計算機アーキテクチャ 152−10
（２００３．３．１０）

DIMMnet ネットワーク接続動作検証環境

濱田芳博[†1] 三橋彰浩[†2] 田邊昇[†4]
天野英晴[†5] 中條拓伯[†3]

メモリバスへ接続する PC クラスタ用 NIC として DIMMnet-1 が試作された．この NIC で利用可能な相互接続網は通信リンクに光通信を用いた RHiNET/SW2 を用いたスイッチによるものである．現在進められている DIMMnet-2 の設計においては，通信リンクに銅線による高速シリアル伝送を用い，スイッチとして商用品を用いることで価格性能比の向上が考えられている．また NIC に複数の入出力ポートを設け，ポート間にルーティング機構を付加することで，スイッチを用いずに小規模のクラスタを構成可能にする機能の付加を考えている．これにより PC クラスタの構成が単純化されるため，システムの使いやすさとコストパフォーマンスが向上される．
本論文においては，これらの機能の検証を行なうための DIMMnet-1 を用いたテストベンチの構成と，これを実現するために必要な FPGA を搭載した PCI ボードについて述べる．

The verification environment of a network for the DIMMnet.

Yoshihiro Hamada ,[†1] Akihiro Mituhashi ,[†1] Noboru Tanabe ,[†4]
Hideharu Amano [†5] and Hironori Nakajo[†1]

DIMMnet-1 is a network interface which is plugged into a memory bus. It is utilized for configuring an environment for high performance computing with a PC cluster. In a prototype of DIMMnet with NICs and RHiNET/SW2, optical links are adopted to configure interconnected network for a cluster. Currently, we have been developing a DIMMnet-2. In this work, we are trying to improve cost performance rather than a previous configuration, thus we have been designing an interconnecting network with a copper wire links and have adopted a commodity network switch. In order to construct a small-scaled cluster, we constructs a directly connected network with the NICs and a router board which has several communication ports to connect to the other nodes.
In this paper, we describe the verification environment of a test-bench of these feature, and PCI router board.

1. はじめに

PC において利用可能なプロセッサの動作速度は年々向上している．PC クラスタ環境においては，この性能向上を並列計算能力向上へ反映させるために，ネットワークの遅延や帯域幅を改善する必要がある．
近年通信リンクとしては，光ファイバと光電気変換モジュールを用いた光伝送やワイヤと差動増幅器による高速シリアル伝送により広帯域化しており，10GigaBitEther や InfiniBand[1] 規格に合わせチャネル当たり 3.125，2.5Gbps が利用可能である．またこれらの規格は複数のチャネルを組み合わせて使用することを前提としており，10GigaBNitEther で 3.125 × 4=12.5Gbps，InfiniBand で 2.5 × 12=30Gbps の帯域が利用可能である．
しかし PC において一般的に使用される I/O バス規格は，PCI2.2 規格の 33MHz/32bits 帯域幅 133MB/s が主流であり，帯域幅のボトルネックとなる．サーバ用途として利用される同規格の 66MHz/64bit 帯域幅 528(MB/s)，PCI-X 1.0 133MHz/64bit 帯域幅 1066(MB/s) を利用することでこのボトルネックを軽減することが可能であるが，クラスタを構成する計算ノードの単価を上げることになるので，価格性能比を下げることになる．
DIMMnet-1[4)5)] は，このボトルネックを解消するため，メモリバスに接続する NIC として試作開発された．この NIC の特徴として細粒度通信への適性と広帯域化可能な通信帯域を備える．試作された NIC は現在 2 種類あり，通信リンクとして光ファイバを利用するもの（光版）と，ワイヤを接続するもの（電気版）がある．これらの NIC の内，クラスタとして利用可能なものは RHiNET/SW2[6)] スイッチを利用で

†1 東京農工大学 工学研究科 電子情報工学専攻
Department of Electrical and Computer Engineering,Graduate school of technology,Tokyo University of Agriculture and Technology
†2 東京農工大学 工学研究科 情報コミュニケーション工学専攻
Department of Computer,Information and communication sciences,Graduate school of technology,Tokyo University of Agriculture and Technology
†3 東京農工大学工学部情報コミュニケーション工学科
Department of Computer, Information and Communication Sciences,Faculty of Technology,Tokyo University of Agriculture and Technology
†4 （株）東芝 研究開発センター
TOSHIBA Corporate Research & Development Center
†5 慶應義塾大学理工学部情報工学科
Department of Information and Computer Science,Faculty of Science and Technology,Keio University

きる光版である．

現在 DIMMnet-1 が持つ特徴を引き継いだ DIMMnet-2 の開発行っている．ここにおいては，通信リンクとしてワイヤを用い，スイッチに商用品を用いることで価格性能比を上げることが考えられている．また付加的な機構として，複数の通信ポートを設けることで直接網を構成し，小規模なクラスタを構成可能にすることが上げられており，これにより PC クラスタの構成が単純化されるため，システムの使いやすさとコストパフォーマンスが向上される．本論文においては，これらの接続網の検証を行なうための DIMMnet-1 を用いたテストベンチの構成と，これを実現するために必要な FPGA を搭載した PCI ボードについて述べる．

2. DIMMnet-1

2.1 2 種類の NIC

DIMMnet-1 はメモリバスに接続する PC クラスタ用通信インタフェースであり，使用する通信 LINK の違いにより 2 種類存在する．光ファイバを利用する光版 NIC と，ワイヤを接続する電気版 NIC である．これらの NIC のコントローラチップには，Martini[7] が用いられる．後述する接続網検証環境では，**図 1** へ示す電気版 NIC を利用する．

図 1 電気版 NIC

2.2 Martini の通信機構

Martini は通信機構として RDMA を備えている．この通信機構を利用するためのプリミティブとして PUSH，PULL が組み込まれている．PUSH はローカルノードのメモリブロックをリモートノードのメモリブロックへ転送し（リモートライト），PULL はリモートノードのメモリブロックをローカルノードのメモリブロックへ転送する（リモートリード）．

また主に DIMMnet-1 向けの通信機構として，1〜8(bytes) までのデータをホストプロセッサが PIO により書き込むことでリモートリード／ライトが行える AOTF や，474(bytes) 以下まで書き込むことでリモートリード／ライトが行える BOTF がある．AOTF においてはパケットヘッダを事前に Martini へ登録しておき，送信時に Martini 内部で送信データに付加してパケット送出を行う．これに対し BOTF においてはユーザプロセスがパケット全体を NIC へ書き込む必要がある．

3. 相互結合網テストベンチ

3.1 PCI ボード構成

3.1.1 構　　成

テストベンチを構成するために必要な PCI ボードは**表 1** へ示す部品で**図 2** の様に構成される．PCI ボードでは電気版 NIC に対する入出力ポート C1，C2 とボード間の入出力ポート P1，P2，P3 を FPGA に接続しており，FPGA には直接網による接続環境と商用スイッチ検証環境の構成のために適当な論理を含ませる．各入出力ポートの構成について以下へ示す．各環境の概要については，3.2 節，3.3 節へ示し，これらの詳細については 4，5 章へ示す．

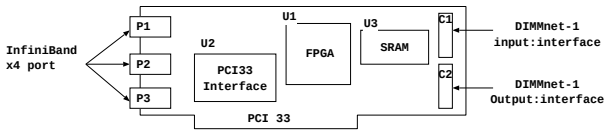

図 2 テストベンチ用 PCI ボード構成

表 1 主要部品

部品番号	部品名	備考
U1	XC2VP40	Xilinx 製 FPGA FF-1152 パッケージ
U2	QL5064	QuickLogic 製 PCI Controller
P1,P2,P3	FCN-268D008-G/2*	富士通コンポーネント製 ソケット (マニュアル実装対応/圧入ポスト付き
C1,C2	MDR10226	3M 製リセプタクル（ストレート型）
U3	SRAM	64bit 幅の同期 SRAM
CableA	FCN-260(4X)	富士通コンポーネント製 オプションでイコライザ可（7m 以下であればイコライザ無しでも伝送可能）
CableB	MDR26P	3M 製部品／ケーブル接続

3.1.2 ボード間入出力ポート

PCI ボード間のリンクは FPGA に搭載されるシリアル伝送 IC を Infiniband で規定される電気レベルで使用する．これは，1 つのモジュールで送受信双方向の接続を行うことが可能である．表 1 へ示す FPGA 内には複数のシリアル伝送 IC が含まれており，これらはチャネルボンドと呼ばれる機能により同調させて動作させることが可能である．これにより InfiniBand における× 1/4/12 の規格に対応可能であり，本 PCI ボードでは× 4 の規格を用いる．

図 3 に送信部，**図 4** に受信部の概略を示す．図中では伝送レートに 2.5(GHz) を使用するため，送信モジュールには外部より 125(MHz) のクロックを入力している．このクロックは送信部内で 20 倍され，伝送レートとして使用される．伝送符号には 8B/10B を用いるため，1bytes(8bits) は 10bits に変換される．送信モジュールへ 4(bytes) の送信データを与える形で用いるとすれば，FPGA の内部論理では 2500(MBits/s)/40(Bits)=62.5(MHz) の周期でデータを与えれば連続してシリアル伝送が行える．受信側におけるデータ受信のタイミングは，8B/10B 符号による自己同期により行われる．取り込まれたデータ

は，シリアル→パラレル変換され ElasticBuffer を介して FPGA 内部のクロックと同期される．

リンク間のケーブルについては表 1 へ示す CableA を用いる．× 4 の規格では双方向の差動信号を用いるため，このケーブルには 8 対のペア線が含まれる．高速シリアル伝送においてはケーブルによるローパスフィルタ効果を解消するため，ケーブル出力側で高周波成分を強調するイコライジングと呼ばれる処理を施すことがある．CableA においてもイコライザをオプションとして付加することが可能であり，ケーブル長が 7m 以上では必須としている．

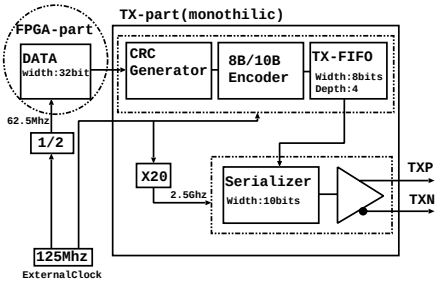

図 3　送信部

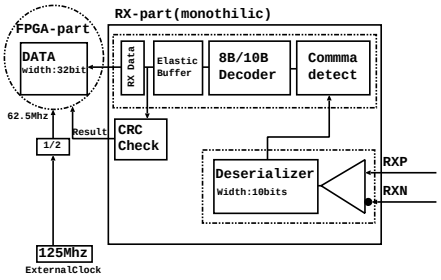

図 4　受信部

3.1.3　PCI ボード-DIMMnet-1 間入出力ポート

PCI ボード-DIMMnet-1 間リンクは FPGA 内に入出力の論理を作成する必要がある．NIC においては通信リンクを伝送クロックに同期した 10(bits) パラレル伝送として実現している．電気レベルは LVDS である．このポートの伝送レートは NIC 側の設定によりいくつか存在する．この設定は SWIF モードと呼ばれ，NIC において利用可能な SWIF モードと伝送クロックを**表 2** へまとめる．これより伝送クロック 400(MHz) の SW2(A) モードでは，PCI ボード側での FPGA による受信は動作クロックの高さより難しいため，125(MHz) の SW2(B) と OIP モードを用いる．表における frame 長と frame 周期は 1 つの転送データの長さと入力間隔を表す．これより，FPGA では SW2 モードでは 31.25(MHz) 毎に 1 データの受信後処理を行い，OIP モードでは 62.5(MHz) 毎に 1 データか，31.25(MHz) 毎に 2 データ分受信後 1 処理を行えれば良いことになる．

リンク間のケーブルについては表 1 へ示す CableB を用いる．このインタフェースでは 13 組の差動信号を用いているため，ケーブル内にも同数のペア線が含まれる．ケーブル長は 40cm である．

表 2　電気版 SWIF モードと伝送クロック

SWIF モード	SW2(A)	SW2(B)	OIP
供給クロック (MHz)	200	62.5	250
伝送クロック (MHz)	400	125	125
frame 長 (bits)	80	80	40
frame 周期 (MHz)	100	31.25	62.5

3.2　直接網による接続環境

直接網による接続環境については**図 5** へ示す．これより各 PC 内部では図 1 へ示す電気版 NIC と PCI ボードが表 1 へ示す CableB で接続される．各 PC 間では PCI ボード同士が表 1 へ示す CableA により接続される．トポロジには双方向のリング網を用いる．この場合 PCI ボードはルータとして振る舞う．

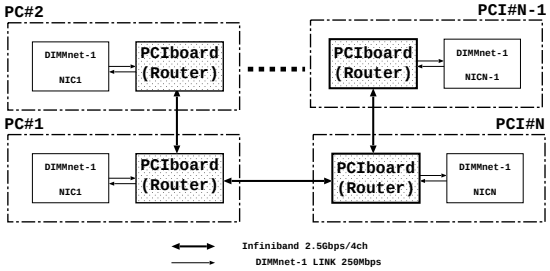

図 5　直接網による接続環境

3.3　商用スイッチ検証環境

商用スイッチの検証環境については**図 6** へ示す．これより，各 PC 内部においては 3.2 節と同様に電気版 NIC と PCI ボードが接続される．各 PC 間には商用スイッチが存在し，PCI ボードと表 1 へ示す CableA により接続される．商用スイッチは InfiniBand 用のものであり，**表 3** へ示す Paceline[3] 社のものが利用可能である．この場合 PCI ボードはパケットコンバータとして振る舞う．

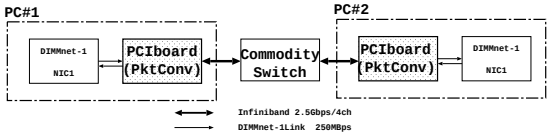

図 6　商用スイッチ検証環境

表 3　InfiniBand 商用スイッチ

	paceline4100
Port-to-Port Latency	160ns
入出力ポート	4X ポート 8 個
MTU	2,048(2K)max
Linear forwarding table	4K
Multicast table size	128

4.　直接網 (リング網) による接続環境詳細

4.1　概　　要

DIMMnet を直接網により接続する環境においては，トポロジとしてはリング網を用いる．この場合，PCI ボード（ルータ）へは NIC や隣接ルータからパケットが入力される．ルータは入力されたパケット中のルーティングヘッダを解釈し，自身宛のパケットである場

合は NIC へパケットを出力し，これ以外は隣接ルータへパケットを出力する．

4.2 FPGA への信号接続

各入出力ポートの FPGA への接続を**図 7** へ示す．NIC との接続は 10 本のデータ線と 2 本の制御線，1 本の伝送クロック線が入出力各々存在する．隣接ルータとの接続は，InfiniBand × 4 のポート 2 組を用いて双方向のリング網を構成するために，各ポートの出力側については 3 本の出力線をデータ線とし，1 本の入力線をルータ間のフロー制御線として用い，入力側については 3 本の入力線をデータ線とし，1 本の出力線をフロー制御線に用いる．

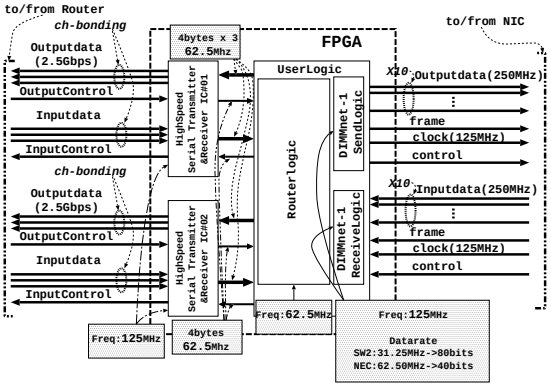

図 7 FPGA への入出力ポート信号接続

4.3 FPGA の論理構成

4.3.1 基 本 構 成

リング網を構成する場合の内部論理は図 8 の様になる．ルータ中各ポートについては 4 つの仮想チャネル[9] を持つものとし，2 つで 1 つの仮想ネットワークとして取り扱う．ルータ間のパケットルーティングはワームホールルーティング[8] により行うものとし，フロー制御についてはクレジットベースにより行う．ルータ-NIC 間でのフロー制御は I/O 線と slackbuffer を用いた stop-and-go によるフロー制御を行う．ルータ間の伝送経路は BitErrorRate が SW2 が持つ 10^{-20} と比較して低下すると考えられるため，伝送データの検査には CRC を用いエラー検出時には再送を行う．ルータ-NIC 間のデータ検査は ECC により行われるので，NIC からの入力時にはルータ側で ECC デコードを行い，NIC への出力時にはエンコードを行う．

4.3.2 SW2 モードにおけるユニキャスト/マルチキャスト

SW2 モード時のパケットヘッダにおけるルーティング情報を**図 9** A) へ示す．このモードではテーブルルーティングが用いらる．スイッチ中においては，パケットの RRID をルーティングテーブルの Index としてパケットの出力ポートを取得しルーティングを行う．送信先 NIC へパケットが到着した時点でこのパケットに対する返答が必要な場合，返送先のルーティング情報にはパケット中の IRID を使用し，この内容を RRID へコピーした後に返答パケットを出力する．

本ルータにおいては NIC が SW2 モードの場合，ルーティング情報を図 9B) の様に用いる．ユニキャストにおいては RRouterID をパケット送信先のルータの ID として扱う，これは PCI ボード毎設定される一意な ID である．R L/R は双方向リングにおいてパケットを左回り，右回りどちらのリングを用いてルーティングを行うかを決定するフラグである．送信先のルータへパケットがルーティングされた場合，ルータは NIC へこれを出力する．パケットの返送路の情報としては IRouterID と I L/R ビットが NIC 内部で RRoutingID と I L/R ビットへコピーされ使用される．

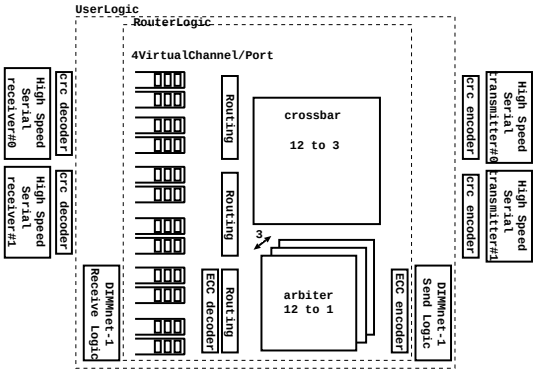

図 8 FPGA 内部論理

マルチキャストは R MultiCast ビットが 1 の時に行われる．この場合，RRouterID は各ルータにおけるマルチキャストテーブルの参照 ID として用いられる．このテーブルは 2bit で構成され，下位ビットが MulticastEnable ビットで，このビットが 1 の場合パケットは NIC へ出力される．また上位ビットは PassEnable ビットで，このビットが 1 の場合パケットは次のルータへルーティングされる．

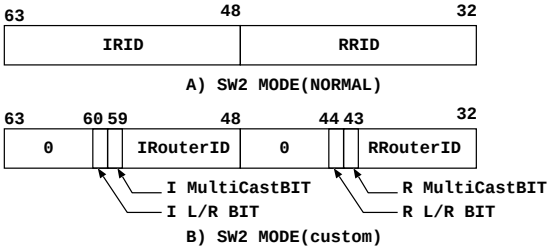

図 9 SW2 モード時ルーティングヘッダ

4.3.3 OIP モードにおけるユニキャスト/マルチキャスト

OIP モード時のパケットヘッダにおけるルーティング情報を**図 10** A) へ示す．このモードの場合ルーティングは rt0〜rt5 で示されるポートへ順次パケットを出力していくソースルーティングで行われる．このモードの場合送信パケットに対する返答のための，返送路のパケットヘッダへの設定はスイッチ（ルータ）が行う様になっている．

本ルータにおいては NIC が OIP モードの場合，ルーティング情報を図 10 B) の様に用いる．ユニキャストにおいては SW2 モードの場合と同様にルーティングされるが，返送路のための IRouterID と I L/R ビットの RRouterID と I L/R ビットへのコピーは送信先ルータがパケットを NIC へ出力する直前に行う．

マルチキャストにおいては，SW2 モードと同様にルーティングされる．

4.3.4 デッドロック回避

SW2 が保証しているネットワークの FIFO 性を保証してデッドロック回避が行えるルーティング方法に

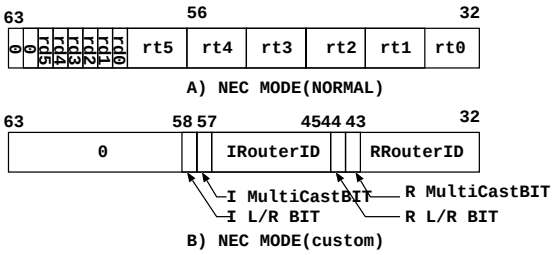

図 10 OIP モード時ルーティングヘッダ

は同スイッチが使用している構造化バッファ/チャネル法[10] があるが，この場合パケットがリンク間を通過する度に 1 つチャネルを上げなければならず，デッドロック回避のために結合網直径+1 の仮想チャネルを用意する必要がある．リング網では結合網直径が接続台数の増加によりリニアに増加するためこのルーティング方法は実用的ではないと考える．そこで極力単純なルーティング方式として 2 つの仮想チャネルを使用して，閉ループを描かない一本のルーティング経路を使用するものとする．リング網においてチャネルを上げる箇所は 1 つだけ存在し，これは PCI ボード上で設定可能にする．

5. 商用スイッチ検証環境詳細

5.1 概　　要

商用スイッチの検証においては，PCI ボードは DIMMnet-1 から入力されたパケットを InfiniBand のパケットへ変換し，スイッチへこれを出力する．スイッチからの入力においてはこの逆操作を行う．

5.2 InfiniBand ネットワーク

5.2.1 ネットワークの規模

InfiniBand によるネットワークの最小単位はサブネットであり，48K-1 のユニキャストアドレスと 16K-1 のマルチキャストアドレスを識別可能である．サブネットはグローバルネットワークの配下に接続され，ルータを介して接続される．グローバルネットワークのアドレスは 128bits あり，上位 64bits に Subnet ID と呼ばれるサブネットの識別子と，下位 64bits に GUID と呼ばれるグローバルネットの識別子を持つ[2]．

DIMMnet が使用するシステムエリアネットワーク (SAN) をこれにより構成する場合にはネットワークの規模はサブネット 1 つで十分である．使用する InfiniBand スイッチ (Paceline4100)[3] においては，表 3 よりユニキャストアドレスは 4K，マルチキャストアドレスは 128 個をそれぞれ識別可能である．

5.2.2 ルーティング

サブネットにおけるルーティングは，スイッチ中において InfiniBand パケットヘッダ中の LocalRouteHeader(LRH) で指定される DestinationLocalID(DLID) を用いて ForwardingTable からパケットを出力するポート番号を得ることによりルーティングされる．ForwardingTable は前述の様に 2 種類あるが，DLID の値によって区別され，0001h〜BFFFh がユニキャストであり，C000h〜FFFEh がマルチキャストとなる．

ユニキャスト用の ForwardingTable の構成は，InfiniBand スイッチ (Paceline4100) においては Linear Forwarding Table(LFT) で実装されているため，DLID を LFT の index として用いて該当するエントリよりパケットを出力するポート番号を得ることになる．

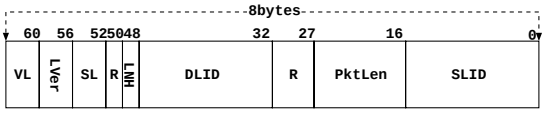

図 11 LRH フォーマット

5.3 FPGA への信号接続

FPGA への信号接続は図 7 とほぼ変わりなく，DIMMnet との接続は同様に 10 本のデータ線と 2 本の制御線，1 本の伝送クロック線が入出力各々存在する．しかしスイッチとの接続は異なり，InfiniBand × 4 のポートを各ポートの出力側について 4 本の出力線全てをデータ線とする，入力側についても 4 本の入力線をデータ線とする．使用するポート数は選択可能とし，ポート数が 1 個の場合は× 4 規格であり，3 個の場合は× 12 の規格となる．

5.4 FPGA に必要な論理

5.4.1 フロー制御

スイッチとのフロー制御は，InfiniBand で規定される LinkLevelPackets フォーマットを用いたクレジットベースのフロー制御により行う．フロー制御の対象は，VirtualLane(VL) と呼ばれる仮想チャネル毎に行う．VL についてはデータ送受信用に 4 組備え，後述する ForwardingTable 等のメンテナンス用に送受信 1 組備える．Infiniband 規格より，メンテナンス用 VL についてはフロー制御は行わない．

5.4.2 パケットフォーマット変換

DIMMnet → InfiniBand のパケットフォーマット変換は，DIMMnet のパケットの前処理を行った後，これを LRH と CRC(32bits) で挟み込むことで行う．パケットの前処理は**図 12** に示すパケットの各ラインについて ECC によるビットエラー検出と訂正を行い，ラインデータのみ取り出す操作である．逆変換は InfiniBand パケットから LRH と CRC(32bits) を取りのぞき，パケットの後処理を行う．これはパケット中に含まれるラインデータにラインタイプを付加し，これらのデータに対して ECC による符号化を行うことで，図 12 に示す DIMMnet のパケットラインへ復元する操作である．

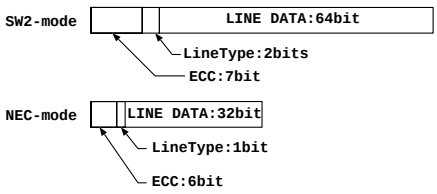

図 12 DIMMnet パケットラインのフォーマット

5.4.3 ForwardingTable

InfiniBand スイッチ (Paceline4100) では，ForwardingTable の管理を行う SubnetManager(SM) を実装している．SM は ForwardingTable の構成のため，SubnetMagement request Packet(SMP) を接続されているデバイスへ送信する．このため，PCI ボードで SMP に対する返答パケットを作成・返送する必要がある．また，この返送パケット中に含まれる値により ForwardingTable が構成されることになるが，この値には，直接網による接続環境で使用した PCI ボード毎に設定される一意な ID を使用する．

5.4.4 再 送 処 理

InfiniBand パケットのエラー検出は CRC により行われ，送信先でのエラー訂正は行われない．このため

エラーとして検出されたパケットについては再送処理を行う必要がある．また CRC の計算にはパケット全体が必要であるため，スイッチ→ PCI ボード→電気版 NIC へのパケットの入力は Store&Forward で動作することになる．これらの間の入出力を極力パイプライン的に動作させるためには，パケットのサイズを小さくする必要がある．このために，送信側 PCI ボードでは１つの DIMMnet パケットを小さなパケットに分割する．これをフレームと呼ぶ．フレームの大きさについては，本テストベンチを用いた実験により決定する．パケットの再送処理については，フレーム単位で行う．

フレームの再送手順については以下に示す．送信側においては送信 VL 数の再送用バッファを備え各 VL 毎にシーケンス番号，ACK を受信したフレーム番号の保持を行う．受信側においては受信 VL 毎にシーケンス番号と，連続受信に成功した回数の保持を行う．

- 送信側状態
  - IDLE 状態
  - 電気版 NIC からのパケット受信状態
    - フレームにシーケンス番号を付加し，これを送信すると共にバッファリングを行う．
    - IDLE 状態へ復帰
  - 受信側よりの ACK 受信状態
    - ACK パケットより，フレーム番号の保持を行う．
    - 受け取ったシーケンス番号以下のフレームについてバッファより削除を行う．
    - IDLE 状態へ復帰
  - 受信側よりの NACK 受信状態
    - ACK を受信しているフレーム番号へ+1 したフレームから，受け取ったシーケンス番号に該当するフレームまでの再送を行う．
    - IDLE 状態へ復帰
- 受信側状態
  - IDLE 状態
  - 送信側よりの正常フレーム受信状態
    - シーケンス番号の検査を行いこれが連続している場合には，正常なフレームとして扱い，連続受信成功レジスタのインクリメントを行う．この値が規定値を越える場合には，送信側へ ACK と共に現在受信したフレームのシーケンス番号の送信を行う．
    - シーケンス番号の検査を行いこれが不連続である場合には，異常フレームとして扱い，この廃棄を行う．送信側へは NACK と共に最後に受信した正常フレームのシーケンス番号にインクリメントした値を送信する．
    - IDLE 状態へ復帰
  - 送信側よりの異常フレーム受信状態
    - フレームの廃棄を行う
    - IDLE 状態へ復帰

6. おわりに

本論文では，今後の DIMMnet-2 の開発に向けて DIMMnet-1 を利用した 2 種類の相互接続網検証用テストベンチとして，直接網 (リング網) による接続環境と商用スイッチ検証環境について述べた．また，これを構成するために必要な PCI ボードと，各環境における PCI ボード上の FPGA 実装する論理の概要について述べた．

ここで述べた PCI ボードはまだ回路の作成途中にある，今後は PCI ボードの完成を優先し，環境の構成と評価を行う予定である．

謝辞

本研究は総務省戦略的情報通信研究開発制度の一環として行われたものである．DIMMnet-1 に関しては新情報処理開発機構が推進してきた RWC(Real World Computing) プロジェクトの並列分散コンピューティング技術研究の一環として開発されたものである．

産業技術研究所の工藤氏，　日立製作所の山本氏，西氏，慶應義塾大学の渡辺氏，元・慶應義塾大学の土屋氏，元・東京農工大学の須田氏，　日立 IT の今城氏，上嶋氏，金野氏，寺川氏，慶光院氏，岩田氏，山本氏，柏原氏，大杉氏，をはじめ MartiniLSI および，DIMMnet-1 の開発に携わった全ての方々に感謝いたします．

また，DIMMnet-2 の開発に関する御議論に御参加頂いている，和歌山大学の国枝教授，上原講師，齋藤講師，慶應義塾大学の大塚氏，伊豆氏，北村氏に感謝します．

参 考 文 献

1) http://www.infinibandta.org/ibta/
2) Tom Shanley,"InfiniBand Network Architecture",MindShare,Inc.
3) http://www.paceline.com
4) N.Tanabe, J.Yamamoto, H.Nishi, T.Kudoh, Y.Hamada, H.Nakajo, H.Amano, "MEMO-net:Network interface plugged into a memory slot.", In CLUSTER2000
5) N.Tanabe, J.Yamamoto, H.Nishi, T.Kudoh, Y.Hamada, H.Nakajo, H.Amano,"On-the-fly Sending:A Low Latency High Bandwidth Message Transfer Mechanism .", In I-SPAN2000,I-SPAN2000,2000
6) 西，他 LASN 用 8Gbps/port 8x8One-chip スイッチ:RHiNET-2/SW，JSPP2000 pp173-180，(May 2000)
7) J.Yamamoto, N.Tanabe, H.Nishi, J.Tuchiya, K.Watanabe, T.Kudoh, H.Amano, "Martini:An ASIC of network interface for high speed network with flexibility.", Japan, 1999
8) W.J.Dally and C.L.Seitz: Deadlock-free message routing in multiprocessor interconnection networks.IEEE Transactions on Computers, 36, 5, pp.547-553(1987).
9) W.J.Dally, "Virtual-Channel Flow Control.", IEEE Trans on parallel and Destributed Systems, Vol.3, No.2, 1992.
10) M.P.Merlin , J.P.Schweitzer, "Deadlock Avoidance in Store-and-Forward Networks-1:Store and Forward Deadlock.", IEEE Trans.on Comm., Vol.COM-28, No.3, pp.345-354, 1980.