(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号

特開平11－275140

(43)公開日　平成11年(1999)10月８日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｈ０４Ｌ　12/56		Ｈ０４Ｌ　11/20	１０２Ｚ

審査請求　有　　請求項の数１　ＯＬ　（全　６　頁）

(21)出願番号　特願平10－71814

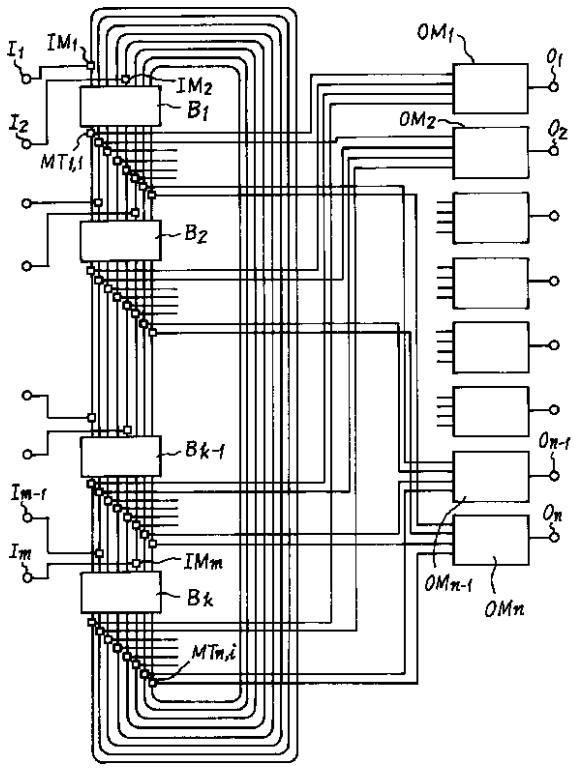
(22)出願日　平成10年(1998)３月20日

特許法第30条第１項適用申請有り　1997年９月24日　社団法人情報処理学会発行の「第55回（平成９年後期）全国大会講演論文集（３）」に発表

(71)出願人　391012327
東京大学長
東京都文京区本郷７丁目３番１号
(72)発明者　斉藤　忠夫
神奈川県横浜市神奈川区松が丘55－４
(72)発明者　相田　仁
神奈川県川崎市宮前区宮崎２－12－１　宮崎台プラザビル203号
(72)発明者　青木　輝勝
埼玉県上福岡市北野１－７－２
(74)代理人　弁理士　杉村　暁秀　（外８名）

(54)【発明の名称】　大容量可変長パケット交換に適したパケットスイッチ方式

(57)【要約】
【課題】　パケットのスイッチング遅延および遅延揺らぎが減少するように適切に構成した大容量可変長パケット交換に適したパケットスイッチ方式を提供する。
【解決手段】　複数ｋ個のバニヤン網を少なくともｎ本の並列ラインによってリング型に接続し、入力ポートを入力モジュールを介して各バニヤン網の入力部に分散して接続し、出力ポートを出力モジュールを介して各バニヤン網の対応する出力部に設けたミスルートタグチェック部に接続し、前記バニヤン網を構成する単位スイッチを、パケットが希望する出力部に出せない場合にはミスルートタグをセットし、このパケットを空いている出力部に出すように構成し、前記ミスルートタグチェック部を、入力されるパケットのミスルートタグをチェックし、これがセットされている場合には次段のバニヤン網側へ、セットされていない場合には前記出力モジュール側へパケットのスイッチングを行うように構成した。



【特許請求の範囲】
【請求項１】　複数m個の入力ポートから複数ｎ個の出力ポートへパケットをスイッチングするパケットスイッチ方式において、複数ｋ個のバニヤン網を少なくともｎ本の並列ラインによってリング型に接続し、前記入力ポートを入力モジュールを介して各バニヤン網の入力部に分散して接続し、前記出力ポートを出力モジュールを介して各バニヤン網の対応する出力部に設けたミスルートタグチェック部に接続し、前記バニヤン網を構成する単位スイッチを、パケットが希望する出力部に出せない場合にはミスルートタグをセットし、このパケットを空いている出力部に出すように構成し、前記ミスルートタグチェック部を、入力されるパケットのミスルートタグをチェックし、これがセットされている場合には次段のバニヤン網側へ、セットされていない場合には前記出力モジュール側へパケットのスイッチングを行うように構成し、前記入力モジュールを、前記入力ポートからのパケットおよびミスルートタグチェック部からのパケットを格納し、これらのパケットを選択的に次段のバニヤン網に入力させるように構成したことを特徴とするパケットスイッチ方式。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、複数の入力ポートから複数の出力ポートへパケットをスイッチングするパケットスイッチ方式に関係するものである。
【０００２】
【従来の技術】共通バス方式はＩＰルータとして現在最も主流となっている方式であり、ほとんどの商用ＩＰルータで用いられている。この共通バス方式のパケットスイッチの簡略図を図１に示す。各入力ポートはラウンドロビン方式によって共通バスの使用権を獲得し、この共通バス上にパケットを送りだす（ブロードキャストする）。出力ポートは、このブロードキャストされたパケットのうち自分宛のものだけを受け取る。したがって、ルータ内のパケット損失を０にするためには、共通バスに（各ポートの伝送速度）×（ポート数）以上の動作速度が必要になる。
【０００３】共通バス方式は、可変長パケットの交換が容易であり、ルータを構成する回路素子数を少なくすることが可能であるなどの利点があるものの、入出力ポート数に比例して共通バスの伝送速度を大きくしなければならない。したがって、この共通バス方式で交換容量が１Ｔｂｐｓ（Tera bit per second ）を越えるような大容量スイッチを実現することは、素子の物理的限界から非常に困難である。
【０００４】バニヤン網は図２に示すように単位スイッチを多段接続することにより大容量スイッチを実現する方式であり、各入力ポートからすべての出力ポートに対して１本ずつの経路が存在している。
【０００５】バニヤン網は網内の各単位スイッチが並列的に動作するため、高速大容量化を実現する方法として適している。しかしその反面、図３に示すように、パケットの同一入線によって内部ブロッキングが発生してしまう問題点がある。なお、図２、３では単位スイッチとして２×２単位スイッチを用いた場合を取り上げたが、４×４、８×８、１６×１６などの単位スイッチを用いても同様の接続が可能である。
【０００６】タンデムバニヤン網はバニヤン網における内部ブロッキングを低減するための一手法であり、図４に示すように複数のバニヤン網を一列に配置したスイッチである。タンデムバニヤン網ではパケットをすべて１段目のバニヤン網（図４の一番左のバニヤン網）に入力させるが、もしバニヤン網で内部ブロッキングが発生した場合には、そのパケットのミスルートタグをセットし、空いている出力ポートに出力させる。そして１段目のバニヤン網を通過後に続けて２段目のバニヤン網に入力させる。一方、１段目のバニヤン網で無事に希望する出力ポートに出られたパケットは、２段目以降のバニヤン網を通らずにそのまま出力バッファに入れられる。
【０００７】以上のように、内部ブロッキングが発生したパケットは希望する出力ポートに出られるまで次段のバニヤン網に入力することができるため、バニヤン網で発生する大部分の内部ブロッキングを回避することが可能である（タンデムバニヤン網において内部ブロッキングが発生するのは１段目から最終段まですべてのバニヤン網でミスルートタグがセットされた場合のみである）。
【０００８】
【発明が解決しようとする課題】タンデムバニヤン網方式は、バニヤン網における内部ブロッキングを低減するための手段として有効である。しかし、タンデムバニヤン網方式では、はじめすべてのセルを１段目のバニヤン網に入力させ、１段目で内部ブロッキングが発生したパケットは引き続き２段目以降のバニヤン網に入力させてゆく。したがって入力トラヒックが大きい場合には、１段目のバニヤン網での内部ブロッキングの発生率が高いため、２段目以降のバニヤン網に入力するパケットの確率がかなり高くなってしまい、パケットのスイッチング遅延および遅延ゆらぎの増加を引き起こす。さらに、遅延ゆらぎの増加に伴い、パケットの順序狂い率も増加してしまうという問題がある。
【０００９】本発明の目的は、上述した従来のタンデムバニヤン網方式における問題を解決し、内部ブロッキングの発生率を抑えることにより、パケットのスイッチング遅延および遅延揺らぎが減少するように適切に構成した、大容量可変長パケット交換に適したパケットスイッチ方式を提供することにある。
【００１０】
【課題を解決するための手段】上記の目的を達成するた

め、本発明によるパケットスイッチ方式は、複数ｋ個のバニヤン網を少なくともｎ本の並列ラインによってリング型に接続し、前記入力ポートを入力モジュールを介して各バニヤン網の入力部に分散して接続し、前記出力ポートを出力モジュールを介して各バニヤン網の対応する出力部に設けたミスルートタグチェック部に接続し、前記バニヤン網を構成する単位スイッチを、パケットが希望する出力部に出せない場合にはミスルートタグをセットし、このパケットを空いている出力部に出すように構成し、前記ミスルートタグチェック部を、入力されるパケットのミスルートタグをチェックし、これがセットされている場合には次段のバニヤン網側へ、セットされていない場合には前記出力モジュール側へパケットのスイッチングを行うように構成し、前記入力モジュールを、前記入力ポートからのパケットおよびミスルートタグチェック部からのパケットを格納し、これらのパケットを選択的に次段のバニヤン網に入力させるように構成したことを特徴とする。

【００１１】本発明によるパケットスイッチ方式においては、図５に示すように複数のバニヤン網をリング状に接続し、入力ポートを各バニヤン網に分散している。これにより、入力トラヒックが分散され、各バニヤン網への見かけ上の入力トラヒック量を低減することができる。例えばバニヤン網の平面数がＮであるとすると、タンデムバニヤン網方式では総トラヒック量が１段目のバニヤン網に入力するのに対し、本方式では各バニヤン網にその１／Ｎのトラヒックを入力させればよいことになる。この分散効果により、各バニヤン網での入力トラヒック量が見かけ上減少するため、１段目のバニヤン網での内部ブロッキング発生率がタンデムバニヤン網と比較して減少し、パケットのスイッチング遅延および遅延ゆらぎを小さくすることができる

【００１２】

【発明の実施の形態】図６は、本発明のパケットスイッチ方式によるパケットスイッチの一実施形態を示すブロック図である。このパケットスイッチ１は、入力ポートI_h（ｈ＝１，２，．．．，ｍ）に結合した入力モジュールIM_hと、バニヤン網B_i（ｉ＝１，２，．．．，ｋ）４と、ミスルートタグチェック部$MT_{h,i}$と、出力ポートO_j（ｊ＝１，２，．．．，ｎ）に結合した出力モジュールOM_jとを具える。この図に示す実施形態においては、ｍ＝ｎ＝８，ｋ＝４としているが、他の数としてもよいことはもちろんである。

【００１３】バニヤン網B_iは、すでに図２に示したように、単位スイッチを多段接続することにより大容量スイッチを実現する。各単位スイッチは、通常のバニヤン網とは異なり、パケットが希望する出力ポートO_jに通じる出力部に出せない場合にはミスルートタグをセットし、空いている出力部にこのパケットを出す。図２では、２×２単位スイッチを用いた場合のバニヤン網を示しているが、４×４、８×８、１６×１６、または３２×３２単位スイッチを用いても同様の多段接続が可能である。バニヤン網B_iの入力部および出力部の数は、少なくとも出力ポートの数ｎだけなければならない。これらのバニヤン網を少なくともｎ本のラインでリング状に接続する。この図においてパケットは、リング状に接続されたバニヤン網を反時計方向に進む。

【００１４】ミスルートタグチェック部$MT_{h,i}$は、１×２スイッチの機能を持つ。入力パケットのミスルートタグをチェックし、これがセットされている場合には次段のバニヤン網B_i側へ、セットされていない場合には出力モジュールOM_j側へパケットのスイッチングを行う。

【００１５】入力モジュールIM_hは、図８に示すように２つのＦＩＦＯ（Ｆｉｒｓｔ　Ｉｎ　Ｆｉｒｓｔ　Ｏｕｔ）バッファ１１および１２とセレクタ１３とを具える。２つのＦＩＦＯバッファ１１および１２はそれぞれ、入力ポートI_hまたはミスルートタグチェック部$MT_{h,i}$から転送されてきたパケットを格納する。セレクタ１３は、２つのＦＩＦＯバッファ１１および１２の一方からパケットを選択し、次段のバニヤン網B_iに入力させる。本実施形態では、入力ポートI_hからのパケットを低優先、ミスルートタグチェック部$MT_{h,i}$からのパケットを高優先と定義しているため、セレクタ１３は、ミスルートタグチェック部$MT_{h,i}$からのパケットがＦＩＦＯバッファ１１に存在する場合にはそのパケットを優先的に次段のバニヤン網B_iに入力させる。

【００１６】また、入力ポートI_h、ミスルートタグチェック部$MT_{h,i}$の双方から１００％の入力トラヒックが存在する場合には、ＦＩＦＯバッファ１１および１２がオーバフローを起こす可能性がある。したがって、入力モジュールIM_hをはじめとするバニヤン網のリング内部のすべての装置を、入力ポートI_hの２倍の伝送速度で動作させる。

【００１７】出力モジュールOM_jは図９に示すように、タイミング調整バッファ２１、共通バス２２、制御部２３および出力バッファ２４から構成されている。タイミング調整バッファ２１は、各バニヤン網B_iから出力したパケットを格納する。制御部２３は共通バス２２を管理し、各タイミング調整バッファ２１にラウンドロビン方式によって共通バス２２の使用権を与える。タイミング調整バッファ２１に格納されたパケットは、この使用権を獲得した時に出力バッファ２４に移される。

【００１８】パケットを入力ポートI_hから入力し、スイッチングして、希望する出力ポートO_jに出力するまでの流れは次の通りである。

【００１９】まずパケットは入力ポートI_hから入力され、入力モジュールIM_h内のＦＩＦＯバッファ１２に格納される。入力モジュールIM_hでは、入力ポートI_hからのパケットよりもミスルートタグチェック部ＭＴ

$_{h,i}$ からのパケットを優先するため、このパケットはミスルートタグチェック部ＭＴ$_{h,i}$ からのパケットがＦＩＦＯバッファ１１内に存在しない場合にのみセレクタ１３を通ってバニヤン網Ｂ$_i$ に入力される。

【００２０】バニヤン網Ｂ$_i$ に入力したパケットは、単位スイッチ間をスイッチングしてゆくが、もし希望する出力ポートＯ$_j$ に通じる出力部が使用中の場合には、パケットヘッダ内のミスルートタグをセットされ、使用されていない出力部に出される。バニヤン網Ｂ$_i$ を通過したパケットは続いてミスルートタグチェック部ＭＴ$_{h,i}$ に渡される。

【００２１】ミスルートタグチェック部ＭＴ$_{h,i}$ では、パケットにミスルートタグがセットされているかどうかをチェックし、セットされている場合には引き続き次段のバニヤン網Ｂ$_i$ に入力させる。一方、ミスルートタグがセットされていない場合には、出力モジュールＯＭ$_j$ に転送される。

【００２２】出力モジュールＯＭ$_j$ に入ったパケットはタイミング調整バッファ２１に格納され、共通バス２２の使用権を獲得するまで待機する。そして共通バス使用権が回ってきたら、パケットは共通バス２２を通じて出力バッファ２４に入れられ、出力ポートＯ$_j$ を経て本パケットスイッチ１から出力される。

【００２３】従来のタンデムバニヤン網方式のパケットスイッチと本方式のパケットスイッチとをシミュレーションにより評価した結果を示す。ここで、使用するバニヤン網を、図１０に示すような４×４単位スイッチを用いた６４×６４バニヤン網とし、ポート伝送速度を１５５．５２Ｍｂｐｓとし、入力負荷を９０％とし、ＩＰデータグラムを２１～１５００ｏｃｔｅｔの可変長とした。

【００２４】図１１にバニヤン網の平面数－パケット損失率特性を示す。図１１では、バニヤン網の段数の増加に伴ってタンデムバニヤン網方式、本方式ともにパケット損失率が減少してゆく様子が示されているが、特に本方式でははじめからトラヒックの分散が行われているため、タンデムバニヤン網方式と比較してパケット損失率が小さくなっている。また、本方式はタンデムバニヤン網方式と比較してパケット損失率の収束が早いため、少ないバニヤン網の段数で高性能化を図ることができる。これは設計の立場に立った場合、大きな長所となるであろう。

【００２５】一方、図１２では、６段のバニヤン網を用いたタンデムバニヤン網方式と本方式に対して、入力パケットが何段目のバニヤン網で出力できたかを示している。

【００２６】図１２では、本方式ではほとんどのパケットが３段目以内のバニヤン網でスイッチングを完了しているのに対し、タンデムバニヤン網方式では５段目のバニヤン網に入力するパケットが１０％近くもあり、本方式と比較して、遅延および遅延ゆらぎが大きくなっていることがわかる。

【００２７】

【発明の効果】本発明によれば、入力ポートをあらかじめ分散しておくことによって、入力トラヒックの分散を図り、各バニヤン網への見かけ上の入力トラヒック量を低減することができる。これにより、例えばバニヤン網の平面数がＮであるとすると、タンデムバニヤン網方式では総トラヒック量が１段目のバニヤン網に入力するのに対し、本方式では各バニヤン網にその１／Ｎのトラヒックを入力させればよいことになる。この分散効果により、各バニヤン網での入力トラヒック量が見かけ上減少するため、１段目のバニヤン網での内部ブロッキング発生率がタンデムバニヤン網と比較して減少し、パケットのスイッチング遅延および遅延ゆらぎを小さくすることができる。したがって本発明によれば、大容量可変長パケット交換に適したパケットスイッチ方式が提供される。

【図面の簡単な説明】

【図１】従来の共通バス方式のパケットスイッチを示す線図である。

【図２】バニヤン網を示す線図である。

【図３】バニヤン網における内部ブロッキングを説明する線図である。

【図４】タンデムバニヤン網を示す線図である。

【図５】本発明による、バニヤン網をリング状に接続したパケットスイッチを図式的に示す線図である。

【図６】本発明のパケットスイッチ方式によるパケットスイッチの一実施形態の構成を示すブロック図である。

【図７】図６のパケットスイッチにおけるミスルートタグチェック部の構成を示すブロック図である。

【図８】図６のパケットスイッチにおける入力モジュールの構成を示すブロック図である。

【図９】図６のパケットスイッチにおける出力モジュールの構成を示すブロック図である。

【図１０】４×４単位スイッチを用いた６４×６４バニヤン網を示す線図である。

【図１１】バニヤン網の平面数－パケット損失率特性を示すグラフである。

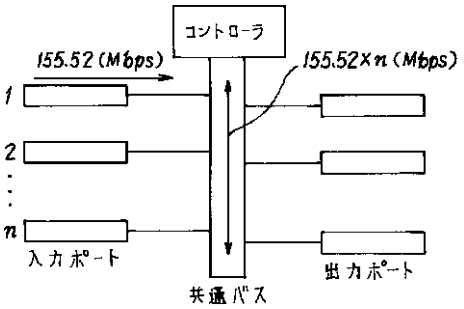
【図１２】入力パケットが何段目のバニヤン網で出力できたかを示すグラフである。

【符号の説明】

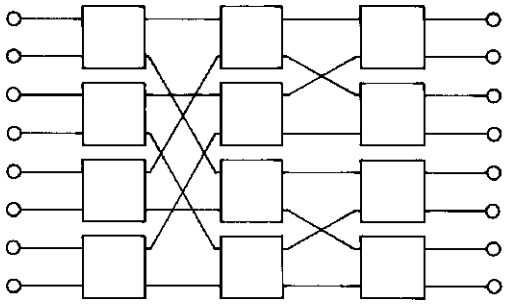
１　パケットスイッチ
１１、１２　ＦＩＦＯバッファ
１３　セレクタ
２１　タイミング調整バッファ
２２　共通バス
２３　制御部
２４　出力バッファ
Ｉ$_h$　入力ポート

IM_h　入力モジュール
B_i　バニヤン網
$MT_{h,i}$　ミスルートタグチェック部
OM_j　出力モジュール
O_j　出力ポート

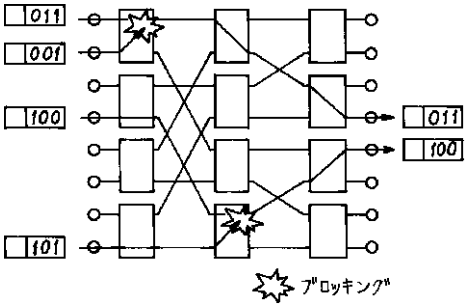
【図１】



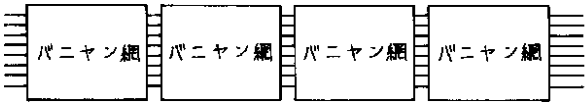
【図２】



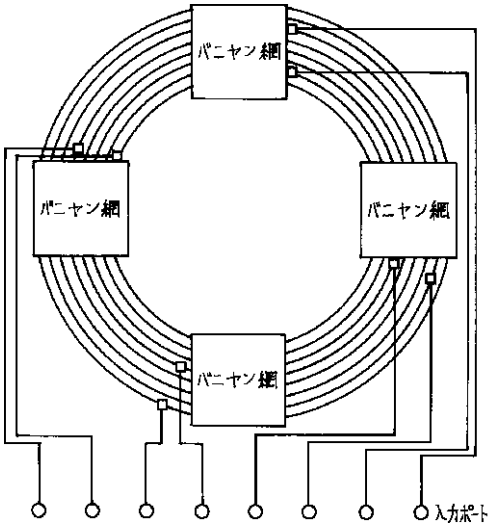
【図３】



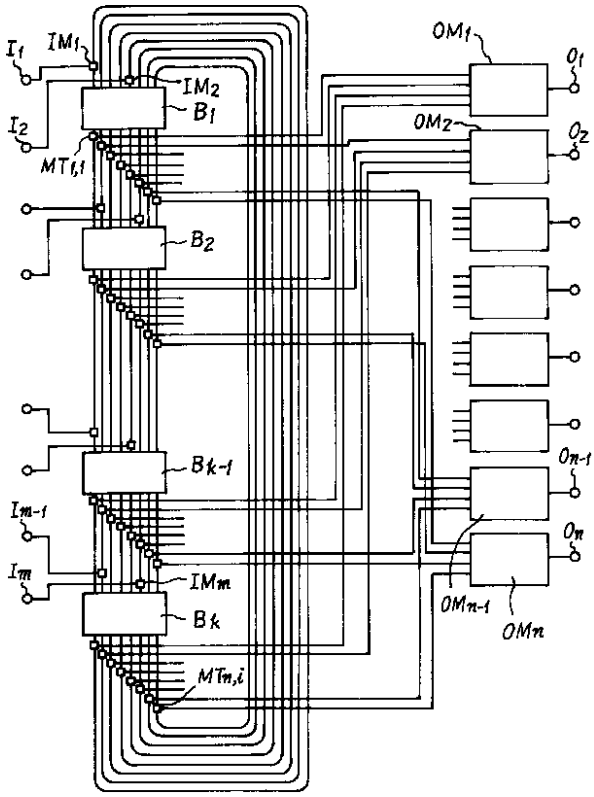
【図４】



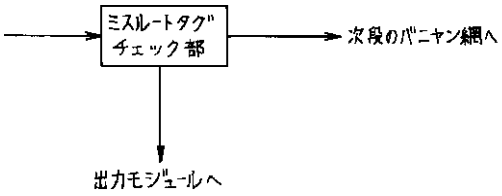
【図５】



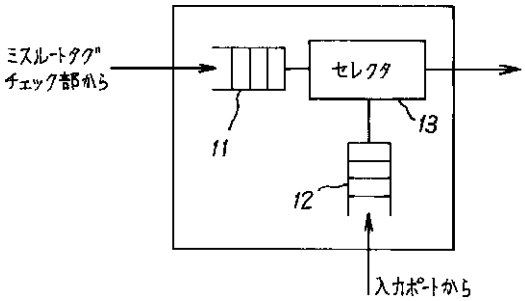
【図６】



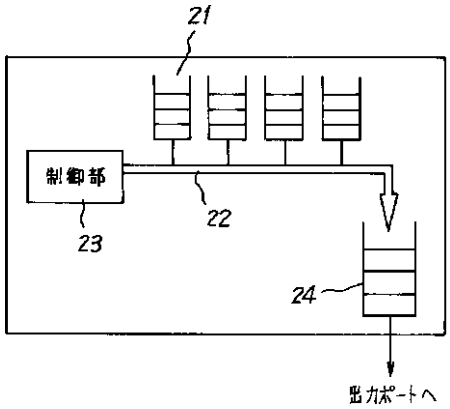
【図７】



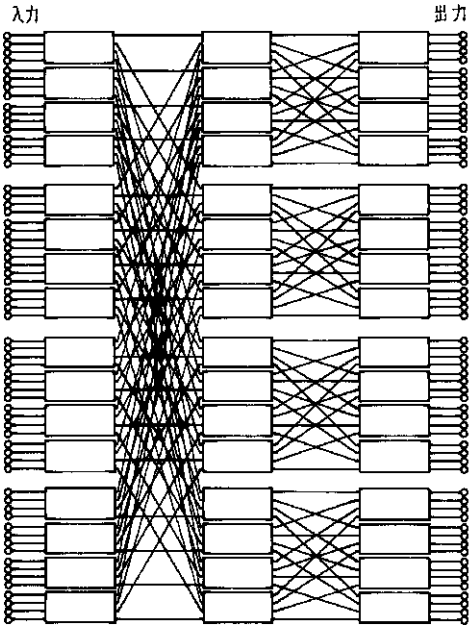
【図８】



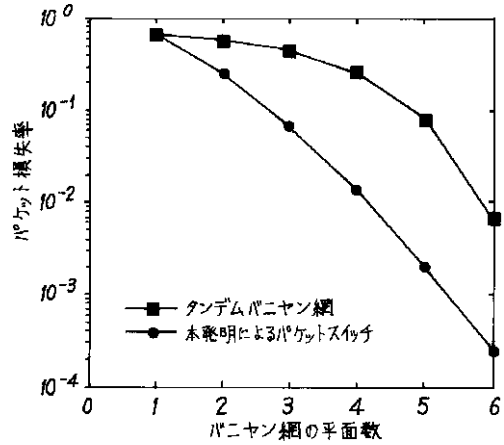
【図９】



【図１０】



【図１１】



【図１２】