# RWC-1の入出用ATMノード

5L－6
廣野 英雄　松岡 浩司　岡本 一晃　横田 隆史　坂井 修一
新情報処理開発機構 つくば研究センタ

## 1 はじめに

我々は、プロセッサ数にして1000規模の超並列計算機RWC-1を研究開発している。RWC-1は、RWCPで研究中の新しいアプリケーション群の実行母体となるため、計算能力はもちろんのこと、入出力にも高い性能が要求されている。このためRWC-1の入出力系には、演算と入出力を並列動作させるための様々な工夫が施され、大量のデータを限られた時間内に転送することができる[1]。特に、入出力専用に高スループットの結合網を設けたことにより、演算用の通信に関わらず入出力データの転送時間を保証することが可能となった。

RWC-1の入出力用結合網は、局所性の利用と実装上の制約から2レベルに階層化されている(図1)。下位階層は、8つの要素プロセッサ(PE)をリングバスで結合した構成をとっており、上位階層はこれを128個、ATMスイッチ網で結合した構成をとっている。ATMスイッチ網には、ディスクシステムや音声・画像インタフェースなどが接続され、下位階層と上位階層はATMノードと呼ばれる通信制御機構によって接続される。

本稿では、ATMノードについて、ハードウェア構成と動作の概略を述べ、RWC-1の入出力の手順を示す。

## 2 ATMノード

RWC-1における入出力動作は、ヘッダとデータ本体からなる転送データをノードからノードへ転送することにより行なわれる。転送データのヘッダには宛先、データ長等が格納されており、データ本体は可変長である。下位階層において、転送データはリングバス上を塊のまま転送される。一方上位階層では、転送データはATMプロトコルに従い、多数のATMセルという形で転送される。

ATMノードは、このような入出力用結合網の上位、下位階層のデータ転送プロトコルの違いに対応するために、転送データの分解・再構成を行う。またATMスイッチ網では輻輳が生じた時にセルの廃棄が行なわれるが、その際の再送制御もATMノードで行なわれる。

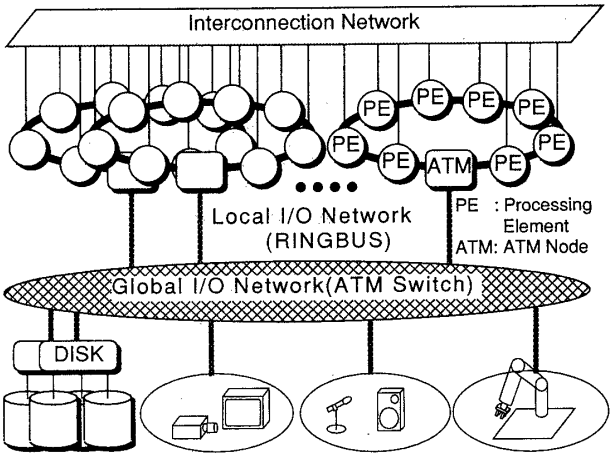


図1: RWC-1の入出力系

### 2.1 ATMノードの構成

ATMノードは、(1)リングバスを通じてPEと通信を行なうRingbusI/F、(2)ATMスイッチ網への送信を行なうATM-S、(3)ATMスイッチ網からの受信を行なうATM-R、(4)ノード全体の制御を行なうControllerの4つの部分からなる(図2)。

ATM-S(-R)は、FIFO型バッファであるRingBuf、2ポートRAMであるATM Buf、ATM Bufの書き込み(読みだし)番地を生成するAdrGen、ATMスイッチ網への送信(ATMスイッチ網からの受信)を行なうATM Send(Receive) I/F、ATM Send(Receive) I/Fの制御を行なうCtlRamからなり、PCIバスを中心に構成されている。

ControllerはPCIバス上の商用CPUを中心としたシステムであり、PCI-PCIブリッジを経由してATM-S、ATM-Rと接続されている。PCI-PCIブリッジを経由したPCIバス同士の接続では、対象となる番地が相手側にある時のみPCIバス間に調停が行なわれ、1つのPCIバス上に接続されているように振舞うが、それ以外の場面では別々のPCIバスとして動作する。これによりATM-S、ATM-R、Controllerはそれぞれ独立した動作が可能となっている。

---

I/O ATM node for the Massively Parallel Computer RWC-1
Hideo HIRONO Hiroshi MATSUOKA Kazuaki OKAMOTO
Takashi YOKOTA Shuichi SAKAI
Tsukuba Research Center, Real World Computing Partnership
Mitsui Bldg.16F, 1-6-1 Takezono, Tsukuba, Ibaraki 305, Japan

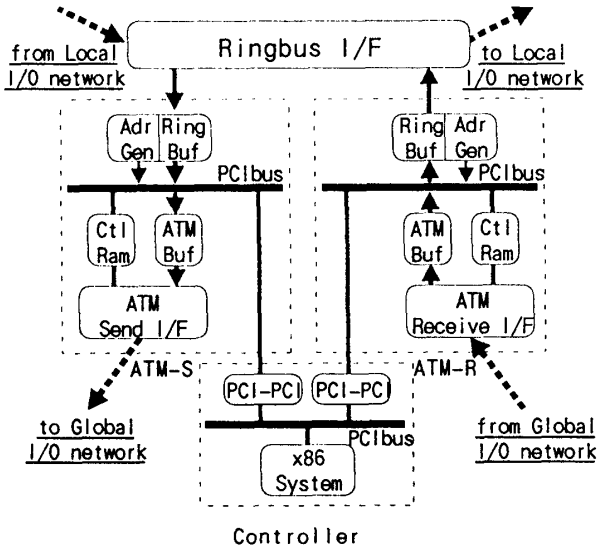


図 2: ATM ノード

## 2.2 ATM スイッチ網への送信動作

ATM-S での送信動作は以下のように行なわれる。

1. PE ノードから送信されたデータは、RingbusI/F を経て RingBuf に蓄えられる。
2. これと同時に ATM-S は Controller に対して転送開始割り込みをかける。
3. Controller は ATM Buf 上の予め確保されている領域の開始番地を AdrGen に設定する。
4. AdrGen は DMA を開始し、RingBuf 上のデータを ATM Buf に書き込む。
5. PE ノードからの転送が完了し、RingBuf が空になると、ATM-S は Controller に対して転送終了割り込みをかける。
6. Controller は ATM Buf 上にある転送データのヘッダから宛先、データ長を読み出し、CtlRam に ATM Buf の読みだし番地、データ長、宛先等を設定する。
7. ATM SendI/F は CtlRam の内容に従い ATM Buf から転送データを読みだし、ATM セルに分解して ATM スイッチ網に送信する。

ATM-S では、(1)RingbusI/F から RingBuf への書き込み、(2)RingBuf から ATM Buf への DMA 転送、(3) ATM セルの送出は重畳化されており、同時に実行可能である。

## 2.3 ATM スイッチ網からの受信動作

ATM-R での受信動作は以下のように行なわれる。

1. ATM スイッチ網から受信した ATM セルは、ATM ReceiveI/F によって、ATM Buf に格納され、再構成される。
2. 1 つの転送データ分の再構成が完了すると、ATM-R は Controller に対して転送開始割り込みをかける。
3. Controller は、CtlRam から転送データの開始番地等を読み出し、AdrGen に読みだし番地、データ長を設定する。
4. AdrGen は DMA を開始し、ATM Buf 上のデータを RingBuf に蓄える。
5. RingBuf 上のデータは RingbusI/F を経てリングバスに出力され、目的の PE ノードに到着する。
6. DMA 転送が終了すると、ATM-R は Controller に対して転送終了割り込みをかける。
7. Controller は次の転送データの用意が出来ていればAdrGen にその読みだし番地、データ長を設定する。

ATM-R では、(1)ATM ReceiveI/F の ATM Buf への書き込み、(2)ATM Buf から RingBuf への DMA 転送、(3)RINGBUS への出力は重畳化されており、同時に実行可能である。

## 3 転送性能

現在の実装では、リングバスの転送幅は 16bit であり、25MHz で動作する。したがって、ATM ノードとリングバス間の転送速度は 50MB/s となる。一方 ATM スイッチ網の転送速度は、現在入手可能なチップセットにより約 20MB/s であるが、将来は高速のチップセットに置き換える予定である。

また、ATM ノードでは転送データを最低 1 個分内部で保持するため遅延が生じる。画像入出力等の時間依存性の高いデータ転送では、この遅延を考慮する必要がある。ただし遅延はデータ長により一意に決定し、また画像の転送周期に比べると十分小さいため、問題はない。

以上、ATM ノード内の各ブロックの独立性・重畳性によってデータ転送および整形は、入出力用結合網に完全に追随した動作を行なうことが可能になった。

## 4 おわりに

本稿では、RWC-1 の入出力用 ATM ノードについて、その構成と動作について述べた。ATM ノードは ATM スイッチ網への送信と ATM スイッチ網からの受信を並列に行なうことが可能である。またそれぞれの動作は重畳化されており、入出力用結合網の上位・下位階層のインタフェースとして機能的にも速度的にも十分なデータ変換能力をもつ。

現在 ATM ノードを試作中である。本ノードは 64 プロセッサからなる RWC-1 テストベッド 2 に実装される。ここで機能検証および性能の検証を行なった後、1024 プロセッサからなる RWC-1 上に実装される。

## 参考文献

[1] 廣野英雄、松岡浩司、岡本一晃、横田隆史、坂井修一、RWC-1 の入出力リングバス、情処計算機アーキテクチャ研究会、SWoPP95(1995),pp.121-128