超並列計算機 RWC-1 の入出力機構とその基礎評価

廣野英雄[†,☆] 松岡浩司[†,☆☆] 岡本一晃[†,☆]
横田隆史[††] 坂井修一[†††,☆☆☆]

超並列計算機において高速な演算処理が実現されると，それに見合った入出力性能が要求される．これを満たすため超並列計算機 RWC-1 では，その要素プロセッサ RICA-1 に DMA 機能を持った入出力インタフェースを備え，演算処理との干渉のない高速入出力機構を実現している．さらに通信のための相互結合網とは独立に，入出力のための入出力網を設け，通信と入出力が干渉することを防いでいる．入出力網は階層化されており，リングによる小規模な下位網と ATM を用いた汎用性が高い上位網からなる．本論文では RWC-1 の入出力機構について，特に下位入出力網の特性を中心に小規模システムでの実機評価を行い，その入出力性能が演算処理によって損なわれることなく十分な性能が得られることを示す．

Basic Performance Evaluation of I/O System on RWC-1

HIDEO HIRONO,[†,☆] HIROSHI MATSUOKA,[†,☆☆] KAZUAKI OKAMOTO,[†,☆]
TAKASHI YOKOTA[††] and SHUICHI SAKAI[†††,☆☆☆]

This paper introduces an I/O system for massively parallel computer RWC-1 and reports its basic performance. The I/O system mainly consists of external I/O devices, an I/O controller on a RICA-1 (Processing Element chip), and its two-layered dedicated interconnection network. RWC-1 can transfer I/O data through the I/O system without interfere with instruction executions. We have evaluated performance of I/O system. The results show I/O system has enough performance for I/O data transfer.

1. はじめに

近年，社会が高度に情報化するにつれて計算機が扱う問題の規模が大きくなってきており，それにともない計算機に求められる性能もますます高くなってきている．こうした中，計算機の性能向上のために計算機アーキテクチャ，特に超並列処理の発展が重要とされている．一方，超並列計算機が大規模な問題を扱うにあたっては，大規模なデータの入出力を実時間で処理することが要求され，従来の入出力機構だけではこれに対応しきれないと考えられる．そこで，演算処理のみならず入出力も並列に処理されるような新しい入出力機構の研究が進められている[1),2)]．

著者らは，要素プロセッサ数にして 1,000 台規模の超並列計算機 RWC-1 を研究開発している．RWC-1 は広範囲のアプリケーションを効率良く実行することを目的とした汎用の超並列計算機であり，特に細粒度の超並列処理を効率良く実現するために，高速の通信機構と高性能な相互結合網を装備している[3)~5)]．同時に，動画像などの大量のデータを決められた時間内に処理するようなアプリケーションに対応するため，演算系の通信と独立した入出力処理を行うことで高い入出力性能を得られる入出力機構を実装している[6)]．

本論文では，8 台の要素プロセッサで構成した小規模なシステムを用いて，RWC-1 の入出力機構についての基礎評価を行い，結果を報告する．以下，2 章で超並列計算機の入出力機構に必要な要件を示し，3 章では RWC-1 の入出力機構について述べる．さらに，4 章では RWC-1 の入出力機構についての評価と考察

† 技術研究組合新情報処理開発機構
Real World Computing Partnership
☆ 現在，三洋電機株式会社
Presently with SANYO Electric Co., Ltd.
☆☆ 現在，日本電気株式会社
Presently with NEC Corporation
†† 三菱電機株式会社先端技術総合研究所
Advanced Technology R&D Center, Mitsubishi Electric Corp.
††† 筑波大学電子・情報工学系
Institute of Information Science and Electronics, University of Tsukuba
☆☆☆ 現在，東京大学
Presently with University of Tokyo

を行う.

2. 超並列計算機における入出力機構の要件

入出力処理は，ディスクや動画像入出力装置などプロセッサの外部にある入出力装置が対象となるため，一般に演算処理とは非同期に発生する．また，このときに転送される入出力データは大容量であることが多い．このため，従来はプロセッサの操作を介さずに入出力装置とメモリとの間で直接データ転送を行うDMA操作が行われてきた．これにより，大規模な入出力処理に対するプロセッサの負荷を軽減することができ，演算性能を低下させずに入出力を行うことが可能となる．超並列計算機においても同様に，要素プロセッサの入出力処理に対する負荷を軽減することは重要であり，個々の要素プロセッサがDMA操作による入出力処理を行うことが望ましい．CP-PACS[7]，JUMP-1[2]，CM-5[8]，Paragon[9]等の超並列計算機においては，それぞれ入出力処理にDMA操作を用いている．

また並列アプリケーションには，要素プロセッサが特定の入出力装置だけでなくすべての入出力装置にアクセスすること，あるいは全要素プロセッサが特定の装置にアクセスすることを前提とするものが数多く存在する[10]．したがって，汎用の超並列計算機では入出力装置を共有資源とし，すべての要素プロセッサがあらゆる入出力装置にアクセスすることを可能としなければならない．そのためには，特定の要素プロセッサのみが入出力装置を持つのではなく，結合網を通じて全要素プロセッサと入出力装置とを直接接続することが重要となる．

しかしこの場合，要素プロセッサと入出力装置の間で行われる入出力データの転送が，演算処理にともなうプロセッサ間通信（以下，単に通信と呼ぶ）と結合網上で衝突する場合が考えられる．入出力データは大容量であることが多いため，結合網上の特定の経路を長時間にわたって塞ぐこととなり，そこを経由する他の通信を阻害する．この通信の遅延はそのまま並列処理のオーバヘッドとなり，計算機全体の演算性能を低下させる．したがって，これを防ぐためには入出力処理と通信を同時に行っても衝突が起こらない，あるいは起こった場合にも性能低下が小さい結合網を考える必要がある．

さらに，超並列計算機ではシステムの規模が大きくなり，実装や検証に関する制約が厳しくなることから，複雑で大規模なハードウェアは得策ではない．したがって，入出力機構を実現する際にはハードウェアコストを小さく単純におさえることも重要である．

以上まとめると，

(1) 演算処理と独立した入出力処理が可能な機構
(2) 入出力と通信との干渉が小さい結合網
(3) コストが小さく単純なハードウェア

が，超並列計算機における入出力機構の要件となる．

3. RWC-1の入出力機構

超並列計算機においては，プロセッサ間の通信・同期によって生じるオーバヘッドをいかに小さくするかが性能向上の鍵である．著者らは通信と演算を高度に融合したアーキテクチャであるRICA（Reduced Interprocessor-Communication Architecture）[11]を提案し，このRICAを実現する超並列計算機RWC-1とその要素プロセッサチップであるRICA-1を独自に開発した[4),12]．本章では前述の超並列計算機における入出力機構の要件をRWC-1でどのように実現したかを述べる．

3.1 入出力用結合網

前章に示した要件(2)である「入出力と通信との干渉が小さい結合網」に関して，RWC-1では特に以下の2点の理由によりその要求が厳しい．

- 演算処理において粒度の細かい通信が頻出するため，結合網上で入出力と通信が干渉した場合に大容量の入出力データが通信を阻害し，演算性能の大幅な低下を招く．
- 動画像など時間依存性が高い入出力データを転送する必要があるため，通信による結合網の混雑度に関係なく入出力データの転送時間を保証しなければならない．

よって，RWC-1では1つの結合網上で入出力と通信をいずれも効率良く行うのは困難であると考え，入出力用の入出力網と通信用の演算網とを独立に用意し，互いに干渉することなく入出力と通信が行えるようにした．

RWC-1の演算網には低レイテンシでランダム転送に強い性質が求められるため，これを解決する独自の網トポロジを採用している[5]．一方，入出力網に求められる性質は動画像データのように大容量のデータを制限時間内に転送するスループットの高さであり，レイテンシやランダム転送の性能は演算網ほど重要でない．そこで，入出力網のトポロジとして

- 次数が小さく少ないピン数で必要な転送容量を実現できる
- ノード内で複雑なスイッチが不必要で高速化しやすい

性質を持つリング網を採用し，要件(3)である「コスト

が小さく単純なハードウェア」でシステムを構築した．また隣接する演算ノードを直接，相互に接続し，リング網上を転送するデータの速度は動画像（秒30フレーム，512×512画素，24 bit フルカラー）の約20 MB/s に十分耐えうるよう50 MB/s とした．RICA-1 ではデータバス幅16 bit，転送クロック25 MHz（システムクロックの1/2）でこれを実現した．

3.2 入出力網の構成

リング網は構成するノード数が少ない場合には良好な転送性能を持つが，ノード数が多くなると網の直径および平均距離が増加して転送性能が悪化するため，1,024台までの演算ノードを想定している RWC-1 にはこのままでは適用しにくい．また，RWC-1 では計算資源を有効に利用するために要素プロセッサをいくつかのグループに分ける空間分割を考えているが[13]，分割領域ではそれぞれの独立性を保証する必要があり，入出力データが境界を越えて他の領域に影響を与えることは許されない．

以上の問題を解決するため，RWC-1 では通信の局所性を利用して要素プロセッサを空間分割の最小構成である8つごとにクラスタ化し，入出力網を上位，下位の2つに階層化した（**図1**）．以下，RWC-1 の下位入出力網をリングバスと呼ぶ．

リングバスには高いスループットが求められているために，複数の演算ノードからの転送要求を処理しリングバス上の転送を効率良く制御する機能，転送の経路が重複しない限り同時並行して転送することが可能な複数同時転送機能，ブロードキャスト機能などを組み込む必要がある．これを実現するために，RWC-1 では演算ノードとは別にリングバス上の転送を制御するアービタを用意した．アービタにはリングバスを構成する全ノードの宛先情報が集められ，即座に複数同時転送やブロードキャストが可能かどうかの判断を行う[6]．

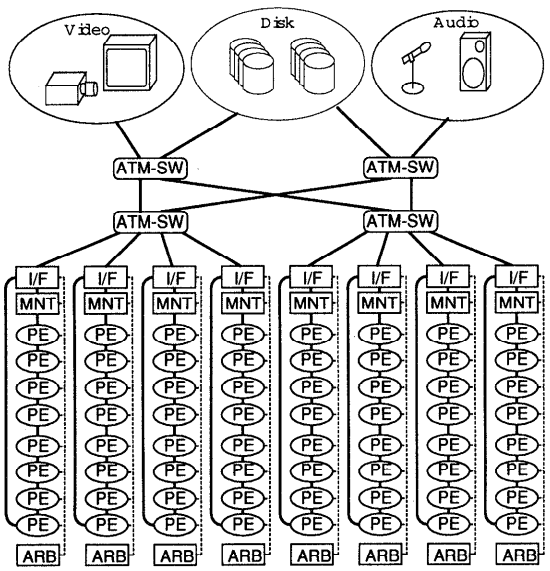


図1 RWC-1 の入出力網
Fig. 1 RWC-1 I/O network.

リングバスには，演算ノードのほかに上位入出力網と下位入出力網を接続する中継ノード（図1中の ATM I/F node）が接続されている．中継ノードはアービタの制御下で要素プロセッサとデータ転送を行い，上位入出力網から受け取ったデータを要素プロセッサに転送し，要素プロセッサから転送されたデータを上位入出力網に送出する[14]．

上位入出力網は，この中継ノードと入出力装置，あるいは中継ノードどうしを相互に接続するものである．空間分割を実現するため，上位入出力網においても異なる分割領域の入出力どうしが干渉することは許されない．そのため上位入出力網には，任意の経路を通るデータ転送どうしが干渉することなく，またシステム構成の柔軟性が高いクロスバスイッチを組み合わせた間接網を用いた[1]．下位入出力網と同様に上位入出力網も高いスループットが必要であり，また実装上入出力装置を離れた場所に設置することが有利であるため，上位入出力網には光接続を用いることとした．さらに規格化された技術を用いると市販の入出力装置やスイッチ等をそのまま使用することができ，汎用性・接続性に優れたシステムを構築できることから，上位入出力網には ATM 規格を採用した．

3.3 RICA-1 の入出力機構

RICA に基づいて高い演算性能を実現する RWC-1 には，それに見合うだけの入出力性能が要求される．また，RWC-1 上で実行されるアプリケーションの中には大容量のファイルや動画像データを制限時間内に転送するなど，高速の入出力処理を要求するものも多い．演算性能を維持したままこれらの入出力処理を行うためには，超並列計算機の入出力機能の要件(1)である「演算処理と独立した入出力処理が可能な機構」が必要である．これを実現するために，RWC-1 でも入出力機構に DMA 機能を導入することにした．RWC-1 は分散メモリ型の並列計算機であるため，演算ノードごとに DMA 操作を独立して行えるよう RICA-1 の入出力機構に DMA 機能を内蔵した（**図2**）．これにより複数の入出力装置と RICA-1 の組で同時に DMA 操作を行うことを可能にした[6]．

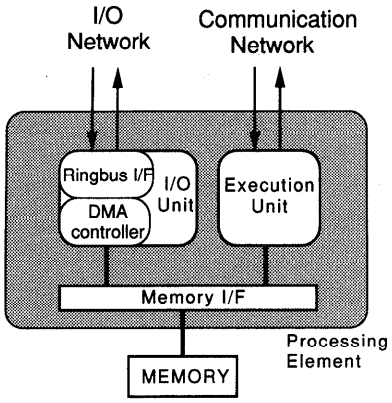


図2 RICA-1の入出力機構
Fig. 2 I/O unit of a processing element.

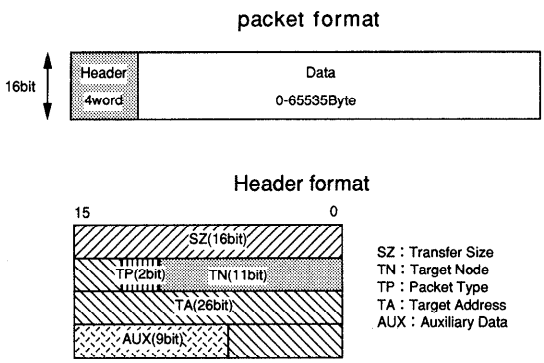


図3 転送データのフォーマット
Fig. 3 I/O packet format.

RICA-1の入出力機構はDMA機能とともに，アービタとの間で必要な情報をやり取りし転送を制御する機能や，高速，低レイテンシを実現するためにデータをパイプライン的に転送する機能を持っている．これらのリングバス・インタフェース機能は演算処理とは完全に独立して動作する．

演算ノードが入力動作を行う場合，入出力装置から読み出されたデータは入出力網を通して演算ノードに転送され，DMA機能により自動的にメモリに書き込まれる．逆に出力動作の場合，RICA-1の専用命令により入出力機構を起動することでデータが自動的にメモリから読み出され，結合網を通して入出力装置に転送される．いずれの場合も演算ノードは割込み通知により転送終了を知ることができる．

RWC-1の入出力は，宛先などが格納されたヘッダと可変長のデータ本体からなる入出力パケットによって行われる（**図3**）．入出力パケットの1回あたりの転送量は大容量の入出力データに対応するため最大64Kバイトとした．

4. 性能評価

本章では，前述のRWC-1の入出力機構が当初の構想どおり超並列計算機の入出力機構の要件を満たしているかという観点から，基本的な性能評価を行う．これまでにハードウェア・エミュレータおよびシミュレータを用いた評価を行っているが[15]，ここでは実機によるより高い精度の評価を行った．今回の評価では，特にRICA-1の入出力機構と下位入出力網に着目し，8プロセッサによる小規模システムで評価した☆．評価項目は以下のとおりである．

(1) 演算ノードがDMA機能を内蔵したことによる効果の検証として，入出力処理の演算性能への影響の評価
(2) 入出力網と演算網を分離したことによる効果の検証として，入出力が通信に与える影響の評価
(3) 入出力網と演算網を分離したことによる効果の検証の第2として，通信が入出力に与える影響の評価
(4) 下位入出力網の検証として，リングバスのデータ転送特性の評価

4.1 入出力処理の演算性能への影響

RICA-1には演算処理機構と独立して入出力処理機構を実装しており両者は並行して動作するが，メモリパスを共有していることやRICA-1からの入出力データ送出の起動にソフトウェア動作が必要なことから，まったく影響を与えないわけではない．そこで，入出力処理が演算処理に対してどの程度の影響を与えるか，逆に演算処理が入出力処理にどの程度の影響を与えるかの評価を行った．

評価では，1つのRICA-1上で入出力処理と演算処理双方が同時にメモリアクセスを行う場合の，競合による入出力性能と演算性能の変化を測定した．演算処理側のメモリアクセスとして，メモリロードあるいはストアを行うスレッドとメモリアクセスを行わないスレッドの2種類を，メモリアクセス量を変えるために実行数の比率を変えながら繰り返し実行した．この際，キャッシュの影響をなくすように，つねにミスヒットあるいはライトバックが起こるようなメモリアクセス・パタンを用いた．一方，入出力処理側のメモリアクセスとして，256バイトまたは64Kバイトの固定長の入出力データを連続して入力または出力した．双方の性能として一定時間内に実行されたスレッド数と入出

☆ なお，現在RWC-1は調整中のため測定はシステムクロック33 MHzで行ったが，評価結果の値は最終的に動作予定の50 MHzとして換算した．

力データの転送速度を計測した（図 4）．

評価の結果を図 5，6，7，8 に示す．横軸はすべて演算処理のメモリアクセス量であり，図 5，6 の縦軸は，スレッド処理性能として一定時間内のスレッド実行数（入出力と衝突しない場合を 1 とする）を示す．また図 7，8 の縦軸はデータ転送速度である．

図 5，6 によると，スレッド処理性能はロード，ストアのいずれの場合にも，メモリアクセス量が多くなるほど低下する．これは演算処理のメモリアクセス要求が入出力処理のメモリアクセス中に衝突した場合，その終了を待つためにスレッド実行時間が長くなるからである．このためメモリアクセス量の増加につれて，入出力と衝突する確率が高くなり，スレッド処理性能が低下する．また，256 バイトデータの入出力は 64 K バイトに比べてメモリパスの占有時間が短いために，メモリアクセス量の増加による性能低下の割合は小さいと考えられる．

スレッド処理性能はロードの場合には単調に減少していくのに対して，ストアではメモリアクセスが 120 MB/s 付近から急激に減少している．これは，ロードでは入出力処理との衝突がそのままスレッド実行時間の増加となるのに対して，ストアは命令実行完了を待たずに次の命令を実行できるため，少ない頻度では入出力処理と衝突しても影響が現れないからであると考えられる．また，ストアにおけるキャッシュのライトバック動作では書き込みと読み出しの 2 回のメモリアクセスが生じるため，ミスヒット時に 1 回のみ読み出しを行うロードに比べてスレッド処理性能の減少の

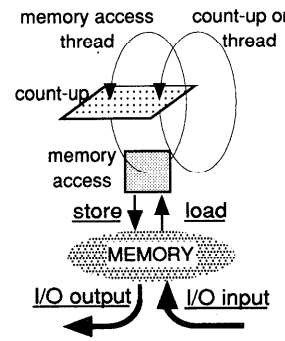


図 4 2 種類のスレッド実行と入出力処理
Fig. 4 Memory access by thread and I/O.

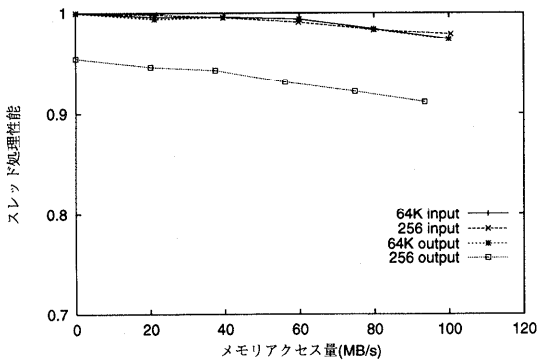


図 5 メモリアクセス量とスレッド処理性能（ロード）
Fig. 5 Performance of thread execution (load).

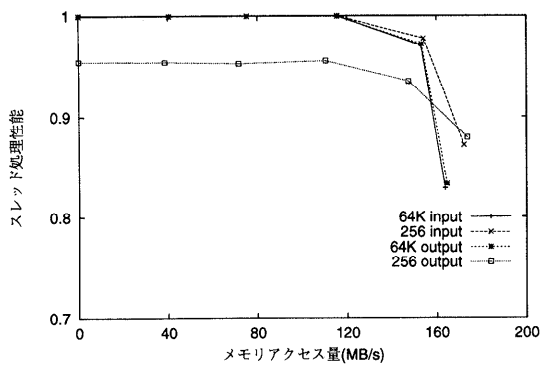


図 6 メモリアクセス量とスレッド処理性能（ストア）
Fig. 6 Performance of thread execution (store).

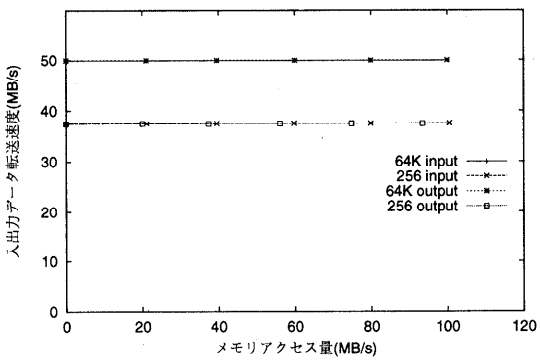


図 7 メモリアクセス量と I/O 転送速度（ロード）
Fig. 7 Performance of I/O (load).

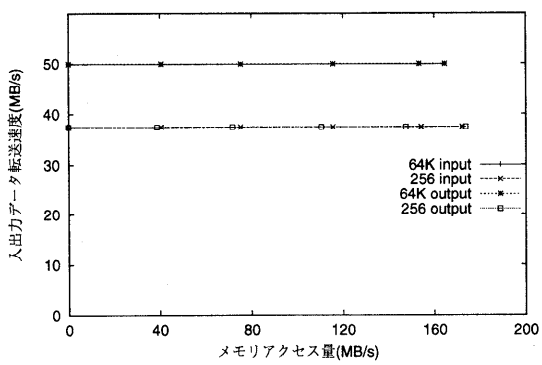


図 8 メモリアクセス量と I/O 転送速度（ストア）
Fig. 8 Performance of I/O (store).

比率が大きい．

さらに，出力時のスレッド処理性能はメモリアクセス量が0の場合にも低下する．これは入力動作の場合はDMA操作のみでソフトウェア処理をともなわないのに対して，出力動作の場合には処理の開始のためにソフトウェア処理が必要であり，1回の出力動作が終了するごとに割込み処理が行われ，スレッドの実行が中断されるからであると考えられる．このオーバヘッドは一定であるため，転送サイズが小さいほど性能の低下は大きくなる．実際，図5，6において，256バイトの出力時の性能低下が64Kバイト出力時に比べて大きくなっている．

図7，8に演算処理にともなうメモリアクセス量と入出力処理性能の関係を示す．ここでは入出力のデータ転送速度が演算処理のメモリアクセス量と種類とにかかわらず変化が見られない．これは，入出力処理においても演算処理時と同様に衝突が起こった場合にはメモリアクセスが待たされるが，メモリバンド幅（400 MB/s）がリングバスの転送速度（50 MB/s）より十分大きく，入出力処理用のバッファにより吸収されるためと思われる．

最後に単純ループによるRICA-1のメモリアクセス量の最大値は，ロード103 MB/s，ストア197 MB/sであり，測定結果の最大値とほぼ等しい．したがって，この評価で得られた値がRICA-1における入出力処理と演算処理の干渉による性能低下の最大値であると考えられる．また，実際の応用ではキャッシュの効果やメモリアクセス以外の処理が増えるため，この評価ほど演算性能が低下することはないと予想される．

以上の結果をまとめると，

- 演算性能が入出力処理により受ける影響は小さく，その性能の低下は最大でも17%にすぎない
- 入出力性能の演算処理による低下は見られない

ことが明らかになった．したがって，RICA-1の入出力機構にDMA機能を導入した効果として，双方の性能を低下させることなく入出力処理と演算処理とを並行して行えることが実証された．

4.2 入出力が通信に与える影響

RWC-1では入出力と通信が相互に干渉するのを防ぐために結合網を入出力網と演算網の2つに分割している．このことがどの程度の効果を上げているかを調べるために，演算網上に演算にともなう通信のための演算パケットだけでなく同時に入出力データを転送するための入出力パケットを混在させ，演算パケットの転送速度への影響を評価した．

まず，演算網上で入出力パケットと演算パケットの経路が重なるように，それぞれの送受信を行う2組の演算ノードを選ぶ（**図9**）．演算パケットとして1ワードのパケットを演算網の最大転送速度の頻度で繰り返し転送し，同時に入出力パケットとして8ワードのパケットを転送し，一定時間内に到着した演算パケット数を測定する．こうして，演算パケット数が入出力パケットの転送によりどの程度減少するかを，入出力側の転送速度を変化させることにより測定した．また，入出力データと演算パケットを同時に転送する同様の評価を，入出力データを入出力網経由で転送した場合にも行った．

評価の結果を**図10**に示す．横軸は入出力パケットによる入出力データの転送速度であり，縦軸は一定時間内に演算網を経由して到着した演算パケットの数であるパケット転送速度を表す．これによると，入出力データの転送速度が大きくなるにつれパケット転送速度は低下し，入出力網の転送速度と同じ50 MB/s付近ではパケット転送速度は2割以上減少する．演算網のデータ転送速度の最大値は8ワードのパケットを用いた場合で246 MB/sであるから，この値は入出力データの転送速度の50 MB/sが演算網全体転送速度に占める割合である20%とほぼ一致する．これは演算

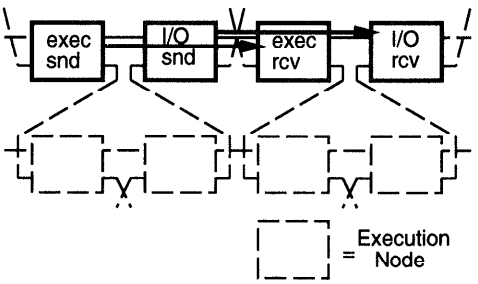


図9 演算網上の入出力と演算パケットの経路
Fig. 9 I/O data path and execution data path.

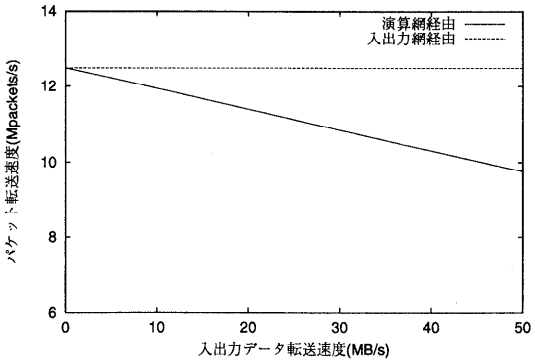


図10 入出力データ転送による通信性能への影響
Fig. 10 Packet transfer with I/O on the same network.

パケット数の低下が演算網の競合が原因であることを示している．

一方，入出力データを入出力網経由で転送した場合にはパケット転送速度の低下は見られない．これにより入出力網を演算網と分離したことで，入出力処理が演算網上の通信にまったく影響を及ぼさないことが確認できた．

4.3 通信が入出力に与える影響

単一の結合網で入出力と通信を行うことは，入出力が通信に影響するだけでなく，逆に通信も入出力を阻害する．そこで，演算網上で入出力データを転送した際に経路上で通信による網の混雑がある場合，それが入出力性能にどのような影響を与えるかを調べた．

まず前節の評価と同様に，1 ワードデータからなる演算パケットと 8 ワードデータからなる入出力パケットの送受信を行う 2 組の演算ノードを転送経路が重なるように選ぶ．入出力パケットの送信側は入出力網の転送速度と等しい 50 MB/s で，演算パケットの送信側も一定の速度（2.78 Mpackets/s）でパケットを送出する．そのうえで，演算ノードの一定時間に処理できる演算パケット数（パケット処理能力）を変化させた場合に，入出力データの転送速度がどのように変化するか調べた．また，入出力データを演算網ではなく入出力網経由で転送した場合も同様に測定した．

評価の結果を**図 11** に示す．横軸は演算ノードのパケット処理能力であり，送信側のパケット転送速度と同じ場合を 1 とする．縦軸は入出力データの転送速度である．

演算網のパケット転送速度は経路上の最もパケット処理能力の低い演算ノードにより制限される．したがって，演算ノードのパケット処理能力が 1 よりも低下すると，演算網上のパケット転送速度も低下する．グラフでは，演算ノードのパケット処理能力が低下してもしばらくは入出力の転送速度は低下しないが，これは演算網の転送能力が低下しても入出力の転送速度が確保できているからである．しかし演算ノードの処理性能がさらに低下すると，演算網の転送速度が入出力に必要な値を下回り入出力の転送速度も急激に低下する．一方，入出力網を通して入出力データの転送を行った場合には転送速度の低下が見られず，演算網の混雑の影響を受けないことが確認された．

4.4 リングバスのデータ転送特性

RWC-1 の入出力網において下位階層に採用したリングバスのデータ転送特性を測定した．RWC-1 では入出力機構を演算ノード間のメモリ-メモリの DMA 転送にも利用できるため，リングバス上では演算ノードと中継ノードとの転送だけでなく，演算ノードどうしの転送も行われる．ここでは，(1) 1 つの演算ノードに他の演算ノードがデータを転送する場合，(2) 1 つの演算ノードが他の 7 つの演算ノードにデータを転送する場合，(3) 8 つの演算ノード間でメモリの内容を相互に交換する complete exchange を行う場合の 3 種類についての転送速度を，転送サイズを変えながら測定した．また，リングバス上で転送経路が重ならない場合には複数同時転送ができ，さらに同じデータを転送する場合にはブロードキャスト機能が利用可能であるため，上記の (2)，(3) の評価ではそれらを利用した場合も測定した．

図 12，13 に評価結果を示す．横軸は転送サイズ，縦軸はリングバス全体の転送速度である．いずれの場合も，転送サイズが大きくなるほど転送速度が大きくなり，転送サイズが 2 K バイトを超えるとリングバスの最大転送速度である 50 MB/s の 90%以上となる．これは，転送にともなうオーバヘッドが転送サイズにかかわらず一定であるためである．また複数同時転送機能を使うことにより，(2) では 7/3 倍，(3) では 4

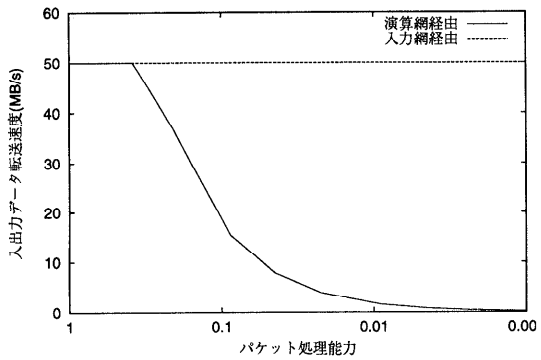


図 11 演算網の混雑による入出力転送性能への影響
Fig. 11 I/O with execution packets on the same network.

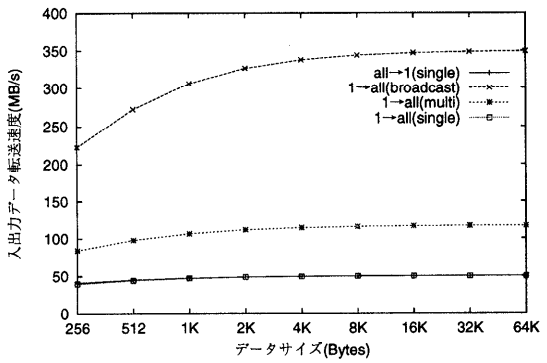


図 12 リングバスのデータ転送性能（all → 1，1 → all）
Fig. 12 Throughput of ringbus (all → 1, 1 → all).

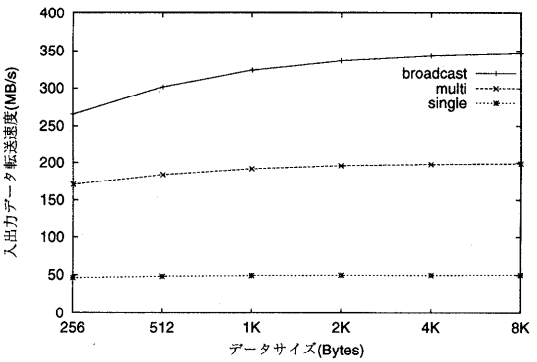


図13 リングバスのデータ転送性能（complete exchange）
Fig. 13 Throughput of ringbus (complete exchange).

倍の転送速度を得ることができ，さらにブロードキャスト機能は一度に7つの演算ノードにデータを転送するため，転送速度を見かけ上7倍にできることが示された．

以上の評価により，入出力や演算ノード間のメモリ転送のような大容量のデータ転送において，リングバスは十分な性能を示すことが実証された．さらに，リングバスの特長である複数同時転送機能やブロードキャスト機能により，大きな性能向上が得られることが明らかになった．

5. おわりに

本論文では，超並列計算機RWC-1の入出力機構について述べ，その基礎評価を行った．RWC-1の入出力機構は要素プロセッサに内蔵され，演算処理と独立して入出力処理を行うためにDMA機能と入出力専用インタフェースを備えている．さらに入出力データの転送が，演算にともなうプロセッサ間通信と干渉することを避けるために，演算網と入出力網とを分離している．この入出力網を空間分割への対応，ハードウェアコストの削減，局所性の利用などの理由から2つに階層化しており，下位階層にはリングバスを，上位階層にはクロスバスイッチによる間接網を採用した．

これらRWC-1の入出力機構の性能を，8台の要素プロセッサで構成した小規模システムを用いて評価した．その結果，要素プロセッサ上での入出力処理に対する演算性能の低下は最大でもわずか17%であり，逆に入出力性能は並行する演算処理の影響をまったく受けないことを確認した．また，入出力網を演算網と分離することにより，他の要素プロセッサの演算処理の効率をまったく落とすことなく入出力データの転送を可能にしており，かつ入出力性能が演算の輻輳の影響を受けないことを確認した．さらに下位階層に採用されたリングバスは大容量のデータ転送においても十分な性能を持ち，複数転送やブロードキャストの機能によりさらに高い転送性能が得られることを確認した．

RWC-1は現在，8～16台構成の小規模なシステム構成で性能評価を行っている．本論文ではこのシステムを用いて，RICA-1の入出力性能，および下位入出力網の基礎評価を行ったが，残る上位入出力網を含めた入出力網全体の評価，動画像処理等を含めた実アプリケーションによる評価については，128台の要素プロセッサシステムであるRWC-1基本部の完成後，改めて行う予定である．

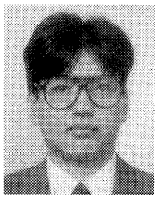
謝辞 本研究の機会を与えていただいた島田潤一RWC研究所長，有益なご指導，ご討論をいただいた佐藤並列分散パフォーマンス研究室長はじめTRC並列分散研究部各研究室の室長および室員の諸氏，ならびに並列応用三洋研究室の諸氏に深く感謝いたします．

参考文献

1) 大西 正，北村 徹，大上靖弘，清水雅久：分散独立型入出力システムのための結合網の構成と評価，情報処理学会研究会報告，ARC-111-6, pp.41–48 (1995).
2) 大谷 智，中條拓伯，金田悠紀夫：超並列計算機JUMP-1における並列I/Oシステムのシミュレーションによる評価，並列処理シンポジウムJSPP '96, pp.283–290 (1996).
3) 坂井修一，岡本一晃，松岡浩司，廣野英雄，児玉祐悦，佐藤三久，横田隆史：超並列計算機RWC-1の基本構想，並列処理シンポジウムJSPP '93, pp.87–94 (1993).
4) 岡本一晃，松岡浩司，廣野英雄，横田隆史，坂井修一：超並列計算機におけるマルチスレッド処理機構と基本性能，情報処理学会論文誌，Vol.37, No.12, pp.2398–2406 (1996).
5) Yokota, T., Matsuoka, H., Okamoto, K., Hirono, H. and Sakai, S.: hMDCE: The Hierarchical Multidimensional Directed Cycles Ensemble Network, *IEICE Trans. Inf. & Syst.* Vol.E79-D, No.8, pp.1099–1106 (1996).
6) 廣野英雄，松岡浩司，岡本一晃，横田隆史，坂井修一：RWC-1の入出力機構と基本性能，情報処理学会研究会報告，ARC-119-34, pp.197–202 (1996).
7) 中村喜三郎，朴 泰介，中村 宏，中田育男，山下義行，岩崎洋一：CP-PACSのアーキテクチャの概要，情報処理学会研究会報告，ARC-108-9, pp.57–64 (1994).
8) Thinking Machine Corporation: The Connection Machine CM-5, Technical Summary (1991).
9) Intel Corporation: Paragon XP/S, Product

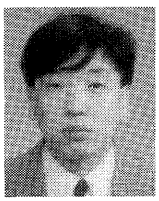
Overview (1991).
10) Oue, Y., Kitamura, T., Ohnishi, K. and Shimizu, M.: Parallel file access for dynamic load balancing on the massively parallel computer, *Proc. Int'l. Symp. on Parallel and Distributed Supercomputing*, pp.179–187 (1995).
11) Sakai, S., Okamoto, K., Kodama, Y. and Sato, M.: Reduced interprocessor-communication architecture for supporting programming models, *Proc. Programming Models for Massively Parallel Computers 1993*, pp.134 143 (1993).
12) 松岡浩司，岡本一晃，廣野英雄，横田隆史，坂井修一：RWC-1の要素プロセッサ—細粒度並列処理機能の強化，情報処理学会研究会報告，ARC-119-43, pp.251–256 (1996).
13) 堀 敦史，石川 裕，小中裕喜，前田宗則，友清孝志：超並列システムカーネル SCore の構想，情報処理学会研究会報告，OS-61-8, pp.57–64 (1993).
14) 廣野英雄，松岡浩司，岡本一晃，横田隆史，坂井修一：RWC-1の入出力用ATMノード，第52回情報処理学会全国大会論文集，pp.(6)153–154 (1996).
15) 廣野英雄，松岡浩司，岡本一晃，横田隆史，坂井修一：超並列計算機RWC-1の入出力処理の評価，並列処理シンポジウムJSPP '97, pp.101–108 (1997).

(平成 9 年 11 月 7 日受付)
(平成 10 年 4 月 3 日採録)



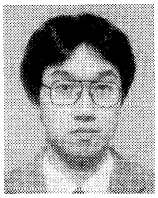
廣野 英雄

1968年生．1991年筑波大学第三学群基礎工学類卒業．同年三洋電機（株）に入社．1992年10月より1998年3月まで（技組）新情報処理開発機構に出向．計算機アーキテクチャ，特に入出力機構の研究に従事．ICCD Outstanding Paper Award（1995年）受賞．電子情報通信学会会員．



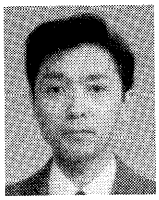
松岡 浩司（正会員）

1961年生．1984年東京工業大学工学部電気電子工学科卒業．1986年同大学院理工学研究科電子物理工学専攻課程修了．同年日本電気（株）に入社．1992年10月より1998年3月まで（技組）新情報処理開発機構に出向，主任研究員．現在，計算機システム全般，特にプロセッサアーキテクチャの研究に従事．ICCD Outstanding Paper Award（1995年）受賞．



岡本 一晃（正会員）

1962年生．1986年慶應義塾大学理工学部電気工学科卒業．同年三洋電機（株）に入社．主にデータ駆動計算機を中心とする並列処理アーキテクチャの研究に従事．1992年10月より1998年3月まで（技組）新情報処理開発機構に出向，主任研究員．超並列アーキテクチャの研究に従事．ICCD Outstanding Paper Award（1995年）受賞．



横田 隆史（正会員）

1960年生．1983年慶應義塾大学工学部電気工学科卒業．1985年同大学院工学研究科電気工学専攻修士課程修了．同年三菱電機（株）に入社．知識処理向けアーキテクチャおよび並列アーキテクチャの研究に従事．1993年12月より1997年3月まで（技組）新情報処理開発機構に出向，超並列アーキテクチャ，相互結合網の研究に従事．ICCD Outstanding Paper Award（1995年）受賞．工学博士（1997年）．電子情報通信学会会員．

坂井 修一（正会員）

1958年生．1981年東京大学理学部情報科学科卒業．1986年同大学院情報工学専門課程修了．工学博士．同年，電子技術総合研究所入所．1991年4月より1年間米国MIT招聘研究員．1993年3月より1996年2月までRWC超並列アーキテクチャ研究室室長．1996年10月より1998年3月まで筑波大学助教授（電子情報工学系）．1998年4月より東京大学助教授（工学系研究科）．計算機システム一般，特にアーキテクチャ，並列処理，スケジューリング問題などの研究に従事．情報処理学会論文賞（1991年），日本IBM科学賞（1991年），市村学術賞（1995年），ICCD Outstanding Paper Award（1995年）等受賞．IEEE，ACM，電子情報通信学会各会員．