

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第3860115号

(P3860115)

(45) 発行日 平成18年12月20日(2006.12.20)

(24) 登録日 平成18年9月29日(2006.9.29)

(51) Int.Cl.

F I

H O 4 L 12/56 (2006.01)

H O 4 L 12/56 1 O O Z

請求項の数 32 (全 22 頁)

(21) 出願番号	特願2002-537075 (P2002-537075)	(73) 特許権者	503146933
(86) (22) 出願日	平成13年10月19日(2001.10.19)		インタラクティブ・ホールディングズ・
(65) 公表番号	特表2004-535690 (P2004-535690A)		エルエルシー
(43) 公表日	平成16年11月25日(2004.11.25)		I N T E R A C T I C H O L D I N G S
(86) 国際出願番号	PCT/US2001/050544		, L L C
(87) 国際公開番号	W02002/034004		アメリカ合衆国ニュージャージー州085
(87) 国際公開日	平成14年4月25日(2002.4.25)		40・プリンストン・ウィリアムストリー
審査請求日	平成16年10月7日(2004.10.7)		ト 62
(31) 優先権主張番号	09/693,357	(74) 代理人	100089266
(32) 優先日	平成12年10月19日(2000.10.19)		弁理士 大島 陽一
(33) 優先権主張国	米国(US)	(72) 発明者	リード、コーク・エス
			アメリカ合衆国ニュージャージー州085
			40・プリンストン・ウィリアムストリー
			ト 62

最終頁に続く

(54) 【発明の名称】 スケーラブルなワームホール・ルーティング集信装置

(57) 【特許請求の範囲】

【請求項1】

相互接続構造であって、

並列データ経路のセットSであって、パケット形式のデータが前記相互接続構造の外部の1以上のデバイスから前記データ経路に進入して、前記セットSの1以上のデータ経路上をデータ出力ポートへとワームホール送信で進むことが可能な、並列データ経路のセットSを有し、

セットSのデータ経路P上のノードAにおいて、該ノードAからのデータの移動は、

前記ノードAからノードBへのデータの移動が前記データ経路Pのデータ出力ポートにより接近する1ステップの移動であるように、データが前記データ経路P上を、ノードAから該ノードAの直後にあるノードBに移動するか、或いは

前記ノードAから、セットSのデータ経路Q上のノードCにデータが移動するかの何れか一方が可能であり、

後者の移動先の前記ノードCから前記データ経路の出力ポートへの距離は、前記ノードBから前記データ経路Pの出力ポートへの距離よりも短いことを特徴とする相互接続構造。

【請求項2】

前記ノードCの直前にある、データ経路Q上のノードDを更に有し、前記ノードDから前記ノードCへのデータ移動が、前記ノードAから前記ノードCへのデータ移動を上回る優先順位を有することを特徴とする請求項1に記載の相互接続構造。

【請求項3】

前記相互接続構造内のトラフィックによってデータの廃棄を生じさせないことを特徴とする請求項1に記載の相互接続構造。

【請求項4】

相互接続構造であって、
複数の送信リングと、
複数の前記送信リング内に配置された複数のノード（102）とを有し、
前記データ送信リングが、最上位レベルLから最下位レベル0まで複数のレベル内に配置されて、
レベルL上のリングが、前記相互接続構造の外側からデータを受信することが可能である1以上のノードを備え、
複数の前記ノードが、ノードA、B、C、X、及びYを含み、
前記ノードBが、レベル0とレベルLとの間のレベルJ上にあり、
前記ノードAが、前記ノードBと同一のデータ送信リング上で前記ノードBの直前にあり、
、
前記ノードCが、前記ノードBと同一のデータ送信リング上で前記ノードBの直後にあり、
、
前記ノードXが、前記レベルJよりも大きいレベルK上にあり、
前記ノードYが、前記レベルJよりも小さいレベルP上にあり、
前記ノードBが、前記ノードA及び前記ノードXからデータを受信することが可能であり、
、
前記ノードBが、前記ノードC及び前記ノードYにデータを送信することが可能であり、
前記データは、データ・メッセージに対してターゲットを特定するヘッダ情報を用いずに自己ルーティングされることを特徴とする相互接続構造。

【請求項5】

前記データは、少なくとも部分的にサービス品質情報に基づいて、自己ルーティングされることを特徴とする請求項4に記載の相互接続構造。

【請求項6】

相互接続構造であって、
複数のデータ送信ライン（104）と、
複数の前記データ送信ラインの上に配置された複数のノード（102）とを有し、
複数の前記ノードが、ノードB及びノードのサブセットSを含み、
前記ノード・サブセットSが、前記ノードBにデータを送信することが可能な1以上のノードを含み、
前記サブセットS内のノードN及び該ノードNに到達するメッセージMに関して、前記ノードBへのメッセージ送信に対して前記ノードNよりも低い優先権を有する前記サブセットS内の別のノードからのメッセージによって、前記ノードNがノードBへの前記メッセージMの送信を決してブロックされないように、前記サブセットS内の前記ノード群がそれらの間で前記ノードBへのメッセージ送信に対して優先権関係を有し、前記メッセージMに対するルーティング情報が存在せずとも複数の前記ノードが前記メッセージをルーティングすることが可能であることを特徴とする相互接続構造。

【請求項7】

前記ノードNによってメッセージのルーティングがなされ得るノード・サブセット T_N と、
前記メッセージMを受信するのが最も望ましいノードから、前記メッセージMを受信するのが最も望ましくないノードまで、前記サブセット T_N の構成要素を順位付けし、前記サブセット T_N のブロックされていない最も望ましい構成要素に前記メッセージMを送信する、ノードNに関連付けられたロジックとを更に有することを特徴とする請求項6に記載の相互接続構造。

【請求項8】

前記Bの直前にあり、前記ノードBへのメッセージ送信に対して最も高い優先権を更に

10

20

30

40

50

有するノードAを有することを特徴とする請求項7に記載の相互接続構造。

【請求項9】

相互接続構造であって、
構造内に配置された複数のノード（102）を有し、
複数の前記ノードが、
ソース・レベルから宛先レベルまでの複数のレベルから成る階層と、
前記レベルに沿って延びる非横断経路内にある複数のノードと、
前記レベルの非横断経路の横断部分内の複数のノードであって、前記相互接続構造が1
つの集信装置になるように、前記ソース・レベルの横断部分内のノード数が、前記宛先レ
ベルの横断部分内のノード数よりも大きくなっており、ノードの前記レベルが前記構造内 10
の前記ノードの位置によって完全に決定される複数ノードと、
レベルLの経路P上のノードNに対してのものを含む、前記構造内で前記ノードを接続す
る複数の相互接続ライン（104）と、
前記レベルLの前記経路P上の第1隣接ノードに接続されたメッセージ入力相互接続ライ
ンと、
前記レベルLの前記経路P上の第2隣接ノードに接続されたメッセージ相互接続ラインと
、
階層内でソース方向のノードからデータを受信するために、前記ノードNのソース方向
の1以上のノードに接続され、且つ／又は階層内で宛先方向のノードにデータを送信する
ために、前記ノードNの宛先方向の1以上のノードに接続された少なくとも1つのメッセー 20
ジ相互接続ラインと、
ソース方向のノードに制御信号を送信するために、前記ノードNのソース方向のノード
に接続され、且つ／又は階層内で宛先方向のノードから制御信号を受信するために、前記
ノードNの宛先方向のノードに接続された少なくとも1つの制御相互接続ラインとを有する
ことを特徴とする相互接続構造。

【請求項10】

前記ノードNが前記レベルLの前記経路P上のメッセージによって占有されているか決
定し、該決定に基づいてソース方向の前記ノードに制御信号を送信し、前記ソース方向の
ノードにおけるメッセージの進行を速めることが可能な、前記ノードNに関連付けられた
ロジックを更に有することを特徴とする請求項9に記載の相互接続構造。 30

【請求項11】

各列がレベル内で非横断経路の横断部分内の複数のノードを相互接続し、少なくとも
1つのメッセージ相互接続ライン上及び少なくとも1つの制御相互接続ライン上のソース方
向及び宛先方向の接続を含むノード間の相互接続網を含んでいる、複数の列を更に有する
ことを特徴とする請求項9に記載の相互接続網。

【請求項12】

前記レベルに沿って延びる非横断経路にそれぞれ接続される複数のFIFOバッファ（1
18）を更に有することを特徴とする請求項9に記載の相互接続構造。

【請求項13】

レベルLの経路P上の前記ノードNに対してのものを含む、前記構造内で前記ノードを 40
接続する複数の前記相互接続ラインは、

前記経路Pのソース方向のノード及び前記レベルL内のノードに接続された第1制御出力
相互接続ラインと、

前記レベルLのソース方向のレベル内のノードに接続された第2制御出力相互接続ライ
ンを更に有することを特徴とする請求項9に記載の相互接続構造。

【請求項14】

レベルLの経路P上の前記ノードNに対してのものを含む、前記構造内で前記ノードを
接続する複数の前記相互接続ラインは、

前記レベルL内のノード及び前記経路Pのソース方向のノードに接続されている第1メッ
セージ入力相互接続ラインと、 50

前記レベルLのソース方向のレベル内のノードに接続されている第2メッセージ入力相互接続ラインと、

前記レベル内のノード及び前記経路Pのソース方向のノードに接続されている制御出力相互接続ラインとを更に有することを特徴とする請求項9に記載の相互接続構造。

【請求項15】

レベルLの経路P上の前記ノードNに対してのものを含む、前記構造内で前記ノードを接続する複数の前記相互接続ラインは、

前記レベルL内のノード及び前記経路Pのソース方向のノードに接続されている第1制御出力相互接続ラインと、

前記レベルLのソース方向のレベル内のノードに接続されている第2制御出力相互接続ラインと、

前記経路Pのソース方向のノードに接続されているメッセージ入力相互接続ラインとを更に有することを特徴とする請求項9に記載の相互接続構造。

【請求項16】

複数の前記ノード間の優先ノードと、非優先ノードであるその他のノードとを更に有し、前記優先ノード及び前記非優先ノードが選択的に相互接続されることを特徴とする請求項9に記載の相互接続構造。

【請求項17】

ソース方向経路上のnノードを宛先方向経路上のmノードに集約するように、相互接続ノードを集団化するロジックを更に有し、

nがmよりも大きく、且つn:mが集約比になっていることを特徴とする請求項9に記載の相互接続構造。

【請求項18】

前記非横断経路内の複数のノードに接続される相互接続ラインのリボン(112)であって、相互接続ラインの該リボンは、前記ソース・レベルから前記宛先レベルまで複数の前記レベルを通過して巻かれ、巻かれる回数が前記ソース・レベルから前記宛先レベルに減少する、相互接続ラインのリボンと、

前記レベルの全ての巻き部分を通過して、前記リボンの横断部分内の前記ノードを接続する複数の列とを更に有することを特徴とする請求項9に記載の相互接続構造。

【請求項19】

前記相互接続ライン・リボンの前記巻き数が、前記ソース・レベルから前記宛先レベルまで、レベル毎に半分減少することを特徴とする請求項18に記載の相互接続構造。

【請求項20】

システムであって、

複数のスイッチ(702)と、

複数の前記スイッチに各々接続される、請求項9に記載の相互接続構造を有する複数の集信装置(700)とを有することを特徴とするシステム。

【請求項21】

相互接続構造であって、

非横断経路内の複数のノードに接続される相互接続ラインのリボン(112)であって、相互接続ラインの該リボンは、前記ソース・レベルから前記宛先レベルまで複数の前記レベルを通過して巻かれ、巻かれる回数が前記ソース・レベルから前記宛先レベルに減少する、相互接続ラインのリボンと、

前記レベルの巻き部分を通過して、前記リボンの横断部分内に渡っている前記ノードに接続されている相互接続ラインによって形成される複数の列とを有することを特徴とする相互接続構造。

【請求項22】

前記ソース・レベルの第1列内の前記ノードに接続されている複数の入力ポートを更に有することを特徴とする請求項21に記載の相互接続構造。

【請求項23】

前記宛先レベルの最後の列内の前記ノードに接続されている複数の出力ポートを更に有することを特徴とする請求項 2 1 に記載の相互接続構造。

【請求項 2 4】

前記レベルに沿って延びる非横断経路にそれぞれ接続されている複数の FIFO バッファを更に有することを特徴とする請求項 2 1 に記載の相互接続構造。

【請求項 2 5】

前記ソース・レベルから前記宛先レベルまでレベル毎に半分減少する、前記相互接続ラインの前記巻数を更に有することを特徴とする請求項 2 1 に記載の相互接続構造。

【請求項 2 6】

前記相互接続構造の中へのメッセージの進入を制御するように、前記相互接続構造の内側のノードから、前記相互接続構造の外側の 1 以上のデバイスに延びている制御ラインを更に有することを特徴とする請求項 2 1 に記載の相互接続構造。 10

【請求項 2 7】

ソース方向経路上の n ノードを宛先方向経路上の m ノードに集約するように、相互接続ノードを集団化するロジックを更に有し、

n が m よりも大きく、且つ $n : m$ が集約比になっていることを特徴とする請求項 2 1 に記載の相互接続構造。

【請求項 2 8】

ノードがレベルの前記経路上のメッセージによって占有されているか決定し、該決定に基づいてソース方向のノードに制御信号を送信し、ソース方向の前記ノードにおけるメッセージの進行を速めることが可能なロジックを更に有することを特徴とする請求項 2 1 に記載の相互接続構造。 20

【請求項 2 9】

複数の前記ノード間の優先ノードと、非優先ノードであるその他のノードとを更に有し、前記優先ノード及び前記非優先ノードが選択的に相互接続されることを特徴とする請求項 2 1 に記載の相互接続構造。

【請求項 3 0】

システムであって、

複数のスイッチ (702) と、

複数の前記スイッチに各々接続される、請求項 2 1 に記載の相互接続構造を有する複数の集信装置 (700) とを有することを特徴とするシステム。 30

【請求項 3 1】

前記相互接続構造を通して移動するデータ・パケットは、該データ・パケット内に含まれるターゲット出力情報に依存しないことを特徴とする請求項 2 に記載の相互接続構造。

【請求項 3 2】

前記データ経路 Q が前記データ経路 P と分離されており、且つ別個のものであることを特徴とする請求項 1 に記載の相互接続構造。

【発明の詳細な説明】

【0001】

(技術分野)

本発明は、集信装置に関し、特に、迅速にデータのルーティングを行い、妨害物を回避することにより情報の流れを向上させ、制限なしに垂直方向にスケラブルで、且つ低レイテンシー及び高処理能力をサポートする集信装置に関する。

【0002】

(背景技術)

通信ネットワーク若しくはコンピューティング・ネットワークは、例えば、金属若しくは光ファイバーのケーブル等の通信媒体経由で物理的に接続される、いくつか若しくは多数のデバイスで構成されている。ネットワーク内に含まれ得るデバイスの一種は、集信装置である。例えば、ラージスケール時分割スイッチング・ネットワークは、中央スイッチン 50

グ・ネットワーク、及びこのスイッチング・ネットワーク内の別のデバイスの入力端子及び出力端子に接続されている一連の集信装置を含んでいてよい。

【0003】

集信装置は通常、ネットワークへのマルチポート接続をサポートするのに用いられる。集信装置は、複数の共有通信ラインに接続された、より少ないライン上に情報を集めるデバイスである。集信装置は、ブロッキング及びデータ損失の発生率を増加させることによって、若しくは記憶のためにバッファ内の情報を呼び出すことによって、相互接続経路の受容能力を本質的に減少させてしまう。

【0004】

大規模な並列コンピューティング及び通信において生じる永続的な問題は、データをプロセッサ及びユーザに移動させる際に生じる。集信装置が実装されたシステム内部においては、受容能力が本質的に減少するため、この問題が悪化する。

【0005】

必要とされているのは、迅速にデータのルーティングを行い、妨害物を回避することにより情報の流れを向上させ、制限なしに垂直方向にスケラブルで、且つ低レイテンシー及び高処理能力をサポートする集信装置である。

【0006】

(発明の開示)

制御信号を用いた制御セル経由の単一ビット・ルーティングの利用を介して、情報集信装置の動作を概ね向上させる相互接続（インターコネクト）構造である。用語「セル」若しくは「制御セル」は、単純なスイッチング要素を示している。用語「ノード」は、1つの構成単位として作動するセル若しくはセル集団を示している。この相互接続構造及び作動方法によって、ワームホール・ルーティング及びメッセージの流れがサポートされる。この構造に進入した任意のパケットが確実に退出するように、この構造に進入するメッセージ・パケットは決して廃棄されない。

【0007】

本発明の一態様によれば、相互接続構造が、非横断部の経路内にある複数ノードに接続される相互接続ラインのリボンを含んでいる。相互接続ラインのこのリボンは、出発レベルから宛先レベルまで、複数のレベル経由で巻かれる。巻かれる回数は、出発レベルから宛先レベルまで減少していく。相互接続構造は、交差部内でレベルの巻付部分を通してリボンを横切るノードを結合する相互接続ラインにより形成される複数の列を、更に含んでいる。相互接続構造上でデータを通信する方法には、データ・パケットを複数の階層レベルを下側にルーティングするための高速、最小ロジックの方法が組込まれている。

【0008】

(発明を実施するための最良の形態)

図1を参照すると、3レベル及びK列を含む、多重レベル32:8集信装置100の形態でのデータ相互接続構造の一例が示されたブロック図である。レベルは、階層的で最下位から0、1、及び2と数えられており、列は左から右に0からK-1まで数えられている。集信装置100は、32の軽負荷データ輸送ラインから入力データを受信し、より重負荷の8ラインにそのデータを渡す。図示されている例では、データ・パケットは、32ポート入力端子120から8ポート出力端子122に、8幅の「リボン」の行の中を進んで行く。このリボン112は、各列における制御アレイ110内に配置された制御セルの8幅セット、及び列Kの右側に配置されたシフト・レジスタ（FIFO）118を含んでいる。最下位の出力レベルはFIFOを含んでいない。リボン112のトポロジーは、図1に図示されるように、理髪店ボール風の巻きとして描写され得る。データは、図2に図示されるパケット200の形式になっている。このリボン112は、レベル2の周囲に4回巻かれ、レベル1の周囲に2回巻かれ、更にレベル0を1回通過する。従って、レベル2は32のセル行及び32のFIFOを有し、レベル1は16のセル行及び16のFIFOを有し、更にレベル0はFIFOなしの8行を有する。リボンの幅は出力ポート数と概ね等しいが、別の構成を用いることもできる。

【0009】

10

20

30

40

50

図1に示される模式的な相互構造接続は、複数のデータ送信ラインから成る8幅リボンを有する。個々のラインは、7つのセグメントに分割されており、各セグメントは1つのメッセージを含むのに十分大きくなっている。8幅リボン中の各データ送信ラインは、一連のノードに接続する。この8幅リボンは、種々の巻きの上にあるノード間にいくつかの相互接続を備えた概ね螺旋構造に巻かれる。データは、ノードからノードへデータ送信ラインに沿って、先入れ先出し動作で進められる。データは、相互接続構造経路でシステム入力ポートからシステム出力ポートに移動する。相互接続構造の上位8ライン内にある列0レベル2内のノードを除き、データ送信ラインL上の各ノードBは、データ送信ライン上に直前のノードAを有する。

【0010】

10

データ送信ラインL上の各ノードBは、データ送信ラインL上にある、若しくは出力ポートである、直後のCを有する。このノードAは、データBへのデータ送信が常に可能である。ノードBは、ノードCへのデータ送信が常に可能である。ノードAからノードBに送信されたデータ、及びノードBからノードCに送信されたデータは、常にブロックされない。

【0011】

いくつかのノードは、リボンのFIFO領域内に配置される。FIFO領域内のノードは、1つだけの入力ポート及び1つだけの出力ポートを有する。

【0012】

いくつかのノードは、データが送信ラインのリボンの外側から進入することを可能にする付加的な入力ポートを有する。

20

【0013】

いくつかのノードは、データがノードに、送信ラインのリボンのずっと上流にあるノードから進入するのを可能にする第2入力ポートを有する。このずっと上流にあるノードは通常、現在のノードと比較して過去のリボンの巻きの上にある。

【0014】

いくつかのノードは、出力ポートのリボン・システムのより近くのリボンへとずっと下側にデータ送信を可能にする1以上の第2出力ポートを有する。複数の出力ポートを備えたノードは、送信ラインのリボンに沿って、直後のノードへのメッセージの送信の試みをするよりもむしろ、メッセージを可能な限りシステム出力ポートへの前方へと送信することを常に試みる関連付けられたロジックを有する。

30

【0015】

複数の入力ポートを備えたノードは、入力ポートに割当てられた優先権を有する。直後からのメッセージは、直後ではないノードからのメッセージを上回る優先権を常に有している。直前にないノードに接続された複数の入力ポートを備えているノードも同様に、データ受信に対して優先順位関係を有している。

【0016】

要約すると、ノードは、可能な限り下側のリボンへのデータ送信を常に試みるが、各ノードBに関連して、ノードBへのメッセージ送信が可能な種々のノードからのメッセージの受信が優先権によって規定されている。最上位の優先順位は、直前のノードに付与される。セットSは、ノードBへのメッセージ送信が可能なノードを含んでいる。優先順位関係は、ノードBへのメッセージ送信に対してセットS内のノードの間で定義される。メッセージの前方への進行を管理する規則は以下の通りである。

40

【0017】

1. ノードNがセットSの構成要素であり、且つメッセージMがノードNに到達する場合、ノードNは、セットS内の別ノードからの、ノードBへの送信に対してノードNよりも優先順位が低いメッセージによって、ノードBへのメッセージ送信がブロックされることは決してない。

【0018】

2. セットS内の各ノードNに対応して、いずれのノードNがメッセージ・ルーティングを行うことが可能であるノードのセットTNが存在している。各ノードNに関連付けられたTN

50

の構成要素は、メッセージMを受信するのが最も望ましいノードNMからメッセージMを受信することが最も望ましくないノードNLの範囲で、順位付けられる。ノードNに関連付けられたロジックは、セットTNのブロックされていない最も望ましい構成要素にメッセージMを送信する。

【0019】

3. ノードBの直前のノードAが、メッセージのノードBへの送信に対して最も高い優先権を有する。

【0020】

図1では、前にある送信ラインのリボン上のノードからのメッセージ受信及びメッセージ送信の能力を備えたノード、即ち、相互接続構造のセグメントを飛越えるノードが、K個の制御セル列内に配置されている。1つだけの入力ポートを備えたノードが相互接続構造のFIFO区間内に配置されている。

10

【0021】

メッセージは、常に1に設定される単一ビット・ヘッダー単一タイミング・ビットが前にあるペイロードを含んでいる。セグメントの各々は、レベル上の行に対応している。1つの行は、その行に広がるK個の制御セルを含んでいる。メッセージは、単一行上のFIFO及びK個の制御セルの内部に適合する必要があるので、このメッセージの長さはFIFO+CKの長さを越えることはできない（ここでCは制御セル内のビット数である）。それゆえに、FIFOの長さは少なくとも、メッセージの最大長からCKを引いた長さである必要がある。システムが多数のメッセージをRの信号に集約させる場合、リボン幅はRである。相互接続構造がL+1レベル及びレベル当たりK列を含む場合、このシステムは、各行に、少なくともメッセージの最大長からCKを引いた長さである、Kの制御セル及びFIFOを有している $R \cdot (2^{L+1} - 1)$ 行を含んでいる。

20

【0022】

一実施例では、メッセージが、列0におけるリボンのセグメントのうちの1つに送られる。このメッセージが列1に移動する際に、このメッセージはこのシステムの実出力ポートのより近くにある別のセグメントへと前方に移動することが可能である。メッセージの第1ビットが送信ラインのリボンの新セグメントに移動するとき、このメッセージが異なる送信ラインに移動してもよいし、若しくはリボン上の同一送信ライン上に留まってもよい。メッセージ・ヘッダが新しい列に到達すると、このメッセージは再度新セグメントへと前方の飛んでもよい。この方法により、メッセージが複数のセグメントに広がってもよいし、最下位レベルがFIFOを有さないことから、全メッセージが集信装置に進入し終わる前にメッセージの第1ビットが集信装置を退出してもよい。

30

【0023】

以下の説明では、用語「パケット」は、通常は直列形式のデータの構成単位を示している。データ・パケットの例には、インターネット・プロトコル（IP）パケット、イーサネット・フレーム、ATMセル、大きいフレーム若しくはデータ・パケットの一部等のスイッチ・ファブリック・セグメント、並列コンピュータのプロセッサ間メッセージ、若しくは長さの上限を有するその他のデータ・メッセージがある。1つのレベルを横切って進むパケットは、制御セル102のK列を通過する。レベルを横切るこのパケットは、1つのセルから、同一レベル上にある次の列の内部のセルに直接、渡される。一例では、同一レベル上のセルの間を横切るパケットに対して、パケットの2ビットが各々のセル内に配置される。システム内でここで説明されたのものと同一のタイミングが用いられてよいが、本説明におけるセル設計は単純化されている。FIFOは、適切なビット数のパケットを含み、それにより列0のアレイにパケットが進入する際に、このヘッダ・ビット202は、同一のレベルを横切るパケットの別のヘッダ・ビットと整列する。この例では、列0に進入するパケットのタイミングが、階層内の上側に配置されているセルから到達するパケットと同期化されるように、パケットが単一行に適合されている。

40

【0024】

タイミング及び制御は、最長パケットのビット長がFIFO内のビット数に列内のビット数の

50

合計を加えたものよりも大きくなってしまわないようにすることで達成される。例えば、各々が2ビットを保持する制御セルで構成される12列を備えた構造内に、400クロック周期に到達する、上位ビットの長さが400であるパケットに対しては、FIFOの長さは376ビットとなる。パケットの第1ビットは、特定レベル上のセルから、次の列内にある同一レベル上のセルへと、2クロック刻み若しくは2クロック・ステップで移動する。パケットの第1ビットは、特定レベル上のセルから、次の列内にある1つ下位のレベル上のセルへと、1クロック刻みで移動する。それゆえに、レベル間のタイミング差異を補正するために、レベル間を接続するFIFOの長さは、1ビット少なくなっている。レベル間のタイミング差異を補正するために、レベル間を接続するFIFOの長さは、1ビット少なくなっている。レベルのタイミングは以下で詳細が説明される。

10

【0025】

最上位レベル内の32のデータ入力ポートのうちの1つから、入力アレイ120の中にこのパケットが挿入される。1つのデータ入力ポートによって、最上位レベル内にある32ラインの各々に供給される。1つのパケットが直列式に各ライン中に挿入される。アレイ120内のリボン112からセルに進入するパケットは、相互接続構造の外側からセルへの進入を試みるパケットを上回る優先権を有している。パケットが構造の上位8行に進入するのをブロックすることが可能な内部パケットは1つもないが、最上位レベルの上位行を横切るパケットは、何らかのブロッキング状態で、「理髪店ポール」の周囲に巻かれて、アレイ120内の最上位レベルの8行下側に再進入することが可能である。構造に進入するパケットが決して廃棄されず、それにより構造に進入する任意のパケットが確実に退出できることが、従来の集信装置を上回る、非常に改善された点である。

20

【0026】

このパケットは、スイッチ・アレイの次列への進行に対して、構造の外側から相互接続構造に進入を試みるパケットを上回る優先権を有する。これらの状況下で、構造への進入が試みられるこのパケットは、進入が禁止される。動作の複数の態様によって、相互接続経由の通行に影響が及ぼされるが、より古いパケットが、より新しいパケットの前に退出する優先権を有する。

【0027】

代替実施例では、パケットが最上位レベルの複数の角にあるリボンに進入することが可能である。最上位レベルのある特定セルに対して、構造内部からセルに進入するパケットの第1ビットと、構造の外側から同一列内にあるセルに進入するパケットの第1ビットとが同時になる。

30

【0028】

換言すると、最上位レベルの32の入力ラインにおいて、構造の外側からのパケットを受信する。32の入力ラインのうち、上位8つの入力ポートがリボンの開始点に接続され、構造内に既に存在するパケットによるブロックは不可能である。残りの24の入力ポートは、ブロックすることが可能である。ブロッキングは、図4A、4B、及び4Cの説明の詳細で考察される。

【0029】

データ輸送リボン内の螺旋の周囲を進行する代わりに、パケットは、階層内で上位レベルのセルから下位レベルのセルへと、より前方に飛ぶことが可能である。この飛越し処理によってリボン内のパケットが集約され、それにより最下位レベルのリボンのセグメントが、最上位レベルのリボンのセグメントよりも平均して多くのメッセージを輸送する点が効果的である。

40

【0030】

所定レベルのパケットが、制御セルアレイ110内の制御セル102に進入する。図4Aを参照すると、パケット P_A 410が制御セルBに進入する場合、制御セルBに関連付けられたロジックによって、セルB経由で同一レベル上の別セルCにルーティングがなされてもよい。或いは、セルBに関連付けられたロジックによって、下位レベルのセルXへとパケット P_A のルーティングがなされてもよく、これは、この集信装置及び動作方法の利用に対する

50

望ましい結果である。

【0031】

あるパケットが、上のライン106若しくは404から、例えばセルB若しくはEから、セルWに進入すると、このパケットは遅延なしで次列のセルXへとライン104上に送信される。それゆえに、パケット P_A がセルBからセルXへルーティングされる場合、パケット P_A はセルWへ進み、次にすぐにセルXに進入する。図4Aと共に図4Cを参照すると、セルBがパケット P_A をライン406に下側にルーティングすると、このパケット P_A は同様にしてすぐにセルXに進入する。図4A及び図4Bに示される図は、セル間のパケット・ルーティングの例と機能的には同等である。組込みの状況によって、他方を上回る一方のルーティング方法の選択に影響が及ぼされる可能性がある。

10

【0032】

パケットの下位レベルへ即座の「降下」若しくは「飛越し」によって、集信装置の相互接続構造の効率的な動作が増進される。基本的には、所定の下位レベルのセルがパケットによって占有されていない場合、即ち、データを受取るために利用可能であり、そのセルが上位レベルの別セルに接続されている場合、上位レベルのセルのパケットが、下位レベルのセル内の空の場所を満たすように降下する。いくつかの降下若しくは飛越しの処理が適切である。制御セル構造によって、降下若しくは飛越し方法が決定される。制御セル構造とデータ全体の流れ及びタイミングとが以下の詳細で説明される。

【0033】

図1は、レベル間の比が2:1である8幅リボン及び3レベルを備えた32:8の集信装置が図示されている実施例であり、それによって、集約比が全体で4:1である集信装置が得られる。別のリボン幅と同様に、別の集信装置の比が実用的であって、望ましくは、提供される入力及び出力ポート数は広範囲である。集約比は、更なるレベルを付加することによって、レベル間の集約比を増大させることによって、もしくは両方によって増加され得る。4:1より大きい全体比が用いられている一実施例では、3よりも多くのレベルを有している。別の実施例において、レベル間で2:1以外の比が用いられてもよい。図1に図示されている制御セルでは、データ輸送ライン及び制御信号輸送ラインの一部だけが図示されている。付加的なラインは、図3A及び3Bで示される構造に関して以下で説明される。

20

【0034】

図2は、パケットの配置が図示されている。パケット200のヘッダ202は、常に1に設定されて、パケットが存在していることを示す単一ビットを含んでいる。パケットの残りは、ペイロード204と呼ばれる。種々のペイロードが図2に例示されている。あるパケットは、任意の退出ポートから相互接続構造100を退出することが可能であり、それによりルーティング情報を含む付加的なヘッダ・ビットが不要となる。

30

【0035】

集信装置100を出た後に、パケットがネットワーク・ルーティング・デバイスに進む場合、ルーティング・デバイスによって利用されるためのルーティング・情報が、ペイロード内にあるデータ・ビットより前になっていてもよい。このルーティング・ビットは、集信装置100のロジックには使用されない。この集信装置100は、常にペイロードの内容204を無視する。

40

【0036】

集信装置（図示せず）の代替実施例では、サービス品質を示す付加的なヘッダ・ビットが用いられてもよい。

【0037】

図3A及び図3Bは、セルの入力ポート及び出力ポートの例が示されている。図示されているポート構成では、セル102は、より上位レベルのセルから2つのデータ入力ライン106上のデータを受信し、且つ下位レベルのセルへと1つのデータ出力ライン106上にデータを送信する。垂直方向接続のこの2-1の比によって、集信装置のデータ集約動作が補助される。セル102は、同一レベル上にあるセルからの1つのデータ入力ライン104と、同一レベル上にあるセルへの1つのデータ出力ライン104とを、更に含んでい

50

る。セル間のデータ入力及び出力ラインに加えて、付加的な相互接続ライン108によって、セル間で制御情報の輸送が行われる。この制御情報は、受信セルからのデータ・パケットをどのようにルーティングするか決定するべく、受信セルのロジックによって用いられる。この制御ラインは、相互接続構造を経由するデータの流れを管理する、位置ベースの優先順位の規則を行使するのに用いられる。

【0038】

ここで説明されるように、第3ノードへのデータ送信を試みる2ノードは、上側の競合を解消するために制御信号を用いて通信を行う。別のネットワークでは、ノードNに進入するパケットは、出力ポートに対して競合する。ここで説明されるシステムでは、パケットがノードの入力ポートに対して競合しており、この競合は制御信号を用いて上流で解消される。上流のデータの流れは、説明されているネットワーク内での固有のトポロジーによって許可されており、それにより2つのパケットが特定の下流のノード入力ポートに対して競合し得る。競合の解消は、ネットワーク内のノードの配置に、少なくとも部分的に基づいている。第3ノードVへのデータ送信に対して、ノードBを上回る優先権を有するノードEは、その優先権を行使するべく直接的に、若しくは間接的にBに制御信号を送信する。図3Aは、1つの制御信号を受信し、1つの制御信号を送信するノードを示している。図3Bは、1つの制御信号を受信し、2つの制御信号を送信するノードを示している。

10

【0039】

2種類のノードの利用は、図4Aに示される相互接続構造の考察の中でより詳細に説明される。

20

【0040】

図3Bは、上位レベルの1つのセルと、同一レベルの1つのセルとを上回る優先権を有するあるセルの一例を示している。

【0041】

図4Aは、全てのデータ及び制御ラインが示されていない、図1に示される集信装置100の一部の拡大図である。セルE及びBは同一レベルにある。セルE及びBの各々を別個のノードとして見ることもできるし、若しくは2つのセルE及びBを単一ノードとして1つにみなすこともできる。

【0042】

集信装置100は、以下の方法で作動する。同一レベル上のセルV、W、及びXに対して、セルVがパケット P_V をセルWに送信し、セルWがパケット P_V の第1ビットをセルXに送信する場合、セルWは、より上位レベルにあるセルEへと制御信号を送信する。セルWはライン104上のセルXにパケット P_V を送信する。セルWは、ライン108上のセルEに制御信号を送信し、それによりセルEが下向きにライン106にデータ送信しないよう通知する。セルEは、ライン402経由で制御信号を送信し、それによりセルBが下側にライン404にデータを送信しないように通知する。セルWがパケット P_V をセルXに送信する場合、セルDからセルEに進入する任意のパケット P_D 412は、セルEのロジックによって、次列のセルFへとルーティングされる。更に、セルAからセルBに進入する任意のパケット P_A が、セルBからセルCへ送信される。

30

【0043】

所定のパケット到達時間に、セルWがセルXにパケットを送信しない場合、セルWがセルXにパケットMを送信しないというセルWの状態が、セルWによってセルEに通知される。

40

【0044】

図4Aを再度参照すると、セルWからセルXに送信されたパケットが1つもないことを通知するために、セルWによってセルEに送信された、制御ライン108上を輸送される制御信号内に、制御情報が含まれている。セルWからセルXへのライン104はパケットによって占有されておらず、データを受取るべく利用可能である。この状態でセルDがパケット P_D をEに送信する場合、セルEは P_D の1ビット・ヘッダ202を読込むことによってパケット P_D の存在を検出する。セルEに関連付けられたロジックは、セルWからの制御信号及びパケット P_D からのヘッダ・ビットを用いて、パケット P_D のセルXへのルーティングをセルW経由にす

50

るかどうか決定する。セルWがセルXに別のパケットを送信するのにデータ・ライン104を現在用いていないので、パケットP_Dは、セルWからセルXへのデータ・ライン104を自由に利用することが可能である。

【0045】

上にあるレベルからセルWに進入する任意のパケットは、常にセルXに直接的に送信される。図示されている集信装置は、セルWに関連付けられたロジックは、セルWと同一レベル上にある別のセルからセルWに到達するパケットのルーティングが可能であるが、上位レベルからセルWに進入するパケットに対しての制御は何ら有していない。上位レベルからのパケットは、セルWを通過して、セルWからセルXへとデータ相互接続ライン104上を進む。セルEによって、パケットP_DがセルWからセルXまでを経由して下側にルーティングされる場合、セルEは次に制御ライン402を介して制御信号をセルBに送信する。この制御信号は、この時にセルBによるライン404上へのパケット送信がブロッキングされるといことが指定される情報を運ぶ。セルEからのブロッキング制御信号が存在すると、セルBはセルAからパケットP_Aを受信した場合に、次に、セルBに関連付けられたロジックによって、パケットP_Aに対してセルBからセルCまで、ライン104経由でのルーティングがなされる。

10

【0046】

ある場合には、特定のパケット到達時間において、セルWに関連付けられたロジックは、セルXへのパケットのルーティングを行わず、且つセルEはパケットをライン106に下側に送信しない。セルBは、ライン404に下側にパケットを自由に送信することができる。セルWは、例えば、単一ビット形式である制御信号を制御ライン108経由で送信し、それにより、セルWがセルE若しくはセルBからパケットの受信がブロッキングされていないことをセルEに対して通知する。セルEは、制御信号をライン402経由で送信して、セルWがセルBからのパケット受信をブロッキングされていないことをセルBに対して通知することによって、セルWからの制御信号及びセルDからの到達するメッセージが存在しないことに応答する。このときに、パケットP_AがセルBに到達すると、セルBは、ライン404を介してセルW経由でセルXへとパケットP_Aを送信する。パケットP_Aは、最初にライン404を通して進み、次にライン104を通る。セルW若しくはセルEからセルXへとルーティングがなされたパケットを運ぶのにライン104が用いられていないので、ライン104はパケットP_Aを自由に運ぶことができる。

20

30

【0047】

図4Aを参照しながら続けると、ノードWは、単一の制御信号輸送ライン108を介して制御信号をノードEに送信するように接続されている。図示された相互接続ラインに加えて、レベルJの上にあるノードは、レベルJ-1上のノードから制御信号を運ぶことが可能な制御信号輸送ライン（図示せず）と、レベルJのノードからレベルJ-1の上にあるノードにパケットを送信することが可能なデータ輸送ライン（図示せず）とを、更に有している。例えば、ノードWは、図3Aに図示された、3つのデータ入力ポート、2つのデータ出力ポート、1つの制御信号入力ポート、及び1つの制御信号出力ポートを含むポート構造を有している。再度、図4Aを参照すると、ノードEがノードWからライン108経由で制御信号を受信している。ノードEは、同一レベル上にあるノードBに制御信号を送信し、更に、ノードEが最上位レベルの上にある場合には、ノードEは、レベルJ+2の上にあるノードにも同様に制御信号を送信する。図示された相互接続ラインに加えて、レベルJ+1上のノードは、最上位レベルでない場合に、レベルJ+2上のノードに接続されている付加的なデータ及び制御の相互接続ラインも有する。例えば、ノードEは、図3Bに図示されるような、レベルJ+2上のノードからの2つのデータ輸送ラインが含まれるポート構造を有する。

40

【0048】

セルBが最上位レベルにない場合、セルBは上位レベルにあるセルに制御信号を送信するための接続ライン（図示せず）に接続される。この制御ラインは、セルWからセルEへの制御ラインと同様の機能を有する。セルBがセルCにパケットP_Aを送信する場合、次にセルBは、上にある次のレベルのセルにブロッキング信号を送信する。

50

【0049】

セルBが列K-1内にあり、且つ上位レベルの上位24行内にある場合には、セルBは、集信装置の外側にあるデバイスから入力データを受信するセルの中へと、FIFOを介してメッセージを送信する位置にいる。同様に構造の外側からメッセージを受信し得る別のセルにメッセージを送信することが可能なセルBのようなセルは、構造の外側にあるデバイスから相互接続構造の中へのメッセージの流れを制御する特別な制御信号を有する。特別な制御信号ライン上の制御信号は、構造の外側にあるデバイスにデータ入力列120に、生じる可能性があるビジー状態を通知する。

【0050】

セルBが最上位レベルの列K-1内にあるが、最上位レベルの上位24行内にはない場合には、セルBは構造の外側に制御信号を送信しない。

10

【0051】

セルEが最上位レベルにない場合、セルEは2つの制御信号を送信する。セルEがセルWにパケットを送信するときに、セルEは制御信号ライン402上をセルBへとブロッキング信号を送信する。セルEがセルFにパケットを送信するときに、セルEは、次の上位レベルへと制御信号ライン403上を上向きにブロッキング信号を送信する。

【0052】

セルEが列K-1内にあり、且つ最上位レベルの上位24行内にある場合には、セルEは構造の外側にあるデバイスから相互接続の中へのメッセージの流れを制御する特別な制御信号を送信する。

20

【0053】

一般的に言うと、階層内で下側にパケットを送信する優先権を有するノードは、図4Bに示されているような2つの制御信号送信ラインを有する。下側へのパケット送信に対する優先権を有さないノードは、制御信号送信ラインを1つだけ有する。

【0054】

図4A及び図4Bは、論理的に同一な相互接続構造の代替実施例を示している。図4Bに示される構造がパケットをセルBからセルXに、最初はセルE経由で、次にセルW経由で送信することが物理的な差異である。メッセージは、セルEがそうでない場合にデータ送信にライン106を用いず、且つセルE及びセルWのいずれもがセルXへのデータ送信にライン104を用いない場合は、メッセージはセルBからセルXへの飛ぶことが可能である。図4A及び図4Bに示されているこの制御構造は同じであり、制御ラインを含んでいる。

30

【0055】

図4Cは、図4A及び図4Bに示されている相互接続構造と論理的に同一である相互接続構造の第3番目の実施例を示している。図4Cでは、セルE及びセルBは直接的にセルXにパケットを送信し、隣接ノードを経由しない。制御ライン構造は、図4A、図4B、及び図4Cに示されている構造の内部のものと同様である。図4A、図4B、及び図4Cは物理的には異なっているが、論理的には同等である。

【0056】

集信装置の動作がうまく成功するかどうかは、少なくとも部分的にはタイミングに依存する。パケットP_Vの第1ビットは、所定のパケット到達時間にノードWに到達する。ノードWに関連付けられたこのロジックは、パケットP_Vの単一ビット・ヘッダ及びレベルJ-1のノードから入力される制御信号に基づいてルーティングの決定を行う。パケットP_VがノードWに存在し、且つ下位レベルにあるノードからブロッキングされない場合、ノードWは下位レベルにパケットP_Vを送信し、更に、上位レベルのノードEに制御信号を送信する。ノードWから制御信号は、そのパケットがノードEに到達するのと同じ時間若しくはそれに近い時間にノードEに到達する。このパケットがノードEに到達する前にパケットがノードWに到達するように、ロジックによってタイミングが制御される。

40

【0057】

図4Cを参照すると、例えば光学的な実施例では、パケットがノードXに到達するためのパケット到達時間が、ロジックによってあらかじめ決定される。ノードEからノードXに送信

50

されるパケットは、パケットがノードWからノードXに送信されるのと同じ時間に、ノードXに到達する必要がある。以下で説明されるように、パケットがノードEに到達する前にパケットはノードWに到達する。それゆえに、パケットがノードEからノードXに進むための時間は、パケットがノードWからノードXに進むための時間より小さくなっている必要がある。光学的な実施例では、ノードWからノードXへの相互接続ラインのファイバーの長さを、ノードEからノードXへの相互接続ラインのファイバーの長さよりも短く選択することによって、タイミングが調整される。この方法によって、到達時間を同期化するべく、上位レベルからのパケットが下位レベルからのパケットに追いつくことが可能である。

【0058】

電氣的な実施例では、同一レベル上にある2ノード間を進むパケットは、2つの1ビット・シフト・レジスタのセルを通過する。レベルを下側に延びているノードは、レジスタ・セルのうちの1つをバイパスしており、それにより図4A、図4B、及び図4Cに示される構造内で、パケット・ビットは、セルEからセルFに2クロック周期サイクルで進み、セルEからセルXに1クロック周期サイクルで進む。

【0059】

図1、図3A、図3B、図4A、図4B、及び図4Cに示される構造は、メッセージを相互接続構造経由で送信することに対して同等の優先権を有している。ノードXの直前のノードWは、ノードXへの送信に対して第1番目の優先権を有する。ノードEは、ノードXへの送信に対して第2番目の優先権を有する。ノードBは、ノードXへの送信に対して第3番目の優先権を有する（ここで、ノードXは、ノードBよりもリボンの出力ポートに近い）。別の実施例では、リボンの出力末端から最も遠くにあるノードが優先権を有する。図5を参照すると、優先権が列に関して変更可能になっており、ノードBがノードAの直前にあり、且つノードAが優先権を有さないセルである場合には、ノードBが優先権を有するセルとなる。

【0060】

図5には、相互接続構造の3列、3レベルの相互接続構造が図示されている。各レベル及び各列には、制御セル102のアレイ110がある。図5に示されている構造内のセルーセル相互接続網は、図4Bに図示されている相互接続網と同様である。セルは、対になってノード502を形成しており、それによりノード対の両セルが次の下位レベルの単一ノードにデータを送信するように配置される。集信装置のより適切な実施例では、制御アレイ110内で行104上のセル102が無作為に配置される。望ましくは、最良の実行配置が選択されるように、複数の無作為配置のソフトウェア・シミュレーションが実行される。色が塗られたボックス504は、ノード内でより高い方優先権を有するセルを示している。色が塗られていないボックスは、低い優先権を有するセルを示している。図示されているノードは各々、高い方の優先権を有する1つのセルと、低い方の優先権を有する1つのセルとを含んでいる。行104に沿って、1つの列において低い優先権を有するパケットが、常に、同一レベルの次列に高い優先権を有するという点が効果的である。

【0061】

ノードが制御セルを1つしか含んでいないと考える場合は、各ノードが同一レベルから1つのデータ入力ラインを有する。最上位レベルにないノードは、上位レベルからの2つのデータ入力ラインを有する。各ノードは、同一レベル上にある1つのノードへのデータ出力ラインを有する。最下位レベル上にないノードは、下位レベルへの付加的なデータ出力ラインを有する。最下位レベル上にない各ノードは、1つの制御入力ポートを有する。最上位レベルにないノードは、1つ若しくは2つの制御信号出力ポートを有する。1つ若しくは複数の入力列におけるノードだけが、外部の入力源への制御信号出力ラインを有する。入力列は、全体クロック信号によって指示される時間にだけ、且つ最上位レベルのノードから対応するブロッキング信号を受信しない場合にだけ、集信装置にパケットを送信することが可能である。入力デバイスは、内部のセルが下位レベルのセルへと下側にパケットをルーティングするために従っている、パケットを集信装置へと「下側に」送信するための同一タイミング及びルーティング規則を用いる。

10

20

30

40

50

【0062】

当業者は、本明細書に記載された基本的な集信装置に対して多くの変形、修正、付加、及び改良を容易に実行することが可能であろう。

【0063】

代替的なトポロジー実施例

上記で説明された集信装置では、行J上のFIFOから出ているラインが、列0に進入する行J-8上のラインに接続すると考えられている。第1の代替的な例では、行J上のFIFOから出ているラインが、同一行J上の列0に進入するラインに接続される。最上位レベルが32のリングを有し、階層内で下側にある次のレベルが16リングを有し、更に、最下位レベルが8リングを有するように、トポロジーが変更される。

10

【0064】

第2の代替的な例では、列K-1とFIFOの列との間で置換が行われる。行JでFIFOを出るラインが、同一行J上の列0に進入するラインに接続される。いくつかの置換では、できあがったトポロジーがレベルの各々に1つのリングを有する。いくつかの場合では、データがレベル0の列k-1における出力ポートから集信装置を出て行くことが常に可能になっている。そのような場合に、図1に示される斬新な「理髪店ボール」構造は、構造の最上部に進入するメッセージがその構造から、固定定数による制限時間内に確実に退出するように作動する点が効果的である。

【0065】

別の場合に、何らかの条件下にあるデータが、集信装置の出力ポートを出て行くのをブロッ

20

【0066】

サービス品質の実施例

ネットワーク及び集信装置の中にサービス品質(QoS)の優先権を組込むための単純な方法の1つは、サービス品質のレベルを指定する1以上の付加的なヘッダ・ビットを含むことである。図4Aを参照すると、QoS優先権の方法が実装されている。パケットP_DがパケットP_Aと同等若しくはそれを上回る優先権を有し、且つセルWがセルXにパケットP_Dを送信しない場合には、パケットP_DがセルXに送信される。しかしながら、パケットP_DがパケットP_Aより低い優先権を有し、且つセルWがセルXにパケットP_Vを送信しない場合には、パケットP_AがセルXに送信される。QoS優先権の実装のために、セルE及びBがQoSヘッダ・ビットの読み込みが可能であり、更にセルBからセルEへの制御ラインによってQoS情報が運ばれる。セルBは、ライン108の退出を介して、若しくは付加的な制御ラインを介して、セルEにQoS情報を送信することが可能である。

30

【0067】

付加的なレベルの実施例

図示されている例では、ノードからのバッファが除外されている。しかしながら、集信装置はバッファのように機能し、従ってバースト・トラフィックを処理することが可能である。例えば、集信装置に進入するメッセージの平均数は8未満であるが、時折8より多いメッセージが集信装置に進入し、集信装置に進入するようにターゲティングされたメッセージが1つもブロックされない可能性がある。バースト・トラフィックを処理する能力は、付加的なレベル、例えばレベル3を、集信装置に付加することによって増進させることが可能である。レベル3は、最上位の32行にデータが進入する64つの行を有してよい。付加的なレベルを含むことによって、全集信装置の有効バッファ・サイズが増大される。別の実施例では、いくつかの付加的なレベルが構造に付加され、それにより、バースト・トラフィックを処理する能力が更に高められる。

40

【0068】

複数入力列の実施例

64:8集信装置を実装するための一方法には、8巻リボンを用いて、64行を備えている付加的なレベル3を付加することが含まれる。代替的な方法では、付加的な入力ポート端子の列がレベル2に付加され、それによりレベル2内の32行で受信されるデータの量は潜在

50

的に2倍にされる。付加的なシフト・レジスタFIFO列がレベル2に付加され、それにより32ライン上のトラフィックが増大される。入力デバイスから集信装置に適用されたメッセージのタイミングは、付加的なFIFO列が巻かれているかどうかによって依存する。メッセージは、集信装置内のノードから入力列にメッセージが進入する際に、集信装置の外部から入力列に進入する。この代替的な方法は、入力チャンネル当たりの平均データ速度がより低く、又、よりバースト状態である適用において有用である。

【0069】

レベル間の異なる相互接続の実施例

各ノードが単一の制御セルを含んでおり、更に、より下位レベルにおける各ノードが、階層内でより上位レベルの1つのノードだけに接続されている。この集信装置は、特定のタイム・ステップにおいて、最下位レベルの上にある特定レベル内のラインの半分だけが、次の下位レベルへと下側にメッセージを送信するように適合され得る。

10

【0070】

一実施例では、4セルにデータ送信することが可能な4つの制御セルが単一ノードに結び付けられる。当業者は、特定レベルの4つのセルが下位レベル上の2セルにデータを送信することが可能なノードを形成するように、この構造を修正することが可能である。

【0071】

図5を再度参照すると、レベル上の2つの制御セルが、階層内で、次の下位レベル上にある単一の制御セルにデータを送信するように配置されている。ノード502には、1つのノード内の2セルによって、下位レベル上の1つのノードへのデータ送信が可能な2:1ノード集信装置の構造が図示されている。機能的な例では、メッセージM1が、列0、レベル1の行2上の制御セルから、列1、レベル1の行2上のセルに送信される。同じ作動時間周期内に、列0内にある、レベル2の行8上の制御セルにメッセージM2が到達する。メッセージM1は、メッセージM2をブロックするように配置され、メッセージM2は同一レベル上に留まる。それゆえに、図5で示されるトポロジー内で、あるレベルの上にある単一メッセージによって、より上位のレベルの上にあるメッセージのブロックが可能である。

20

【0072】

図6A及び図6Bを参照すると、セルが、例えばノード602内に分類され、相互接続ライン604上の次の下位レベル上の2セルにデータを送信するように、1つのレベルの上に4セルが配置されている。各レベル上のノードは、4つの制御セルを含んでいる。レベル2の上の最も左側にあるノードの4つのセルの各々は、レベル1の上にある2セルのいずれかにデータを送信することが可能である。図6Aに示されている一例では、ノードN内の4セルが、ノードPのセルQ及びRにデータを送信することが可能である。同様にして、ノードM内の4セルは、ノードP内のセルS及びTにデータを送信することが可能である。あるレベルにあるメッセージで、上位レベルにあるメッセージをブロックすることが可能なものは1つも存在せず、それにより、1つのレベルから次の下位レベルへの処理能力が増大され得る。図6Aは、1つのノード内の4セルが下位のレベルの2セルのいずれかにデータを送信することが可能な、4:2ノード集信装置の構造が図示されている。別の実施例では、例えば図6Bに示されているように、ノード610内のセル数を増大させるようにロジックを付加されてよい。例えば、ノードT及びU内の8セルが組合わせて制御され、それにより、8:4ノード集信装置を形成するように、ノードV及びWのいずれか若しくは両方の内部にある4セル全てにデータが送信される。ノード当たりでより多いロジックを費やし処理能力が増大された設計では、より複雑なノードが用いられる。

30

40

【0073】

図7を参照すると、ブロック図によって、いくつかのチップからより大きなネットワークを形成するように複数の集信装置及び複数のスイッチを用いているシステムが図示されている。ある代替実施例では、ねじれ形状の立方体の中により小さなネットワークが配置されている。図7に図示されているシステムによって、ねじれ形状立方体構造が改良されている。

【0074】

50

実際には、図示されている複数チップ・モジュールを利用して、極めて大きなネットワークが構築されてよい。例えば、128のスイッチ・チップ及び128の集信装置チップが組合わされてよい。例えば、128のチップの各々が64のオンライン入力ポートと64の3ライン出力ポートとを含んでいてよい。この組合わせによって、 64^2 の入力ライン及び 64^2 の出力ラインを備えた単一チップが形成される。図7に図示されている例は、図示されている目的に対して有用である、ずっと小さいシステムである。実際的には、より大きなシステムが構築されてよい。

【0075】

図7に示されているネットワーク702及び704は、3つのレベルと複数の列を有している。各レベルは4つの行を有している。レベル0上にある列のうちの3つは、出力ポートを含んでいる。図示されているシステムでは、チップ702の各々は、アドレス0に送られる3つの出力ライン、アドレス1に送られる3つの出力ライン、アドレス2に送られる3つの出力ライン、及びアドレス3に送られる3つの出力ラインを有している。最下部の集信装置は、アドレス0に接続されている全12出力ラインを受取る。3つの別の集信装置は、アドレス1、2、及び3に適切に接続されている12の入力ラインを各々受取る。

10

【0076】

別のネットワークは種々の時間において、異なる列からデータを送信する。図7に示されるシステムでは、異なる列からのメッセージが適切な遅延ラインFIFO（図示せず）を通過しており、それにより全ての列からのメッセージが同時に集信装置700に到達する。

【0077】

20

集信装置700は、レベル0が4行を有し、レベル1が8行を有し、更に、レベル2が16行を有するように、3つのレベルの上に4幅リボンの行を有している。レベル2の上では、16行の下位の12行が、スイッチ702から、集信装置700の入力ポート（図示せず）においてデータを受信するように配置されている。

【0078】

データが非常に混雑していることが予想される適用においては、付加的な層が集信装置700に付加されてよい。集信装置700は、12チャンネル出力ラインからのデータを、第1列内の各チップから第2列内の4つの入力ラインの上に集約させる。この集信装置は更に、ホット・スポットの可能性を減少させるように、時間内にこのデータを広げる、若しくは配信する。集信装置700からのデータは、スイッチ・チップ704の第2列の中に適用される。

30

【0079】

集信装置モジュール700とスイッチ・チップ704との間でFIFOが1つも必要とされないように、メッセージは同期して集信装置700を退出する。メッセージは、スイッチ704から出現して、集信装置モジュール706の第2列の中への進入のために、時間内にメッセージを整列させるようにFIFO（図示せず）を通過する。集信装置チップ706の各々は、4つの集信装置を含んでいる。集信装置チップ706内にある4つの集信装置の各々は、1行の幅を備えたリボンを有する。

【0080】

第2列の集信装置は、レベル0が1行を有し、レベル1が2行を有し、且つレベル2が4行を有する、3つのレベルで設計される。最上位レベルの4行のうちの3行は、入力データを受信することが可能である。同様に、バースト・トラフィックに対しては、集信装置チップ706内で集信装置に付加行が確実に付加される。

40

【0081】

制御ライン（図示せず）によって、上流のチップに下流のチップのデータ・ブロック状況を通知するように、下流のチップから上流のチップに制御信号が供給される。例えば、スイッチ704が集信装置700からデータを受信することが不可能である場合、このデータは集信装置700の最上位の行のうちの1つに転送される。集信装置700の下位の12行だけが上流のスイッチ702からメッセージを受信するので、上位の4行は常に、データ受信を自由に行うことができる。

50

【0082】

最上位のスイッチに戻る制御ラインは、多数の場所から来ていてよい。一実施例では、バッファは集信装置800とスイッチ804との間に配置されてよい。バッファが能力レベルを越えて満たされると、満杯のバッファに供給を行ったスイッチ802の出力ポートに、制御信号が選択的に送信される。或いは、トラフィックが混雑しているときには、列内の上位レベルに配置されたノードから左側に制御信号を出ることが可能であるが、トラフィックが混雑していないときにメッセージを受信しないように、集信装置の内部からの制御信号が発信されてもよい。この制御信号が、集信装置内のトラフィックに基づいて、スイッチ800の特定の出力ポートをブロックしてもよい。

【0083】

10

フィードバックされるメッセージの最大数がリボン幅と等しく、且つリボン幅が集信装置の最上位の行に常に開かれているので、全ての大きさの集信装置に対して、集信装置の最上位行へのブロックされたメッセージのフィードバックは、常に成功する。

【0084】

本発明は、種々の実施例に関して説明がなされたが、これらの実施例は例示的なものであり、本発明の範囲がそれらに制限されるものではないことを理解されたい。説明された実施例に対して多くの変形、修正、付加及び改良が可能である。例えば、当業者によって、本明細書に記載されている構造及び方法を提供するのに必要なステップが容易に実装可能であり、又、処理のパラメータ、材料、及び寸法は例のためだけに与えられたものであり、本発明の範囲内で、修正と同様に所望の構造を得るために変形がなされてもよいことを理解されたい。本発明で説明される実施例の変形及び修正が、本明細書の請求項で示される本発明の範囲及び精神から離れることなしに、本明細書に示された説明に基づいてなされてよい。例えば、当業者によって、第1及び第2の品質サービス方法が、本明細書で説明されたものとは別の相互接続構造に同様に適用されてよい。

20

【0085】

本明細書の請求項では、特に断らない限りは、全てのものは「1以上」を示している。

【図面の簡単な説明】

【図1】 図1には、多重レベル32：8の集信装置の形態であるデータ相互接続構造の一例が示されている。

【図2】 図2は、パケットの形態で相互接続構造経由して運ばれるデータ・フォーマットが示されている。

30

【図3A】 図3Aは、入力ポート及び出力ポートの種々の構成を含む、図1に示される相互接続構造の内部で用いられ得るセルの例が示されている。

【図3B】 図3Bは、入力ポート及び出力ポートの種々の構成を含む、図1に示される相互接続構造の内部で用いられ得るセルの例が示されている。

【図4A】 図4Aは、図1に示された構造等の相互接続構造の内部で用いられるのが適切なセル間の相互接続構造の複数の例が示されている模式的なブロック図である。

【図4B】 図4Bは、図1に示された構造等の相互接続構造の内部で用いられるのが適切なセル間の相互接続構造の複数の例が示されている模式的なブロック図である。

【図4C】 図4Cは、図1に示された構造等の相互接続構造の内部で用いられるのが適切なセル間の相互接続構造の複数の例が示されている模式的なブロック図である。

40

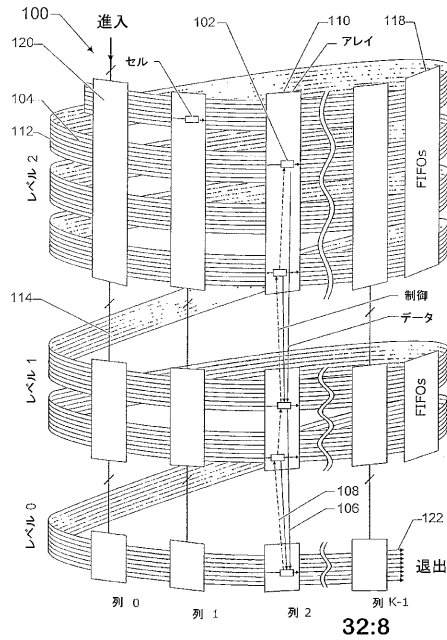
【図5】 図5は、レベルに関して変更可能である、相互接続構造内のセルの優先権が図示されている模式的なブロック図である

【図6A】 図6Aは、相互接続されたセルのノードへの分類の例が示されている模式的なブロック図である。

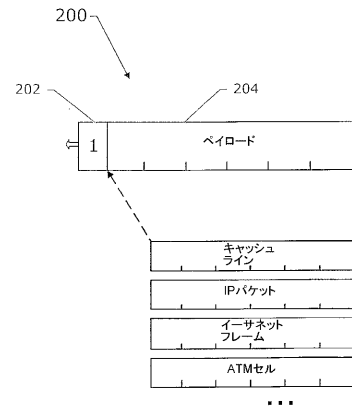
【図6B】 図6Bは、相互接続されたセルのノードへの分類の例が示されている模式的なブロック図である。

【図7】 図7は、複数の集信装置を用いているシステムが図示された模式的なブロック図である。

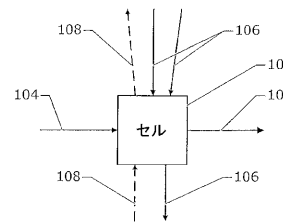
【図 1】



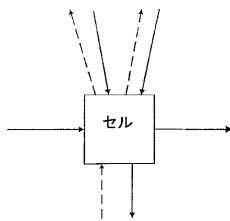
【図 2】



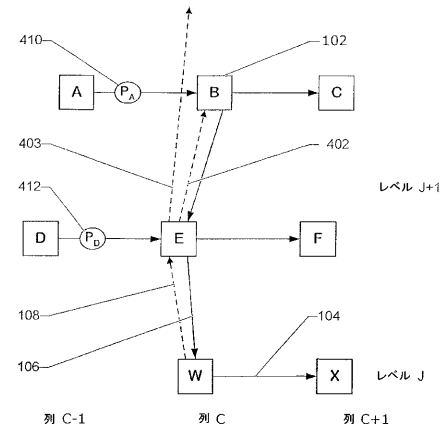
【図 3 A】



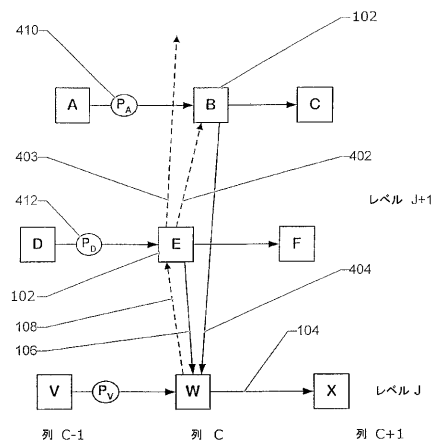
【図 3 B】



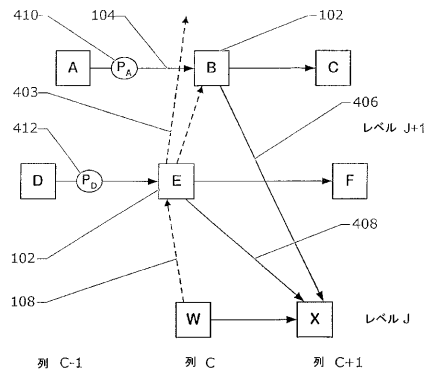
【図 4 B】



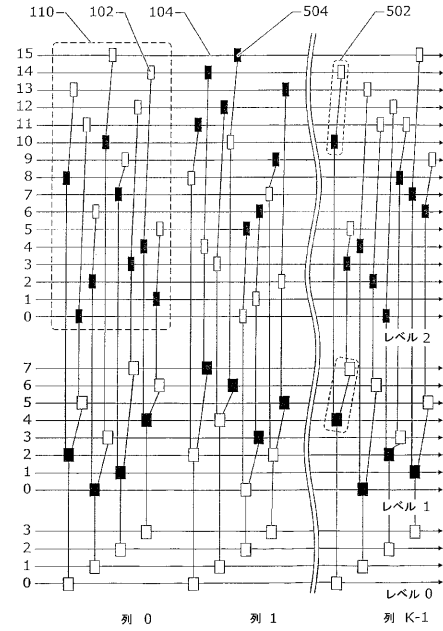
【図 4 A】



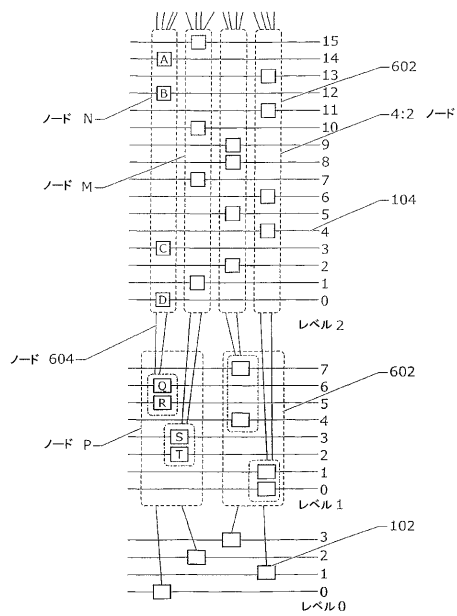
【図 4 C】



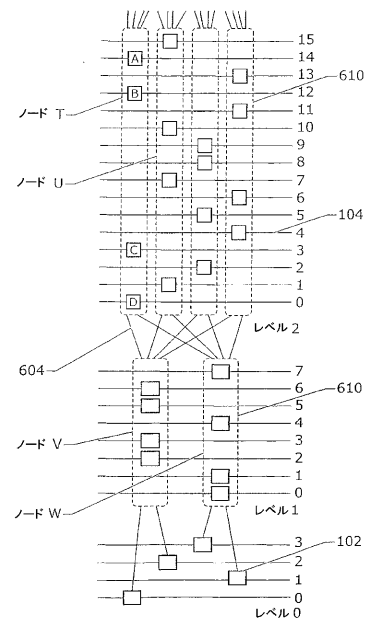
【図 5】



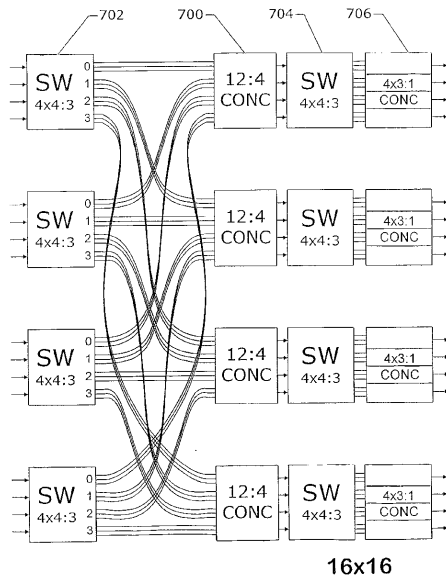
【図 6 A】



【図 6 B】



【図 7】



フロントページの続き

(72)発明者 ヘッセ、ジョン

アメリカ合衆国カリフォルニア州94038・モスビーチ・ステットソンストリート 827

審査官 菊地 陽一

(56)参考文献 特開平11-289341 (JP, A)

特開平08-328981 (JP, A)

米国特許第05649108 (US, A)

(58)調査した分野(Int.Cl., DB名)

H04L 12/56