

リアルワールドコンピューティングにおける 光インターコネクション

松岡浩司*・横田隆史**・西村信治***

実世界に対応した情報処理（リアルワールドコンピューティング）、例えば、動画像の認識や、その結果を用いた高速に移動する物体の制御では、行列によって記述された問題を高速に解く従来の科学技術計算とは違い、実世界から得た情報を行列の形に加工する（前処理と呼ばれる）、あるいは、こうした前処理を経ずに限られた時間内に問題を解くことが直接的な課題となる。

問題を解くための時間そのものを短くするためには、問題を細かな部分に分割し、より多くのプロセッサで並列に処理すればよい。ところが、プロセッサ数を多くすれば多くするほど、プロセッサ間での負荷の不均衡が生じ、並列処理の効率が低下してしまう。この傾向は実世界に対応した不規則なデータ構造をもった問題を扱う場合に特に顕著である。しかし、処理の単位をさらに細かくすれば、スケジューリングの機会が豊富となるので、負荷を分散させることが逆に容易になる可能性がある。

以上のような観点から、筆者らは超並列計算機 RWC-1^{1,2)}を開発し、細粒度並列処理の有効性の検証を行った。本稿では、まず、実世界に対応した情報処理を行う細粒度並列計算機を工学的な観点から概説し、その中における光インターコネクションの意義について述べる。

1. 並列計算機

限られた時間内に問題を解くため、プロセッサ数を多くしようとする、システムは物理的に大きくなり、例えば、リモートデータをアクセスした場合、システムの物理的な大きさに応じた大きな待ちが生じるようになる。この待ちが直接的にシステムの性能に影響を及ぼさないようにすることが、大規模な並列計算機において満足すべき重要な要件のひとつになる（この要件を満足できないと、ある程度の数を超えてプロセッサ数を増やしても、待ち時間

が長くなり、逆に性能が低下してしまう）。RWC-1はマルチスレッド処理を採用することによってこの要件を満足している。マルチスレッド処理では、待ちが生じた時点で処理の単位であるスレッドを切り替え、常にプロセッサが問題を解くために働くようにする。これによって、結果的に問題を解くために要する時間を短縮する。待ちを隠蔽するために必要とする数のスレッドがあれば、待ち時間の長さにシステムの性能が左右されることはない（他にもプリフェッチにより回避する方法などがあるが、ここでは説明を省略する）。

従来の並列計算機の多くは、基本的には、多数のプロセッサを接続することを前提としていないので、リモートアクセスなどの遅延の増大がシステムの性能に影響を及ぼす場合が多い。この場合、光インターコネクションを挿入すると、無視できない程度にシステムの性能が低下する可能性があった。したがって、従来は、短い配線で接続された複数のプロセッサからなるクラスターを光インターコネクションで接続する例（^光対間接続の例）があったに過ぎない。今回筆者らは、光インターコネクションの遅延そのものを小さくし、さらに、遅延の増大をマルチスレッド処理により隠蔽することにより、基本となった電気インターコネクションをもつシステムと同等の性能をもつ光実装並列計算機を実現した^{3,4)}。

2. 光インターコネクション

並列計算機では各プロセッサにデータが分散して配置されるため、プロセッサ間で頻繁にデータが交換される。このため、プロセッサ間の結合網には高いスループット（と、できるだけ短い遅延）が要求される。ところが、このスループット（単位時間あたりに転送するデータの量）の面でも、物理的な大きさが問題となり、電気インターフェースを用いた場合、ある上限を超えようとする、非常に困難な問題に直面する。例えば、RWC-1では、使用する配線材、コネクタ等を特別に考慮したが、150 MHzで10 m程度の距離しか信号を伝搬させることができない。これは主に高周波損失によるもので、より太い線材を使えば長距

* 日本電気(株) C&Cメディア研究所 (〒216-8555 川崎市宮前区宮崎台 4-1-1)

E-mail: matsuo@ccm.cl.nec.co.jp

** 三菱電機(株)先端技術総合研究所 (〒661-0001 尼崎市塚口本町 8-1-1)

*** RWCP光インターコネクション日立研究所 (〒185-8601 国分寺市東恋ヶ窪 1-280 (株)日立製作所中央研究所内)

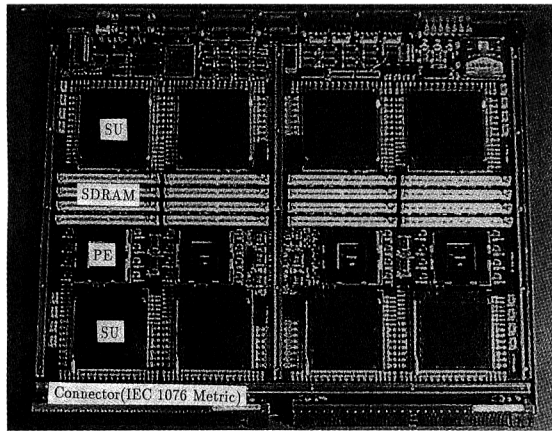


図1 プロセッサボード (電気実装).

離 (あるいは, 高いデータ転送周波数) の転送が可能となるが, コネクタの局数が少なくなり, システム全体としてみたときのスループットが低下してしまう。

光実装 RWC-1 では 12 芯のリボンファイバーを用いた光インターコネクションを採用している。光ファイバーでは損失をほとんど無視することができ, デバイスが動作する最大の周波数で任意の距離のデータ転送が可能である。現状ではコネクタ等の機構部品による制約が多いが, 光ファイバー自体は非常に細いものであるから, コネクタの局数をさらに多くできる可能性がある。将来的には, 電気インターフェースでは実現することができないスループットを実現できる可能性がある (現状でも, 2 倍程度のスループットの向上が可能)。図 1 に電気インターフェースを採用したプロセッサボードの外観を, 図 2 に光インターフェースを採用したプロセッサボードの外観を示す。コネクタの幅が小さくなっていることが確認できる。電気版では, 信号の伝搬距離が長いいため, 差動信号インターフェースを採用しているが⁵⁾, 光実装 RWC-1 では, 近傍に配置された光インターコネクションモジュールとの接続なので小振幅信号インターフェースを採用しており, LSI (SU; switching unit) を小型化でき, 実装面積を削減することができた。この削減した部分に光インターコネクションモジュールを実装しているため, 基板もほぼ同じ大きさである。

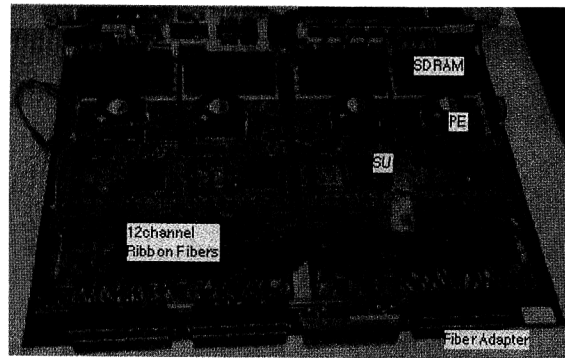


図2 プロセッサボード (光実装).

RWC-1 は電気インターフェースを前提に設計が進められたシステムであり, 光インターコネクションを採用したことによって性能が向上したわけではないが, 実装上の優位性などを実際に確認することができた。筆者らは, ここで得られた知見を基に, 光インターコネクションでなければ実現できないスループットのプロセッサ間結合網を有するシステムを開発し, 光インターコネクションの優位性を性能面でも実証する予定である。

文 献

- 1) T. Yokota, H. Matsuoka, K. Okamoto, H. Hirono and S. Sakai: "hMDCE: The hierarchical multidimensional directed cycles ensemble network", *IEICE Trans. Inf. & Syst.*, **E79-D** (1996) 1099-1106.
- 2) 松岡浩司, 岡本一晃, 廣野英雄, 佐藤三久, 横田隆史, 坂井修一: "超並列要素プロセッサ RICA-1 の高速メッセージハンドリング機構", 並列処理シンポジウム (JSPP '98) (1998) pp. 79-86.
- 3) 西村信治: "超並列計算機 RWC-1 における光実装技術", *エレクトロニクス実装学会誌*, **1** (1998) 180-185.
- 4) S. Nishimura, H. Inoue, S. Hanatani, H. Matsuoka and T. Yokota: "Synchronized parallel optical interconnection for the massively parallel computer RWC-1," *Technical Digest of OC (Optical Computing) '98* (1998) pp. 536-540.
- 5) T. Yokota, H. Matsuoka, K. Okamoto, H. Hirono, A. Hori and S. Sakai: "A prototype router for the massively parallel computer RWC-1," *Proc. Int. Conf. on Computer Design (ICCD' 95)* (Austin, Texas, 1995) pp. 279-284.

(1998年9月4日受理)