



SCHWEIZERISCHE EIDGENOSSENSCHAFT
BUNDESAMT FÜR GEISTIGES EIGENTUM

51 Int. Cl.³: G 06 F
G 06 F

13/00
9/00

Erfindungspatent für die Schweiz und Liechtenstein

Schweizerisch-liechtensteinischer Patentschutzvertrag vom 22. Dezember 1978



12 PATENTSCHRIFT A5

11

628 167

21 Gesuchsnummer: 2088/78

22 Anmeldungsdatum: 27.02.1978

30 Priorität(en): 16.03.1977 US 778290

24 Patent erteilt: 15.02.1982

45 Patentschrift
veröffentlicht: 15.02.1982

73 Inhaber:
International Business Machines Corporation,
Armonk/NY (US)

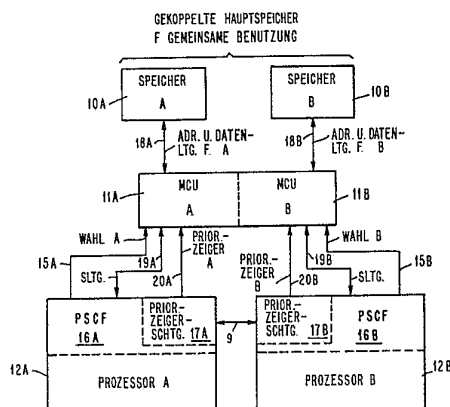
72 Erfinder:
William David Connors, Poughkeepsie/NY (US)
Dale Milton Junod, Highland/NY (US)

74 Vertreter:
Dipl.-Ing. Wolfgang Möhlen, c/o IBM Corp.,
Rüschlikon

54 **Prozessoreinrichtung zur Prioritätssteuerung der Zugriffe zum gemeinsam benutzten Arbeitsspeicher einer Mehrprozessoranlage.**

57 Die Prozessoreinrichtung umfasst eine Prioritätszeigerschaltung (17A,17B) zur Ausgabe eines Zeigersignals, wenn dem betreffenden Prozessor (12A,12B) Zugriff zum Arbeitsspeicher (10A,10B) gewährt ist. Zugriffsgewährung erfolgt auf Anforderung des Prozessors nach Auswertung intern auftretender Speicheranforderungssignale durch eine lokale Prioritätsschaltung. Letztere leitet die intern ausgewählte Anforderung der Prioritätszeigerschaltung zu, wobei der Speicherzugriff im nächsten Speicherzyklus einsetzt, in dem ein anderer mit Speicherzugriff operierender Prozessor keine Anforderung mehr vorrätig hat. Die Gewährung des Zugriffs löst ein Prioritätszeigersignal und ein Speicherwahlsignal aus, welche durch eine Übertragungseinrichtung (11A,11B) die Datenverbindungen (18A,18B) zwischen Prozessor (12A,12B) und Arbeitsspeicher (10A,10B) steuern. Liegen Speicheranforderungen eines weiteren Prozessors der Anlage vor, dann bricht der Zugriff des einen ab, sobald ihm während eines Speicherzyklus die notwendige Anforderung fehlt. Verlangen nach einer Pause zwei Prozessoren gleichzeitig Zugriff, dann erhält jener den Vorrang, der als letzter bereits Zugriff hatte.

Damit ist eine Prozessoreinrichtung geschaffen, die eine dynamische Zuordnung des Zugriffs erlaubt, so dass zu verschiedenen Zeiten die Priorität von einem Prozessor zu einem anderen ohne eine vorgegebene Ordnung unter den Prozessoren verschoben wird.



PATENTANSPRÜCHE

1. Prozessoreinrichtung zur Prioritätssteuerung der Zugriffe zum gemeinsam benutzten Arbeitsspeicher (10A, 10B) einer Mehrprozessoranlage, wobei jeder Prozessor (12A, 12B) mit mehreren Registern (56, 57, 58) zur Speicherung lokaler Zugriffsanforderungen, mit einer lokalen Prioritätsschaltung (64) zur Verwertung dieser Zugriffsanforderungen, sowie mit einer Übertragungseinrichtung (11A, 11B) zur Steuerung der Datenverbindungen (15A, 19A, 20A; 18A; 15B, 19B, 20B; 18B) zwischen Prozessor (12A, 12B) und Arbeitsspeicher (10A, 10B) in beiden Richtungen ausgerüstet ist, dadurch gekennzeichnet, dass eine Prioritätszeigerschaltung (17A, 17B) vorgesehen ist und mit dem Ausgang (75A) der genannten lokalen Prioritätsschaltung (64) in Verbindung steht und dass die Ausgangsleitung (20A, 20B) der Prioritätszeigerschaltung (17A, 17B), auf der das Auftreten eines Zeigersignals den dem betreffenden Prozessor (12A, 12B) gewährten Zugriff zum Arbeitsspeicher (10A, 10B) anzeigt, an die genannte Übertragungseinrichtung (11A, 11B) geführt ist.

2. Prozessoreinrichtung nach Patentanspruch 1, dadurch gekennzeichnet, dass die genannte Prioritätszeigerschaltung (17A) erste Schaltmittel (76A, 77A, 78A) zur Übernahme und Weitergabe lokal ausgegebener Zugriffsanforderungssignale, mehrere logische Schaltglieder (81A, 82A, 83A) zur Verarbeitung der weitergegebenen Zugriffsanforderungssignale und ein erstes bistabiles Schaltelement (87A) umfasst, das an die kombinierten Ausgänge der genannten logischen Schaltglieder (81A, 82A, 83A) angeschlossen ist, dass ein zweites bistabiles Schaltelement (88A) mit dem Ausgang des ersten bistabilen Schaltelementes (87A) und mit einer ein verzögertes Taktsignal zuführenden Leitung sowie mit der genannten Ausgangsleitung (20A) zwecks Ausgabe des Prioritätszeigersignals für den betreffenden Prozessor (12A) verbunden ist, dass zudem eine Rückkopplungsleitung von dieser Ausgangsleitung (20A) des zweiten bistabilen Schaltelementes (88A) an Eingänge von wenigstens zwei der logischen Schaltglieder (82A, 83A) zurückgeführt ist und dass schliesslich vom Ausgang des ersten bistabilen Schaltelementes (87A) über einen Inverter (91) eine Verbindungsleitung (92) an Eingänge entsprechender logischer Schaltglieder (82B, 83B) wenigstens eines weiteren Prozessors (12B) geführt sind.

3. Prozessoreinrichtung nach Patentanspruch 2, dadurch gekennzeichnet, dass die lokale Prioritätsschaltung (64) mehrere Ausgänge für die internen Zugriffsanforderungssignale des Prozessors besitzt, wobei aber zu gegebenem Zeitpunkt jeweils nur ein solches Signal verfügbar ist, dass für jeden Ausgang der lokalen Prioritätsschaltung (64) ein weiteres logisches Schaltglied (66, 67, 68) vorgesehen und je ein Eingang desselben mit einer Rückleitung (87A-1) verbunden ist, deren anderes Ende an den Ausgang des ersten bistabilen Schaltelementes (87A) der Prioritätszeigerschaltung (17) geschlossen ist, ferner dass die Ausgänge der weiteren logischen Schaltglieder (66, 67, 68) je an den Eingang dritter bistabiler Schaltelemente (71, 73, 74) geführt und deren Ausgänge über ein ODER-Glied (70) an eine Speicherwahlleitung (15A) zur Ausgabe eines Arbeitsspeicherwahlsignals geschlossen sind und dass schliesslich die Speicherwahlleitung (15A) mit der genannten Übertragungseinrichtung (11A) zwecks Steuerung des Durchschaltens von Datenleitungen (18, 19) zwischen dem angeforderten Arbeitsspeicher und dem anfordernden Prozessor (12) verbunden ist.

Die Erfindung betrifft eine Prozessoreinrichtung zur Prioritätssteuerung der Zugriffe zum gemeinsam benutzten Arbeitsspeicher einer Mehrprozessoranlage, wobei jeder Prozessor mit mehreren Registern zur Speicherung lokaler Zugriffsanforde-

rungen, mit einer lokalen Prioritätsschaltung zur Verwertung dieser Zugriffsanforderungen, sowie mit einer Übertragungseinrichtung zur Steuerung der Datenverbindungen zwischen Prozessor und Arbeitsspeicher in beiden Richtungen ausgerüstet ist.

Über die Prioritätsbestimmung des Speicherzugriffs innerhalb eines Prozessors und auch bei mehreren Prozessoren in einem Mehrprozessorsystem gibt es bereits viele Publikationen. Der Wettbewerb um den gleichzeitigen Speicherzugriff durch mehrere Prozessoren in einer Mehrprozessoranlage kann auf verschiedene Art gelöst werden. Eine solche Lösung arbeitet mit vorgegebenen Prioritäten unter mehreren Prozessoren, so dass bei einem gleichzeitigen Zugriff der Prozessor mit der höchsten Priorität zuerst den Zugriff erhält. Dies ist eine Art statischer Priorität, sie ändert sich nicht. Eine andere Lösung besteht darin, die Zugriffsgewährung zwischen mehreren Prozessoren in einem Mehrprozessorsystem abzuwechseln, so dass jeder Prozessor eine bestimmte Phase der wechselnden Zyklen eines Taktgebers zugeteilt erhält. Wenn ein Prozessor während seines Zyklus keine Anforderung stellt, dann bleibt der Zyklus auch durch andere Prozessoren ungenutzt. Das ist eine andere Form einer vorgegebenen Priorität.

Die Aufgabe der vorliegenden Erfindung besteht in der Schaffung einer Einrichtung, durch welche den Prozessoren einer Mehrprozessoranlage die Zugriffe zum gemeinsam benutzten Arbeitsspeicher unter Vermeidung von im voraus festgelegten Prioritäten gewährt werden. Der Zugriff soll dabei nach Bedarf solange gewährt sein, als ein Prozessor laufend und ohne Auslassung eines Speicherzyklus Zugriffsanforderungen stellt, wodurch der Wirkungsgrad der Speicherbelegung erhöht wird. Die genannte Einrichtung ist derart auszuführen, dass sie in jeden Prozessor einer Mehrprozessoranlage integriert werden kann, unabhängig davon, ob die Prozessoren selbständig mit eigenen Speichern oder im Verbund bei Benutzung eines gemeinsamen Arbeitsspeichers operieren.

Diese Aufgabe wird durch die im kennzeichnenden Teil des Patentanspruches 1 aufgeführten Merkmale gelöst.

Mit Hilfe der vorliegenden Erfindung wird nun die Priorität nicht auf der bisher gebräuchlichen statischen Basis zugunsten eines bestimmten Prozessors in einer Mehrprozessoranlage festgelegt, sondern es wird eine dynamische Zuordnung des Zugriffs vorgesehen, die zu verschiedenen Zeiten die Priorität von einem Prozessor zu einem anderen ohne eine vorgegebene Ordnung unter den Prozessoren verschiebt.

Mit Hilfe der Erfindung wird folgende, im Ausführungsbeispiel gezeigte Arbeitsweise ermöglicht:

Jeder Prozessor in einem Mehrprozessorsystem steuert den Speicher mit einem Bündel von Speicheranforderungen an. Ein Bündel dauert so lange, wie ein Prozessor lokal eine Speicheranforderung für jeden folgenden Speicherzyklus stellen kann. Das Bündel endet, sobald ein Prozessorzyklus ohne gestellte Speicheranforderung abläuft. Ein Bündel kann aus einer und bis zu mehreren Dutzend Speicheranforderungen bestehen. Während eines Bündels wird zum gemeinsam benutzten Arbeits- oder Hauptspeicher durch einen Prozessor so rasch zugegriffen, als es die Zykluszeit des Speichers gestattet, wobei eine Störung durch einen anderen Prozessor nicht erlaubt ist. Bei Beendigung eines Bündels erhält ein anderer Prozessor, der eine Speicheranforderung verfügbar hat, die Priorität während des anschliessenden Zyklus, so dass beim Umschalten kein Speicherzyklus verloren geht. Dieser Prozessor arbeitet dann mit der höchsten Priorität weiter, bis er sein Bündel durch Auslassen eines Zyklus beendet, wenn er während eines Zyklus keine Anforderung abgibt. Wenn in einem Speicherzyklus kein Prozessor eine Anforderung macht und während eines nächsten Zyklus zwei Prozessoren gleichzeitig eine Anforderung abgeben, erhält derjenige die Priorität, der zuletzt die Priorität für ein Bündel hatte. So

wird die Priorität zwischen den Prozessoren hin- und her verschoben, während sie den Hauptspeicher in Anforderungsbündeln ansteuern.

Durch die Erfindung wird der Wirkungsgrad des Mehrprozessorbetriebes dadurch verbessert, dass ungenutzte Speicherzyklen vermieden werden, während Anforderungen vorliegen. Herkömmliche Prioritätssteuerungen für Mehrprozessorsysteme haben einen niedrigen Wirkungsgrad, weil sie aufgrund ihres Prinzips der Prioritätszuordnung zu verschiedenen Prozessoren im allgemeinen Speicherzyklen unbenutzt lassen. Der höchste Wirkungsgrad für ein Mehrprozessorsystem wird erreicht, wenn die Summe der Speicherzugriffsraten der Prozessoren im System ungefähr 100% der Speicherzugriffsraten des Hauptspeichers beträgt. Wenn die Prozessoren identisch sind, beträgt die optimale Speicherzugriffsraten für jeden Prozessor ungefähr 100/N%, worin N die Anzahl der Prozessoren im System ist. Auch bei identischen Prozessormodellen ändert sich natürlich die Speicherzugriffsraten des Prozessors mit den einzelnen programmierten Aufgaben. In einem Mehrprozessorsystem mit zwei grossen Prozessoren wurde festgestellt, dass die interne Zugriffsraten eines jeden Prozessors 12 bis 15% und die seines Kanals durchschnittlich bei 2% liegt, so dass sich für den einzelnen Prozessor eine Rate von 14 bis 17% der maximalen Hauptspeicherrate ergibt.

Ein Ausführungsbeispiel der Erfindung ist in den Zeichnungen erklärt und wird anschliessend näher beschrieben.

Es zeigen:

Fig. 1 in einem Blockdiagramm ein Mehrprozessorsystem mit zwei erfindungsgemäss ausgerüsteten Prozessoren,

Fig. 2A die Teile einer Prozessor-Speichersteuerung, welche für die Speicheranforderung und -wahl benutzt werden,

Fig. 2B zwei Prioritätszeigerschaltungen und ihre Verbindungen im Ausführungsbeispiel der Erfindung,

Fig. 3 und 4 Speicher-Vermittlungsschaltungen in zwei Übertragungseinheiten eines Mehrprozessorsystems, welche die Datenübertragungen zwischen den zwei Prozessoren und dem gemeinsam benutzten, aus der engen Kopplung der Speichereinheiten eines jeden Prozessors im System entstandenen Hauptspeicher steuern und

Fig. 5 Zeitdiagramme mit Beispielen der Zeiteinteilung im Ausführungsbeispiel der Erfindung.

Das in Fig. 1 gezeigte Mehrprozessorsystem ist mit zwei Speichereinheiten 10A und 10B ausgerüstet, die voneinander getrennt in den entsprechenden Prozessoren 12A und 12B angeordnet sind. Die Speichereinheiten 10A und 10B werden als eine festgekoppelte Hauptspeichereinheit behandelt, die von mehreren Prozessoren in einem Mehrprozessorsystem gemeinsam benutzt wird. Die adressierbare Grösse dieser Hauptspeichereinheit ist gleich der Summe ihrer separaten Adressbereiche. So können z.B. die unteren Adressen im festgekoppelten Speichersystem in der Speichereinheit 10A des Prozessors 12A vorgesehen werden und die oberen Adressen in der Speichereinheit 10B des Prozessors 12B. Davon abweichend kann man die durchgehenden Adressen auch auf lokale Speicherbereiche (LSU) der zwei Speichereinheiten 10A und 10B in jeder beliebigen Ordnung aufteilen.

Fig. 1 zeigt in einem Blockdiagramm ein Mehrprozessorsystem mit den beiden Prozessoren 12A und 12B, die über ihre Übertragungseinheiten (MCU) 11A und 11B mit den Speichereinheiten 10A und 10B verbunden sind. Adress- und Datensammelleitungen 19A bzw. 19B verbinden die Prozessor-Speichersteuerung (PSCF) 16A und 16B in den Prozessoren 12A und 12B mit den entsprechenden Übertragungseinheiten 11A und 11B. Die Bahnen der Adress- und Datensammelleitungen werden dann vervollständigt und laufen von den Übertragungseinheiten 11A und 11B zu den entsprechenden Hauptspeichereinheiten 10A und 10B über die Adress- und Datensammellei-

tungen 18A bzw. 18B. Prioritätszeigerschaltungen 17A und 17B sind in den Prozessoren vorgesehen. Leitungen 9 verbinden die Prioritätszeigerschaltungen 17A und 17B miteinander. Die Prioritätszeiger-Ausgangsleitungen 20A und 20B verbinden die entsprechenden Prioritätszeigerschaltungen mit den Übertragungseinheiten 11A und 11B, um die Datenübertragung auf den Adress- und Datensammelleitungen vom gewählten Prozessor an das gemeinsame Speichersystem zu leiten. Die Ausgangssignale der Prioritätszeigerschaltungen 17A und 17B leiten entsprechend auch Wahlschaltungen in der Prozessor-Speichersteuerung 16A bzw. 16B mit entsprechenden Steuersignalen auf den Prozessor-Wahlleitungen 15A und 15B, die die Datenübertragung vom gemeinsamen Speichersystem an einen der gewählten Prozessoren 12A oder 12B auf der entsprechenden Adress- und Datensammelleitung schalten.

Fig. 2A zeigt im einzelnen die Speicheranforderungs- und Wahlsteuerung in der Prozessor-Speichersteuerung 16A im Prozessor 12A.

Die Prozessor-Speichersteuerung in Fig. 2A enthält konventionelle Schaltungen zur Erzeugung von Speicheranforderungen, Pufferschaltungen und Prioritätszuweisungsschaltungen, wie sie in einem einzelnen Prozessor untergebracht sind. Diese Schaltungen umfassen eine lokale Prioritätsschaltung 64, die alle Prioritätsanforderungen innerhalb des Prozessors bewertet, so dass sie zu einer gegebenen Zeit nur eine Zugriffsanforderung zum Speicher auf einer Leitung 75A für den Prozessor 12A ausgibt. Eine entsprechende Prozessor-Speichersteuerung 16B im Prozessor 12B gibt ihre Einzelanforderungen auf der Leitung 75B parallel zur Leitung 75A aus.

Die festgekoppelte Speicheranordnung gestattet somit einem jeden Prozessor den Zugriff zu einem wesentlich grösseren Speichersystem, als er sonst adressieren könnte, wenn seine Zugriffe nur auf seinen eigenen Hauptspeicher beschränkt wären. Jeder der Prozessoren 12A oder 12B hat daher Zugriff zu den Adressen in beiden Hauptspeichereinheiten 10A und 10B, so dass die Systemsteuerprogramme und Daten in den beiden Einheiten 10A und 10B nicht dupliziert werden müssen. Jeder Prozessor besitzt ausserdem die doppelte Verfügbarkeit beider Hauptspeichereinheiten, falls eine von beiden ausfallen sollte.

Wenn beide Speichereinheiten 10A und 10B fest zu einem Hauptspeicher gekoppelt sind, können beide Prozessoren nicht mehr gleichzeitig Speicherzugriffe zu ihren entsprechenden Speichern 10A und 10B vornehmen. Die erfindungsgemässe Mehrprozessor-Prioritätsanordnung kann jedoch die seriellen Zugriffe für beide Prozessoren ungefähr auf der Höhe der Summe der separaten Zugriffsraten für die Prozessoren zu nicht gekoppelten separaten Speichereinheiten 10A und 10B halten. Dabei ist jedoch vorausgesetzt, dass die Zugriffsraten zum festgekoppelten Hauptspeicher höher ist als die Summe der Speicheranforderungsraten beider Prozessoren. Die heutige Hauptspeichertechnologie besitzt eine Speicherzugriffs-Bandbreite, die eine mehr als doppelt so hohe Speicherzugriffsraten der durchschnittlichen Zugriffsraten grosser Systeme erlaubt.

Ein Konflikt über den Zugriff zum Hauptspeicher des Mehrprozessorsystems liegt nur vor, wenn zwei oder mehr Prozessoren gleichzeitig Zugriff zum Speicher fordern. Durch die Konfliktlösung gemäss vorliegender Erfindung wird die Leistung des gesamten Mehrprozessorsystems verbessert. Der Konflikt wird dadurch gelöst, dass einem adressierenden Prozessor jeder Hauptspeicherzyklus so lange zur Verfügung gestellt wird, als dieser Prozessor Anforderungen präsentiert. Kein anderer darf diesen Prozessor stören. Ein anderer Prozessor erhält Zugriff für seine ausstehenden Anforderungen zum Speicher, sobald ein adressierender Prozessor nicht mehr kontinuierlich in jedem folgenden Speicherzyklus Anforderungen abgibt. Die Anforderungspuffer eines adressierenden Prozessors sind normalerweise am Ende seines Zugriffsbündels, wenn er

die Priorität verliert, leer. Der Prozessor und seine Kanaleinheiten können trotzdem ohne Verlust des Wirkungsgrades weiterarbeiten, da diese Operationen die Anforderungspuffer füllen und das nächste Anforderungsbündel des Prozessors vorbereiten. Herkömmliche Prioritätssteuerungen für Mehrprozessorsysteme konnten die Zugriffspriorität von einem adressierenden Prozessor abschalten, während er noch gepufferte Anforderungen verfügbar hatte und für nachfolgende Speicherzyklen lokale Speicherbereiche (LSU) benötigte. So konnte der adressierende Prozessor unterbrochen werden, bevor seine Anforderungspuffer wirklich geleert waren. Das führte oft dazu, dass die Anforderungspuffer bald voll waren und der Prozessor erst weiterarbeiten konnte, wenn sie durch eine später erteilte Zugriffspriorität geleert wurden. Diese bisherige Arbeitsbehinderung wird durch die vorliegende Erfindung vermieden.

In Fig. 2A ist gezeigt, dass Zugriffsanforderungen des Prozessors und der E/A-Geräte auf herkömmliche Art an eine lokale Prioritätsschaltung 64 gegeben werden. Wie in einer herkömmlichen Zentraleinheit empfängt die lokale Prioritätsschaltung 64 gepufferte Zugriffsanforderungen des E/A-Kanales durch ein UND-Glied 63 und Speicherzugriffsanforderungen des Prozessors durch ein UND-Glied 62, gepuffert durch eines der Speicheradressregister (STAR) 57 oder 58. Sie empfängt auch Prozessor-Abrufanforderungen über das UND-Glied 61, gepuffert durch ein Abruf-Adressregister (FAR) 56. Die Prozessor-Anforderungen werden durch die Adressschalter 52 an das Abruf-Adressregister 56, das Speicher-Adressregister 57 oder das Speicher-Adressregister 58 gegeben. Jede Anforderung durch das Abruf-Adressregister 56 gilt für die Übertragung eines Blocks bestehend aus vier Doppelwörtern mit je acht Bytes in den Cache-Speicher 53. Eine Anforderung vom Speicheradressregister 57 oder 58 gilt der Speicherung eines Doppelwortes in den Hauptspeicher. Jede Kanalanforderung wird in einem Kanalpuffer 51 gepuffert, der bis zu 24 ausstehende Kanalzugriffsanforderungen puffern kann. Es wird zu einem Zeitpunkt jeweils eine der 24 Kanalanforderungen durch ein UND-Glied 63 weitergeleitet. Jedes Signal der UND-Glieder 61, 62 und 63 wird nur an die lokale Prioritätsschaltung 64 weitergeleitet, wenn der angeforderte lokale Speicherbereich (LSU) für die Anforderung zur Verfügung steht. Dieser lokale Speicherbereich kann für jede der gegenwärtig durch die UND-Glieder für die lokale Prioritätsschaltung 64 eingegebenen drei Anforderungen eine andere sein. Die lokale Prioritätsschaltung 64 wählt dann eine dieser drei Anforderungen als ihre Ausgabe und gibt die erste Priorität den E/A-Anforderungen, die zweite Priorität den Anforderungen aus dem Speicheradressregister (STAR) und die dritte Priorität den Anforderungen aus dem Abruf-Adressregister (FAR). Auf diese Weise vergibt die lokale Prioritätsschaltung 64 die lokale Priorität und kann genehmigte Anforderungen mit derselben hohen Geschwindigkeit wie die Hauptspeicher-Zyklusraten ausgeben, solange eines der UND-Glieder 61, 62 oder 63 kontinuierlich Zugriffsanforderungen an die lokale Prioritätsschaltung 64 gibt und so ein Anforderungsbündel am Ausgang der lokalen Prioritätsschaltung 64 erzeugt wird. Diese Anforderungen werden sequentiell über ein ODER-Glied 72 auf die Ausgangsleitung 75A für genehmigte Anforderungen gegeben. Ein Merkmal eines Prozessors besteht also darin, dass er auf längere Zeit nicht kontinuierlich Anforderungen an die lokale Prioritätsschaltung 64 mit der Speicherzugriffsrate geben kann, weil nach kurzer Zeit entweder die erforderlichen lokalen Speicherbereiche (LSU) nicht mehr zur Verfügung stehen oder die Anforderungspuffer leer sind. Die relativen Raten zwischen dem gemeinsam benutzten Hauptspeicher und einem Prozessor in Fig. 1, betrachtet aus dem Blickwinkel der lokalen Prioritätsschaltung 64, liegen so, dass auf lange Sicht die durchschnittliche Prozessorrates unter 50% der Hauptspeicherzugriffsrate liegt. In der Praxis liegt üblicherweise

diese Prozessorrates nicht über 30% der Speicherzugriffsrate. Solange die Summe der Durchschnittsraten aller Prozessoren 100% der Hauptspeicher-Zugriffsrate nicht übersteigt, kann der Hauptspeicher alle Anforderungen für alle Prozessoren ohne nennenswerte Überlastung annehmen. Für Prozessoren mit gleicher Rate in einem Mehrprozessorsystem liegt das optimale Ratenverhältnis pro Prozessor bei der dem Prozessor gewährten durchschnittlich höchsten Zugriffsrate, dividiert durch die höchste Speicherzugriffsrate. Dieses Verhältnis sollte ungefähr 1 durch N sein, wobei N die Anzahl von Prozessoren im Mehrprozessorsystem ist.

Die vorliegende Erfindung erreicht also einen maximalen Wirkungsgrad für ein Mehrprozessorsystem dadurch, dass sie jedem Prozessor den Speicherzugriff zur Befriedigung aller ausstehenden, in seinen Registern 56, 57, 58 und dem Kanalpuffer 51 gespeicherten Anforderungen gestattet, solange ein lokaler Speicherbereich verfügbar und in der Lage ist, die kontinuierliche Ausgabe der Anforderungen von der lokalen Prioritätsschaltung 64 auf deren Ausgangsleitung 75 zu stützen. Es kann vorkommen, dass eine gegebene Anforderung nicht in die lokale Prioritätsschaltung 64 eingegeben wird, weil der angeforderte lokale Speicherbereich im Hauptspeicher mit einer anderen Anforderung belegt ist und eine andere Anforderung an die lokale Prioritätsschaltung 64 erhält dann zu dieser Zeit Zugriff. Wenn keine Anforderung an die lokale Prioritätsschaltung 64 während eines Prozessorzyklus eingegeben wird, wird die zusammenhängende Folge genehmigter Anforderungen auf der Ausgabeleitung 75A unterbrochen.

Nach Art der vorliegenden Erfindung erfüllt der Hauptspeicher nur Anforderungen von einem Prozessor und nimmt keine Anforderungen von einem anderen Prozessor an, solange eine zusammenhängende Folge genehmigter Anforderungen auf der Ausgangsleitung 75A vorliegt. In einem Mehrprozessorsystem kann aber diese Folge gewährter Anforderungen von einem Prozessor nicht lange anhalten. Das Ende der einanderfolgenden Zyklen von Anforderungen von der lokalen Prioritätsschaltung 64 definiert das Ende eines Bündels von Zugriffsanforderungen. Dann kann ein anderer Prozessor den Hauptspeicher solange adressieren, bis sein Anforderungsbündel beendet ist.

Die vorliegende Erfindung erzielt ihren Wirkungsgrad durch eine Mehrprozessor-Prioritätsschaltung, die den Fluss von Speicherzugriffen durch einen Prozessor solange nicht unterbricht, als dieser den Fluss bei jedem Speicherzyklus aufrecht erhalten kann. Sobald er das nicht mehr kann, verliert er die Priorität an einen anderen Prozessor, der einen zusammenhängenden Zugriffsfluss aufrecht erhalten kann. So geht durch Prioritätsumschaltung für den Zugriff zu dem durch mehrere Prozessoren im Mehrprozessorsystem gemeinsam benutzten Hauptspeicher kein Speicherzyklus verloren.

Fig. 2B zeigt im einzelnen die Prioritätszeigerschaltungen 17A und 17B und ihre gegenseitige Verbindung. In den Prioritätszeigerschaltungen 17A und 17B wird entschieden, welcher der gleichzeitig auf den Ausgangsleitungen 75A und 75B Anforderungen abgebenden Prozessoren für den nächsten Speicherzugriff zum gemeinsam benutzten Speicher bestehend aus den Speichereinheiten 10A und 10B ausgewählt wird.

In der Prioritätszeigerschaltung 17A werden genehmigte Anforderungen vom Prozessor 12A auf der Leitung 75A empfangen und durch den Schalter 76A festgehalten, der durch jede genehmigte Anforderung vom Prozessor 12A für einen Prozessorzyklus neu eingestellt wird. Ein entsprechender Schalter 76B in der Prioritätszeigerschaltung 17B wird durch jede Anforderung von Prozessor 12B auf der Leitung 75B für einen Zyklus eingestellt.

In der Prioritätszeigerschaltung 17A koordinieren drei UND-Glieder 81A, 82A und 83A für den Prozessor 12A die genehmigten Anforderungen, die gegenwärtig von den Prozes-

soren 12A und 12B abgegeben werden. Die Prioritätszeigerschaltung 17B hat entsprechend identische UND-Glieder 81B, 82B und 83B, die die Priorität für den Prozessor 12B koordinieren.

In der Prioritätszeigerschaltung 17A entsprechen die A- und B-Anforderungsleitungen dem Ausgang des Schalters 76A bzw. Verzweigungen der Leitung 75B. Die Inverter 77A und 78A speisen Komplementleitungen für die Anforderungen A bzw. B. Die Eingänge zum UND-Glied 81A empfangen die A-Anforderung und das B-Anforderungs-Komplement, während die Eingänge zum UND-Glied 82A die A-Anforderung und die B-Anforderung empfangen. Die Eingänge zum UND-Glied 83A empfangen das A-Anforderungs-Komplement und das B-Anforderungs-Komplement. Ein ODER-Glied 86A empfängt die Ausgänge aller UND-Glieder 81A, 82A und 83A sowie einen Ausgang von einem ODER-Glied 84A, das aktiviert wird, wenn eine seiner drei Eingangsleitungen ein Signal überträgt. Diese Eingangsleitungen sind die Leitungen 84A-1, 84A-2 und 84A-3. Die Leitung 84A-1 wird aktiviert, wenn die beiden Speichereinheiten 10A und 10B gemäss den Angaben eines Konfigurationsregister 80A nicht als gemeinsam benutzter Hauptspeicher eines Mehrprozessorsystems verbunden sind, so dass jeder Prozessor nur mit seiner eigenen Speichereinheit arbeitet und die Speichereinheit des anderen Prozessors nicht adressieren kann.

Ein Signal auf der Leitung 84A-2 zeigt an, dass der Strom zum Prozessor B abgeschaltet ist und deshalb Prozessor A kontinuierlich Priorität für den gemeinsamen Hauptspeicher erhalten sollte. Ein Signal auf der Leitung 84A-3 zeigt an, dass der Prozessor kein Teil eines Mehrprozessorsystems ist, daher nur Zugriff zu seiner eigenen Hauptspeichereinheit 10A hat und die andere Einheit 10B nicht adressieren kann.

Bei dem in Fig. 2A gezeigten Taktgeber 50 handelt es sich um eine herkömmliche Einrichtung, welche Trigger- und Schaltertaktimpulse für Signale liefert, die gemäss Fig. 5 durch die Prozessorschaltungen geleitet werden. Der Taktgeber 50 gibt Trigger-Taktsignale C(S) und C(R) ab, die die Einschaltung bzw. Rückstellung der Trigger (T) im Prozessor zeitlich steuern (vgl. Fig. 2B). In ähnlicher Weise steuern die Schalter-Taktsignale C(S) und C(R) zeitlich Einschaltung und Rückstellung der selbsthaltenden Schalter (L) im Prozessor. Die Trigger (T) und Selbsthalteschalter (L) können identische bistabile Schaltungen benutzen, werden aber an unterschiedliche Taktsignalleitungen angeschlossen.

Ein Trigger (T) 87A in Fig. 2B wird während der Trigger-taktzeit C(S) durch ein Signal auf der Ausgangsleitung des ODER-Gliedes 86A eingeschaltet und ein Schalter (L) 88A wird einen halben Zyklus später während der Schaltertaktzeit C(S) durch ein Ausgangssignal vom Trigger 87A eingeschaltet. Der Trigger 87A und der Schalter 88A werden weniger als einen Zyklus später nach ihrer jeweiligen Einschaltung durch entsprechende Rückstelltaktimpulse C(R) zurückgestellt. Jeder Trigger (T) und jeder Schalter (L) wird daher für weniger als einen Taktzyklus aktiviert, und zwar um einen halben Taktzyklus gestaffelt. Während seiner aktiven Zeit gibt der Schalter 88A auf der Leitung 20A ein Signal als Prioritätszeiger A an die Übertragungseinheit (MCU) 11A ab.

Die Signale von Trigger 87A und Schalter 88A veranlassen den nächsten Hauptspeicherzugriff durch den Prozessor 12A. Das Signal auf der Leitung 20A aktiviert die Übertragungseinheit 11A zur Verbindung der Speicheradress-Sammelleitungen und Datensammelleitungen mit dem Prozessor 12A zwecks Übertragung vom Prozessor an den Hauptspeicher. Wenn der Trigger 87A eingeschaltet ist (einen halben Zyklus vor dem Schalter 88A) aktiviert er eine Prioritätsleitung 87A-1. Diese ist an die Prozessor-Speichersteuerung (PSCF) 16A zurückgeführt und leitet für Übertragungen vom Hauptspeicher an den Pro-

zessor 12A ein Signal über die Prozessorwahlleitung 15A.

Wenn der Trigger 87A eingeschaltet ist, liefert er auch ein Ausgangssignal an die Verbindungsleitung 92 der Prioritätszeigerschaltungen über einen Inverter 91 an UND-Glieder 82B und 83B in der Prioritäts-Zeigerschaltung 17B, um diese daran zu hindern, dem anderen Prozessor 12B die Priorität einzuräumen.

Der Ausgang des Schalters 88A liefert ein Signal, das an die UND-Glieder 82A und 83A zurückgeführt wird, so dass diese jede nächstfolgende lokale Speicheranforderung von der Prozessor-Speichersteuerung 16A während des laufenden Schalterzyklus durchlassen. Mit dieser Rückkopplungsschaltung wird jede nächste lokale Anforderung, die von der Prozessor-Speichersteuerung 16A während eines Triggerzyklus abgegeben wird, während eines Schalterzyklus in den Trigger 87A gesetzt, damit die Prioritäts-Zeigerschaltung 17A bei nachfolgenden Zyklen kontinuierlich A-Prioritäts-Zeigersignale auf die Leitung 20A gibt. Das Triggersignal auf der Leitung 92 sperrt die Eingänge zum Trigger 87B und unterdrückt so Zeigersignale auf der Leitung 20B.

Sobald jedoch ein Schalterzyklus für den Prozessor 12A abläuft, in dem von der Prozessor-Speichersteuerung 16A keine lokale Anforderung kommt, wird der Schalter 76A nicht eingestellt und das Ausgangssignal «keine A-Anforderung» des Inverters 78B wird in der anderen Prioritäts-Zeigerschaltung 17B aktiviert. Ausserdem wird durch das ODER-Glied 86A kein Signal zum Einschalten des Triggers 87A gegeben und der Ausgang 92 des Inverters 91 stellt dann die Schaltungen 82B und 83B in der Prioritäts-Zeigerschaltung 17B ein.

Wenn das Signal «keine Anforderung A» aktiv ist, wird das UND-Glied 81B eingestellt, wodurch es anhängige lokale Anforderungen des anderen Prozessors 12B weiterleitet, für die der Schalter 76B gesetzt wurde. Das Signal einer anhängigen Anforderung wird durch das UND-Glied 81B geleitet, um den Trigger 87B einzuschalten, und dann wird der Schalter 88B einen halben Zyklus später eingestellt, um ein B-Prioritäts-Zeigersignal auf der Leitung 20B zu erzeugen. Wenn die Prioritäts-Zeigerschaltung 17B einmal ein Zeigersignal liefert, werden die Tore 81A, 82A und 83A in der Zeigerschaltung 17A gesperrt, um die Zeigerschaltung 17A daran zu hindern, die jetzt der Zeigerschaltung 17B eingeräumte Priorität zu stören. Die Prioritäts-Zeigerschaltung 17A wird dadurch gesperrt, dass im Tor 81A das Signal «keine Anforderung B» und in den Toren 82A und 83A das A-Prioritäts-Zeigersignal von der Leitung 20A abgeschaltet werden. Die Schaltung 17A kann erst wieder Priorität bekommen, wenn ein Zyklus auftritt, in dem der Schalter 76B nicht mehr eingestellt ist und dadurch anzeigt, dass vom Prozessor 12B keine Anforderung vorliegt. Eine anhängige lokale Anforderung, durch die der Schalter 76A während des gesperrten Zustandes der Prioritäts-Zeigerschaltung 17A eingestellt wurde, erlangt für den Prozessor 12A keine Priorität. Der Prozessor 12B hält seine Priorität so lange, als der Schalter 76B bei jedem folgenden Zyklus durch Anforderungen des Prozessors 12B eingestellt wird.

Zu einem gegebenen Zeitpunkt kann also nur eine der beiden Prioritäts-Zeigerschaltungen 17A oder 17B ein Ausgangssignal auf der Leitung 20A bzw. 20B abgeben. Zwischen den Ausgängen der Prioritäts-Zeigerschaltungen 17A und 17B besteht also eine Flipflop-Wirkung.

Das Prioritäts-Zeigersignal auf der Leitung 20A bzw. 20B steuert im Mehrprozessorsystem die Wahl desjenigen Prozessors, der als nächster Zugriff zum gemeinsam benutzten Hauptspeicher erhält. Die Signale auf den Leitungen 20A und 20B schalten dazu die Adress- und Datenleitungen in den Übertragungseinheiten 11A und 11B. Die Fig. 3 und 4 zeigen im einzelnen, wie diese Übertragungseinheiten die Adress- und Datensammelleitungen unter Steuerung von Signalen auf den Prioritäts-Zeigerleitungen 20A bzw. 20B umschalten. Fig. 3

zeigt die Durchschaltung der Datenleitung von einem Prozessor zum gemeinsam benutzten Hauptspeicher und Fig. 4 die Durchschaltung der Datenübertragungen vom gemeinsam benutzten Hauptspeicher zu den Prozessoren A oder B.

In den Figuren 3 und 4 hat jede Übertragungseinheit Schaltungen für jede Bitposition in den Adress- und Datensammelungen für beide Richtungen zwischen Hauptspeicher und jedem Prozessor. Die Speicher-Vermittlungsschaltung 11A-1 in Fig. 3 besteht aus den UND-Gliedern 21 und 22 und einem selbsthaltenden Schalter 24, der durch ein Ausgangssignal von einem der UND-Glieder 21 oder 22 eingestellt wird. Ein Eingang zum UND-Glied 21 empfängt eine Bitposition der Adress- oder Datensammelung vom Prozessor 12A, während das UND-Glied 22 eine entsprechende Bitposition der Adress- oder Datensammelung vom Prozessor 12B empfängt. Das UND-Glied 21 ist mit einem weiteren Eingang an die A-Prioritäts-Zeigerleitung 20A angeschlossen und das UND-Glied 22 mit einem weiteren Eingang an das komplementäre Signal der Leitung 20A vom Inverter 23. Der Ausgang des Schalters 24 auf der Sammelleitung 18A-1 wird an die Hauptspeichereinheit 10A gegeben. Sobald also einer der Prozessoren 12A oder 12B eine Zugriffsanforderung zum Speichern oder Abrufen von Daten zur Speichereinheit 10A bewilligt bekommt, werden die Adresse für diese Anforderung und die Daten (wenn es eine Speicheranforderung ist) durch das UND-Glied 21 bzw. 22 an eine Adresse in die gemeinsam benutzte Speichereinheit 10A geleitet.

Die Speicher-Vermittlungsschaltung 11B-1 besteht aus den UND-Gliedern 26 und 27, dem Inverter 28 und dem selbsthaltenden Schalter 29. Die beiden UND-Glieder 26 und 27 empfangen entsprechende Bitpositionen der Adress- und Datensammelungen von den Prozessoren 12A bzw. 12B. Die B-Prioritäts-Zeigerleitung 20B liefert auch ein Signal an das UND-Glied 27 und ein komplementäres Signal über den Inverter 28 an das UND-Glied 26. Die Schaltung 11B-1 arbeitet genauso wie die Schaltung 11A-1 und leitet eine Anforderung von einem Prozessor an eine Adresse der gemeinsam benutzten Speichereinheit 10B.

Die Speicher-Vermittlungsschaltungen 11A-2 und 11B-2 in Fig. 4 besorgen die Zugriffsübertragung für jede Bitposition in der Datensammelung vom Hauptspeicher zu einem anfordernden Prozessor. Antworten vom Hauptspeicher bringen eine Verzögerung von mehreren Prozessorzyklen zwischen der Anforderungsempfangszeit durch den Hauptspeicher und der Antwortzeit mit sich. Ein Zeitverzögerungssignal für die Speicherantwort wird von einem Zeitgeber 30 oder 35 für die entsprechenden Prozessoren erzeugt.

Die Prioritäts-Zeigerleitungen 20A und 20B steuern indirekt die Übertragung durch die Schaltungen 11A-2 und 11B-2 über die in Fig. 2A gezeigten Prozessor-Speichersteuerungen

16A und 16B. In Fig. 2A wird z.B. ein Signal auf der Zeigerleitung 20A als Signal auf der A-Prioritätsleitung 87A-1 vom Trigger 87A einen halben Zyklus früher an die UND-Glieder 66, 67 oder 68 abgegeben, von denen eines durch eine ausgegebene Anforderung von der lokalen Prioritätsschaltung 64 eingeschaltet wird. Das aktivierte UND-Glied 66, 67 oder 68 stellt dann einen der Schalter 71, 73 oder 74 zur Erzeugung des Wahlsignales auf der Leitung 15A durch das ODER-Glied 70 ein.

In Fig. 4 ist gezeigt, wie die Prozessor-Wahlleitung 15A den Eingang zum Zeitgeber 30 speist, während ein entsprechendes Wahlsignal auf der Leitung 15B von der Prozessor-Speichersteuerung 16B den Eingang zum Zeitgeber 35 speist. Eine auf der Prioritätszeigerleitung 20A signalisierte Anforderung wird somit in den Zeitgeber 30 und eine auf der Prioritätszeigerleitung 20B signalisierte Anforderung in den Zeitgeber 35 der anderen Übertragungseinheit eingegeben. Das Wahlsignal wird durch die Zeitgeber 30 bzw. 35 verzögert, bis die Datenantwort vom Speicher auf den Sammelleitungen 18A-2 bzw. 18B-2 ausgegeben wird. Das Ausgangssignal eines Zeitgebers wird an eines der Tore 31 oder 32 in der Schaltung 11A-2 oder eines der Tore 36 oder 37 in der Schaltung 11B-2 eingegeben, abhängig davon, welcher der beiden Prozessoren die Anforderung stellt und welche Speichereinheit den Zugriff gibt. Die Schalter 34 bzw. 39 leiten die Zugriffsantwort an den anfordernden Prozessor zurück.

Die Prioritäts-Zeigerschaltungen 17A und 17B arbeiten noch, wenn der Anforderungseingang vom anderen Prozessor beim UND-Glied 82 nicht mehr vorliegt, d.h., das Signal «Anforderung B» kann vom Tor 82A und das Signal «Anforderung A» vom Tor 82B abgeschaltet sein. Entweder Tor 82A oder das Tor 82B halten die Einschaltbedingung für ihren Trigger 87 und den selbsthaltenden Schalter 88 solange aufrecht, als lokale Anforderungen von ihrem entsprechenden Prozessor in aufeinanderfolgenden Taktzyklen empfangen werden.

Fig. 5 zeigt Beispiele der lokalen Zeiteinteilung für einen Prozessor, wenn er seine Priorität erhält. Die Trigger-Taktzeit und die Schalter-Taktzeit weisen einen halben Zyklus Abstand voneinander auf.

Die oben beschriebene Erfindung lässt sich auf ein Mehrprozessorsystem mit N Prozessoren anwenden, wenn das oben für zwei Prozessoren beschriebene Ausführungsbeispiel erweitert wird. Das geschieht durch Erhöhung der Anzahl von UND-Gliedern in jedem Satz 81 bis 83 der Prioritäts-Zeigerschaltung 17 zur Berücksichtigung aller Kombinationen der Original- und Komplementsignale aller Anforderungsschalter 76. In gleicher Weise braucht jede Speicher-Vermittlungsschaltung, wie sie in den Fig. 3 und 4 gezeigt ist, ein zusätzliches UND-Glied für jeden im Mehrprozessorsystem vorgesehenen zusätzlichen Prozessor.

FIG. 1

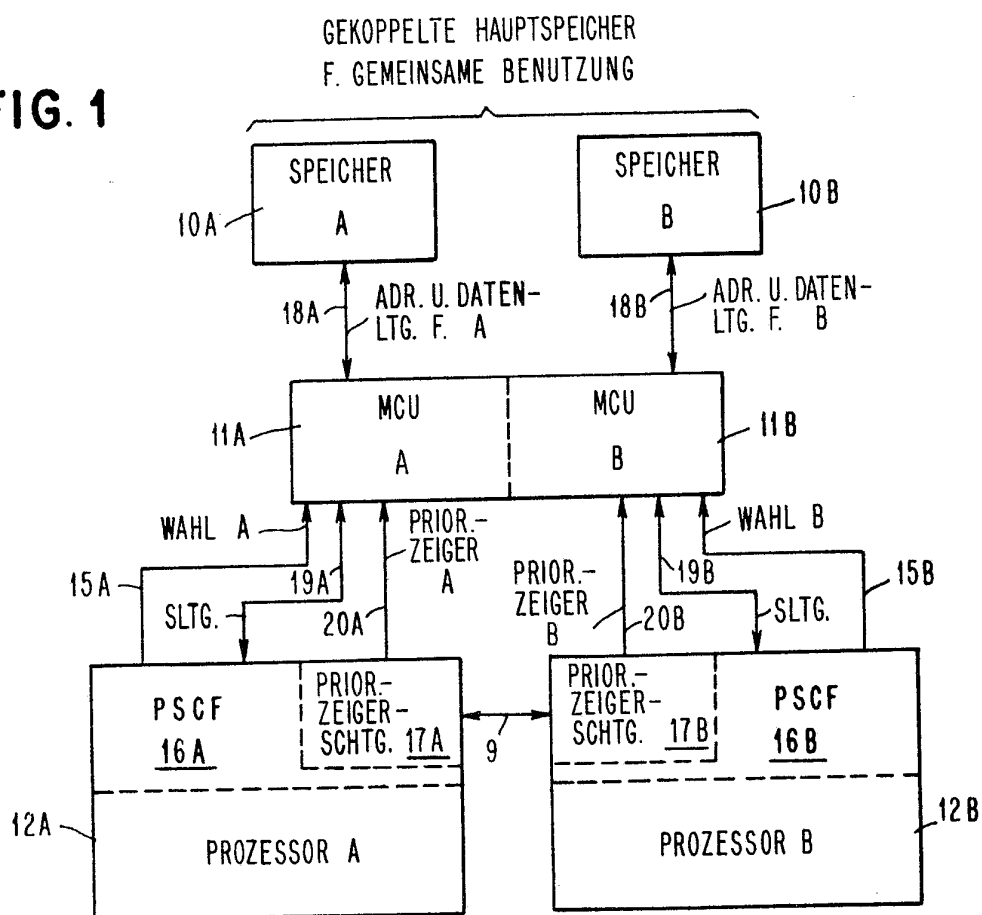


FIG. 5

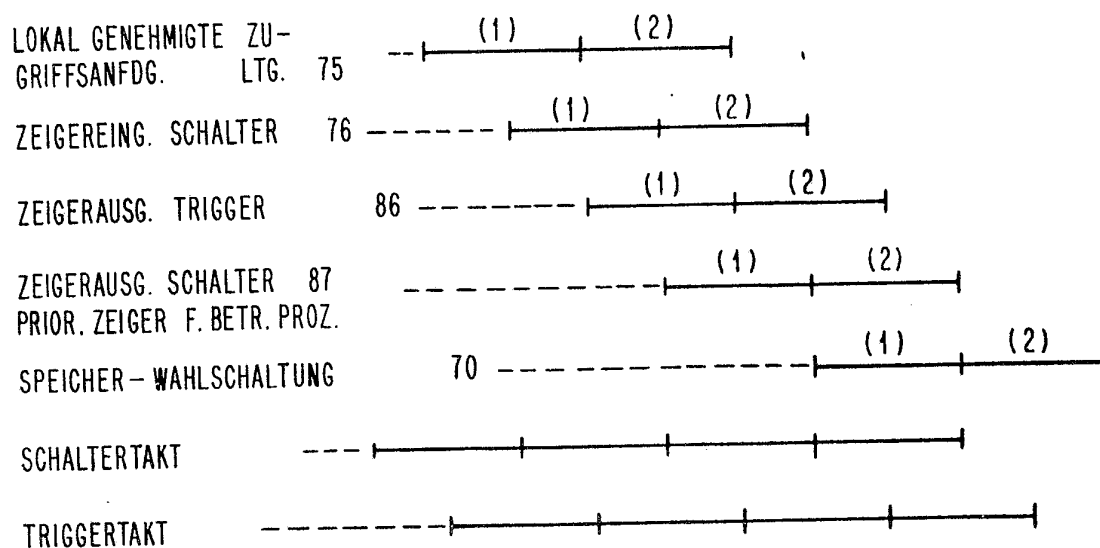


FIG. 2B

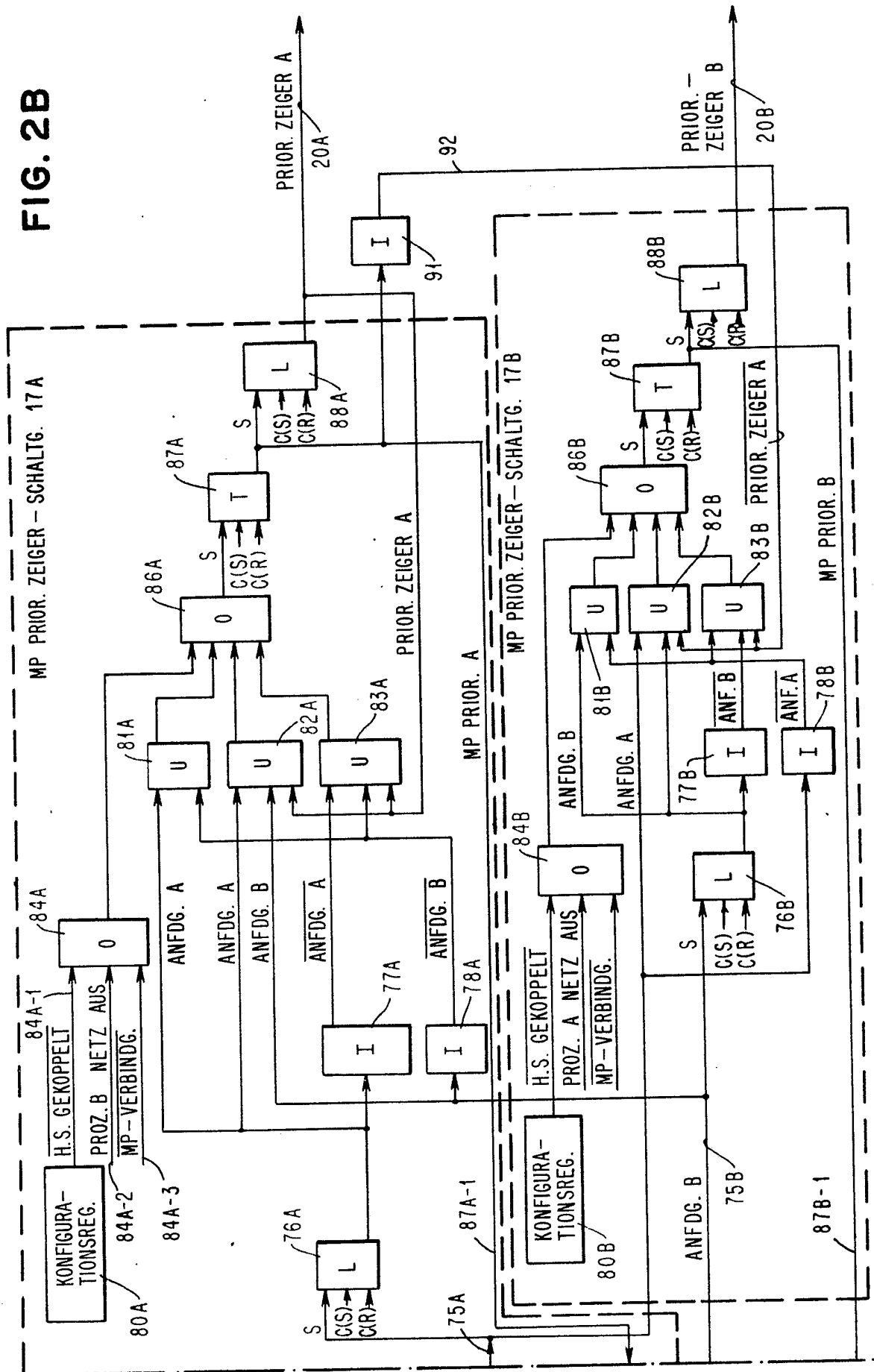


FIG. 3

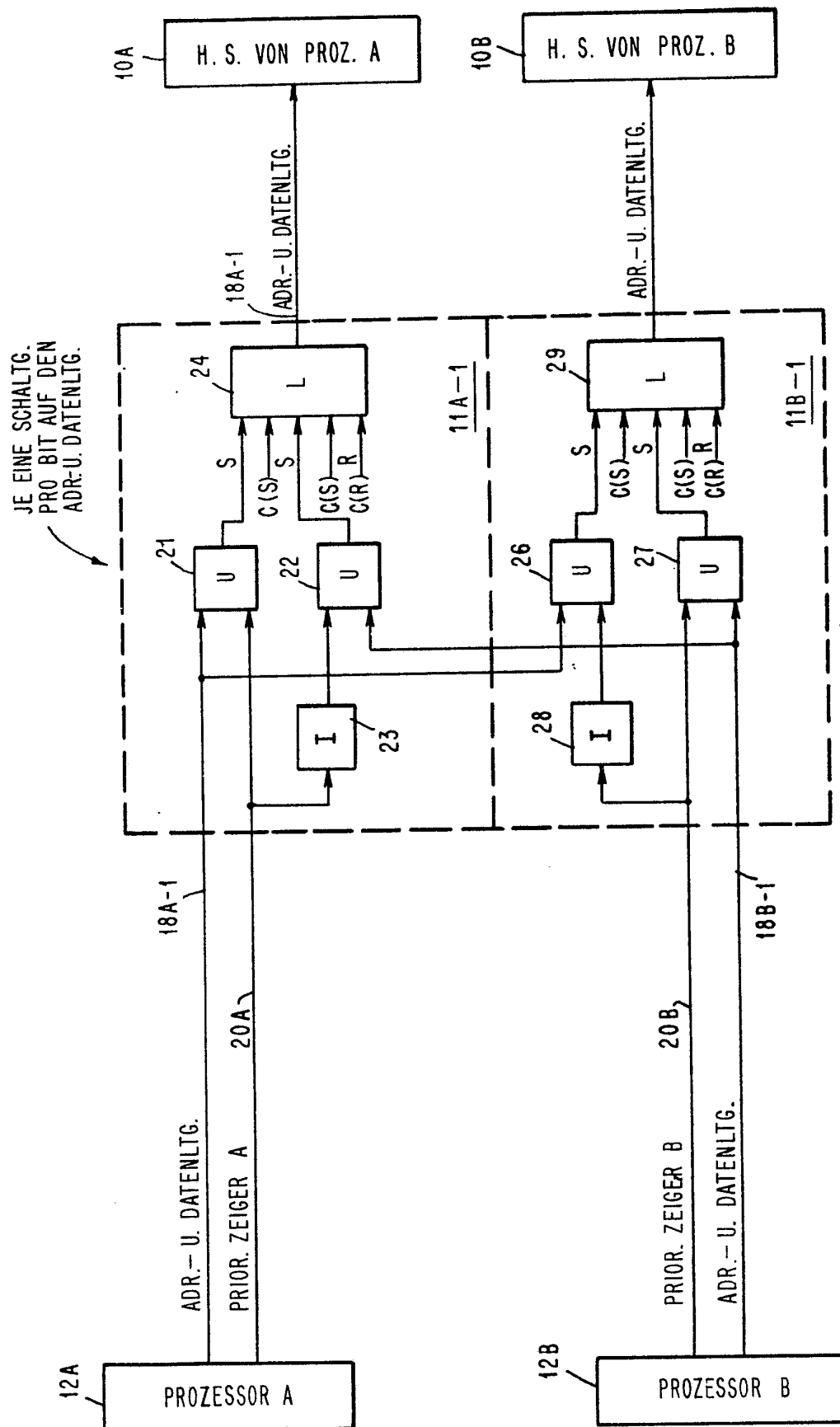


FIG. 4

