespeichert. Bei einem Plattenarray werden die identischen Daten beispielsweise auf zwei getrennten Platten in dem Plattenarray bereitgestellt. Das Spiegelverfahren weist die Vorteile einer hohen Leistungsfähigkeit und einer hohen Datenverfügbarkeit aufgrund der Duplexspeichertechnik auf. Das Spiegelverfahren ist jedoch auch relativ teuer, da es die Kosten des Speicherns von Daten effektiv verdoppelt.

[0005] Bei dem zweiten bzw. „Paritäts"-Verfahren wird ein Teil des Speicherbereichs verwendet, um redundante Daten zu speichern, die Größe des redundanten Speicherbereichs ist jedoch geringer als der zum Speichern der ursprünglichen Daten verwendete verbleibende Speicherbereich. Bei einem Plattenarray, das fünf Platten aufweist, könnten beispielsweise vier Platten verwendet werden, um Daten zu speichern, wobei die fünfte Platte für das Speichern von redundanten Daten reserviert ist. Das Paritätsverfahren ist vorteilhaft, da es kostengünstiger ist als das Spiegelverfahren; es weist jedoch auch geringere Leistungsfähigkeits- und Verfügbarkeitscharakteristika im Vergleich zu dem Spiegelverfahren auf.

[0006] Ein Problem, auf das man beim Entwurf von Plattenarraydatenspeichersystemen stößt, betrifft die Aufgabe des Hinzufügens von mehr Speicherplattenlaufwerksvorrichtungen, um die Speicherkapazität des Plattenarrays zu erhöhen. Herkömmliche Plattenarrays erfordern den folgenden langwierigen Prozess. Zuerst werden die derzeit auf dem Plattenarray gespeicherten Daten in eine gänzlich separate Speichervorrichtung (z.B. ein Zweirollen-Bandspeichersystem) kopiert. Als Nächstes werden die neuen Speicherplatten zu dem Plattenarray hinzugefügt. Das gesamte Plattenarray wird anschließend dahingehend umkonfiguriert, die neuen Speicherplatten zu enthalten. Schließlich werden die Daten von der separaten Speichervorrichtung in das neu konfigurierte Plattenarray zurückkopiert. Ungünstigerweise ist das Plattenarray während dieses Prozesses nicht zugänglich, und die Daten, die vorübergehend an ein separates Sicherungssystem transferiert wurden, stehen nicht zur Verfügung.

[0007] Die EPA-4,851 110 betrifft das logische Partitionieren bzw. Unterteilen eines Redundantes-Array-Speichersystems, das als redundantes Array von kostengünstigen Platten (RAID – redundant array of inexpensive disks) konfiguriert sein kann. Die logische Organisation der Daten in Redundanzgruppen unterschiedlicher Redundanzebenen kann automatisch verändert werden, ohne die normalen Zugriffsvorgänge anzuhalten.

[0008] Das Dokument EP-A-0 654 736 stellt den Stand der Technik gemäß Artikel 54(3) EPÜ dar. Es bezieht sich auf dynamisch erweiterbare Speichereinheitsarraysysteme. Auf ein Erhöhen der Anzahl von Speichervorrichtungen hin liest eine Datenneuordnungsfunktion Daten aus den ursprünglichen Speichereinheiten in einen Cache-Speicher der Plattenarraysteuervorrichtung ein. Dieser Cache-Speicher ist separat von den jeweiligen Plattenlaufwerken.

[0009] Die vorliegende Erfindung beruht auf der Aufgabe des Bereitstellens eines Verfahrens zum Vergrößern der Speicherkapazität in einem Datenspeichersystem, das ein Plattenarray von mehreren Speicherplatten aufweist, in denen die Verfügbarkeit von Daten aufrechterhalten wird.

[0010] Diese Aufgabe wird durch ein Verfahren gemäß Anspruch 1 gelöst.

[0011] Weitere Verbesserungen sind in den Unteransprüchen definiert.

[0012] Diese Erfindung kann ferner auf eindeutige Weise auf eine Verwendung bei hierarchischen Plattenarraydatenspeichersystemen zugeschnitten sein, die Daten gemäß unterschiedlicher Redundanzebenen bzw. RAID-Ebenen speichern. Dieses Verfahren umfasst zuerst ein Abbilden des physischen Speicherraums in einen ersten virtuellen Speicherraum, der eine Mehrzahl von RAID-Bereichen aufweist. Jeder RAID-Bereich kann entweder als ein erster oder ein zweiter RAID-Bereich-Typ konfiguriert sein, wobei ein erster RAID-Bereich-Typ Daten gemäß einer ersten RAID-Ebene (z.B. der RAID-Ebene **1** oder Spiegelredundanz) hält und ein zweiter RAID-Bereich-Typ Daten gemäß einer zweiten RAID-Ebene (z.B. RAID-Ebene **5** oder Paritätsredundanz) hält. Diese RAID-Bereiche erstrecken sich über mehrere Speicherplatten in dem Plattenarray und weisen eine gleiche Größe aufweisende Speicherregionen von jeder Speicherplatte auf. Der erste virtuelle Speicherraum kann anschließend in einen zweiten virtuellen Speicherraum abgebildet werden, der mehrere virtuelle Blöcke aufweist. Dieser zweite virtuelle Speicherraum ist die Speicherungsansicht, die durch den Benutzer oder Host definiert und demselben präsentiert wird.

[0013] In diesem Fall werden, wenn eine oder mehrere Speicherplatten zu dem hierarchischen Plattenarray hinzugefügt werden, Daten, die in einem ausgewählten RAID-Bereich gespeichert sind, zu einem anderen Bereich einer gleichen oder einer größeren Speicherkapazität bewegt. Das Abbilden des ersten in den zweiten virtuellen Speicherraum wird aktualisiert, um die Datenbewegung widerzuspiegeln. Der physische Speicherraum, der dem ausgewählten RAID-Bereich entspricht, wird anschließend in einen erweiterten RAID-Bereich neu abgebildet, der alle Speicherplatten, einschließlich der neuen Platten, überspannt. Dieser Prozess wird wiederholt, bis alle RAID-Bereiche in dem hierarchischen Plattenarray dahingehend erweitert wurden, dass sie Regionen auf den neuen Speicherplatten umfassen. Das Abbilden des zweiten virtuellen Speicherraums in den ersten wird kontinuierlich modifiziert, um die Erweiterung der RAID-Bereiche in dem virtuellen Speicherraum widerzuspiegeln.

[0014] Die Daten bleiben während dieses Kapazitätsvergrößerungsprozesses verfügbar, um die neue Platte zu assimilieren. Die Daten sind weiterhin auf dem Plattenarray vorhanden und zugänglich und werden nicht an eine andere separate Speichervorrichtung transferiert. Nur die Daten, die sich infolge eines Bewegens bzw. eines Verschiebens in einem Umabbildungsprozess befinden, werden für einen Zugriff gesperrt, aber auch nur vorübergehend. Die Verfahren dieser Erfindung gewährleisten somit auf effektive Weise, dass die Daten dem Benutzer jederzeit zur Verfügung stehen.

BESCHREIBUNG DER ZEICHNUNGEN

[0015] Unter Bezugnahme auf die folgenden beigefügten Zeichnungen, in denen Beispiele gezeigt sind, die den besten Modus zum Praktizieren der Erfindung verkörpern, werden nachfolgend bevorzugte Ausführungsbeispiele der Erfindung beschrieben.

[0016] Fig. 1 ist ein schematisches Blockdiagramm eines Speichersystems gemäß dieser Erfindung;

[0017] Fig. 2 ist eine schematische Veranschaulichung, die eine Datenspeicherung auf RAID-Ebene **1** zeigt;

[0018] Fig. 3 ist eine schematische Veranschaulichung, die eine Datenspeicherung auf RAID-Ebene **5** zeigt;

[0019] Fig. 4 ist eine schematische Veranschaulichung einer Speicherabbildungsanordnung dieser Erfindung, bei der zwei virtuelle Speicherräume bezüglich eines physischen Speicherraums abgebildet werden;

[0020] Fig. 5 ist eine schematische Veranschaulichung eines Speicherraums auf mehreren Speicherplatten und demonstriert ein Bewegen von Daten von einem ganzen RAID-Bereich zu einem anderen Speicherraum, so dass der ursprüngliche RAID-Bereich dahingehend erweitert werden kann, Raum auf einer neuen Platte zu umfassen.

[0021] Fig. 6 ist ein Flussdiagramm, das die methodischen Schritte zeigt, die damit verbunden sind, eine Speicherkapazität durch Hinzufügen neuer Platten zu vergrößern und dabei die Datenverfügbarkeit aufrechtzuerhalten, gemäß einem bevorzugten Ausführungsbeispiel.

[0022] Fig. 7 ist eine schematische Veranschaulichung eines Speicherraums auf mehreren Speicherplatten und veranschaulicht einen anfänglichen Schritt bei einem weiteren bevorzugten Verfahren dieser Erfindung, der ein Umorganisieren und Packen von Daten innerhalb eines Streifens betrifft, um den Streifen gleichzeitig dahingehend zu vergrößern,

ein neues Segment von einer neuen zusätzlichen Platte zu umfassen.

[0023] Fig. 8 zeigt den Speicherraum der Fig. 7, demonstriert jedoch einen Schritt, der an den in Fig. 7 Gezeigten anschließt.

[0024] Fig. 9 zeigt den Speicherraum der Fig. 7, demonstriert jedoch einen Schritt, der an den in Fig. 8 Gezeigten anschließt.

[0025] Fig. 10 zeigt den Speicherraum der Fig. 7, demonstriert jedoch einen Schritt, der an den in Fig. 9 Gezeigten anschließt.

[0026] Fig. 11 zeigt den Speicherraum der Fig. 7, demonstriert jedoch einen Schritt, der an den in Fig. 10 Gezeigten anschließt.

[0027] Fig. 12 zeigt den Speicherraum der Fig. 7 und demonstriert das allgemeine Format kontinuierlicher Schritte, die anschließend an den in Fig. 11 gezeigten Schritt unternommen werden.

[0028] Fig. 13 ist ein Flussdiagramm, das die methodischen Schritte des Vergrößerns der Speicherkapazität durch Hinzufügen neuer Platten gemäß der Veranschaulichung in Fig. 7–Fig. 12 zeigt.

AUSFÜHRLICHE BESCHREIBUNG DER ERFINDUNG

[0029] Fig. 1 zeigt ein Datenspeichersystem **10**, das gemäß dieser Erfindung aufgebaut ist. Vorzugsweise ist das Datenspeichersystem **10** ein Plattenarraydatenspeichersystem, das folgende Merkmale umfasst: ein hierarchisches Plattenarray **11**, das eine Mehrzahl von Speicherplatten **12** aufweist, eine Plattenarraysteuerung **14**, die mit dem Plattenarray **11** gekoppelt ist, um einen Transfer von Benutzerdaten zu und von den Speicherplatten **12** zu koordinieren, und ein RAID-Verwaltungssystem **16**.

[0030] Für die Zwecke dieser Offenbarung ist eine „Platte" jegliche nicht-flüchtige, direkt zugreifbare, wiederbeschreibbare Massenspeichervorrichtung, die die Fähigkeit aufweist, ihre eigenen Speicherausfälle zu erfassen. Sie umfasst sowohl sich drehende magnetische und optische Platten als auch Festzustandsplatten, oder nicht-flüchtige elektronische Speicherelemente (z.B. PROMs, EPROMs und EEPROMs). Der Begriff „Plattenarray" umfasst eine Sammlung von Platten, die Hardware, die erforderlich ist, um sie mit einem oder mehreren Hostcomputern zu verbinden, und eine Verwaltungssoftware, die verwendet wird, um den Betrieb der physischen Platten zu steuern und sie der Host-Betriebsumgebung als eine oder mehrere virtuelle Platten zu präsentieren. Eine „virtuelle Platte" ist eine abstrakte Entität, die durch die Verwaltungssoftware in dem Plattenarray verwirklicht ist.

[0031] Der Begriff „RAID" (Redundant Array of Independent Plattes, redundantes Array aus unabhängigen Platten) bedeutet ein Plattenarray, bei dem ein Teil der physischen Speicherkapazität verwendet wird, um redundante Informationen über Benutzerdaten, die auf dem Rest der Speicherkapazität gespeichert sind, zu speichern. Die redundanten Informationen ermöglichen eine Regeneration von Benutzerdaten für den Fall, dass eine der zu dem Array gehörenden Platten oder der Zugriffspfad zu derselben ausfällt. Eine ausführlichere Erläuterung von RAID-Systemen findet sich in einem Buch mit dem Titel The RAIDBook: A Source Book for RAID Technology, das am 9. Juni 1993 von RAID Advisory Board, Lino Lakes, Minnesota, veröffentlicht wurde.

[0032] Die Plattenarraysteuerung **14** ist über einen oder mehrere Schnittstellenbusse **13**, beispielsweise eine Kleincomputer-Schnittstelle (SCSI – small computer system interface) mit dem Plattenarray **11** gekoppelt. Das RAID-Verwaltungssystem **16** ist über ein Schnittstellenprotokoll **15** wirksam mit der Plattenarraysteuerung **14** gekoppelt. Ferner ist das Datenspeichersystem **10** über einen I/O-Schnittstellenbus **17** mit einem Hostcomputer (nicht gezeigt) gekoppelt. Das RAID-Verwaltungssystem **16** kann als separate Komponente verkörpert sein oder kann in der Plattenarraysteuerung **14** oder in dem Hostcomputer konfiguriert sein, um eine Datenverwaltungseinrichtung zum Steuern von Plattenspeicherungs- und von Zuverlässigkeitsebenen sowie zum Transferieren von Daten zwischen verschiedenen Zuverlässigkeits-Speicherstufen bereitzustellen. Diese Zuverlässigkeits-Speicherebenen sind, wie nachfolgend beschrieben wird, vorzugsweise Spiegel- oder Paritäts-Redundanzebenen, können jedoch auch eine Zuverlässigkeits-Speicherebene ohne jegliche Redundanz umfassen.

[0033] Die Plattenarraysteuerung **14** ist vorzugsweise als duale Steuerung implementiert, die aus der Plattenarraysteuerung A **14a** bezeichnet ist, und aus der Plattenarraysteuerung B **14b** bezeichnet ist, besteht. Die dualen Steuerungen **14a** und **14b** verbessern die Zuverlässigkeit, indem sie für den Fall, dass eine Steuerung funktionsuntüchtig wird, eine fortlaufende Sicherung und Redundanz liefern. Die Verfahren dieser Erfindung kann jedoch auch mit einer einzigen Steuerung oder mit anderen Architekturen praktiziert werden.

[0034] Das hierarchische Plattenarray **11** kann als unterschiedliche Speicherräume, einschließlich seines physischen Speicherraums und eines oder mehrerer virtueller Speicherräume, charakterisiert werden. Diese verschiedenen Speicheransichten sind durch Abbildungstechniken verbunden. Beispielsweise kann der physische Speicherraum des Plattenar-

rays in einen virtuellen Speicherraum abgebildet werden, der Speicherbereiche gemäß den verschiedenen Datenzuverlässigkeitsebenen darstellt. Manche Bereiche in dem virtuellen Speicherraum können einer ersten Zuverlässigkeitsspeicherebene zugewiesen sein, beispielsweise Spiegel- oder RAID-Ebene **1**, und andere Bereiche können einer zweiten Zuverlässigkeitsspeicherebene zugewiesen sein, beispielsweise Paritäts- oder RAID-Ebene **5**. Die verschiedenen Abbildungstechniken und virtuellen Plätze, die RAID-Ebenen betreffen, werden nachfolgend ausführlicher beschrieben.

[0035] Das Datenspeichersystem **10** umfasst einen Speicherabbildungsspeicher **21**, der eine durchgängige Speicherung der virtuellen Abbildungsinformationen liefert, die verwendet werden, um verschiedene Speicherräume ineinander abzubilden. Der Speicherabbildungsspeicher befindet sich außerhalb des Plattenarrays und vorzugsweise in der Plattenarraysteuerung **14**. Die Speicherabbildungsinformationen können durch die Steuerung oder das RAID-Verwaltungssystem kontinuierlich oder periodisch aktualisiert werden, während sich die verschiedenen Abbildungskonfigurationen unter den verschiedenen Ansichten ändern.

[0036] Vorzugsweise ist der Speicherabbildungsspeicher **21** als zwei nicht-flüchtige RAMs (Direktzugriffsspeicher) **21a** und **21b** verkörpert, die in jeweiligen Steuerungen **14a** und **14b** angeordnet sind. Ein beispielhafter nicht-flüchtiger RAM (NVRAM) ist ein batteriegepufferter RAM. Ein batteriegepufferter RAM verwendet Energie von einer unabhängigen Batteriequelle, um im Falle eines Verlustes von Leistung an das Datenspeichersystem **10** die Daten über einen bestimmten Zeitraum in dem Speicher beizubehalten. Ein bevorzugter Aufbau ist ein selbstauffrischender batteriegepufferter DRAM (dynamischer RAM).

[0037] Die dualen NVRAMs **21a** und **21b** sorgen für eine redundante Speicherung der Speicherabbildungsinformationen. Die virtuellen Abbildungsinformationen werden dupliziert und gemäß Spiegelredundanztechniken in beiden NVRAMs **21a** und **21b** gespeichert. Auf diese Weise kann der NVRAM **21a** dafür reserviert sein, die ursprünglichen Abbildungsinformationen zu speichern, und der NVRAM **21b** kann dafür reserviert sein, die redundanten Abbildungsinformationen zu speichern. Bei einem alternativen Aufbau kann ein gespiegelter Speicherabbildungsspeicher unter Verwendung eines einzelnen nichtflüchtigen RAMs mit einem ausreichenden Platz, um die Daten im Duplikat zu speichern, konfiguriert sein.

[0038] Wie in Fig. 1 gezeigt ist, weist das Plattenarray **11** mehrere Speicherplattenlaufwerksvorrichtungen **12** auf. Beispielhafte Größen dieser Speicherplatten betragen ein bis drei Gigabytes. Die Speicherplatten können unabhängig voneinander mit mechanischen Buchten verbunden oder von denselben getrennt sein, die eine Schnittstellenbildung mit dem SCSI-Bus **13** liefern. Bei einer Implementierung ist das Datenspeichersystem mit zwölf aktiven mechanischen Buchten entworfen. Vier SCSI-Busse werden verwendet, um eine Schnittstelle zwischen diesen Buchten und der Plattenarraysteuerung **14** zu bilden (d.h. ein Bus pro 3 mechanischen Buchten). Falls die aktiven Buchten vollständig beladen sind, weist das Datenspeichersystem eine beispielhafte kombinierte Kapazität von 12 bis 36 Gigabytes auf. Die Plattenarraysteuerung **14** erkennt die Speicherplatten **12** unabhängig davon, in welche Bucht sie eingesteckt sind. Das Datenspeichersystem **10** ist entworfen, um ein „heißes Einstecken" zusätzlicher Platten in verfügbare mechanische Buchten in dem Plattenarray, während das Plattenarray in Betrieb ist, zu ermöglichen. Die Hinzufügung neuer Speicherplattenvorrichtungen wird nachfolgend unter Bezugnahme auf Fig. 5–Fig. 13 ausführlicher beschrieben.

[0039] Die Speicherplatten **12** in dem Plattenarray **11** können zu Erläuterungszwecken vom Konzept her so erklärt werden, dass sie in einer Spiegelgruppe **18** aus mehreren Platten **20** und einer Paritätsgruppe **22** aus mehreren Platten **24** angeordnet sind. Die Spiegelgruppe **18** stellt eine erste Speicherposition oder einen ersten RAID-Bereich des Plattenarrays dar, die bzw. der Daten gemäß einer ersten bzw. Spiegelredundanzebene speichert. Diese Spiegelredundanzebene wird auch als RAID-Ebene **1** betrachtet. Die RAID-Ebene **1** bzw. oder ein Spiegeln von Platten bietet die höchste Datenzuverlässigkeit, indem sie insofern einen Eins-Zu-Eins-Schutz bietet, als jedes Datenbit dupliziert und in dem Datenspeichersystem gespeichert wird. Die Spiegelredundanz ist schematisch durch die drei Paare von Platten **20** in Fig. 1 dargestellt. Ursprüngliche Daten können auf einem ersten Satz von Platten **26** gespeichert werden, während dupliziert, redundante Daten auf dem gepaarten zweiten Satz von Platten **28** gespeichert werden.

[0040] Fig. 2 veranschaulicht die Speicherung von Daten gemäß der RAID-Ebene **1** noch ausführlicher. Die vertikalen Spalten stellen einzelne Platten dar, von denen Platten **0**, **1**, **2** und **3** veranschaulicht sind. Der in diesem Plattenarray aus vier Platten enthaltene physische Speicherraum kann zu mehreren Streifen konfiguriert sein, wie sie durch die horizontalen Zeilen veranschaulicht sind. Ein „Streifen" erstreckt sich über die Speicherplatten und ist aus zahlreichen gleich großen Segmenten an Speicherraum gebildet, wobei ein Segment jeder Platte in dem Array zugeordnet ist. Das heißt, dass ein Segment der Abschnitt eines Streifens ist, der sich auf einer einzelnen Platte befindet. Jeder Streifen hält eine vorbestimmte Datenmenge, die über die Speicherplatten verteilt ist. Manche Segmente eines Streifens werden für ur-

sprüngliche Daten verwendet, während andere Segmente für redundante Daten verwendet werden.

[0041] Bei diesem Beispiel einer Spiegelredundanz (RAID-Ebene **1**) werden Daten, die auf der Platte **0** in dem Segment **0** des Streifens **0** gespeichert sind, dupliziert und auf der Platte **1** in dem Segment **0'** des Streifens **0** gespeichert. Desgleichen werden Daten, die auf der Platte **2** in dem Segment **5** des Streifens **2** gespeichert sind, zu dem Segment **5'** des Streifens **2** auf der Platte **3** gespiegelt. Auf diese Weise wird jede Dateneinheit dupliziert und auf den Platten gespeichert. Der Redundanzentwurf der Fig. 2 ist zu Erläuterungszwecken vorgesehen. Die redundanten Daten müssen nicht sauber in demselben Streifen plaziert werden, wie gezeigt ist. Beispielsweise könnten Daten, die auf der Platte **0** in dem Segment **2** des Streifens **1** gespeichert sind, dupliziert werden und auf der Platte **3** in dem Segment T' des Streifens S plaziert werden.

[0042] Unter erneuter Bezugnahme auf Fig. 1 stellt die Paritätsgruppe **22** der Platten **24** eine zweite Speicherposition bzw. einen zweiten RAID-Bereich dar, in der bzw. dem Daten gemäß einer zweiten Redundanzebene, z.B. RAID-Ebene **5**, gespeichert werden. In dieser erläuternden Veranschaulichung von sechs Platten sind ursprüngliche Daten auf der fünften Platte **30** gespeichert, und redundante „Paritäts"-Daten sind auf der sechsten Platte **32** gespeichert.

[0043] Fig. 3 zeigt einen Paritäts-RAID-Bereich-Entwurf etwas ausführlicher. Ähnlich dem Entwurf des Spiegel-RAID-Bereichs der Fig. 2 kann der physische Speicherraum der Platten **0**, **1**, **2**, **3** zu mehreren gleich großen Streifen konfiguriert sein. Bei diesem veranschaulichten Beispiel werden Daten gemäß der RAID-Ebene **5** gespeichert, und die redundanten Daten werden in den Segmenten gespeichert, auf die durch den Buchstaben P Bezug genommen wird. Die redundanten P-Segmente speichern die Parität der anderen Segmente in dem Streifen. In dem Streifen **0** speichert das redundante P-Segment auf der Platte **3** beispielsweise die Parität der Platten **0**, **1** und **2**. Die Parität für jeden Streifen wird durch eine bestimmte Funktion berechnet, beispielsweise eine Exklusiv-ODER-Funktion, die durch das Symbol „⊕" dargestellt wird. Die Paritäten für die ersten vier Streifen (wobei die tiefgestellte Zahl den entsprechenden Streifen darstellt) lauten wie folgt:

P_0 = Segment **0** ⊕ Segment **1** ⊕ Segment **2** = Platte **0** ⊕ Platte **1** ⊕ Platte **2**

P_1 = Segment **3** ⊕ Segment **4** ⊕ Segment **5** = Platte **0** ⊕ Platte **1** ⊕ Platte **3**

P_2 = Segment **6** ⊕ Segment **7** ⊕ Segment **8** = Platte **0** ⊕ Platte **2** ⊕ Platte **3**

P_3 = Segment **9** ⊕ Segment **10** ⊕ Segment **11** = Platte **1** ⊕ Platte **2** ⊕ Platte **3**

[0044] Eine Paritätsredundanz ermöglicht eine Regeneration von Daten, die auf einer der Platten nicht mehr zur Verfügung stehen. Wenn beispielsweise die Daten in dem Segment **5** nicht mehr zur Verfügung stehen, kann ihr Inhalt aus den Segmenten **3** und **4** und den Paritätsdaten in dem Segment P ermittelt werden. Eine Paritätsspeicherung ist kostengünstiger als eine Spiegelspeicherung, ist jedoch auch weniger zuverlässig und weist eine geringere Leistungsfähigkeit auf.

[0045] Die Plattenanordnung der Fig. 1 ist zu Konzeptionszwecken vorgesehen. In der Praxis weist das Plattenarray **11** einfach eine Mehrzahl von Platten **12** auf, die in der Lage sind, Daten gemäß einer Spiegel- und Paritätsredundanz zu speichern. Aus dem durch alle Platten **12** gelieferten, zur Verfügung stehenden Speicherraum wird ein Abschnitt dieses Speicherraums einer Spiegelredundanz zugewiesen, und ein weiterer Abschnitt wird einer Paritätsredundanz zugewiesen. Vorzugsweise sind die Platten **12** konfiguriert, um mehrere gleich große Speicherregionen (die bei Fig. 4 mit dem Bezugszeichen **35** benannt sind) zu enthalten, wobei einzelne Regionen mehrere Segmente aufweisen. Die Regionen sind gruppiert, um bei einer virtuellen Ansicht des Speicherraums RAID-Bereiche zu bilden. Zusätzlich wird eine weitere (durch einen Host definierte) Ansicht eines Speicherraums dem Benutzer oder Host präsentiert, so dass die RAID-Bereiche und die Datenredundanz-Speichertechniken für den Benutzer oder Host transparent sind. Diese Merkmale werden nachstehend unter Bezugnahme auf Fig. 4 ausführlicher erläutert.

[0046] Das Datenspeichersystem **10** verwaltet die „Migration bzw. Bewegung" von Daten zwischen Spiegel- und Paritätsspeicherschemata. Die Verwaltung beider Redundanztypen wird durch das RAID-Verwaltungssystem **16** (Fig. 1) koordiniert. Das RAID-Verwaltungssystem **16** verwaltet die zwei unterschiedlichen Typen von RAID-Bereichen in dem Plattenarray als Speicherhierarchie, wobei die Spiegel-RAID-Bereiche ähnlich einem Cache für die Paritäts-RAID-Bereiche agieren. Das RAID-Verwaltungssystem **16** verschiebt und organisiert die Daten zwischen dem Spiegel- und dem Paritäts-RAID-Bereich gemäß einem definierten Leistungsfähigkeitsprotokoll, und verwaltet die Daten ansonsten auf ebensolche Weise. Der Vorgang des Bewegens von Daten zwischen dem Spiegel- und dem Paritäts-RAID-Bereich wird als „Migration" bezeichnet.

[0047] Das Datenspeichersystem **10** versucht, die für die Leistungsfähigkeit kritischeren Daten in die Spiegel-RAID-Bereiche zu plazieren, da dies die höchste Leistungsfähigkeit und Zuverlässigkeit liefert. Die Leistungsfähigkeitsprotokolle, die durch das RAID-Verwaltungssystem **16** implementiert sind, umfassen eine von zwei bevorzugten Migrationsrichtlini-

en. Gemäß der ersten Migrationsrichtlinie, die als „Zugriffshäufigkeit" bekannt ist, werden die Daten in dem hierarchischen Plattenarray, auf die am häufigsten zugegriffen wird, in dem Spiegel-RAID-Bereich **18** beibehalten. Daten, auf die weniger häufig zugegriffen wird, werden in dem Paritäts-RAID-Bereich **22** gehalten. Gemäß einer zweiten Migrationsrichtlinie, die als „Zugriffsaktualität" bekannt ist, werden die Daten, deren Zugriff am wenigsten lange zurückliegt, in dem Spiegel-RAID-Bereich **18** beibehalten, während die Daten, deren Zugriff länger zurückliegt, in dem Paritäts-RAID-Bereich **22** gespeichert werden. Es können auch andere Leistungsfähigkeitsprotokolle verwendet werden. Idealerweise sind solche Protokolle auf der Basis der spezifischen Computeranwendung und der Erfordernisse des Benutzers definiert.

[0048] Zusätzlich „stimmt" das RAID-Verwaltungssystem **16** die Speicherressourcen eines Datenspeichersystems gemäß einer Funktion von zwei Parametern „ab". Größe der physischen Speicherkapazität und Größe der vorhandenen Menge an Benutzerdaten, die in dem Datenspeichersystem gespeichert sind. Anfänglich werden alle Daten in Spiegel-RAID-Bereichen gespeichert, da dies die höchste Leistungsfähigkeit und Zuverlässigkeit liefert. Während mehr Daten zu dem Datenspeichersystem hinzugefügt werden, werden die Daten zwischen Spiegel-RAID-Bereichen und Paritäts-RAID-Bereichen bewegt, um die Leistungsfähigkeit und Zuverlässigkeit zu optimieren. Während sich das Datenspeichersystem an eine volle Kapazität annähert, werden im Rahmen des Bemühens, alle durch den Benutzer gestellten Anforderungen zu erfüllen und trotzdem durch Redundanz eine Zuverlässigkeit zu liefern, immer mehr Daten in Paritäts-RAID-Bereiche bewegt. Dementsprechend liefert das Datenspeichersystem dieser Erfindung eine maximale Flexibilität und Anpassung. Es erfordert nicht, dass der Benutzer einen spezifischen Speicherbetriebszustand auswählt, sondern kann sich statt dessen an jegliche durch den Benutzer an das System gestellte Forderung anpassen.

[0049] Fig. 4 veranschaulicht ein Speicherabbilden des verfügbaren Speicherraums des Datenspeichersystems **10** als mehrere Lagen eines abgebildeten virtuellen Speicherraums. Jedes vertikal verlängerte Rechteck in dem Diagramm stellt eine Ansicht des physischen Speicherraums dar. Bei diesem Diagramm wird auf den physischen Speicherraum **34** durch zwei virtuelle Speicheransichten **40** und **50** Bezug genommen. Der physische Speicherraum **34** wird durch vier Platten (z.B. Platten **12** in Fig. 1) dargestellt, die durch die Bezugszeichen **0**, **1**, **2** und **3** benannt sind. Die den Platten zugeordneten vier Rechtecke stellen eine Ansicht des physischen Speicherraums dar, bei der die Platten **1**, **2** und **3** eine ungefähr gleiche Speicherkapazität aufweisen und die Platte **0** eine etwas geringere Speicherkapazität aufweist. Beispielhafte Speicherkapazitäten für derartige Platten betragen 1 bis 3 Gigabytes. Der Speicherraum **34** ist in Bereiche A0, A1, A2 usw. unterteilt. Einzelne Bereiche enthalten zahlreiche Streifen, im Bereich A0 z.B. Streifen O-Q. Einzelne Bereiche enthalten auch zahlreiche Regionen **35**. Die Regionen **35** bestehen vorzugsweise aus einer ausgewählten Anzahl von Segmenten einer gleichmäßigen Größe auf jeder Speicherplatte, so dass die Regionen über das gesamte Plattenarray eine gleiche Größe aufweisen. Eine beispielhafte Größe einer Region **35** ist ein Megabyte.

[0050] Der Speicherraum der Platten wird in eine erste virtuelle Zwischenansicht **40** der RAID-Ebene des physischen Speicherraums **34** abgebildet. Diese erste virtuelle Ansicht ist vom Konzept her ein Satz von RAID-Bereichen, die auf eine zweite Anwendungsansicht abgebildet werden können, die einen angrenzend adressierbaren Speicherraum darstellt. Die physische Konfiguration und RAID-Ansichten des Speicherraums entziehen sich der Anwendungsansicht.

[0051] Der RAID-Bereich-Speicherraum **40** ist die Speicheransicht, die den Spiegel- und Paritätsspeicherraum identifiziert. Beispielsweise kann ein RAID-Bereich **42** einen Spiegel-RAID-Bereich von M Zuweisungsblöcken **43** darstellen, während der RAID-Bereich **44** einen Paritäts-RAID-Bereich von N Zuweisungsblöcken **45** darstellen kann. Die Zuweisungsblöcke **43** und **45** weisen vorzugsweise eine gleiche Größe auf, wobei eine beispielhafte Größe 64 Kilobytes beträgt. Diese RAID-Bereiche beziehen sich auf entsprechende physische Bereiche A0, A1, A2 usw. auf dem physischen Speicherraum **34**. Beispielsweise können sechzehn 64 K-Zuweisungsblöcke **43** oder **45** an der virtuellen RAID-Ansicht auf eine einzige 1M-Region **35** abgebildet werden.

[0052] Die Spiegel- und Paritäts-RAID-Bereiche können den gesamten Speicherraum **34** des Plattenarrays einnehmen, müssen aber nicht. Dementsprechend kann während bestimmter Anwendungen ein unbenutzter und unbezeichneter Speicherraum vorliegen, der keinem bestimmten RAID-Bereich entspricht. Ein solcher Speicherraum kann jedoch in einen Spiegel- oder Paritäts-RAID-Bereich umgewandelt werden. Ferner ist zu beachten, dass die RAID-Bereiche in der Abbildung in benachbarte Bereiche auf dem Plattenarray abgebildet werden, wobei jede einem RAID-Bereich zugeordnete Region an derselben physischen Adresse auf jeder Speicherplatte angeordnet ist. Die RAID-Bereiche können alternativ auch zu nicht-benachbarten Bereichen auf dem Plattenarray abgebildet werden.

[0053] Der in den RAID-Bereichen verfügbare Speicherraum wird in eine zweite virtuelle Vorderes-Ende-Ansicht **50** der Anwendungsebene abgebildet, die

eine Speicheransicht ist, wie sie durch den Benutzer oder das Hostanwendungsprogramm definiert und demselben präsentiert wird. Bei einer Betrachtung durch den Benutzer oder das Hostanwendungsprogramm kann die virtuelle Ansicht **50** der Anwendungsebene eine einzige große Speicherkapazität darstellen, die den verfügbaren Speicherraum auf den Speicherplatten **12** angibt. Der virtuelle Speicherraum **50** präsentiert eine Ansicht eines linearen Satzes von gleich großen virtuellen Speicherblöcken **52** und **53**, die einzeln als 0, 1, 2, ... J – 1, J, J + 1, ..., L – 1, L, L + 1, ..., usw. bezeichnet werden. Die virtuellen Blöcke **52** und **53** weisen vorzugsweise dieselbe Größe auf wie die Zuweisungsblöcke in dem RAID-Bereich **40**, wobei eine beispielhafte Größe 64 Kilobytes beträgt. Der virtuelle Blockspeicherraum **50** wird in der durch die RAID-Bereiche **40** dargestellten Ansicht durch eine Tabelle von Bezugnahmen oder Zeigern (wie sie durch Pfeile **54** dargestellt sind) auf Zuweisungsblöcke, dargestellt. Die virtuellen Blöcke **52** und **53** bei der virtuellen Anwendungsansicht **50** sind somit über die in der virtuellen Blocktabelle beibehaltenen Zeiger den Zuweisungsblöcken **43** und **45** an der virtuellen RAID-Ansicht **50** zugeordnet. Es gibt mindestens zwei Arten von RAID-Bereichen, auf die von der virtuellen Blocktabelle verwiesen werden kann: Spiegel und Parität.

[0054] Das RAID-Verwaltungssystem **16** kann die Konfiguration der RAID-Bereiche über den physischen Speicherraum dynamisch verändern. Die Anzahl von RAID-Bereichen für jeden Typ kann je nach der Menge von Benutzerdaten, die in dem System gespeichert sind, und je nach der Größe des physischen Plattenspeicherraums erhöht oder verringert werden. Folglich befinden sich das Abbilden der RAID-Bereiche in der virtuellen Ansicht **40** der RAID-Ebene auf die Platten und das Abbilden der virtuellen Front-End-Ansicht **50** auf die RAID-Ansicht **40** allgemein in einem Veränderungszustand. Der Speicherabbildungsspeicher in den NVRAMs **21a** und **21b** (Fig. 1) behält die aktuellen Abbildungsinformationen bei, die durch das RAID-Verwaltungssystem **16** verwendet werden, um die RAID-Bereiche auf die Platten abzubilden, und behält die Informationen bei, die verwendet werden, um zwischen die zwei virtuellen Ansichten abzubilden. Während das RAID-Verwaltungssystem die Abbildungen der RAID-Ebene dynamisch verändert, aktualisiert es auch die Abbildungsinformationen in dem Speicherabbildungsspeicher, um die Veränderungen zu reflektieren.

[0055] Die Migrationsoperation des Speichersystems **10** wird nun unter Bezugnahme auf Fig. 1 und Fig. 4 beschrieben.

[0056] Für die Zwecke einer fortführenden Erläuterung nehmen virtuelle Blöcke **53** des virtuellen Speicherraums **50** der Anwendungsebene Bezug auf zugeordnete Zuweisungsblöcke **45** in dem Paritäts-RAID-Bereich **44**, die in dem Bereich A1 des physischen Speicherraums **34** gespeichert sind. Derartige virtuelle Blöcke **53** können als „virtuelle Paritätsblöcke" bezeichnet werden, während die zugeordneten Zuweisungsblöcke **45** als „Paritätszuweisungsblöcke" bezeichnet werden. Desgleichen nehmen virtuelle Blöcke **52** Bezug auf zugeordnete Zuweisungsblöcke **43** in dem Spiegel-RAID-Bereich **42**, die in dem Bereich A0 des physischen Speicherraums **34** gespeichert sind. Derartige virtuelle Blöcke **52** können hierin als „virtuelle Spiegelblöcke" bezeichnet werden, während die zugeordneten Zuweisungsblöcke **43** als „Spiegelzuweisungsblöcke" bezeichnet werden.

[0057] Um Daten von einem RAID-Bereich zu einem anderen zu bewegen, wird allgemein ein virtueller Block, der einem Zuweisungsblock eines Typs der ersten RAID-Ebene (z.B. Spiegel oder Ebene **1**) zugeordnet ist, ausgewählt. Anschließend wird ein nicht benutzter Zuweisungsblock, der einen Typ der zweiten RAID-Ebene (beispielsweise Parität oder Ebene **5**) darstellt, lokalisiert. Falls ein nicht benutzter Zuweisungsblock nicht lokalisiert werden kann, wird einer erzeugt. Als nächstes werden Daten von dem Zuweisungsblock, der zuvor dem ausgewählten virtuellen Block zugeordnet war, an den nicht benutzten Zuweisungsblock transferiert, der bewirkt, dass die Daten einer Redundanzebenenänderung unterzogen werden. Beispielsweise werden Daten, die einst gemäß der Spiegelredundanz gespeichert wurden, nun gemäß einer Paritätsredundanz gespeichert, oder umgekehrt. Als abschließenden Schritt wird die **Abb.** 54 des virtuellen Speicherraums **50** der Anwendungsebene auf den virtuellen Speicherraum **40** der RAID-Ebene modifiziert und aktualisiert, um die Datenverschiebung widerzuspiegeln. Der ausgewählte virtuelle Block, der zuvor einem Zuweisungsblock des Typs der ersten RAID-Ebene zugeordnet war, nimmt nun über einen aktualisierten Zeiger Bezug auf einen Zuweisungsblock des Typs der zweiten RAID-Ebene, der die migrierten Daten enthält. Jegliche Abbildungsveränderung, die während dieser Transaktion auftritt, würde in dem Speicherabbildungsspeicher **21** aktualisiert werden.

[0058] Die fortdauernde Erörterung liefert eine ausführlichere Erklärung des Bewegens von Daten zwischen Spiegel- und Paritätsspeicherbereichen gemäß bevorzugten Verfahren und Sequenzen dieser Erfindung. Zum Bewegen von Daten von einem Paritäts- zu einem Spiegelspeicher wird die folgende Sequenz verwendet:

1. Das RAID-Verwaltungssystem lokalisiert einen unbenutzten Spiegelzuweisungsblock **43** in einem Spiegel-RAID-Bereich **42**.
2. Falls keiner gefunden wird, erzeugt das RAID-Verwaltungssystem einen Spiegelzuweisungsblock (nachfolgend erläutert).
3. Das RAID-Verwaltungssystem hebt neue Spei-

cheranforderungen an den zu bewegenden virtuellen Block zeitweilig auf.
4. Das RAID-Verwaltungssystem wartet, bis alle aktiven Datenspeicherungsanforderungen an den virtuellen Block abgeschlossen sind.
5. Die Daten von dem Paritätszuweisungsblock **45**, der dem virtuellen Block **53** zugewiesen ist, werden in einen vorübergehenden Speicherpuffer eingelesen.
6. Die Daten werden anschließend in den bei Schritt **2** gewählten Spiegelzuweisungsblock **43** geschrieben.
7. Die virtuelle Blocktabelle wird modifiziert, um auf die neue Position der Daten in dem Spiegelzuweisungsblock **43** zu verweisen.
8. Die zeitweilig aufgehobenen Speicheranforderungen werden wiederaufgenommen.

[0059] Gemäß der obigen Prozedur bewegte sich ein virtueller Block **53** in der virtuellen Ansicht **50** der Anwendungsebene von einer Paritäts- zu einer Spiegelspeicherung. Relativ zu der virtuellen Zwischenansicht **40** bewegten sich Daten von einem Paritätszuweisungsblock **45** in dem Paritäts-RAID-Bereich **44** zu einem Spiegelzuweisungsblock **43** in dem Spiegel-RAID-Bereich **42**. In dem physischen Speicherraum bewegten sich Daten von dem Bereich A1 zum Bereich A0.

[0060] Falls kein unbenutzter Spiegelzuweisungsblock lokalisiert werden kann (Schritt 1 oben), versucht das RAID-Verwaltungssystem die folgende bevorzugte Sequenz von drei Techniken. Erstens versucht das RAID-Verwaltungssystem, einen unbenutzten (und somit unbenannten) RAID-Bereich zu lokalisieren, der in einen Spiegel-RAID-Bereich umgewandelt werden kann, ohne die Systemschwelle eines unbenutzten Speichers der RAID-Ebene zu verletzen, die benötigt wird, um zu garantieren, dass eine Migration immer stattfinden kann. Falls dies fehlschlägt und das System mehr als die reservierte Menge an unbenutztem Speicher der RAID-Ebene aufweist, bewegt das System Daten innerhalb des Paritätsspeichers, um unbenutzte Paritätszuweisungsblöcke zu unbenutzten RAID-Bereichen zu sammeln. Falls diese Migration einen unbenutzten RAID-Bereich ergibt, der, wie oben erwähnt, in einen Spiegel-RAID-Bereich umgewandelt werden kann, wandelt das System ihn in einen Spiegel-RAID-Bereich um. Andernfalls bewegt das System alternativ dazu Daten von einem Spiegel- zu einem Paritätsspeicher, verdichtet den Spiegelspeicher und wandelt einen unbenutzten Speicher der RAID-Ebene in einen Paritätsspeicher um, bis das System den unbenutzten Speicher der RAID-Ebene ausreichend für die Lokalisierung eines unbenutzten Spiegelzuweisungsblocks oder eine Umwandlung eines unbenutzten RAID-Bereichs in einen Spiegel-RAID-Bereich erhöht. Da Spiegelzuweisungsblöcke mehr physischen Speicherraum belegen als Paritätszuweisungsblöcke, führt diese letzte Technik zu einer Nettoerhöhung des Umfangs an unbenutztem Speicher der RAID-Ebene.

[0061] Das Erzeugungs-/Umwandlungsprotokoll, das verwendet wird, um unbenutzte Spiegelzuweisungsblöcke zu lokalisieren und einzurichten, ist vorteilhaft, da es dem RAID-Verwaltungssystem ermöglicht, die Speicherzuweisung gemäß der Menge an Benutzerdaten und der Größe des physischen Speicherraums zwischen Paritäts- und Spiegelbereichen selektiv einzustellen. Da die Datennutzung und die Speicherkapazität variieren, verwendet das RAID-Verwaltungssystem eine oder mehrere der obigen drei Techniken, um die Datenmenge, die in einem Spiegelspeicher gehalten wird, zu maximieren.

[0062] Das RAID-Verwaltungssystem versucht, die Situation zu vermeiden, bei der eine Speicheranforderung warten muss, bis die Platzschaffungssequenz einen unbenutzten Spiegelzuweisungsblock ergibt, indem sie während einer Leerlaufzeit unbenutzte RAID-Bereiche erzeugt. In manchen Situationen können Speicheranforderungen während der Platzschaffungssequenz jedoch zeitweilig aufgehoben werden. Das RAID-Verwaltungssystem konfiguriert den virtuellen Blockspeicherraum derart, dass der virtuelle Platz geringer ist als die RAID-Ansicht. Dies gewährleistet, dass für eine Migration oder andere Zwecke ein freier Platz reserviert wird, der gleich zumindest einem RAID-Bereich ist. Auf diese Weise ergibt die Sequenz von Techniken immer einen unbenutzten Spiegelzuweisungsblock.

[0063] Um Daten von einem Spiegel- zu einem Paritätsspeicher zu bewegen, wird die folgende Sequenz verwendet:
1. Das RAID-Verwaltungssystem wählt einen virtuellen Block aus **52** aus, um gemäß einer Migrationsrichtlinie, beispielsweise Zugriffsaktualität oder Zugriffshäufigkeit, von einem Spiegel- zu einem Paritätsspeicher zu bewegen.
2. Das RAID-Verwaltungssystem lokalisiert einen unbenutzten Paritätszuweisungsblock **45** in einem Paritäts-RAID-Bereich **44**.
3. Falls kein solcher Block gefunden werden kann, wird ein für eine Migration reservierter Platz gemäß den oben beschriebenen Erzeugungstechniken in einen Paritäts-RAID-Bereich umgewandelt.
4. Neue Speicheranforderungen an den zu bewegenden virtuellen Block werden zeitweilig aufgehoben.
5. Das RAID-Verwaltungssystem wartet, bis alle aktiven Speicheranforderungen an den virtuellen Block abgeschlossen sind.
6. Daten werden von dem Spiegelzuweisungsblock **43**, der dem virtuellen Block **52** zugeordnet ist, in einen vorübergehenden Speicherpuffer gelesen.
7. Die Daten werden in den gewählten Paritätszu-

weisungsblock **45** geschrieben.
8. Die virtuelle Blocktabelle wird modifiziert, um auf die neue Position der Daten in dem Paritätszuweisungsblock **45** zu verweisen.
9. Datenanforderungen an den virtuellen Block werden wiederaufgenommen.

[0064] Die zwei oben aufgeführten, aus neun Schritten bestehenden Sequenzen liefern Beispiele dessen, wie das Speichersystem dieser Erfindung arbeiten kann, um Daten zwischen zwei verschiedenen Redundanzebenen zu bewegen.

[0065] Unter erneuter Bezugnahme auf Fig. 1 umfasst das Datenspeichersystem **10** ferner einen Satz **31** von Reservespeicherplatten **33**. Das Datenspeichersystem **10** ist dazu entworfen, ein „heißes Einstecken" („hot plug") von zusätzlichen Platten in verfügbare mechanische Buchten in dem Plattenarray zu ermöglichen, während das Plattenarray im Betrieb bleibt. Das Datenspeichersystem wird zu diesem Zeitpunkt nicht abgeschaltet. Um die Speicherkapazität des Datenspeichersystems zu erhöhen, werden einfach eine oder mehr Reserveplatten **33** in verfügbare Buchten eingesteckt, um mit dem SCSI-Bus **13** eine Schnittstelle zu bilden. Die Steuerung **14** erkennt die neue(n) Platte(n) und beginnt die Prozesse dieser Erfindung, um den zusätzlichen Speicherraum in nutzbare RAID-Bereiche zu assimilieren.

[0066] Verfahren zum Erhöhen der Speicherkapazität durch heißes Einstecken einer oder mehrerer Speicherplatten werden nun unter Bezugnahme auf Fig. 5–Fig. 13 beschrieben. Die Verfahren werden vorzugsweise während Leerlaufzeiten des Datenspeichersystems im Hintergrund durchgeführt.

[0067] Fig. 5 und Fig. 6 zeigen ein Verfahren gemäß dieser Erfindung, bei dem ganze RAID-Bereiche einzeln dahingehend umkonfiguriert werden, die neue(n) Speicherplatte(n) zu beinhalten. In Fig. 5 wird eine neue, zusätzliche Platte, die mit Bezugszeichen **4** versehen ist, zu dem Plattenarray hinzugefügt. Es können mehr als eine Speicherplatte hinzugefügt werden; der Einfachheit der Erläuterung halber wird bei diesem Beispiel jedoch lediglich eine Platte hinzugefügt. Die neue Platte ist anfänglich leer. Die neue Platte kann dieselbe oder eine andere Kapazität aufweisen als andere vorhandene Speicherplatten in dem Plattenarray.

[0068] Unter Bezugnahme auf Fig. 6 besteht ein möglicher anfänglicher Schritt **100** darin, Speicherraum auf dem derzeit konfigurierten Plattenarray zu reservieren. Vorzugsweise ist für diesen und andere Zwecke immer ein gewisser Speicherraum in dem Plattenarray verfügbar. Beispielsweise reserviert das RAID-Verwaltungssystem **16** zum Zweck einer Datenmigration zwischen RAID-Bereichen, wie sie oben unter Bezugnahme auf Fig. 4 beschrieben wurden, einen Raum, der äquivalent zu einem oder mehreren RAID-Bereichen ist. Die Größe des reservierten Raums wird nicht als Teil der für Benutzerdaten zur Verfügung stehenden Kapazität berechnet, um zu gewährleisten, dass das System Raum zum Bewegen von Daten zwischen Speicherbereichen hat. Dies erfolgt durch Einstellen einer Schwellenbedingung in der Steuerung, die eine Verwendung des reservierten Raums für irgendeinen anderen Zweck als eine Datenbewegung oder -migration hemmt oder blockiert.

[0069] Der zur Verwendung während der Prozesse des Hinzufügens neuer Speicherplatten reservierte Raum kann als Teil der normalen Aufgaben des RAID-Verwaltungssystems **16** oder durch die Steuerung **14** oder auf andere Weise verwaltet werden. Der reservierte Raum weist eine Kapazität auf, die gleich derjenigen eines vorhandenen RAID-Bereichs oder größer als dieselbe ist. Auf diese Weise weist der reservierte Raum eine Größe auf, die ausreichend ist, um zumindest dieselbe Datenmenge zu halten, die in einem RAID-Bereich gespeichert ist. Eine Raumreservierung kann in der ersten bzw. virtuellen RAID-Ansicht, der zweiten bzw. virtuellen Benutzer-Ansicht oder einer Kombination aus beiden dargestellt werden.

[0070] Bei Schritt **102** werden eine oder mehr Speicherplatten zu dem Plattenarray hinzugefügt, z.B. Platte **4** in Fig. 5. Bei Schritt **103** wird der zweite virtuelle Speicherraum **50** dahingehend modifiziert, den neuen Speicherraum, der sich durch die Hinzufügung der Speicherplatte(n) ergibt, widerzuspiegeln. Somit erkennt der Host unmittelbar, dass mehr Raum zu dem Datenspeichersystem hinzugefügt wurde. Jedoch kann der Host eine Verzögerung bezüglich eines anfänglichen Zugriffs auf den neuen Plattenraum erleben, bis über die folgenden RAID-Bereich-Erweiterungstechniken ausreichend Raum integriert wurde, um eine Host-Anforderung unterzubringen.

[0071] Bei Schritt **104** wird ein vorhandener RAID-Bereich ausgewählt. Ein vorhandener RAID-Bereich ist einer, der derzeit Daten enthält, jedoch noch nicht dazu konfiguriert wurde, den Raum auf der neuen Platte **4** zu integrieren. Bei Schritt **106** werden alle in dem ausgewählten RAID-Bereich enthaltenen Daten an einen der folgenden zwei Orte bewegt bzw. verschoben: (1) einen bereits erweiterten RAID-Bereich, der den neuen Plattenraum umfasst, oder (2) einen anderen Raum auf dem Speicherplattenarray, der den Raum umfasst, der bei Schritt **100** reserviert wurde. Fig. 5 zeigt die Bewegung von Daten von einem ausgewählten Bereich A1 (der einem RAID-Bereich **42** in der virtuellen RAID-Ansicht entspricht; siehe Fig. 4) zu dem reservierten Raum im Bereich AV. Der reservierte Speicherraum ist in der Darstellung in einem zusammenhängenden physischen Bereich AV enthalten, der zu Erläuterungszwe-

cken in einen einzigen zusammenhängenden RAID-Bereich abgebildet wird. Jedoch ist der reservierte Speicherraum üblicherweise auf nicht-zusammenhängende Weise in dem ganzen Plattenarray verteilt und wird in mehrere RAID-Bereiche abgebildet.

[0072] Der physische Speicherraum, der dem ausgewählten Bereich A1 entspricht, wird dann dahingehend einer Umabbildung unterzogen oder umkonfiguriert, die neue Speicherplatte **4** zu umfassen (Schritt **108**). Diese Umabbildung erzeugt einen erweiterten RAID-Bereich, der die mehreren Speicherplatten **0**–**3** und die neue zusätzliche Speicherplatte **4** überspannt. Der erweiterte RAID-Bereich umfasst eine Region einer gleichen Größe (die aus mehreren Segmenten gleicher Größe besteht) von jeder Platte in dem Array (d.h. Regionen **35a**–**35d**), einschließlich der neuen Platte (d.h. Region **35e**). Ein Umabbilden vorhandener RAID-Bereiche in erweiterte RAID-Bereiche (Schritte **104**–**108**) kann während Leerlaufzeiten des Datenspeichersystems in einem Hintergrundprozess durchgeführt werden.

[0073] Speicherraum, der in dem RAID-Betriebszustand noch nicht benannt wurde, wächst automatisch dahingehend, die neue Plattenkapazität zu umfassen. Dieser Raum wird nach Bedarf in einen RAID-Bereich umgewandelt. Für den Fall, dass zusätzliche erweiterte RAID-Bereiche benötigt werden, um die Speicherungsanforderung eines Benutzers zu berücksichtigen, wird ungenutzter Speicherraum lokalisiert und anschließend in einen erweiterten RAID-Bereich, der den neuen Plattenraum umfasst, abgebildet. Man beachte, dass dieser neu erweiterte RAID-Bereich verwendet werden kann, um Daten mittels Schritt **106** von vorhandenen RAID-Bereichen zu empfangen.

[0074] Nach Abschluss dieser Datenbewegung bzw. -verschiebung werden die Speicherabbildungsinformationen in den Speicherabbildungsspeicher **21** sowohl für den ersten als auch für den zweiten virtuellen Speicherraum aktualisiert, um die Datenbewegung und die Erweiterung des RAID-Bereichs in dem hierarchischen Plattenarray bei Schritt **114** widerzuspiegeln. Im Einzelnen werden die Speicherabbildungsinformationen aktualisiert, während die Daten in jedem virtuellen Block während des Erweiterungsprozesses bewegt bzw. verschoben werden. Der Prozess indexiert weiterhin jeweils immer einen RAID-Bereich zu einem Zeitpunkt durch das Plattenarrayspeichersystem, bis alle RAID-Bereiche dahingehend erweitert, wurden, den neuen Plattenraum zu umfassen (Schritt **116**).

[0075] Gemäß den Verfahren dieser Erfindung stehen die Daten dem Benutzer weiterhin jederzeit zur Verfügung.

[0076] Das RAID-Verwaltungssystem **16** löst jegliche Konflikte, die während des Prozesses an der Schicht **50** des zweiten virtuellen Speicherraums entstehen. Im Einzelnen löst das RAID-Verwaltungssystem **16** Konflikte zwischen Datenzugriffsanforderungen an die virtuellen Blöcke und der Bewegung der virtuellen Blöcke. Die Konfliktlösung ermöglicht, dass Daten jederzeit verfügbar bleiben. Das RAID-Verwaltungssystem handhabt Konflikte mittels vieler verschiedener Techniken. Eine beispielhafte Technik besteht darin, einen Zugriff auf die virtuellen Blöcke, die dem ausgewählten RAID-Bereich zugeordnet sind, vorübergehend auszusetzen. Eine weitere Technik, die als „Nicht-Prioritätsgebung" (non-preemption) bekannt ist, nutzt eine ununterbrochene Verwendung einer Systemressource, um jegliche zu einem Konflikt führende Aktivität zu vermeiden. Es wird verhindert, dass Konflikte entstehen, weil die Ressource zu jeglichem Zeitpunkt immer nur einer Aufgabe gewidmet ist. Eine weitere beispielhafte Technik besteht darin, jegliche zu einem Konflikt führende Aktivität, die daraus resultiert, einen Zugriff auf Daten zu versuchen, die gerade bewegt bzw. verschoben werden, zu protokollieren, und später zurückzugehen, um diesen Konflikt zu lösen, nachdem die Daten verschoben wurden.

[0077] Dieses Verfahren kann bezüglich einer Nicht-RAID-Umgebung verallgemeinert werden. Der physische Speicherraum des Plattenarrays wird zu mehreren Streifen, z.B. Streifen O-Q in **Fig. 3** und **Fig. 4**, statt RAID-Bereichen konfiguriert. Die Streifen erstrecken sich über alle Platten und umfassen üblicherweise ein einzelnes Segment von jeder Platte, obwohl Streifen mit mehr als einem Segment pro Platte möglich sind. Ein Abschnitt eines physischen Raums kann – muss aber nicht – dafür reserviert sein, Daten, die in dem Streifen gehalten werden, vorübergehend zu speichern. Falls dies der Fall ist, weist der Abschnitt des Raums eine ausreichende Größe auf, um eine gleiche oder größere Menge von in einem Streifen gespeicherten Daten zu speichern.

[0078] Ein Streifen wird zur Erweiterung ausgewählt, und eine Zuweisung in diesem Streifen wird gehemmt. Alle Daten in dem ausgewählten Streifen werden zu einer anderen Speicherposition in dem Plattenarray bewegt bzw. verschoben. Der den Streifen enthaltende physische Speicherraum wird anschließend in einen erweiterten Streifen zum Speichern von Daten umkonfiguriert, der das Segment auf der neuen Speicherplatte umfasst. Somit überspannt der erweiterte Streifen alle Speicherplatten, einschließlich der neuen zusätzlichen Speicherplatten. Anschließend wird eine Zuweisung in dem erweiterten Streifen wiederaufgenommen. Dieser Prozess wird Streifen um Streifen fortgesetzt, bis alle Streifen dahingehend konfiguriert wurden, die neuen Platten zu umfassen. Während des Prozesses wird eine Verfügbarkeit von Daten für alle nicht-ausgewählten

Streifen aufrechterhalten.

[0079] Fig. 7–Fig. 13 zeigen ein weiteres bevorzugtes Verfahren gemäß dieser Erfindung, bei dem Segmente während des Umkonfigurationsprozesses innerhalb eines Streifens umorganisiert und gepackt werden. Anders als bei dem oben beschriebenen Verfahren werden Daten anfänglich innerhalb des ausgewählten Streifens bewegt bzw. verschoben und nicht zu einer neuen Position außerhalb des ausgewählten Streifens. Außerdem kann bei diesem Verfahren eine während des Betriebs erfolgende Umkonfiguration durchgeführt werden, ohne Raum auf dem vorhandenen Speicherarray zu reservieren.

[0080] Fig. 7–Fig. 12 veranschaulichen den Prozess auf einem Plattenarray, das anfänglich aus vier Platten **0**–**3** besteht. Der physische Speicherraum des Plattenarrays wird zu mehreren Streifen zum Speichern einer vorbestimmten Datenmenge konfiguriert (Schritt **200** in Fig. 13). Jeder Streifen erstreckt sich über jede Speicherplatte **0**–**3** und umfasst ein Segment aus derselben. Jeder Streifen enthält Daten und redundante Paritätsinformationen P.

[0081] Eine Vorabbildung, die die Streifenkonfiguration oder das Speicher-Layout des physischen Speicherraums und der darin gespeicherten Daten darstellt, wird durch das Datenspeichersystem erzeugt und unterhalten (Schritt **202**). Die Vorabbildung kann als Tabellen verkörpert sein, die in dem Speicherabbildungsspeicher **21** in der Plattenarraysteuerung **14** unterhalten und aktualisiert werden (Fig. 1). Alternativ dazu kann die Vorabbildung einen Satz von arithmetischen Gleichungen aufweisen, die in das RAID-Verwaltungssystem **16** hineinentworfen sind. Diese Gleichungen leiten Abbildungsinformationen auf der Grundlage der Anzahl von Speicherplatten in dem Plattenarray ab.

[0082] Eine neue zusätzliche Platte **4** wird anschließend zu dem Plattenarray hinzugefügt (Schritt **204**). Die neue Platte ist anfänglich leer. Das RAID-Verwaltungssystem **16** erzeugt eine Nachabbildung, die eine neue Streifenkonfiguration oder ein neues Speicher-Layout des physischen Speicherraums des Plattenarrays nach einem Integrieren der zusätzlichen Speicherplatte **4** darstellt (Schritt **206**). Diese Nachabbildung kann desgleichen als Tabelle in dem Speicherabbildungsspeicher **21** gehalten werden. Alternativ dazu kann die Nachabbildung durch die arithmetischen Gleichungen erzeugt werden, wenn der Parameter der neuen Anzahl von Platten, die nun in dem Plattenarray konfiguriert sind, gegeben ist.

[0083] Bei Schritt **208** wird eine Gruppe A von Segmenten definiert. Vorzugsweise besteht die Gruppe A aus einem erweiterten Streifen, der ein Segment von jeder Speicherplatte, einschließlich der neuen zusätzlichen Speicherplatte, aufweist. Bei Fig. 7 besteht Gruppe A aus Segmenten **0**, **1**, **2**, P und einem neuen Segment N in Platte **4**. Als Nächstes wird eine Gruppe B von Segmenten definiert (Schritt **210**). Gruppe B umfasst Segmente, die nicht in der Gruppe A vorliegen. Gruppe B besteht vorzugsweise aus genügend Segmenten, um einen erweiterten Streifen zu füllen. Bei dem Beispiel der Fig. 7–Fig. 12 weist die Gruppe B anfänglich dieselbe Größe auf wie Gruppe A, einschließlich Segmenten von einem einzelnen erweiterten Streifen. Zu einem späteren Zeitpunkt in dem Prozess wird Gruppe B jedoch vergrößert, um Segmente von zwei erweiterten Streifen zu umfassen. Demgemäß können die Segmente in Gruppe B in demselben physischen Streifen vorliegen, müssen aber nicht. Nach einer Definition der Gruppen A und B beginnt das RAID-Verwaltungssystem eine Handhabung bzw. Bewältigung in Bezug auf Konflikte, die während des Erweiterungsprozesses entstehen könnten.

[0084] Zu Beginn des Prozesses könnte es nötig sein, dass Daten in den Segmenten der Gruppe A von einer ersten Datenanordnung in den Segmenten gemäß der Vorabbildung zu einer zweiten Datenanordnung in den Segmenten gemäß der Nachabbildung umorganisiert werden müssen (Schritt **213**). Das heißt, dass die Daten, die unter der Vorabbildung angeordnet sind, bezüglich der Nachabbildung ausgewertet werden, um zu bestimmen, ob die Daten nach der Erweiterung an derselben Stelle verbleiben oder umorganisiert werden müssen. Nach mehreren Iterationen in dem Prozess besteht Gruppe A jedoch gänzlich aus leeren Segmenten, und somit gibt es keine Daten zu umorganisieren.

[0085] Demgemäß wird bei Entscheidungsschritt **211** bestimmt, ob alle Segmente in der Gruppe A leer sind. Falls dies der Fall ist, wird der Umorganisierungsschritt nicht durchgeführt, und Daten werden einfach von den Segmenten in Gruppe B in die leeren Segmente in Gruppe A bewegt bzw. verschoben. Umgekehrt gilt, dass, wenn die Segmente leer sind, der nächste Entscheidungsschritt **212** bestimmt, ob die Daten angesichts der Nachabbildungskonfiguration derzeit in ihrer richtigen Reihenfolge vorliegen. Falls dies der Fall ist, wird der Umorganisierungsschritt erneut vermieden. Wenn die Daten dagegen nicht in der richtigen Reihenfolge vorliegen, wird der Umorganisierungsschritt **213** durchgeführt. Diese Schritte werden nachstehend unter Bezugnahme auf das bei Fig. 7–Fig. 12 gezeigte Beispiel ausführlicher beschrieben.

[0086] Für dieses Beispiel sei angenommen, dass die Vorabbildung eine Datenkonfiguration auf den Platten **0**–**3**, wie in Fig. 7 gezeigt, aufwies. Ferner sei angenommen, dass die Nachabbildung eine Speicherkonfiguration enthielt, die die neue Platte **4** beinhaltet, wie in Fig. 11 gezeigt.

[0087] Man beachte, dass die Daten in den Segmenten der Gruppe A der Fig. 7 bereits in der gewünschten Reihenfolge vorliegen. Es ist keine Umorganisation notwendig. Stattdessen werden Daten von dem Segment **3** der Gruppe B einfach verschoben, um das Segment P in der Gruppe A, das zuvor Paritätsinformationen enthielt, zu füllen. Wenn er sich durch das Flussdiagramm der Fig. 13 arbeitet, kehrt der Entscheidungsschritt **211** zu einem „Nein" zurück, das angibt, dass die Segmente in der Gruppe A nicht leer sind, und der Entscheidungsschritt **212** kehrt zu einem „Ja" zurück, das angibt, dass sich die Daten in der richtigen Reihenfolge befinden. Somit wird der Umorganisierungsschritt **213** umgangen, und der Fluss setzt sich direkt zum Schritt **214** fort, wo die Daten von einem Segment in der Gruppe B verschoben werden, um ein leeres Segment in der Gruppe A zu füllen. Bei Schritt **216** werden Paritätsinformationen für die Daten in den Segmenten der Gruppe A wieder hergestellt und in das neue Segment N auf der neuen Platte **4** platziert. Etwaige Konflikte bezüglich der virtuellen Blöcke, die den Segmenten in den Gruppen A und B zugeordnet sind, die während dieses Prozesses entstanden, werden anschließend gelöst.

[0088] Bei Fig. 8 wird Gruppe A dahingehend umdefiniert, ein geräumtes Segment V, ein Datensegment **4**, ein Paritätssegment P, ein Datensegment **5** und ein neues Segment N auf jeweiligen Platten **0**–**4** zu umfassen (Schritt **208**). Gruppe B wird ebenfalls dahingehend umdefiniert, Segmente **6**, P, **7**, **8** und N auf jeweiligen Platten **0**–**4** zu umfassen. Man beachte, dass sich die Daten in den Segmenten der Gruppe A nicht in der richtigen Reihenfolge befinden und somit bei Schritt **213** umorganisiert werden. Die Daten in dem Segment **4** werden von der Platte **1** zu dem geräumten Segment V auf der Platte **0** verschoben, und anschließend werden die Daten in dem Segment **5** von der Platte **3** zu dem nun geräumten Segment auf der Platte **1** verschoben.

[0089] Es ist zu beachten, dass Daten während der Umorganisierung auch zu dem neuen Segment N auf Platte **4** innerhalb der Gruppe A verschoben werden könnten. Beispielsweise könnten, wenn dies die Nachabbildung vorgibt, Daten von dem Segment **4** in Gruppe A zu einem ungenutzten neuen Segment P auf der zusätzlichen Speicherplatte **4** verschoben werden.

[0090] Nachdem die Daten bei Schritt **213** innerhalb der Gruppe A umorganisiert wurden, werden Daten von Gruppe B verschoben, um die leeren Segmente zu füllen. Gemäß der Verwendung in diesem Dokument bedeuten „leere" Segmente solche Segmente, die geräumt wurden (mit V bezeichnet), Segmente, die zuvor Paritätsinformationen speicherten (mit P bezeichnet) und neue Segmente, die durch die neue Platte bereitgestellt werden (mit N bezeichnet). In Fig. 8 werden Daten in dem Segment **6** auf der Platte **0** verschoben, um das leere Segment P auf der Platte **2** zu füllen; und Daten in dem Segment **7** auf der Platte **2** werden bewegt, um das neue Segment N auf der Platte **4** zu füllen. Die Paritätsinformationen werden wieder hergestellt und in das geräumte Segment auf der Platte **3** platziert. Wiederum werden etwaige Konflikte bezüglich der virtuellen Blöcke, die den Segmenten in den Gruppen A und B zugeordnet sind, gelöst.

[0091] Fig. 9 zeigt die dritte Iteration durch das Flussdiagramm der Fig. 13, wo Gruppen A und B umdefiniert werden, um den dritten bzw. vierten Streifen abzudecken. Diese Iteration verwendet erneut eine Kombination einer Datenumorganisierung innerhalb der Gruppe A und ein Datenpacken oder -füllen mit Daten, die von der Gruppe B geliefert werden. Die Paritätsinformationen werden für den neuen Datensatz regeneriert/erneut erzeugt und in dem Segment auf der Platte **2** gespeichert.

[0092] Fig. 10 zeigt die vierte Iteration. Gruppe A wird dahingehend umdefiniert, die Segmente in dem vierten Streifen abzudecken. Man beachte, dass Gruppe A nun alle leeren Segmente enthält. In der Gruppe A gibt es keine Daten wiederzuerkennen. Demgemäß umgeht Schritt **211** in dem Flussdiagramm der Fig. 13 den Umorganisierungsschritt **213** und geht direkt zu dem Füllschritt **214** über. Der Umorganisierungsteilprozess wird somit lediglich am Anfang durchgeführt und wird nach nur einigen wenigen Iterationen rasch beendet. In der Tat endet der Umorganisierungsschritt nach der N-1ten Iteration, wobei N gleich der Anzahl von Speicherplatten in dem Array vor dem Hinzufügen der neuen Platte(n) ist. Bei diesem Beispiel ist N gleich Vier, und somit wird der Umorganisierungsteilprozess nach der dritten Iteration nicht fortgesetzt.

[0093] Gruppe B wird dahingehend umdefiniert und vergrößert, die Segmente in dem fünften und sechsten Streifen abzudecken. Die Vergrößerung erfolgt, weil zum Füllen aller leeren Segmente in der Gruppe A mehr Daten verwendet werden, als durch einen einzigen Streifen beherbergt werden können.

[0094] Fig. 11 veranschaulicht die siebte Iteration. Bei dieser Iteration beginnt sich zwischen den Gruppen A und B eine Lücke ungenutzter Segmente zu bilden. Diese Lücke entsteht, weil Daten von einem kleineren Speicherraum in Gruppe B und darunter in einen größeren erweiterten Raum über der Gruppe A verschoben werden. Die Lücke beginnt sich zwischen der 2N-1ten Iteration zu bilden, wobei N die Anzahl von Speicherplatten in dem Array vor dem Hinzufügen der neuen Platte(n) ist. Die Lücke wächst, während der Prozess fortgesetzt wird.

[0095] Der Prozess endet, wenn die Gruppe B das

untere Ende des Plattenarrayspeicherraums erreicht. An diesem Punkt hat die Lücke ihre größte Größe erreicht und enthält alle verfügbaren ungenutzten Segmente. Ferner befinden sich alle ungenutzten Segmente unter den benutzten Segmenten in dem Plattenarray. Dies ist vorteilhaft, da es die Verwendung der arithmetischen Gleichungen (wenn sie zum Abbilden implementiert sind) erleichtert, da die Gleichungen nun mit der neuen Anzahl von Platten in dem Plattenspeicherarray geladen werden können.

[0096] Fig. 12 veranschaulicht den allgemeinen Fall des Prozesses nach vielen Iterationen. Daten sind in Segmente G, G-1, G-2, G-3 usw. über der Gruppe A gepackt, um den neuen Speicherraum auf der Platte **4** zu umfassen. Dagegen verbleiben Daten in und unter der Gruppe B in den Segmenten H, H + 1, H + 2 usw., wie sie vor dem Hinzufügen der neuen Platte existierten. Die Segmente auf der neuen Platte **4** sind leer. Die Gruppe A besteht aus gänzlich leeren Segmenten. Die Gruppe B besteht aus Daten, die in die Gruppe A gefüllt werden sollen. Die Lücke zwischen den Gruppen A und B enthält viele leere Segmente E.

[0097] Der Prozess wird Streifen um Streifen fortgesetzt, bis alle Streifen dahingehend erweitert wurden, den neuen Plattenraum zu umfassen. Danach wird der dem Host zur Verfügung stehende maximale Raum erhöht, um eine Erweiterung, die sich aus dem Hinzufügen der neuen Platte(n) ergibt, widerzuspiegeln. Das obige Verfahren kann auch in dem hierarchischen Plattenarray, das verschiedene RAID-Bereiche aufweist, verwendet werden. In diesem Fall werden Regionen in RAID-Bereichen innerhalb eines ausgewählten RAID-Bereichs umorganisiert und anschließend gemäß den obigen Techniken mit Daten von Regionen außerhalb des ausgewählten RAID-Bereichs gepackt. Derartige Regionen können in einem großen Volumen gehandhabt werden, werden jedoch vorzugsweise in ihre mehreren Streifen getrennt, so dass die Streifen in jedem RAID-Bereich Streifen um Streifen vergrößert werden, bis der gesamte RAID-Bereich auf die oben unter Bezugnahme auf Fig. 10 beschriebene Weise abgeschlossen ist.

[0098] Die Verfahren dieser Erfindung sind vorteilhaft, da sie ein heißes Einstecken zusätzlicher Speicherplatten zum Vergrößern der Speicherkapazität ermöglichen und dabei die Verfügbarkeit der Daten aufrechterhalten. Indem der neue Raum Bereich um Bereich umkonfiguriert wird, können die Daten weiterhin auf dem Plattenarray vorhanden und zugänglich sein und müssen nicht an eine separate Speichervorrichtung transferiert werden, wie dies bei Verfahren des Standes der Technik getan wird.

Patentansprüche

1. Ein Verfahren zum Vergrößern der Speicherkapazität in einem Datenspeichersystem, wobei das Datenspeichersystem ein Plattenarray (**11**) von mehreren Speicherplatten (**12**) aufweist, wobei die Speicherplatten einen physischen Speicherraum definieren; wobei das Verfahren folgende Schritte umfasst:
Konfigurieren des physischen Speicherraums (**34**) des Plattenarrays (**11**) zu mehreren Streifen zum Speichern einer vorbestimmten Datenmenge, wobei sich einzelne Streifen über mehrere Speicherplatten (**0**–**3**) in dem Plattenarray (**11**) erstrecken und ein Segment der gleichen Größe von jeder der einzelnen Speicherplatten in dem Plattenarray aufweisen;
Hinzufügen zumindest einer zusätzlichen Speicherplatte (**4**) zu dem Plattenarray (**11**);
dadurch gekennzeichnet, dass dies folgende Schritte aufweist:
Auswählen eines Streifens;
Bewegen von Daten von zumindest einem Segment in dem ausgewählten Streifen zu einer ungenutzten Stelle auf dem Plattenarray; und
Umkonfigurieren des physischen Speicherraums (**34**), der den ausgewählten Streifen enthält, zu einem erweiterten Streifen zum Speichern von Daten, wobei der erweiterte Streifen die mehreren Speicherplatten (**0**–**3**) und die neue, zusätzliche Speicherplatte (**4**) überspannt, wobei der erweiterte Streifen ein Segment von jeder der einzelnen Speicherplatten und ein Segment von der neuen, zusätzlichen Speicherplatte aufweist.

2. Ein Verfahren gemäß Anspruch 1, das ferner folgende Schritte aufweist:
Reservieren eines Abschnitts des physischen Speicherraums (**34**) in dem Plattenarray (**11**) einer ausreichenden Größe, um zumindest eine in einem einzelnen Streifen gespeicherte Datenmenge zu speichern; und
Bewegen aller Segmente von dem ausgewählten Streifen zu dem Reserveabschnitt.

3. Ein Verfahren gemäß Anspruch 1, bei dem der Schritt des Bewegens von Daten ferner ein Umorganisieren von Segmenten in dem Streifen umfasst, indem Daten von einem Segment zu entweder (**1**) einem anderen Segment in dem ausgewählten Streifen oder (**2**) einem Segment von der neuen zusätzlichen Platte, das in den Streifen aufgenommen werden soll, bewegt werden.

4. Ein Verfahren gemäß Anspruch 1, das ferner folgende Schritte umfasst:
Bereitstellen einer Vorkartierung, die die Konfiguration des physischen Speicherraums (**34**) des Plattenarrays anzeigt, vor dem Schritt des Hinzufügens der zumindest einen zusätzlichen Speicherplatte (**4**) zu dem Plattenarray; und
Erstellen einer Nachkartierung, die eine neue Konfiguration des physischen Speicherraums des Plattenarrays nach Integration der zusätzlichen Speicherplatte anzeigt.

5. Ein Verfahren gemäß Anspruch 4, bei dem das Umkonfigurieren des physischen Speicherraums folgende Schritte umfasst:
Definieren einer Gruppe A von Segmenten, die zumindest einen erweiterten Streifen umfassen sollen, wobei der erweiterte Streifen ein Segment von den einzelnen Speicherplatten (**0**–**3**) und ein Segment der neuen, zusätzlichen Speicherplatte (**4**) umfasst;
Definieren einer Gruppe B von Segmenten, die Segmente umfassen sollen, die man in der Gruppe A nicht findet;
Umorganisieren der Daten in den Segmenten der Gruppe A von einer ersten Datenanordnung in den Segmenten gemäß der Vorkartierung zu einer zweiten Datenanordnung in den Segmenten gemäß der Nachkartierung, wobei die Gruppe A von Segmenten zumindest ein leeres Segment aufweist; und
Bewegen von Daten von zumindest einem Segment in Gruppe B, um das zumindest eine leere Segment in Gruppe A zu füllen.

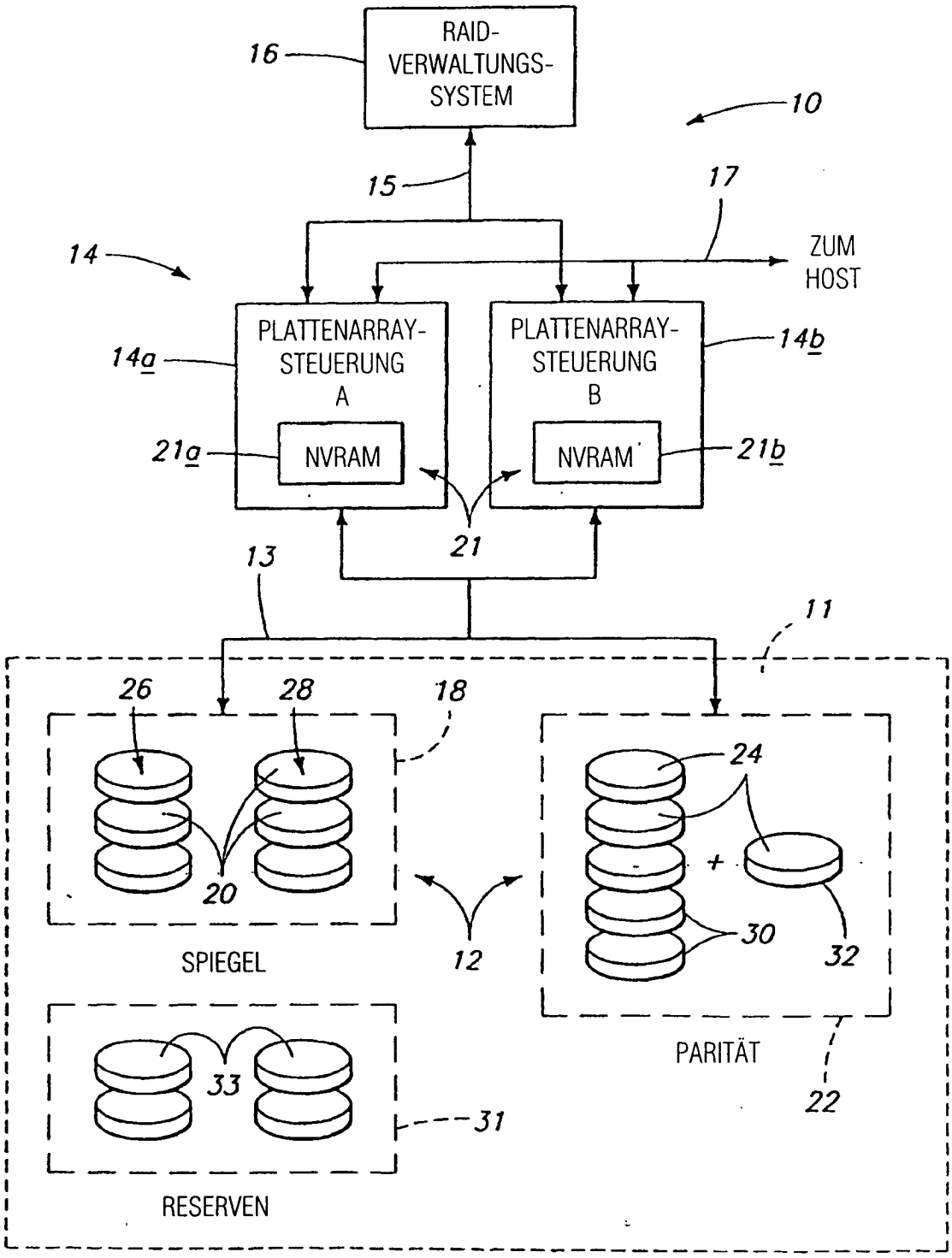
6. Ein Verfahren gemäß Anspruch 5, das ferner gekennzeichnet ist durch:
Wiederherstellen einer Parität auf der Basis der in den Segmenten der Gruppe A enthaltenen Daten.

7. Ein Verfahren gemäß Anspruch 5, das ferner dadurch gekennzeichnet ist, dass der Umorganisierungsschritt nicht fortgesetzt wird.

8. Ein Verfahren gemäß Anspruch 5, das ferner dadurch gekennzeichnet ist, dass Konflikte zwischen einem Zugriff auf die Daten in den Segmenten der Gruppen A und B und einer Bewegung dieser Daten gelöst werden.

Es folgen 9 Blatt Zeichnungen

Anhängende Zeichnungen



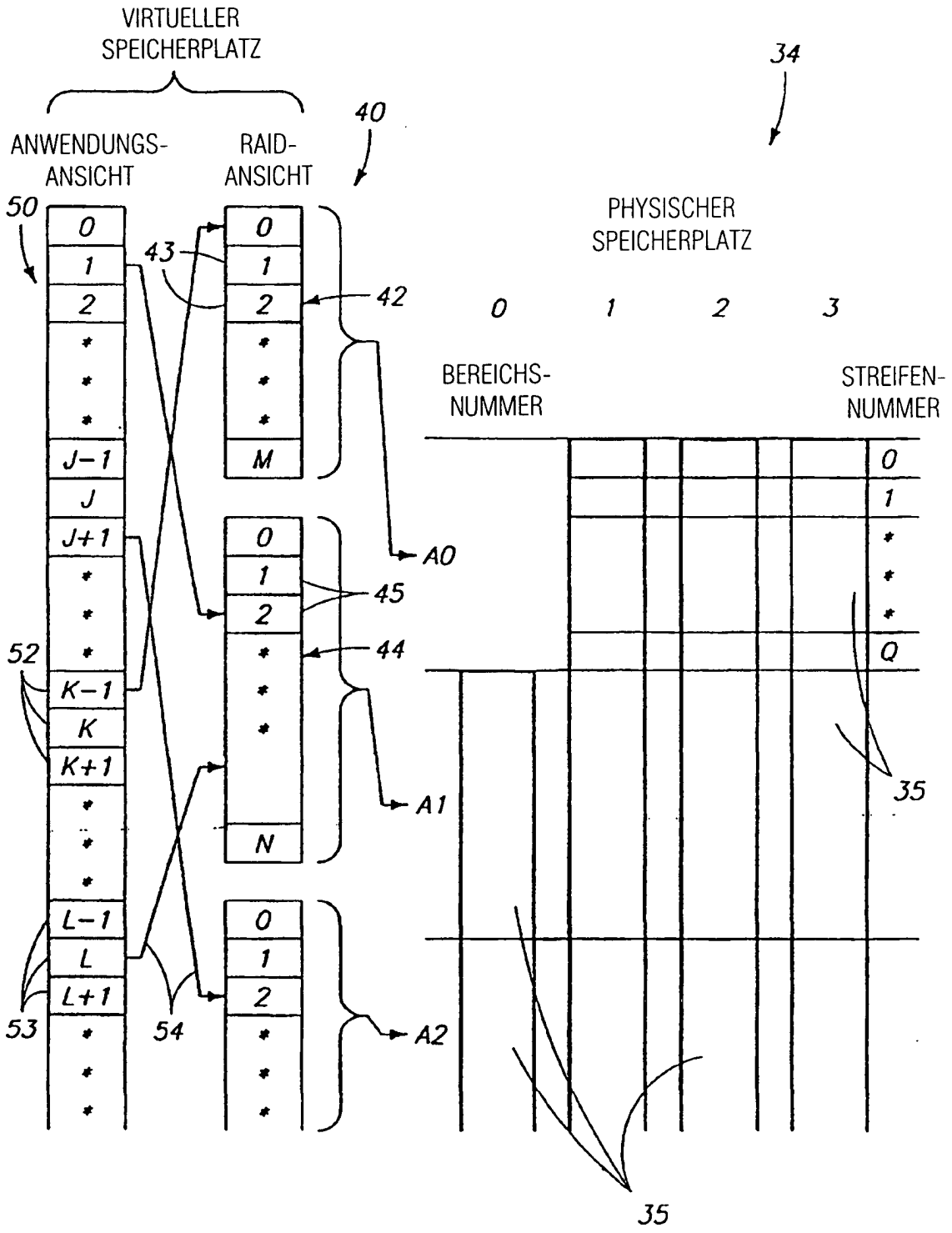
FIGUR 1

PLATTEN				STREIFEN-NUMMER
0	1	2	3	
0	0'	1	1'	0
2	2'	3	3'	1
4	4'	5	5'	2
6	6'	7	7'	3
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
T−1	T−1'	T	T'	S

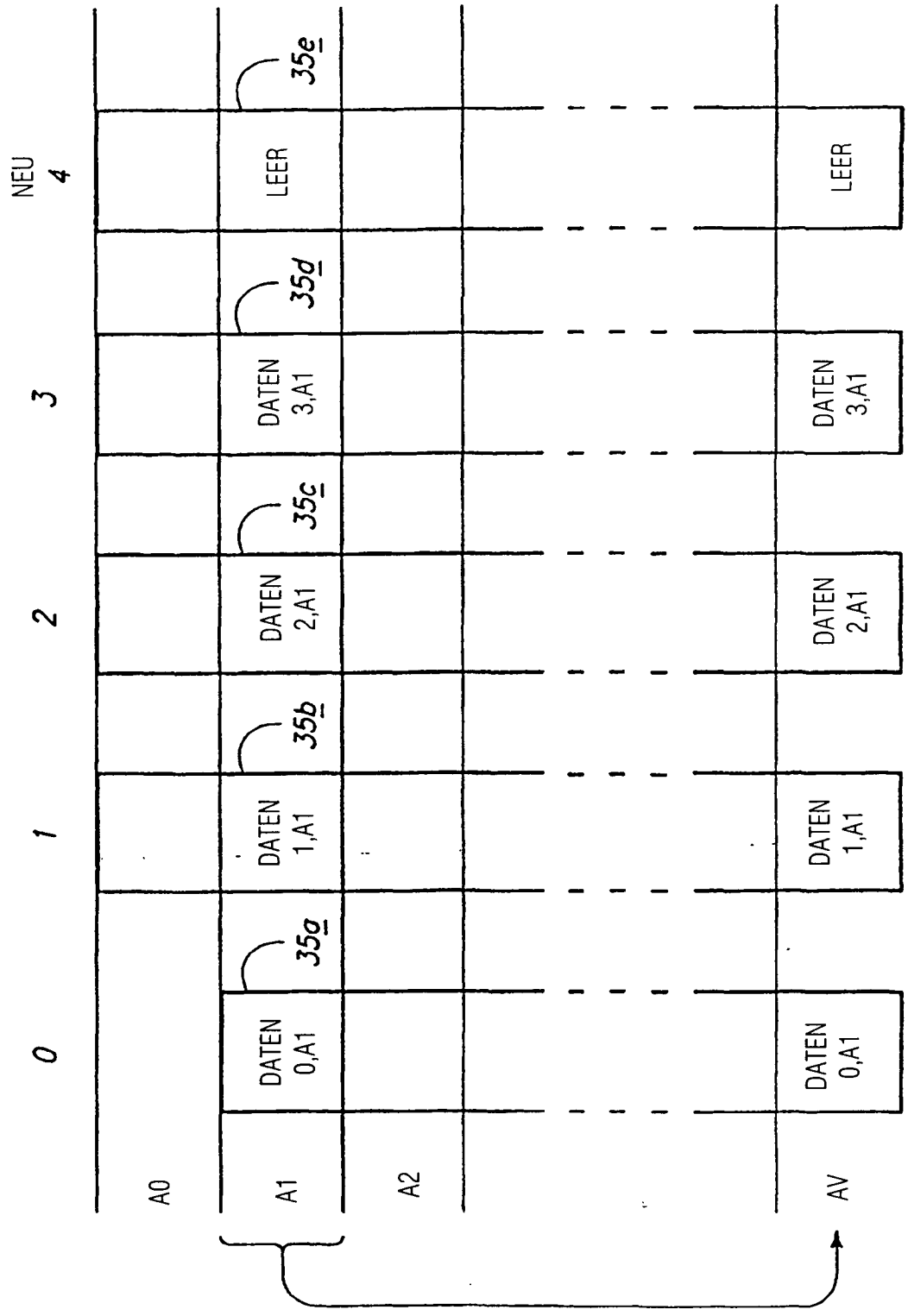
FIGUR 2

PLATTEN				STREIFEN-NUMMER
0	1	2	3	
0	1	2	P	0
3	4	P	5	1
6	P	7	8	2
P	9	10	11	3
*	*	*	*	*
*	*	*	*	*
*	*	*	*	*
P	R−2	R−1	R	Q

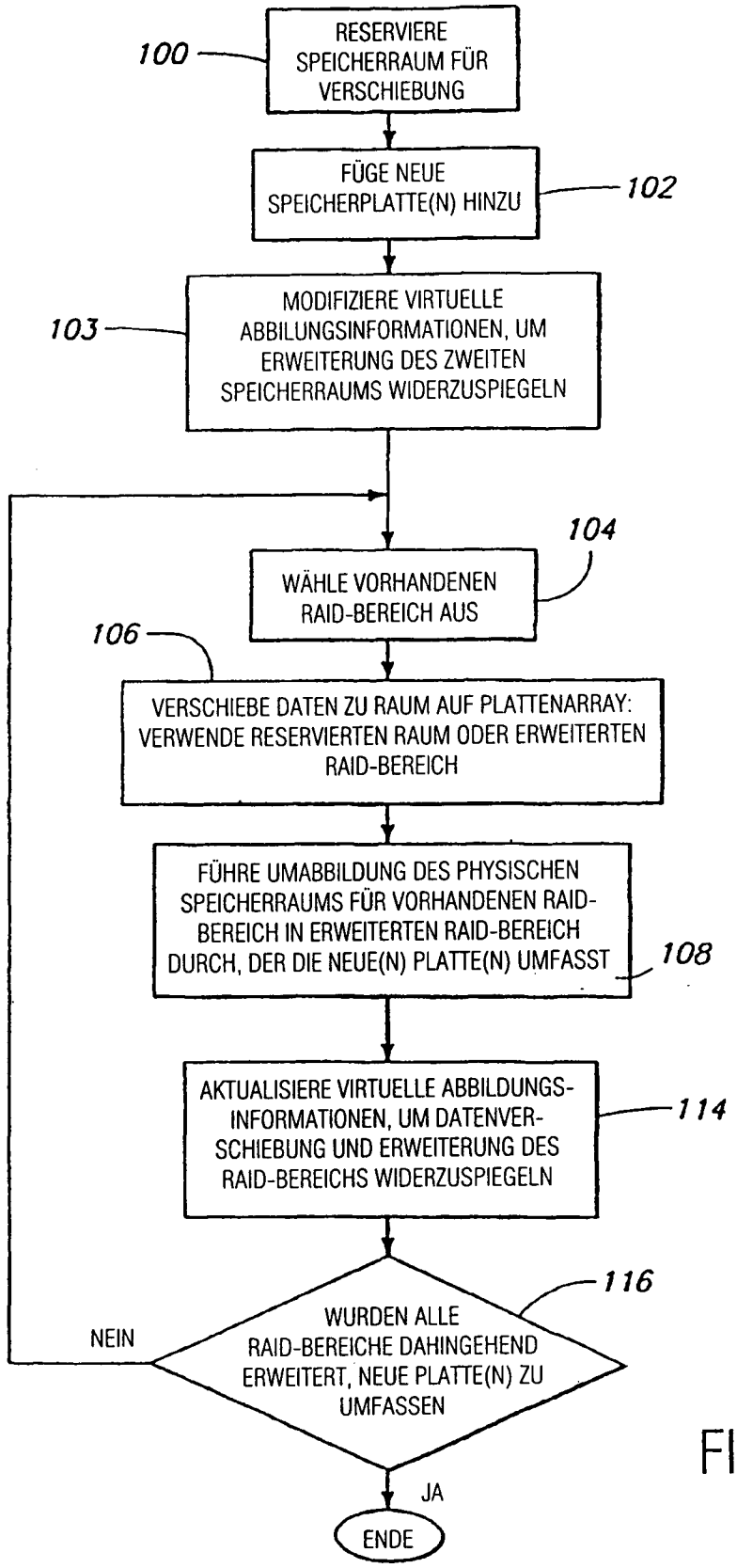
FIGUR 3


VIRTUELLER SPEICHERPLATZ
ANWENDUNGS-ANSICHT
RAID-ANSICHT
40
34
50
PHYSISCHER SPEICHERPLATZ
0
1
2
3
BEREICHS-NUMMER
STREIFEN-NUMMER
43
42
45
44
52
53
54
A0
A1
A2
35
35
0
1
2
*
J-1
J
J+1
K-1
K
K+1
L-1
L
L+1
M
N
Q

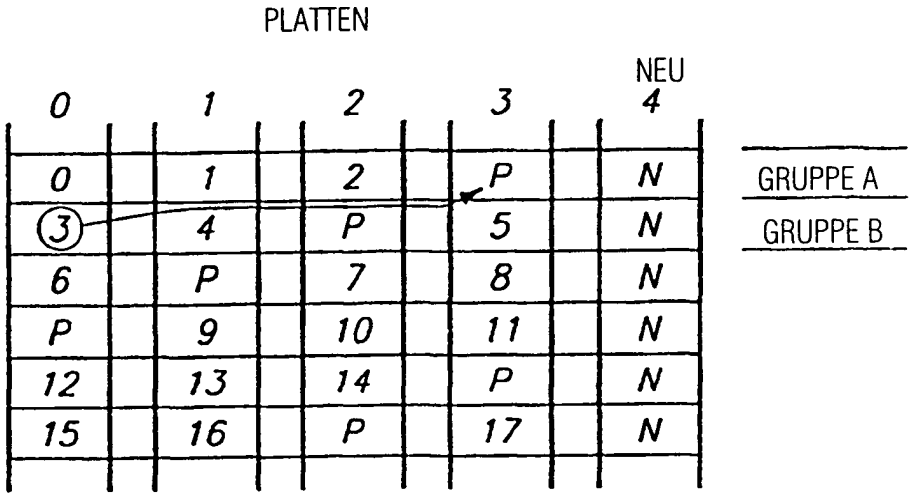
FIGUR 4


NEU
0
1
2
3
4
A0
A1
A2
AV
DATEN 0,A1
DATEN 1,A1
DATEN 2,A1
DATEN 3,A1
LEER
35a
35b
35c
35d
35e

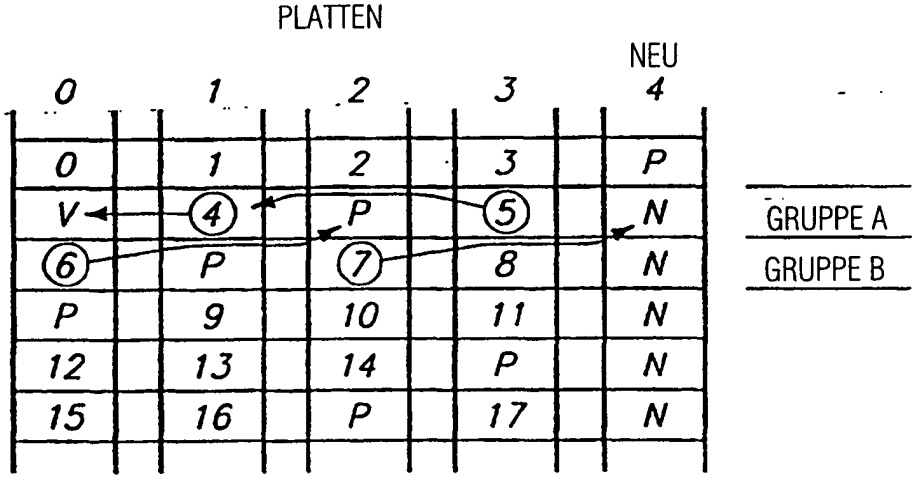
FIGUR 5


100
RESERVIERE SPEICHERRAUM FÜR VERSCHIEBUNG
FÜGE NEUE SPEICHERPLATTE(N) HINZU
102
103
MODIFIZIERE VIRTUELLE ABBILUNGSINFORMATIONEN, UM ERWEITERUNG DES ZWEITEN SPEICHERRAUMS WIDERZUSPIEGELN
104
WÄHLE VORHANDENEN RAID-BEREICH AUS
106
VERSCHIEBE DATEN ZU RAUM AUF PLATTENARRAY: VERWENDE RESERVIERTEN RAUM ODER ERWEITERTEN RAID-BEREICH
FÜHRE UMABBILDUNG DES PHYSISCHEN SPEICHERRAUMS FÜR VORHANDENEN RAID-BEREICH IN ERWEITERTEN RAID-BEREICH DURCH, DER DIE NEUE(N) PLATTE(N) UMFASST
108
AKTUALISIERE VIRTUELLE ABBILDUNGS-INFORMATIONEN, UM DATENVER-SCHIEBUNG UND ERWEITERUNG DES RAID-BEREICHS WIDERZUSPIEGELN
114
116
WURDEN ALLE RAID-BEREICHE DAHINGEHEND ERWEITERT, NEUE PLATTE(N) ZU UMFASSEN
NEIN
JA
ENDE

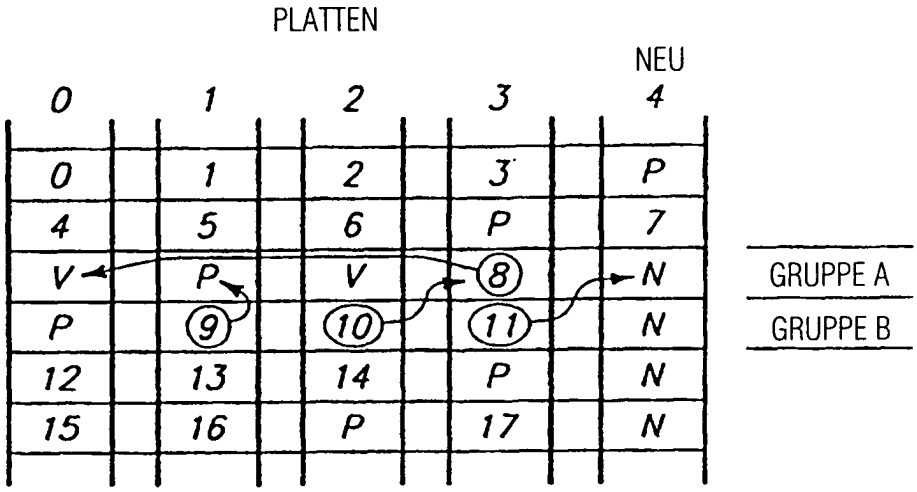
FIGUR 6



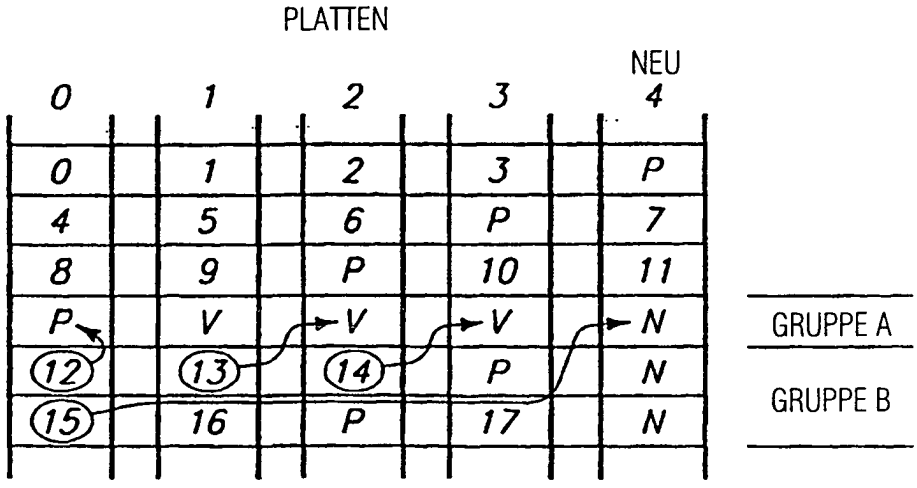
FIGUR 7



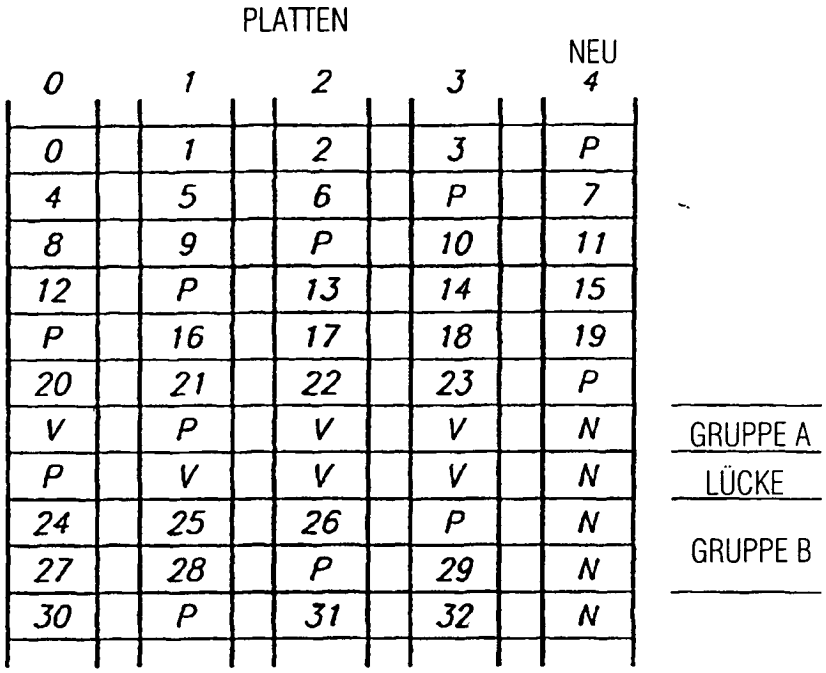
FIGUR 8



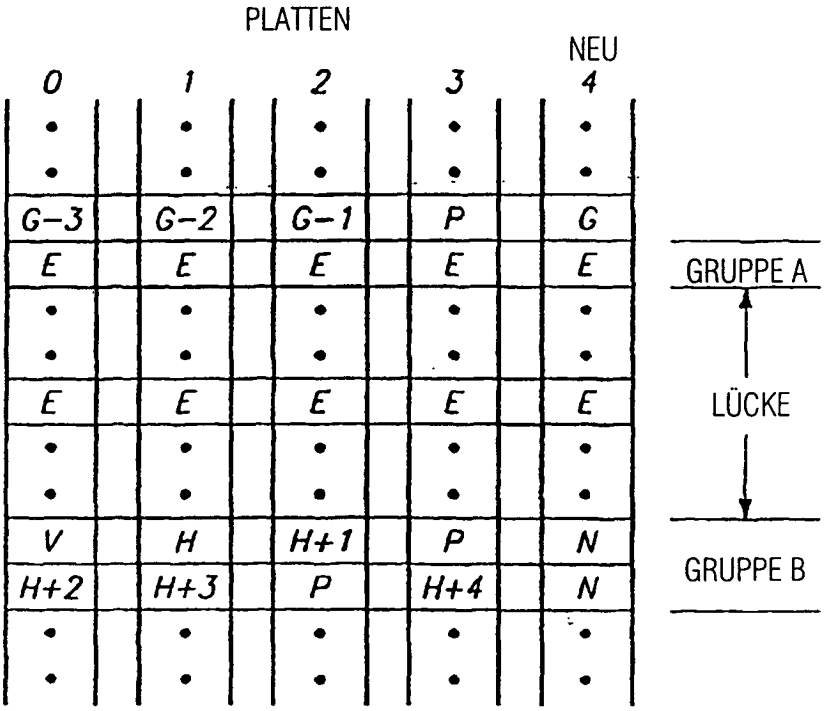
FIGUR 9



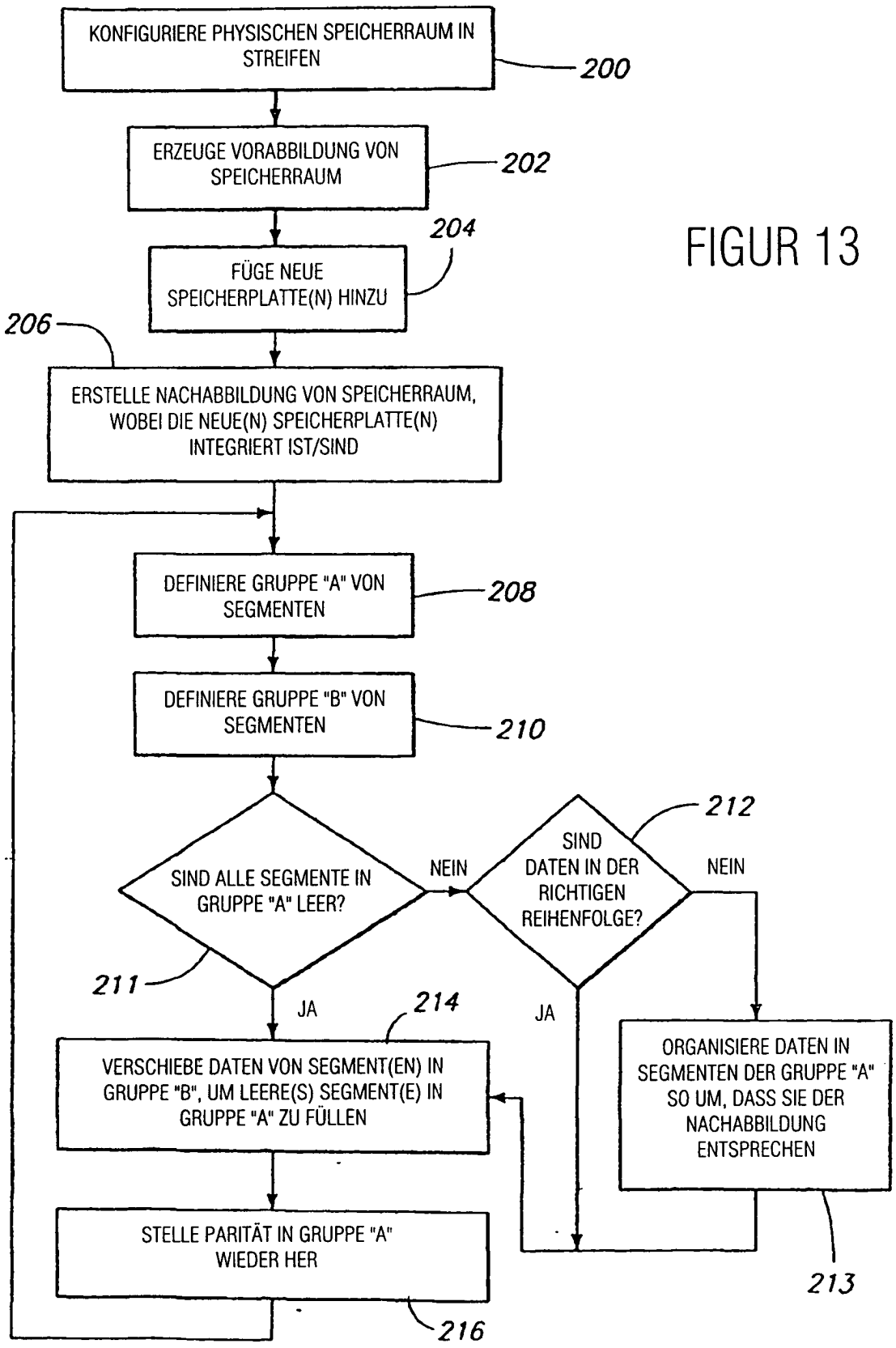
FIGUR 10



FIGUR 11



FIGUR 12


KONFIGURIERE PHYSISCHEN SPEICHERRAUM IN STREIFEN
200
ERZEUGE VORABBILDUNG VON SPEICHERRAUM
202
204
FÜGE NEUE SPEICHERPLATTE(N) HINZU
206
ERSTELLE NACHABBILDUNG VON SPEICHERRAUM, WOBEI DIE NEUE(N) SPEICHERPLATTE(N) INTEGRIERT IST/SIND
DEFINIERE GRUPPE "A" VON SEGMENTEN
208
DEFINIERE GRUPPE "B" VON SEGMENTEN
210
SIND ALLE SEGMENTE IN GRUPPE "A" LEER?
NEIN
212
SIND DATEN IN DER RICHTIGEN REIHENFOLGE?
NEIN
211
JA
214
JA
VERSCHIEBE DATEN VON SEGMENT(EN) IN GRUPPE "B", UM LEERE(S) SEGMENT(E) IN GRUPPE "A" ZU FÜLLEN
ORGANISIERE DATEN IN SEGMENTEN DER GRUPPE "A" SO UM, DASS SIE DER NACHABBILDUNG ENTSPRECHEN
213
STELLE PARITÄT IN GRUPPE "A" WIEDER HER
216

FIGUR 13