社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2005－ARC－164 (22)
2005／8／4

DIMMスロット装着型デバイスDIMMnet-2の改良方針

田邊 昇[†] 羅 徴哲[††] 濱田芳博[††]
中條拓伯[††] 北村 聡[†††] 宮代具隆[†††]
宮部保雄[†††] 天野英晴[†††]

筆者らはパーソナルコンピュータ (PC) のメモリスロットに装着されるプリフェッチ機能を有するメモリモジュール兼 PC クラスタ構築用ネットワークインタフェースである DIMMnet-2 を開発してきた．これは最高性能やコストは追求しない機能確認用のプロトタイプであった．未知なる商品候補の価値を十分示すためにはさらなる改善が必要がある．よって，将来の ASIC ベースでの製品化を展望しつつ，DDR2 スロットに装着される DIMMnet-3 を FPGA ベースで開発する．本論文では，DIMMnet-2 における問題点を列挙し，DIMMnet-3 におけるその解決策について述べる．主記憶データベースに対する Wisconsin ベンチマークへの適用を通し，DDR と DDR2 の違いや周波数，CAS レイテンシやメモリチャネル本数が想定アプリ性能に与える影響に関する評価結果を示す．

Strategy of improvement for a DIMMnet-2 Device Plugged into a Memory Slot

Noboru Tanabe,[†] Zhengzhe Luo,[††] Yoshihiro Hamada,[††] Hironori Nakajo,[††]
Akira Kitamura,[†††] Tomotaka Miyashiro,[†††] Yasuo Miyabe[†††]
and Hideharu Amano [†††]

The authors have developed a DIMMnet-2 plugged into a memory slot of personal computer (PC). It is not only a network interface but also a memory module with prefetching functions. This is a prototype not for achieving maximum performance and low cost but for confirming functional design. In order to showing high value for a canditate for new product, more improvements are needed. Therefore, we decided to develop FPGA based DIMMnet-3 plugged into dual DDR2 DIMM slots with ambition for future ASIC based commercialization. In this report, we show problems on a DIMMnet-2 and its solutions on a DIMMnet-3. We report the effects on performance of Wisconsin benchmark on main memory database by the difference between DDR and DDR2, frequency, CAS latency and number of memory channels.

1. はじめに

キャッシュアーキテクチャをベースとするパーソナルコンピュータ (PC) 等の単体性能向上や，コストパフォーマンスの高い PC クラスタ構築を目的として，筆者らは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や高機能メモリモジュールといったデバイスを提案してきた．

さらに，SDR 型 DIMM スロットに装着可能なネットワークインタフェースである DIMMnet-1 プロトタイプ [1] や，DDR 型 DIMM スロットに装着可能なネットワークインタフェース兼高機能メモリモジュール DIMMnet-2 プロトタイプ [2][3][4][5][6][7][8] を作成し，有効性や実現性を示してきた．特に，DIMMnet-2 プロトタイプは，ASIC 化を念頭に，性能やコストは重視せず，ハイエンドな FPGA により実装され，機能検証を中心に開発された．現在，PC1600 の DDR 型メモリスロットに装着されて，ネットワークインタフェースとしても，高機能メモリモジュールとしても，基本機能が動作している．

メモリスロットをホストインタフェースとすることにより，ムーアの法則に従って高速化を遂げる CPU に見合った性能向上をネットワークインタフェースやキャッシュアーキテクチャの欠点をカバーするデバイスの性能向上を確保することが可能である．このような利点とは表裏一体の関係として，本方式は改良せずに放置しておくとすぐに陳腐なハードウェアになってしまうという欠点が存在し，継続的に改良を加えていかねばならない宿命を背負っている．

一方，近年の LSI 開発費用は，90nm のプロセスではマスク 1 枚 1 億円とも言われるレベルまで高騰しており，限られた試作用の研究開発予算では ASIC 化を行なうことは容易ではなくなって来た．試作用の安価な ASIC 開発手段も存在するが，DIMMnet-2 のような信号ピン数が多い案件には対応できない．

PC のマザーボードも最近はデュアルチャネルのメモリバスが主流となり，周波数も向上した結果 PC1600 等の 100MHz 台前半の DDR メモリスロットに対する BIOS 設定ができないものが増加してきている．ビット単価が最安のメモリの規格は現時点では DDR 型であるが，DDR 型から DDR2 型へ徐々に移行が進むことは確実視されている．

以上を鑑みて，筆者らは現時点での ASIC 化を行なわない範囲で，目立ってきた DIMMnet-2 プロトタイプにおける未対応の欠点の克服や，動作速度や量産性を向上させる改良を加えた DIMMnet-3 プロトタイプを作成し，並列処理の実験や，コンパイラ等のソフトウェア環境の充実を図りつつ，商品化可能性を高める手立てを講ずることとした．

本報告では，第二章で DIMMnet-2 プロトタイプの概要とその開発状況を述べ，第三章で改良すべき課題を検討し，第四章でその改良方針を述べる．第五章では改良項目のうちメモリ規格や周波数や CAS レイテンシの設定変更やメモリチャネル本数に伴う主記憶データベースに対する Wisconsin ベンチマークの性能の変化に関する評価を述べ，第六章でまとめる．

† (株) 東芝，研究開発センター
Corporate Research and Developpment Center, Toshiba
†† 東京農工大学
Tokyo University of Agriculture and Technology
††† 慶應義塾大学
Keio University

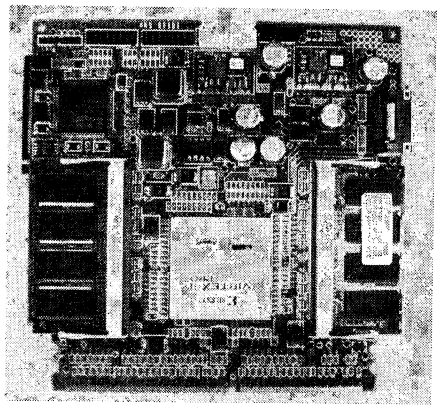

図 1 DIMMnet-2 試作基板

2. プロトタイプの開発状況

図1に試作基板の概観を示す．DIMMnet-2のプロトタイプ基板は，コントローラ部に Xilinx 社の FPGA である Virtex-II Pro XC2VP70-7FF1517C を用いている．この FPGA は InfiniBand，Fibre Channel や 10Gigabit Ethernet に対応した高速シリアル I/O インタフェースである RocketIO トランシーバを搭載しており，これを利用して InfiniBand スイッチ (4X:10Gbps) に接続する．

基板上にはノート PC 用の汎用メモリである 200pin DDR SO-DIMM を 2 枚搭載する．これは通信用のバッファに使用するほか，ホスト PC のデータ記憶領域として使用する．現在，256MByte の SO-DIMM を 2 枚搭載しているが，将来的にはホスト PC のメモリスロット 1 本当たりに搭載可能な最大メモリ容量多くの SO-DIMM を搭載し，大規模な分散共有メモリシステムを構築することを視野に入れている．

コントローラ部に FPGA を用いていることから，高い動作周波数での動作が困難であるため，FPGA を 100MHz で動作させ，PC-1600 の規格での動作に対応させている．

2.1 コントローラ部の開発状況

2.1.1 通信処理

コントローラ部の基本的な開発は完了しており，基板上の SO-DIMM から読み出したデータを InfiniBand のケーブルを通してリモートノードへ転送，またはリモートノードの SO-DIMM のデータを読み出すことが可能となっているほか，コントローラ内部のバッファに書き込んだデータをパケットとしてネットワークに送出する Block On The Fly(BOTF) 通信も可能である．基本通信性能の評価の結果，商用の PC クラスタ向けネットワークの中でも特にレイテンシの低いネットワークである Quadrics Network(QsNET) よりも低いレイテンシでの通信が可能であることが確認されている [6)7)]．

現在，スイッチから転送される制御パケットの扱いに関して検討を重ねている段階であり，性能の測定には達していないものの，InfiniBand スイッチを介した通信も可能である．

また，通信におけるデータの受信領域を送信側ではなく，受信側が決定する通信手法の実装も行っている．一般に，データの受信領域を送信側が指定する手法では，実際に通信を行う前に，受信領域を通知するための通信が必要となるが，受信側が受信領域を指定する本手法では，受信領域の通知が必要無くなるという利点がある．現在，受信領域を一つのリングバッファとして用いることで，この機構を実装し，シミュレーションによる動作も確認されているが，全ての送信ノードに対して一つのバッファを用いているため，受信側では受信した順番にデータを読み出すことになる．

そこで，各送信ノードごとに異なるバッファを確保する手法を実装しており，受信領域のためのアドレス管理手法を検討している段階である．

2.1.2 不連続な領域に対するアクセス

通信処理以外の機能として，SO-DIMM に格納されたデータに対してストライドアクセスやリストアクセスといった不連続領域に対するアクセスを搭載しており，現在，ストライドアクセスの性能評価を行っている段階である [8)]．リストアクセスなど，他の不連則領域に対するアクセス機構は実装を進めている段階である．

この不連続領域に対するアクセスは，リモートノードに対しても可能であるように設計してあり，リモートストライドストアなどが実行可能であるが，未だ，性能評価には到っていない．

3. 改良すべき課題

本章では，現状の DIMMnet-2 プロトタイプにおける改良すべき課題について述べる．

3.1 低い量産性

3.1.1 FPGA の値段がネック

FPGA を用いた DIMMnet-2 プロトタイプにおいては，チップ間配線による性能劣化や，チップ間の情報伝達に関する設計工数増加の回避のために，設計当時に入手可能と見られた最高グレードかつ最大規模の FPGA1 チップでの実装を行なった．

Infiniband のインタフェースもこの FPGA に付属する数 Gbps 級のシリアル I/O ハードマクロ (RocketIO) を用いて実装したため，ここで用いられたデバイスは非常に高価なものとなってしまい，基板の試作価格の 8 割以上を占めることとなった．この種の FPGA の利用件数がこの時点ではあまり多くなかったことが原因と思われる．こうして基板の試作価格も予想を大幅に上回るものとなり，このまま多数作成して PC クラスタ環境を行なうには研究予算上無理が生じることとなった．

また，DIMMnet-2 プロトタイプにおいて用いられたのは 130nm プロセスの FPGA であり，現時点での生産の主流は 90nm プロセスの FPGA に移行しつつあるため，130nm プロセスの FPGA の価格は下げ止まってしまっており，今後も価格低下はあまり望めない状況にある．

3.1.2 量産時の数量不足

量産時のコストに関しては，どれだけ多くの枚数の基板がどれくらい長く生産されるかによって決まってくる．Myrinet や Infiniband や QsNET など PC クラスタ専用の NIC は，高性能計算ユーザーに利用者が限られるため，市場規模が小さい．このためボードも必然的に高価なものとなってしまう．

これに対して，DIMMnet-2 では PC クラスタ専用の NIC であるとともに，高機能メモリモジュールとしての機能を有する．このため，予測される市場規模は Myrinet 等の NIC よりは大きなものを想定することができる．しかし，いずれにしても高性能計算ユーザーに利用者が限られるため，市場規模はさほど大きなものではない．

ASIC 化していくためには最低でも月産 1000 枚程度の生産が 1 年以上見込める必要が生じるが，高性能計算ユーザーだけではその生産量には及ばないことが予想される．

よって，将来 ASIC 化によって将来の DIMMnet の高性能化と低価格化を実現していくためには，生産量を劇的に増やす利用方法の開拓が望ましい．

3.2 部品入手性の悪さ

DIMMnet-2 設計スタート時には，DDR 型のメモリチャネルが 1 本からなるマザーボードが主流であった．メモリチャネルの周波数も 100MHz(PC1600 相当) から

133MHz(PC2100) が主流であった.

設計・試作を進めて時間が経つうちに，マザーボードの主流はデュアルチャネル 200MHz(PC3200) に移ってしまい，さらに DDR2 デュアルチャネルをサポートするチップセットやマザーボードが市販されている．DDR・100MHz(PC1600 相当) という現状の DIMMnet-2 プロトタイプの装着可能な BIOS 設定を有するマザーボードが減少の傾向をみせている．

完成時点でのマザーボードの入手性を向上させるための規格の選択を時期に合わせて適切に行なうとともに，新規格への対応が容易になるような構造的工夫がなされることが望ましい．

3.3 ASIC への移行困難性

DIMMnet-2 プロトタイプは，ホストインタフェースとして DIMM1 本分，外部メモリインタフェースとして DIMM2 本分の外部接続が必要である．よって，使用する信号ピン数が多く，これが FPGA のパッケージを大きくさせ，価格を押し上げる原因となった．一方，試作用の安価な ASIC 開発手段も存在するが，DIMMnet-2 のような信号ピン数が多い案件には対応できない．ASIC 開発コストへの影響の観点からも，ASIC 単価の観点からもピン数が多い点を緩和することが望ましい．

DIMMnet-2 プロトタイプにおいて ASIC 化を阻害する要因としてはさらに，現在使用している FPGA がピン互換の ASIC を提供していないため，ASIC 化は必ず基板再設計を伴うという点が挙げられる．

さらに，DIMMnet-2 プロトタイプにおいてはマザーボード上でのノースブリッジと DIMM スロット間のデータ線のねじれをマザーボードの品種に応じて FPGA のコンフィギャレーションデータを準備することで対応していた．よって，この方法は固定配線が必要な ASIC にはなじまないため，このまま ASIC へ移行することはできないため別の解決方法が必要である．

3.4 周波数の低さ

現状の DIMMnet-2 はベースクロック周波数が 100MHz で動作しているため，メモリ規格は DDR200 すなわち PC1600 のメモリとして動作する．この周波数は PC3200 が主流の現時点では遅いと言わざるをえず，この設定を可能なマザーボードが徐々に市場から消えつつある．

周波数の向上を阻害しているクリティカルパスは，主に 2 箇所存在し，DIMM 型のホストインタフェースにおけるアドレス分配部と，外部 SO-DIMM インタフェースにおける連続したアドレスへのリクエスト時に SO-DIMM へのコマンドを連結する機能に関する部分である．いずれも CAS レイテンシ設定の緩和を行なわないと解決が難しい．

また，前述の ASIC 向けのロジックによるねじれ解消は CAS レイテンシ設定に無関係ではないので，将来の DIMMnet の CAS レイテンシを緩和することは必須と考えられる．

3.5 搭載メモリ容量の少なさ

DIMMnet-2 プロトタイプにおいては基板に搭載している外部メモリが 256MB の SO-DIMM2 枚のみであった．機能検証の目的としてはこれでも十分であったが，本来 DIMMnet-2 はマザーボードごとに規定されている DIMM スロットの装着可能容量上限を超えるメモリ容量を提供できる．

また，コスト的には SO-DIMM は同容量の通常サイズ DIMM と比べ，やや割高となる傾向があった．同時期に入手できるモジュール当たりの最大の容量的にも 2 倍の差がある．

3.6 規格変更時の影響が大きい

現状の DIMMnet-2 はホストインタフェースとして用いているメモリスロットの規格が変更になれば，基板も FPGA 内ロジックも全面的に再設計を余儀なくされる．

メモリスロットインタフェースを採用する限り，2,3 年に一度は主流になりそうなメモリスロット規格に合わせて設計変更は必要になってくるが，Time to market を実現していくためには，その際の設計変更が最小限で収まるような工夫がなされることが望ましい．

3.7 シングルチャネルメモリバスのみ

現状の DIMMnet-2 ではデュアルチャネルへの対応は考慮されていない．現状のままの DIMMnet-2 を 2 枚 PC に装着しなければならないとすると，コスト面でも問題がある．

一方，現在のマザーボードの主流はデュアルチャネルに移ってしまっている．FB-DIMM は複数チャネルで構成することが容易である上，DDR または DDR2 ベースになるのでバンド幅面から複数チャネルで用いられることは必然である．さらに数年後，XDR-DRAM や DDR3 などのより大きなバンド幅を提供できるメモリへの移行が進んだ場合も，モジュール当たりの容量限界と，ポイント to ポイント接続の必然性から，この傾向はかなりの長期間に渡って続く可能性が高い．

3.8 基板のサイズ・機械的支持の問題

現状の DIMMnet-2 では，1 チップの FPGA で作成しているため，基板を分離することができず，DIMM スロットに通常より相当大きな基板を装着しなければならない．このため，1U サーバなどの薄型筐体には入らない上，特別な機械的支持構造も準備しなければならなかった．

4. 改良方針

本章では，現状の DIMMnet-2 プロトタイプにおける改良すべき課題をどのように DIMMnet-3 プロトタイプにおいて解決するかについて述べる．

4.1 Si ディスクドライバによるコモディティ化

量産性の向上による劇的なコストダウンをはかるべく，今後は DIMMnet をシリコンディスクとして用いられるようなドライバを開発する予定である．これは DIMMnet-2 プロトタイプ上でも実験が可能なのでその上での開発を行なう．

従来のシリコンディスクは MS-DOS の RAMDISK.SYS ドライバや IO DATA 社の RamPhantom[9] のように主記憶上にソフト的に作るか，GIGABYTE 社の iRAM[10] のように I/O バスの先に配置されていた．前者はあまり大容量を確保すると主記憶を圧迫するため逆効果になる．後者は I/O インタフェースのバンド幅や遅延に律速されるため，主記憶をアクセスするのに比べれば性能低下は免れない．

しかし，DIMMnet の場合は主記憶が配置されるメモリスロット上に配置されるため，ページフォルト時にアクセスされる Windows の pagefile や Linux の swap をこの上に配置すれば，ページフォルトが発生しても主記憶をアクセスするのと殆ど変わらない．

さらに，通常 4GB までしか装着できない安価なマザーボード上でもあたかもサーバー機のように大量の主記憶が載っているかのような体感速度が得られるものと思われる．Windows の起動時にも HDD アクセスが抑制されるため高速に立ち上がる．64bit 化された Windows が普及するとともに大容量のメモリをうまく活用するアプリケーションが登場するようになれば，高価かつ高騒音なサーバー機を用いずとも，その恩恵を受けられるようになる．

また，技術計算を個人が日常的に用いている PC 上で計算することが多い研究所などでは，シリコンディスクドライバ付きの DIMMnet を用いれば，Linux 上で技術計算している時だけでなく，Windows で立ち上げた時にもプログラムを書き換えることなくアプリの高速化や，PC 起動時間の大幅な短縮などメリットが生じることになり，普及を促進するものと考える．

このように，並列処理とは縁遠かったユーザーが購買層として期待できるため，PC クラスタ用 NIC でしかない場合に比べ，大幅な低コスト化が可能である．

4.2 DDR2 への対応

現時点での PC 用メモリは DDR 型が大半を占めている．DDR2 型は販売はされているもののビット単価において劣る

上，CPU の FSB がボトルネックであるため多くの PC 上では DDR 型に対する性能上のメリットが現時点では顕著ではない．しかし，この状況は 2006 年頃には逆転すると見られており，市場の主流は DDR2 型に移行するのは時間の問題である．対応可能なマザーボードの入手性も DDR2 の方が先行きが明るい．

よって，DIMMnet-3 においてはホストインタフェースは DDR2 型 DIMM にすることにした．DDR2 型では電圧が 1.8V に低下し，ピンで消費される電力が減るとともに，動作周波数上でもメリットがある．

なお，DDR2 の場合は最低の周波数が 200MHz(PC2-3200) であるため，DIMMnet-3 のコアロジックも最低でもこれに追随できなければならない。したがって，ASIC 化しなければ達成不可能が確実になった場合は DDR を採用する．

一方，DIMMnet-3 上に搭載するメモリについては，等間隔またはランダムアクセス時の性能とビット単価の兼ね合いから，今後の市場の推移を見つつ，慎重に選ぶこととする．

DIMMnet-2 および 3 が主に高速化をもくろむアプリケーションはキャッシュが効きにくい不連続アクセスが多発するアプリケーションであるため，余分なデータを DRAM 内でフェッチすることが少ない DDR 型の方が同じ周波数であれば実行性能が高いと予想される．そこで本報告においては後半で，DDR の場合と DDR2 の場合で周波数によりアプリケーションの性能がどの程度変化するか，実験評価する．

4.3 搭載 DIMM 枚数および容量拡大

DIMMnet-3 においては搭載可能なメモリ容量を大幅に拡張する．具体的には，1,2 年後には入手可能と思われる 4GB の DIMM に対応できるロジックを作成し，ボード上には 4 枚の DIMM スロットを搭載するものとする．基板面積の関係から SO-DIMM で実装しなければならない場合はモジュール当たり 2GB までになるが，それでも 8GB のメモリ容量となるので現状のサーバー機にも匹敵するメモリを搭載可能ということになる．搭載容量が多くなるためエラーの発生確率も上がることが予想されるため，ECC にも対応することを目標とする．

4.4 90nm 世代 FPGA の利用

周波数の向上と低コスト化をはかるため，DIMMnet-3 においては 90ns 世代の LSI を用いる．予算の関係から ASIC は用いることが困難なため FPGA を用い，具体的には XILINX 社の Virtex4FX または LX と，Altera 社 StratixII を候補として検討する．

入手できる時期や調達可能価格との兼ね合いから，両社のデバイスのどちらかを採用する予定である．

まず，XILINX 社の Virtex4 シリーズの場合は RocketIO がサポートされている FX が現在の DIMMnet-2 で用いられている VirtexIIPro からの移行が最も簡単である．予定価格も DIMMnet-2 で用いられている品種より若干ゲート数を落とせば一桁ダウンを見込めるためコスト的なメリットは大きい．ただし，入手性についてはあと数ヶ月の変化に注意が必要と考える．また RocketIO を持たない LX は入手性には問題はないと思われるので，これを用いる場合は別チップの市販 SERDES か，小規模な Virtex4 FX または VirtexIIPro により Infiniband のインタフェースを作成することになる．Virtex4 シリーズの場合はピン互換性のある ASIC マイグレーションは提供されておらず，ASIC ベースで量産する場合には基板の再設計も必要である．周波数的には，既に Virtex4 による DDR2-533(266MHz) での実装実績があることが判っているため，注意深く設計すれば最低 200MHz でのホストインタフェース動作は不可能ではないと考えている．

一方，Altera 社 StratixII の場合は RocketIO に相当する数 Gbps クラスのシリアル I/O マクロがサポートされている 90nm 世代の FPGA が提供されていない．このため上記の Virtex4 LX で実装する場合と同様に別チップの市販 SERDES か，小規模な Virtex4 FX または VirtexIIPro により Infiniband のインタフェースを作成することになる．StratixII のメリットは HardCopyII というピン互換性がある ASIC マイグレーションが提供されてる点にあり，この基板のまま月産 1000 枚以上の量産に移行する場合は開発期間やコストの面でメリットがある．さらに HardCopyII の場合は XILINX 社の EasyPath と異なり，移行による高速化が期待できる．約 50%の高速化が可能と言われているため，それが正しければ FPGA で PC2-400 が動作できるなら，HardCopyII 移行時には PC2-600 が動作できることが期待できる．ただし，Virtex4 に比べ，現状ではコストが高いように見受けられる．

4.5 CAS レイテンシの延長容認

ASIC 化や EasyPath 移行が可能な新ねじれ解消論理を導入しつつ，周波数の向上をはかるため，DIMMnet-3 においては CAS レイテンシのデフォルト設定からの延長を容認するものとする．DDR と DDR2 のマザーボード上での BIOS 設定選択肢上の傾向では，DDR2 の方が高周波への対応を念頭に CAS レイテンシが大きい選択肢が増えている傾向にある．例えば DDR と DDR2 の両方をサポートしているマザーボード上では，DDR では CAS レイテンシは 3 までしか設定できないが，DDR2 では 5 まで設定できる．よって DDR2-400 では CAS レイテンシ 4 または 5 という設定が可能である．

ただし，CAS レイテンシの延長はアプリケーション性能の低下が予想されるため，本報告では後半でこの点について実験検証を行ない，許容できるものか否かを検討する．

4.6 ASIC 対応可能なねじれ解消

DIMMnet-3 においては一般ユーザによる不特定なマザーボード上での利用や，量産時の ASIC 化または EasyPath 化を想定し，DIMMnet-2 のように FPGA のコンフィギャレーションによる切り換えは用いず，マルチプレクサベースでマザーボード上のデータ線ねじれ解消を行なうものとする．ねじれ検出はホストから 1 ビットだけ異なる 64bit データを 64 回書き出し，それが DIMM 上のどのビットに現れたかをホストにエコーバックする．この際，ねじれ状況が判別するまではメモリバスは全ビット共通の 1 本のシリアル通信路とみなして，ホストへのエコーバックデータの転送を行なう．判定結果に応じてマルチプレクサを切り替えることによって，対応するチップセットを搭載するあらゆるマザーボードに対応できるようにする．

4.7 デュアルチャネルメモリバス対応

DIMMnet-3 においては昨今のマザーボードの動向および近い将来でのその傾向の予測を踏まえ，デュアルチャネル対応を行なう．その際，ホストには二本のメモリチャネルに同等の SPD(タイミングおよび容量) 情報を有する DIMM が装着されているように見せかけることでデュアルチャネルロックステップモードで動作可能とし，DIMMnet-3 を装着することで，姫野ベンチマークに代表される連続アクセスが多用されるメモリバンド幅ネックになりがちなアプリケーションでの性能低下を極力食い止めるものとする．

両方のチャネルに装着されるボードを完全に対称にして，外部メモリやその制御部を両方に搭載する方法もある．その場合は両方の制御部が発生したアドレスが自分のモジュールにない場合は廃棄する必要がある．またコスト的には 2 倍になってしまい高くつく．

このため，DIMMnet-3 においては少なくとも片方にはホストインタフェースのみを搭載した基板を装着するものとする．ホストインタフェースのみ搭載する基板上には Read ウィンドー，Write ウィンドー，制御レジスタ，LLCM(Low Latency Common Memory) といった DIMMnet-2 上ではホストの仮想空間上にマップされ DRAM としてアクセスされていたハードウェアが搭載される．

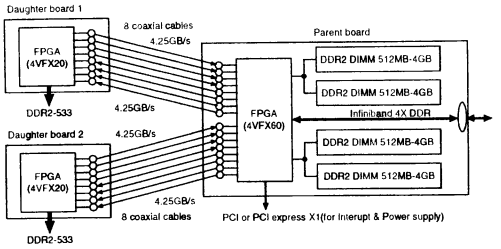

図 2 DIMMnet-3 の概略構造

4.8 ホスト I/F 部と外部メモリ部の別基板化

DIMMnet-3 においては基板サイズが 1U 等の薄型筐体になじまない点や機械的支持構造が必要だったという問題点の解決と，デュアルチャネル対応する際のコスト低下と，ホストインタフェースとなるメモリスロット規格の変更への対応容易性を考慮し，ホストインタフェース部と，外部メモリを搭載する部分を分離し，別基板として実装することを目指す．図 2 に DIMMnet-3 の概略構造を示す．

特に，RocketIO などの数 Gbps クラスの I/O を有する FPGA がリーズナブルな価格帯で適切な時期に利用可能であった場合には，ホストインタフェース部基板 (子基板) と，外部メモリ搭載基板 (親基板) の間の情報転送には数 Gbps クラスの I/O を複数用いて少ないピン数と可能な限り高いバンド幅で接続する．例えば，Virtex4 FX を双方の基板で用いる場合には，8.5Gbps の RocketIO を 4 組ずつと同軸ケーブル 8 本を用いて 4.25GB/s のバンド幅を有する全二重通信路で両基板間を接続するという設計が有望と思われる．論理回路がデュアルチャネル構成の際には 3 枚の基板にそれぞれ 1 個ずつ搭載された 3 チップの FPGA に分割されるため，単価が安い中小規模の FPGA(例えば子基板には 4VFX20，親基板には 4VFX60) を利用可能となる．

外部メモリを搭載する親基板を PCI または PCI express スロットに装着することで，機械的強度の向上や，薄型筐体への対応や，通常サイズの DIMM の利用や，PCI または PCI express スロット経由の割り込みの実装や，スタンバイ電源を利用した DRAM の記憶保持が可能となると考えている．

5. 性能評価

5.1 評価環境

表 1 に性能評価を行ったマシンの仕様を示す．この環境は単にシミュレーションを行った機械という意味合いだけでなく，HOKKE'05 の論文 [5] における評価方法を用いているために評価対象そのものを構成する．つまり，キャッシュラインサイズやキャッシュ容量はもちろんのこと，メモリバンド幅や命令の CPU 内部での実装の細部に至るまで評価対象そのものを構成し，提案メモリモジュールと組み合わせた際の性能を左右する．これらを仮想的に変更することができない代わりに，通常のシミュレータを用いた評価手法に比べて高速性と精度が高い．

評価アプリケーションは HOKKE'05 の論文 [5] と同様で，Wisconsin ベンチマークの検索対象データを主記憶上に構築した C 言語で記述した Original のプログラムと，それを DIMMnet-2 を用いて実行させる場合の最適化適用状況を変えたプログラムで，ハード上のパラメータを変更して実行時間を測定した．最適化適用状況としては Simple(単純に DIMMnet-2 で動くようにしたもの)，PW2(プリフェッチ Window を 2 枚用いてソフトウェアパイプライニングを適用したもの)，Unroll(ループアンローリングを適用したもの)，NoCLF(CLFLUSH 命令を使用しない場合の参考値) の組合せで 6 種類測定した．タプルサイズは 140 バイト，タプル数が 650,000 のデータベースに対して最小値を検索するクエリーで評価した．

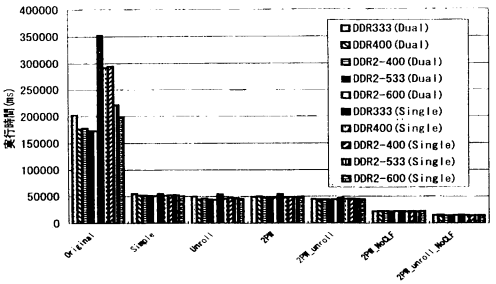

図 3 Wisconsin ベンチマークの最小値検索クエリーの実行時間における DDR と DDR2 の性能差 (タプル長 140 バイト，タプル数 650000)

表 1 評価環境

マザーボード	ASUStek P5GDC-V Deluxe
チップセット	i915G
CPU	Pentium®4(Prescott)
FSB 周波数	800MHz
コア周波数	3.0GHz
L1 キャッシュ容量	16KB
L2 キャッシュ容量	2MB
L1 キャッシュラインサイズ	64B
L2 キャッシュラインサイズ	128B
メモリ種類	PC3200 or PC2-3200
メモリバス本数	1 or 2
メモリ容量	2GB
OS	Linux 2.4.26
コンパイラ	gcc 3.2.2
最適化オプション	-O3

5.2 DDR1 と DDR2 の差

Wisconsin ベンチマークの最小値検索クエリーのタプル長 140 バイト，タプル数 650,000 における実行時間における DDR と DDR2 の性能差を，CAS レイテンシ 3(DDR333,DDR400,DDR2-400) または 4(DDR2-533,DDR2-600) に固定でいくつか周波数を変えて測定した結果を図 3 に示す．

その結果，DDR2 は同じ周波数であるならば若干 DDR より性能が低下するが，その低下率はあまり大きなものではなかった．FPGA による実装で周波数が 200MHz(PC3200 相当) で留まった場合にも，ダメージは少ない．この結果から，1.2 年後におけるマザーボードや搭載メモリモジュールの入手性が高いことが予想される DDR2 とするメリットの方が優先されるべきと考えている．

DDR は 200MHz(PC3200 相当) 以上の周波数は規格上ない．これに対して DDR2 はより高周波化を指向した規格であり，周波数を上昇させることでこのロスを補うことができる可能性がある．本アプリケーションのように不連続アクセスが主体のものでは，プリフェッチビット数が DDR から DDR2 になることで 2 から 4 に延びたことによる性能低下が強く現れると思われるアプリケーションであるが，周波数を 533MHz，600MHz と上昇させることによる性能向上は実験結果からはアプリケーション性能に反映されることがデュアルチャネルの場合は読み取れなかった．シングルチャネルでは読み取れた．これは FSB が PC3200(200MHz)2 本分相当しかなくボトルネックとなったためであると考える．

その点を考慮して，CPU 側の FSB が改善されていくならば，ASIC 化や高速グレードの FPGA 利用により，DDR2 で周波数を向上させることが性能上も有利と考えている．

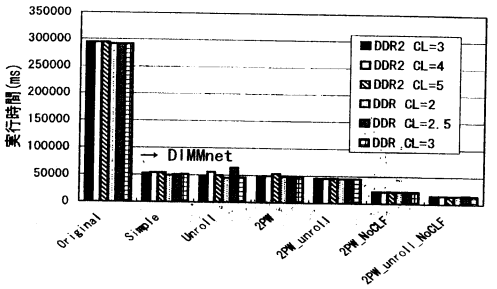

図 4 Wisconsin ベンチマークの最小値検索クエリーの実行時間における CAS レイテンシ延長の影響 (タプル長 140 バイト, タプル数 650000, 周波数 200MHz, シングルチャネル時)

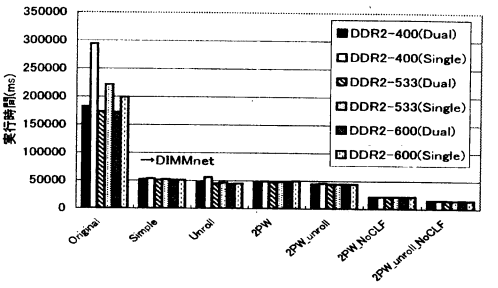

図 5 Wisconsin ベンチマークの最小値検索クエリーの実行時間におけるチャネル本数の影響 (タプル長 140 バイト, タプル数 650000, DDR2 で CAS レイテンシは 4)

5.3 CAS レイテンシ延長の影響

DDR と DDR2 において周波数 200MHz 固定で CAS レイテンシを変えて測定した前記と同様のクエリーの実行時間の結果を図 4 に示す.

その結果, CAS レイテンシに対してこのアプリケーションは鈍感であることが判った. デュアルチャネルの場合もシングルチャネルの場合と傾向は同じであった. 別途行なった実験から NAS CG についても同様の傾向が見えていた. よって, DIMMnet が加速対象とする主なアプリケーションでは CAS レイテンシを多少犠牲にしても影響は少ないと考える.

5.4 チャネル本数の影響

シングルチャネルになるように DIMM を装着した場合とデュアルチャネルになるように装着した場合で, DDR2 で CAS レイテンシ 4 に固定し, 周波数を変動させた場合の前記と同様のクエ リーの実行時間を測定した結果を図 5 に示す.

その結果, このアプリケーションの場合は, とりわけ DIMMnet を用いなかった場合にチャネル本数や周波数を上げてメモリバンド幅が上がると素直にアプリケーションの性能に反映される. デュアルチャネルの場合は周波数に対して性能は鈍感である. これは FSB ネックのためであることを示唆している結果と考える. DIMMnet を用いた場合は本アプリケーションの場合は DIMMnet によってほぼ常時 L2 キャッシュがヒットで L1 がミスしている状況であるため, L1 ミス時のバンド幅によって性能が制限されており, デュアルチャネル化で増えたメモリバンド幅がこれ以上の性能向上に結び付かなかったものと考えられる. L1〜L2 のバンド幅ではなくメモリが依然としてネックであるようなアプリケーションではデュアル化の効果が表れるはずと考えるが, その確認は今後の課題である.

6. おわりに

本報告では, DIMMnet-2 プロトタイプの概要とその開発状況を述べ, 改良すべき課題を検討した. さらにそれらの改良方針を述べた. 改良項目のうちメモリ規格や周波数や CAS レイテンシの設定変更やメモリチャネル本数に伴う Wisconsin ベンチマークの性能の変化に関する評価を述べた.

その結果, メモリ規格は DDR2 にしても同じ周波数なら不連続アクセスが多用されるアプリケーションでも性能低下が少ないことが判った. CAS レイテンシは性能にはほとんど鈍感であり, 主目的とするアプリケーションを中心に考える場合は CAS レイテンシを延長しても良いと考える.

メモリチャネルの本数については DIMMnet を用いない場合はデュアルの方が性能上有利だが, 周波数を DDR2 にして上げても性能向上はあまり見られなかった. これは FSB ネックの影響である. 一方, 本アプリケーションはシングルチャネルの DIMMnet で十分に高速化され, メモリネックではなくなっているためメモリチャネルの本数や周波数の影響は少なかった. よりメモリネックの度合が高いアプリケーションでの評価は今後の課題である.

上記で得られた知見や方針をもとに, コスト的にも性能的にも改善された DIMMnet-3 プロトタイプを開発し, 並列処理の実験まで行なっていくことが今後の課題である.

謝辞 本研究は総務省戦略的情報通信研究開発推進制度の一環として行われたものである. DIMMnet の開発に関する議論にご参加いただいている慶應義塾大の西講師, 渡辺氏, 大塚氏, 伊沢氏, 東京農工大の並木助教授, 荒木氏, 池田氏, 柴田氏, 森氏, 立命館大の国枝教授, 和歌山大の齋藤講師, 横浜国立大の土肥名誉教授, 箱崎氏, 安藤氏, 日立 IT の上嶋氏, 今城氏, 岩田氏に感謝いたします.

参 考 文 献

1) 田邊, 山本, 今城, 上嶋, 濱田, 中條, 工藤, 天野 : "DIMM スロット搭載型ネットワークインタフェース DIMMnet-1 の試作", 情報処理学会 HPC 研究会 (SWoPP2001), Vol.2001, No.77, pp.99-104, Jul. 2001.
2) 田邊, 濱田, 中條, 天野 : "メモリスロット装着型ネットワークインタフェース DIMMnet-2 の構想", 情報処理学会計算機アーキテクチャ研究会, 2003-ARC-152, pp.61-66, Mar. 2003.
3) 田邊, 中武, 箱崎, 土肥, 中條, 天野 : "プリフェッチ機能付きメモリモジュールによる不連続アクセスの連続化", 情報処理学会計算機アーキテクチャ研究会, 2004-ARC-157, pp.139-144, Mar. 2004.
4) 田邊, 安藤, 箱崎, 土肥, 中條, 天野 : " プリフェッチ機能を有するメモリモジュールによる PC 上での間接参照の高速化", 先進的計算基盤システムシンポジウム SACSIS 2005, pp.17-24, May.2005.
5) 田邊, 箱崎, 安藤, 土肥, 中條, 天野 : " メモリモジュール上での等間隔アクセス連続化の効果", 情報処理学会アーキテクチャ研究会, 2005-ARC-162, Mar.2005.
6) 北村, 濱田, 宮部, 伊澤, 宮代, 田邊, 中條, 天野 : "DIMMnet-2 ネットワークインタフェースコントローラの設計と実装", 先進的計算基盤システムシンポジウム SACSIS 2005, pp.293-300, May.2005.
7) 宮部, 北村, 濱田, 宮代, 伊澤, 田邊, 中條, 天野 : "DIMMnet-2 低遅延通信機構の実装と評価", 情報処理学会アーキテクチャ研究会, 2005-ARC-163, May.2005.
8) 宮代, 宮部, 伊澤, 北村, 箱崎, 田邊, 中條, 天野 : "DIMMnet-2 ネットワークインタフェースにおけるプリフェッチ機構の実装と評価" 情報処理学会アーキテクチャ研究会, 2005-ARC-163, May.2005.
9) IO DATA : "RAM ディスク作成・管理ソフト RamPhantom", http://www.iodata.jp/prod/memory/list/2004/ramphantom/
10) 西田 "COMPUTEX TAIPEI 2005 レポート", DOS/V Power Report, pp.25-27, Aug. 2005