

高速性と柔軟性を併せ持つネットワークインタフェース用チップ: Martini

山本 淳二[†] 田邊 昇[†] 西 宏章[†]
土屋 潤一郎^{††} 渡辺 幸之介^{††} 今城 英樹^{†††}
上嶋 利明^{†††} 金野 英俊^{†††} 寺川 博昭^{†††}
慶光院 利映^{†††} 工藤 知宏[†] 天野 英晴^{††}

我々は、フロア内やビル内に設置された計算機間で低レイテンシで高バンド幅な通信を実現するネットワーク RHINET および MEMOnet の開発を行なっている。本報告では、RHINET および MEMOnet のネットワークインタフェースのコントローラチップである Martini について述べる。Martini は、ユーザレベルのゼロコピー通信 (OS をバイパスしたユーザプロセスのメモリ空間間のリモート DMA) をハードウェアによりサポートすると共に、チップ内部のコアプロセッサにより柔軟な通信を実現する。

Martini: An ASIC of network interface for high speed network with flexibility

JUNJI YAMAMOTO,[†] NOBORU TANABE,[†] HIROAKI NISHI,[†]
JUN'ICHIRO TUCHIYA,^{††} KOUNOSUKE WATANABE,^{††}
HIDEKI IMASHIRO,^{†††} TOSHIAKI UEJIMA,^{†††} HIDETOSHI KINNO,^{†††}
HIROAKI TERAKAWA,^{†††} TOSHITERU KEIKOUIN,^{†††}
TOMOHIRO KUDOH[†] and HIDEHARU AMANO^{††}

We are developing low latency high bandwidth networks called RHINET and MEMOnet which can connect computers distributed in one or more floors of building. In this report, a controller ASIC of the network interfaces for both of the networks called "Martini" is introduced. Martini supports OS-bypassed user-level zero-copy remote DMA operations by hardwired logic. The on-chip core processor supports various complicated communication functions such as lock and barrier operations.

1. はじめに

近年、高性能 PC を多数用いて並列処理を行なういわゆるクラスタコンピューティングが注目されている。高速な通信を提供するクラスタでは、低レイテンシで高バンド幅な通信を実現するために、OS をバイパスしユーザプロセスが直接ネットワークインタフェースにアクセスして通信を行なうユーザレベル通信、およびユーザプロセスのメモリ空間をネットワークインタフェースが直接アクセスすることにより、ホストプロセッサによるメモリコピーを行わないゼロコピー通信が広く用いられている。これらの機能は、ネットワークインタフェースがサポートする必要があり、ネット

ワークインタフェースの機能や速度によって通信性能が左右される。

Martini は、これらの機能を 1 チップで実現するネットワークインタフェースコントローラチップである。レイテンシとバンド幅が要求される単純なデータ転送はハードウェアのみによりサポートし、ロックやバリア、同期通信 (isend/ireceive) などの機能はチップ内に実装されたコアプロセッサにより実現する。モジュール単位のパイプライン化と代行機能により、コアプロセッサは、ハードウェアの一部を動作させながら、処理に介入することが可能であり、柔軟なソフトウェア/ハードウェア処理分担が可能となる。

Martini は、種々のホストインタフェースとネットワークインタフェースを持つ。ホストインタフェースとしては、我々が開発している RHINET および DIMMnet のネットワークインタフェースカードとして用いるため、64bit/66MHz の PCI バス (RHINET 用) および DIMM スロット (DIMMnet 用) 用のインタフェースを装備する。さらに、Martini は RHINET のネットワー

[†] 新情報処理開発機構

Real World Computing Partnership

^{††} 慶應義塾大学

Keio University

^{†††} (株) 日立インフォメーションテクノロジー

Hitachi Information Technology

クスイッチである RHiNET-2/SW または RHiNET-3/SW に 8Gbps または 10Gbps のバンド幅を持つ光インタコネクションモジュールを介して接続される。近距離、低レイテンシ用には、2.5Gbps の電氣的なネットワーク用インタフェースも用意されている。本稿では Martini の構成と予備評価について報告する。

2. RHiNET と DIMMnet-1

最近、かなり高い性能を持つ PC が、ビル内やフロア内に分散して配置されているが、これらは常時稼働しているわけではない。これらの計算機を高速低遅延なネットワークで接続し、余剰計算能力を活用することができれば専用の PC クラスタシステムと同等の能力を得ることができる。我々は、このような計算機群を接続するためのネットワークとして PCI に装着する RHiNET⁴⁾ およびメモリスロットに装着する DIMMnet³⁾ を開発している。100m を越す距離の計算機を接続するため、両者共に高速光ネットワークを用いる。このためのスイッチとして、8Gbps のリンクを用い window ベースのフローコントロールを持つ RHiNET-2/SW⁵⁾ の開発に成功し、現在 10Gbps のリンクを用い credit ベースのフローコントロールを持つ RHiNET-3/SW⁶⁾ を開発し、さらに高い信頼性と長距離転送を目指している。この光ネットワークに対応するインタフェースは、CPLD を用いたボード上に実装してきたが、転送速度の向上と高機能化に追従することが困難となった。現在、Myrinet¹⁾ 用のネットワークインタフェース、Quadrics 社の QsNET²⁾ など PCI に接続するネットワークインタフェースは商用化されているが、性能、機能の点で、RHiNET、DIMMnet の両者の要求を満足することは困難である。そこで、この両者を目的とした新しい構成のネットワークインタフェースチップとして Martini を提案し、開発を行うこととした。

2.1 RHiNET ネットワークインタフェース

RHiNET ネットワークインタフェースは、PCI バスに装着されるネットワークインタフェースである。ユーザプロセスが OS を介さずに直接ネットワークインタフェースにアクセスして通信を起動するユーザレベル通信、ネットワークインタフェースがユーザプロセスのメモリ空間に直接データを読み書きするゼロコピー通信を実現する。ユーザレベル通信を実現するために、NIC 内部にユーザプロセスの仮想アドレスから物理アドレスへのアドレス変換を行なうための TLB を持つ。TLB に格納される属性によりプロセス間の保護を実現する。ゼロコピーを行なう際には、通信に関わる領域はあらかじめピンダウンされて、メモリに割り付けられていることが保証されている必要がある。アプリケーション開発は message passing (MPI) ベース、共有メモリ (Open/MP) ベースの両方のスタ

イルで記述でき、専用の LINUX のライブラリおよびデバイスドライバを用いる。

これらの機能を実現するため、RHiNET では、プリミティブを、window と呼ばれるメモリ領域に起動に必要なデータを書き込むことにより起動する。window はホストプロセッサの仮想メモリシステムにより、その window を使用できるユーザプロセスのみが書き込むことができるように保護されている。まずユーザプロセスが送信データとパケットヘッダ情報を window に書き込む。さらに window 中の特定のアドレス (キックアドレス) に対しての書き込みを行なうとパケットが生成され通信が起動される。パケットには、その window にアソシエイトされた PGID が埋め込まれ、これにより、ユーザプロセスが変造できないプロテクション情報が付加される。

RHiNET ネットワークインタフェースの接続される PCI バスは連続転送時に高性能が得られるよう設計されている。そこで、RHiNET ではネットワークインタフェースがホストメモリから連続転送、すなわち DMA による処理を基本としている。また、ホストが処理の終了の確認で PCI デバイスに対するポーリングを行わないよう、ネットワークインタフェースがホストメモリ上の指定された変数に終了ステータスを書き込むよう設計している。Martini では、リモートライトである PUSH とリモートリードである PULL を DMA による基本機能として提供する。これらの基本プリミティブは TLB 参照等を含めて完全にハードウェア制御で時刻する必要がある。

以上のように RHiNET では DMA を基本とするため、Martini は強力な DMA 転送機能を持つ必要がある。しかし一方で、少量のデータの送信の場合は、ネットワークインタフェースが改めて PCI トランザクションを起動する DMA よりもホストが直接データを書き込む PIO による通信が適している。これらの要求を満足するため、Martini は PIO によるデータ転送を直接転送する BOTF(Block On The Fly) 機構を設ける必要がある。BOTF は、push などの通信のプリミティブを書き込むものと同様の window に対する書き込みによりパケットを生成し、送信する。

2.2 DIMMnet-1 ネットワークインタフェース

PCI バスは 64bit/66MHz で用いても、3Gbps 程度の実効バンド幅しか確保できない。また、PCI バスのレイテンシは大きく、RHiNET-2/SW や RHiNET-3/SW のように遅延の小さなスイッチを用いてネットワークを構築しても、PCI バスのレイテンシによって通信のレイテンシが大きくなってしまふ。

メモリスロットにネットワークインタフェースを装着すれば、これらの問題を解決することができる。このような、メモリスロットに装着されるネットワークインタフェースのクラスを MEMOnet と呼ぶ³⁾。DIMMnet-1 は、DIMM ソケットに装着されるネットワークイン

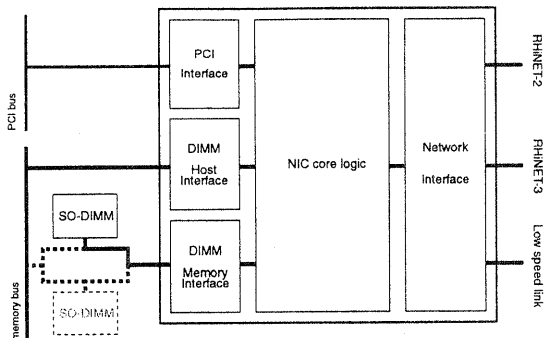


図 1 Martini ブロック図

タフェースである。DIMMnet-1は、RHiNET ネットワークインタフェース同様のユーザレベル/ゼロコピー通信のサポートに加え、メモリスロットに装着されることによるレイテンシの小ささを最大限に利用するために、ネットワークインタフェースへの単一の書き込みでメッセージを送信できる AOTF(Atomic On The Fly) などの通信機構を持つ。

3. Martini の構成

3.1 Martini の全体構成

多くの Network Interface 同様、Martini は図 1 に示すように、外部とのインタフェースを行うインタフェース部と転送制御を行うコア部により構成される。

しかし、Martini は、他の Network Interface に比べて、様々なインタフェースを持ち、しかもこれらの間で高速かつ多機能な転送を行う必要がある。この要求を満足するため、以下の特徴を持つ。

- モジュール間の転送操作を可能な限り標準化し、モジュール間で独立かつ任意の組み合わせでデータ転送を可能にする。
- 制御部は、細かいモジュールに分割してパイプライン化し、それぞれのモジュール単位の処理を、コア CPU が代行可能とする。代行しているモジュール以外のモジュールはコア CPU と同時動作が可能である。
- コア CPU は、ハードウェアモジュールの状態遷移を制御するレベルまでハードウェア処理に密着した制御が可能である。
- 4 系統の独立したチャネルを全転送に対して設け、基本的にバッファを予約してから要求を発生する方針を取ることで、デッドロックを防ぐ。

3.2 コア部

Martini は、ハードウェア制御コアと、コア CPU が共同して制御を行う。コア CPU は、TLB ミス等の例外処理や複雑なプリミティブの単独実行の他、部分代行機能により、必要に応じて様々なレベルでのハードウェア/ソフトウェア共同作業が可能である。

また、コア CPU は、一部の資源には、ハードウェア制御部を動かしつつアクセス可能であり、ハードウェアとソフトウェアの並行処理を行うことも可能である。

3.2.1 コア CPU

コア CPU は R3000 と命令レベルで互換性のある独自設計の 32bit RISC であり、ハードウェア制御用にバスを 64bit に拡張してある。キャッシュは持たず、チップ内部のローカルメモリに命令、データ全てを格納して動作する。基本的な 5 段パイプラインから構成され、浮動小数点演算機能を持たない。

3.2.2 ハードウェア制御コア

ハードウェア制御コアは、コア CPU の介在なしで、基本プリミティブである push/pull を独立に実行可能である。push/pull はそれ自体は単純な処理だが、任意のサイズについて可能にするためには、データのアラインメントを行い、必要な大きさの packets 転送要求を発生し、ページ境界においてテーブル参照のやり直しを行うなど、かなり複雑な処理が要求される。また BOTF 転送をサポートする必要もある。そこで、ハードウェア制御コアを要求を発生する要求発生部と、要求に基づき DMA 転送を行う DMA 転送部に分離し、パイプライン処理を行う。

前者は、複雑な制御回路を持った複数のモジュールから構成し、これらがパイプライン的に動作し、モジュール単位でコア CPU が代行することを可能とする。コア CPU による代行処理は、モジュールからの要求によっても可能であるし、コア CPU の方から積極的に代行要求を出すことも可能である。

一方、DMA 転送部の目的は、高速大容量の転送であり、標準化されたインタフェース間で、独立に複数経路での DMA 転送が可能な構成になっている。この部分の転送はコア CPU とは独立に実行される。

3.2.2.1 要求処理部

要求処理部は、図 2 に示すように、ホスト側からの要求を処理する Initiator Controller とネットワーク側からの要求を処理する Remote Controller から構成され、これらはアドレス変換テーブルを共有する。ホスト側からの要求は、Window と呼ばれる小規模のメモリに転送プリミティブに関する情報を書き込むことで発生される。最後の番地に書き込むと処理がキックされ、この Window 番号からテーブル参照が行われ、転送に必要なグローバル ID 等が得られる。この情報と転送開始番地から、さらに物理アドレスが TLB により参照され、パケットと DMA 転送要求が発生される。pull プリミティブについては、コマンドのみのパケット転送が行われるが、push プリミティブで転送開始サイズが大きい場合、パケットと DMA 要求はネットワークに応じたサイズに分離され、ページ境界では TLB の再参照が行われる。Window に書かれたデータ解析部と物理アドレス TLB 参照に関する一連の処理を行う部分はパイプライン化され、独立にコア

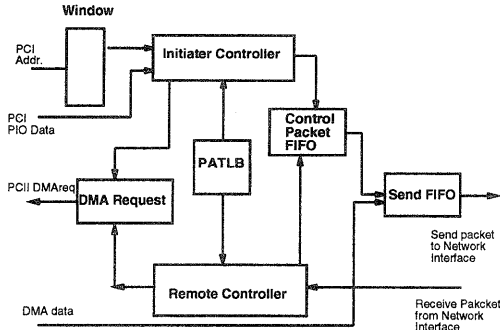


図 2 要求処理部の構造
Fig. 2 Structure of request path

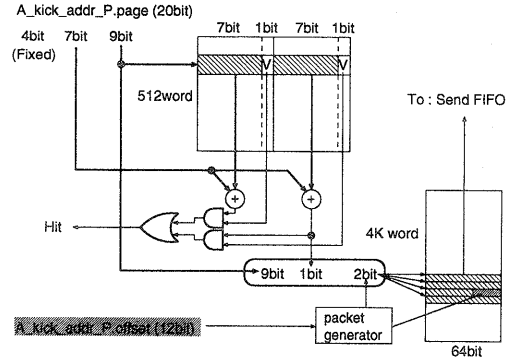


図 3 ヘッダ TLB の構造
Fig. 3 Structure of header TLB

CPU による代行が可能である。

Remote Controller は、ネットワーク側からのパケットの受信により起動され、受信側のアドレス参照を行うための TLB 参照と物理アドレス変換用の TLB 参照を連続して行い、プリミティブの種類に応じた DMA 転送要求と、パケット生成要求を発生する。Initiator Controller 同様、モジュール単位でパイプライン化されており、コア CPU による代行が可能になっている。

Martini は内部に、要求、転送共に、独立した 4 つのチャンネルを持っており、Initiator Controller と Remote Controller は、それぞれ専用のチャンネルを用いることによりデッドロックを回避する。

3.2.2.2 DMA 転送部

DMA 転送部は、要求処理部からの要求に基づき、PCI インタフェース、DIMM インタフェース、ネットワークインタフェース、CPU 内部メモリ間の DMA 転送を行う。要求されたモジュールが重ならないければ同時に転送を行うことも可能である。DMA 転送は完全にハードウェア制御され、要求発生部、コア CPU とは独立に動作し、その転送速度は遅い方のインタフェースによって制限される。

3.2.2.3 AOTF 送信制御部

Martini は、ホストでの 1 命令のメモリアクセスで、1 つのパケットを生成する極めて低遅延な送信を実現するために、AOTF (Atomic on-the-fly) 送信制御部を有する。高速処理の実現のため、これらの操作は、通常の要求処理部とは別のモジュールにより取り扱われる。

AOTF 送信制御部はトランザクション FIFO と図 3 に示したヘッダ TLB から構成される。

AOTF はホストプロセッサが DIMM I/F に 1 ワード書き込みを行うことで実行が開始される。まずはじめに、AOTF 送信部は DIMM ホストインタフェース部から渡されたアドレスとデータをトランザクションキューに格納する。この操作により、DIMM の書き込みタイミングと内部の処理時間との差を吸収する。

次に、ヘッダ TLB を利用して、トランザクションキューの先頭にあるアドレスの上位からパケットヘッダに使用するヘッダシードを連想する。ヘッダ TLB でミスヒットした場合は割り込みを利用してコア CPU にリフィル処理を行い、終了後、AOTF 処理を続ける。ヘッダシードには、パケットヘッダに必要なほとんどの情報が含まれるが、リモートのアドレスの下位とデータは含まれていない。この 2 つの情報はトランザクションキューから得、パケットの対応する部分に書き込まれる。以上の処理でパケットが生成される。生成されたパケットは DMA や BOTF で作られるパケットと同じ形式であるため、受信側は AOTF と他の処理を区別する必要は無く、AOTF においても様々な機能が使用できる。

AOTF の送信でホストプロセッサが関わる部分は最初に 1 ワード書き込むことだけである。ヘッダの生成も含め大半の処理は Martini 内部で並列に処理されるため、AOTF は BOTF と比べても非常に低レイテンシな通信である。

さらに、全ヘッダシードを保持するヘッダバッファの内容はシステムコール経由で設定されるため、DMA/BOTF と異なり、リモートのアドレスとして物理アドレスを指定することを可能にしている。物理アドレスを使用することで、受信側でのアドレス変換をスキップすることができ、さらにレイテンシを小さくすることが可能である。

3.3 インタフェース部

3.3.1 Host インタフェース

RHiNET では PCI インタフェース部、DIMM-net では Host DIMM インタフェース部が Host とのインタフェースとなる。PCI インタフェース部は 33MHz/66MHz, 32bit/64bit のいずれのモードでも動作するが、入出力部の電圧レベルは 3.3V に限られる。

DIMM Host インタフェース部は基本的には

PC100/PC133 の規格に従い、DIMM と同様に振る舞う。そのため、ホストプロセッサからは DIMM ソケットに接続されたメインメモリの一部として認識される。

また、DIMM の拡張規格で DIMM ソケットに DSP などの機器を接続するための規格である PEMM 規格に従って動作するモードも持つ。しかし、現在の一般的な PC で PEMM 規格に従ったマザーボードが無いため、Martini は基本的な DIMM の規格だけでも動作できるように設計されている。

Martini のコア部はホストプロセッサのクロックとの整合性を鑑み、PCI へ接続された時には 133MHz で、PC100 規格の DIMM として接続された時には 100MHz で、PC133 規格では 133MHz で動作する。

3.3.2 DIMM メモリインタフェース

Martini ではチップ外部に、1GB までの SO-DIMM を接続することができる。この SO-DIMM は RHiNET ネットワークインタフェースとして動作時には各種ワーキングメモリとして利用される。

DIMMnet-1 ネットワークインタフェースとしての動作時は SO-DIMM を 2 枚利用し、2バンク構成とする。それぞれをバススイッチ経由でホスト側、Martini 側に接続し、ホストと Martini が同時に送信データと受信データの書き込みを可能としている。必要に応じてバンクを反転させることで、SO-DIMM をダブルバッファとして利用する。

3.3.3 外部ネットワークインタフェース

Martini には、外部ネットワークインタフェース部を介してスイッチへのリンクが接続される。スイッチへのリンクには以下の 3 種類がある。

- RHiNET-2/SW 対応リンク (8Gbps)
- RHiNET-3/SW 対応リンク (10Gbps)
- 低速インタフェース (2.5Gbps)

NIC コア部と外部ネットワークとの接続には SWIF と呼ばれるインタフェースを介する。Martini は外部ネットワークとして、RHiNET-2/SW、RHiNET-3/SW および低速インタフェースと結合可能であるため、SWIF では、それぞれについてプロトコルや転送周波数、bit 数の変換を受け持つ。

ネットワーク入力/出力クロックドメインは、RHiNET-2/SW に接続される場合には 100MHz、RHiNET-3/SW に接続される時には 135MHz となる。入力クロックドメインは入力信号のクロックに同期し、出力クロックドメインはネットワークインタフェースカード上の発振器から提供される出力用クロックに同期する。

コアから SWIF にわたされた出力は、まず FIFO でコアクロックドメインから、ネットワーク出力クロックドメインに受け渡される。ECC または CRC コードの付加など必要な操作を行なった後、マルチプレクサを経て、高速出力信号ピン (LVDS 信号) からパケッ

表 1 計測条件
PCI bus 64bit, 66MHz
Martini clock 133MHz

表 2 Martini 処理時間 (送信)

処理内容	時刻 (ns)
PCI プリミティブ情報書き込み開始	0.00
PCI プリミティブ起動開始	150.00
LB トランザクション開始	162.75
PCI データ受渡開始	189.00
LB データ受渡開始	215.25
LB DMA 要求	275.25
PCI PCI バス要求	309.00
NIF ヘッド生成	327.74
Net ヘッド送信	365.25
LB トランザクション開始	387.75
PCI 送信データ受渡開始	429.00
NIF 送信データ受渡開始	470.25
Net データ送信開始	500.25

PCI: PCI バストランザクション

LB: PCI I/F - core 間トランザクション

NIF: core - Net.I/F 間トランザクション

Net: ネットワークトランザクション

トとして出力される。

外部から高速入力信号ピン (LVDS) に入ってきたパケットは、デマルチプレクサを経て、ECC、CRC などの必要な処理が行なわれた後 FIFO によってコアクロックドメインにのせかえられる。その後パケットヘッダ等の解析が行なわれ、コア部に受け渡すための仮想チャンネルバッファに投入される。

SWIF は、他にも以下の特徴を持つ。

- RHiNET-2/SW 用の window ベース、RHiNET-3/SW 用の credit ベースのフローコントロールをそれぞれサポートする。
- RHiNET-2,3/SW は仮想ネットワークと呼ばれる、相互に不干渉なネットワークを内部に複数設ける機能があり、これを Martini 内部の 4 つのチャンネルを割り当てることが可能である。

4. 予備評価

本章では Martini の性能をシミュレーション結果を交えて評価する。

4.1 DMA 転送

4.1.1 RHiNET ネットワークインタフェース

RHiNET でリモート DMA 書き込みである PUSH を実行した場合の送信側の処理時間を表 2 にしめす。測定条件は表 1 に示す。受信は表 3 に示す。

表 2 に表 3 を加えた 781.85ns が Martini での総処理時間となる。実際は、他にマルチプレクサ/デマルチプレクサの遅延 (約 16ns×2)、ネットワークの遅延 (5ns/m) が加わる。結果として、総遅延は約 814ns にネットワークの遅延を加えた値と予想される。以上

表 3 Martini 処理時間 (受信)

処理内容	時刻 (ns)
Net パケット到着	0.00
NIF クロック乗せ変え他	60.15
core アドレス変換他	142.85
PCI データ転送準備	281.60

は 66MHz,64bit の PCI の場合だが、33MHz,32bit PCI では約 1588.85ns になる。

一方、32bit,33MHz PCI の Myrinet を用いた PC クラスタ上の PM では約 $7.5\mu\text{s}$ である⁷⁾。 $7.5\mu\text{s}$ のうちのソフトウェア処理時間の約 $1\mu\text{s}$ を除いても、Martini が低レイテンシであることがわかる。

この差の原因は Myrinet はプロセッサが制御を行っているのに比べ、Martini がハードウェア制御だけで処理がされるためである。

4.1.2 DIMMnet-1 ネットワークインタフェース

Martini では、ホストのアクセスから内部にデータを出し始めるまでに、4 clock (133MHz で 30ns) である。また、DMA 要求からデータ転送開始までは、ライト DMA で 10clock(同 75ns)、リード DMA で 17clock(127.8ns) である。

これらの値から DIMMnet-1 ネットワークインタフェースでのリモート DMA 転送でのレイテンシを求めると、約 756ns で通信が完了する。PCI に比べ 58ns 早い、この主な差は PCI バスでのトランザクションを開始するまでの手間に比べ、DIMM インタフェースは手順が簡単で高速なためである。

4.2 AOTF

本節では DIMMnet-1 に固有の AOTF のレイテンシについて評価する。表 4 に AOTF 固有の処理についての処理内容と時間を示す。この表から、AOTF に固有な送信処理には 75.2ns かかることがわかる。パケットとしてネットワークに出るためにはさらに SWIF を通過する必要があり、これは DMA と同様に 30ns かかる。さらに、マルチプレクサの通過に 16ns かかるため、結果として、AOTF の送信処理には合計で 121.2ns かかる。これは DMA 通信の 756ns に比べると約 1/6 倍のレイテンシであり、AOTF の低レイテンシ性が明らかである。

次に、受信側の評価について述べる。アドレス変換をスキップし、内蔵の共有メモリがターゲットの場合、SWIF を通過するのに必要な 52.6ns とメモリに書き込むために必要な 7.52ns があれば受信処理が可能である。パケット到着からでは、デマルチプレクサのレイテンシ 16ns を加えて、76.12ns で受信処理が終了する。

以上のように、AOTF を用いると非常に高速なデータ通信が可能となる。

4.3 諸元

表 5 に Martini の諸元を示す。

表 4 AOTF 送信レイテンシ

処理内容	時刻 (ns)
ホストからの書込み	22.56
トランザクションキュー処理	15.04
ヘッダ TLB 参照	22.56
転送サイズ判定	7.52
送信バッファハンドシェイク	7.52

表 5 Martini の諸元

ASIC	日立 DDC 製 0.14 μm
メモリ総量	538KB
論理総ゲート数	1200KGates(予定)
内部動作周波数	100MHz/133MHz
PCI 動作周波数	33MHz/66MHz
DIMM 動作周波数	100MHz/133MHz

5. まとめ

本稿では、クラスタ環境において低レイテンシ・高バンド幅な通信を実現するために、1 チップでユーザレベルでのゼロコピー通信をサポートする Martini を紹介し、その機能について述べた。

シミュレーションによる予備評価では DMA による PCI-to-PCI で 814ns、DIMM-to-DIMM で 756ns となり、Myrinet などと比べ 1/2.8 倍~1/4.3 倍のレイテンシで通信が可能であることが分かった。更に、AOTF は 197.32ns で通信が完了し、フラグなどのシングルワード転送では非常に低レイテンシな通信が可能である事が分かった。

現在、Martini は実装中であるが、2000 年度末にチップが完成し、システムに組み込みテストを開始する予定である。

参考文献

- 1) <http://www.myri.com/>
- 2) <http://www.quadrics.com/>
- 3) 田辺昇, 山本淳二, 工藤知宏: メモリスロット搭載型ネットワークインタフェース DIMMnet-1 における細粒度通信機構, 情報処理学会研究報告, 2000-ARC-137, pp.65-70, 2000
- 4) Tomohiro Kudoh et al.: RHiNET: A network for high performance parallel processing using locally distributed computers, IWIA 99, 1999
- 5) Hiroaki Nishi et al.: RHiNET-2/SW: One-chip switch ASIC for a local area system network, Hot Interconnect, 2000
- 6) 上野龍一郎, 稲沢 悟, 西 宏章, 工藤知宏, 天野英晴: 光インタコネクトを使った高速転送におけるフロー制御手法, 情報処理学会研究報告, 2000
- 7) 手塚 宏史, 堀 敦史, Francis O'Carroll, 石川 裕: RWC PC Cluster II の構築と性能評価, Hokke '98, 1998