

光ファイバ接続による共有メモリについての考察

7M-2

岡本 弘, 佐藤 明行, 田浦 元治, 西井 龍五

三菱電機(株) 情報電子研究所

1 はじめに

計算機システムの、高い信頼性を必要とする分野においては、冗長構成によりシステムの信頼性を向上させる場合がある。例えば、複数の計算機を用いて、多重系システムを構築し、故障した計算機の機能を他の計算機が継承する手法がある。機能の継承方法として、メモリやディスクを計算機間で共有し、情報を伝達する方法があり、高速処理を必要とする場合に共有メモリを用いる。共有メモリと計算機間は、通常、電気的に結合するため、接続距離の制約上、各計算機を同一フロアに集中設置する。

しかしながら、局所的災害(建屋の火災など)が発生した場合は、全ての計算機が障害を受け、システム全体の障害にまで至る危険性がある。

本稿では、工場の構内などのローカルエリアにおいて、複数の計算機を別の建屋に設置し、耐障害性にすぐれた計算機システムを構築するために、長距離接続可能な光伝送結合を検討する。まず、計算機および共有メモリ間を、光ファイバにて長距離接続した場合の問題点を明らかにする。次に、主な検討項目として、共有メモリの構成とアクセスの高速化を取り上げ、それぞれ考察を加える。

2 共有メモリの特徴

共有メモリの主な特徴を以下にまとめる。

- (1) 複数の計算機からアクセス可能なメモリである。
- (2) ディスクなどのI/O装置と異なり、メモリ空間の一定領域を占有し、ロード/ストア命令などにより、バイトやワードなどの単位で、ランダムにアクセス可能である。
- (3) 計算機間の同期が必要なアプリケーションのために、排他制御のメカニズムをもつ。

また、共有メモリが必要とする要件はいくつかあり、主要な項目を以下に示す。

・メモリとしての一貫性

共有メモリは、ローカルメモリとの差を意識することなく、透過的にアクセス可能であることが必要である(例えば、ローカルメモリに対して有効な命令は、共有メモリに対しても有効であることなど)。

・ブロック転送

ストリング処理などの場合、共有メモリおよびローカルメモリ間のブロック転送が生じる。したがって、連続

アドレスに対する大量データを、高速に転送する必要がある。

3 光伝送による共有メモリの接続

計算機と共有メモリ間を光伝送結合した場合、以下の利点がある。

- (1) 低損失であるため長距離伝送が可能である。
- (2) 広帯域であるため大容量高速伝送が可能である。
- (3) 銅線で問題となる誘導ノイズなどの影響を受けない。
- (4) 細径かつ軽量のためケーブル配線工事が容易となる。

以上の利点により、冒頭にて述べたとおり、計算機の設置建屋を別にした多重系計算機システムが構築可能となる。

しかしながら、電気的結合に比べて長距離伝送が可能となるがゆえに、光ファイバの伝播遅延時間が無視できない。例えば、ローカルエリア内の計算機システムを考え、1000メートルの光ファイバを用いた場合、光ケーブルをデータが往復する時間は、約10マイクロ秒を要する。したがって、ランダムアクセスやブロック転送に対して、共有メモリの高速アクセスをいかに実現するかが重要な検討課題となる。

4 検討と考察

4.1 共有メモリの構成

共有メモリの構成について検討する。光ファイバ接続による共有メモリの主要な構成と、その考察を以下に示す。

(1) 分散-リング型共有メモリ

他の計算機からもアクセス可能な共有メモリを、各計算機ごとに内蔵し、それぞれの計算機をリング型の光ファイバにて結合した構成である。図1に2台の計算機を結合した場合を示す。

共有メモリ専用のユニットを必要とせず、コンパクトな構成とすることができる。また、共有メモリが各計算機に分散しているため、危険分散という意味でも有効な構成である。

しかし、一つの計算機が他の計算機の内部バスをアクセスするため、ある計算機の共有メモリに対するアクセス時の障害が、その共有メモリを内蔵する計算機全体に影響を及ぼす可能性がある。また、リング型の光伝送結合については、FDDI(Fiber Distributed Data Interface)などでも採用されている一般的構成であるが、計算機から共有メモリへのアクセス経路が必ずしも最短距離ではなく、伝播遅延の見地から得策な構成とはいえない。

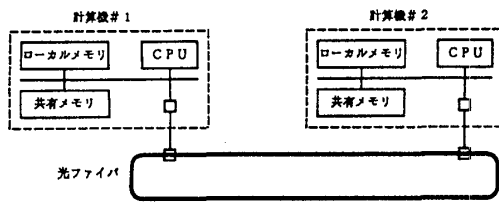


図 1: 分散-リング型共有メモリ

(2) 集中-ポイントツーポイント型共有メモリ

共有メモリを独立させ、各計算機との間をポイントツーポイントで光伝送結合した構成である。図 2 に、二重化した共有メモリに、2 台の計算機を結合した場合を示す。共有メモリが別ユニットであるため、共有メモリにアクセスした時の障害が他の計算機自体に影響を及ぼさない。また、各計算機と共有メモリ間を最短距離で結合するため、光ファイバの伝播遅延時間を最小にとどめることができる。

しかしながら、光結合形態がポイントツーポイントであるため、計算機と共有メモリを接続するハードウェアの規模が増大する。

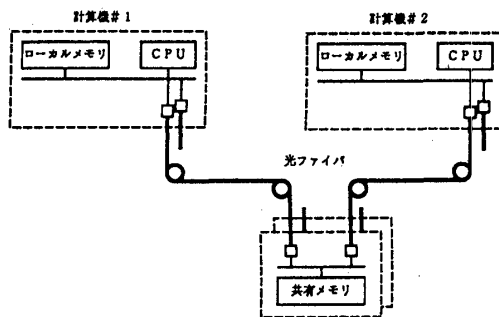


図 2: 集中-ポイントツーポイント型共有メモリ

以上の検討により、前章にて指摘した光伝送結合の伝播遅延時間の問題を考慮すると、共有メモリの構成としては、図 2 の構成が適している。

4.2 共有メモリアクセスの高速化

高速ランダムアクセスを前提とした上で、ブロック転送に対する高速化の手法を検討する。

共有メモリへのストア動作については、共有メモリに FIFO を持つことにより、アクセスを高速化することができる¹⁾。

共有メモリからのフェッチ動作については、共有メモリからローカルメモリへの、連続アドレス転送の場合を検討する。この場合、光ファイバによる伝播遅延時間のため、一般的に共有メモリからのフェッチ動作に時間を

要し、転送性能が低下する。対策として、アクセスした次のアドレスの内容を、事前に光ファイバ経由で読むことにより（以下、先読み制御と呼ぶ）、データ転送を高速化することができる。しかしながら、アクセスするたびに先読み制御を実行すると、先読み制御の実行と、ランダムアクセス（連続しないアドレスへのアクセス）の競合タイミングが頻発し、ランダムアクセスに要する時間が増加する。したがって、直前にアクセスしたアドレスを保持しておき、連続したアドレスに対するアクセスが認識できた時のみ、以降、先読み制御を開始するアルゴリズムとする。以下に、共有メモリのフェッチ動作のアルゴリズムを示す。

```
cm_fetch()
{
  if (req_address == next_address){
    /* 連続アドレスの判定 */
    if (mode) /* 2 度目以降の連続アドレス検出 */
      send_buff_data(); /* CPU にデータ送信 */
    else{ /* 初めての連続アドレス検出 */
      mode = 1; /* 先読み制御モードのセット */
      cm_fetch_req(); /* 共有メモリにデータ要求 */
      send_buff_data();
    }
    cm_fetch_req();
  }
  else{ /* 不連続アドレス検出 */
    mode = 0; /* 先読み制御モードのクリア */
    cm_fetch_req();
    send_buff_data();
  }
  next_address_gen(); /* 次アドレスの生成 */
}
```

この方式によると、連続したアドレスの 3 番目以降のアクセスから高速転送となる。なお、不連続アドレスへのアクセスを検出した時点で、先読み制御は停止する。

本アルゴリズムは、高速ブロック転送の発生頻度が圧倒的に多い用途、例えば共有メモリをファイルシステムとして扱う用途などでは有効な方法といえる。

5 おわりに

本稿では、ローカルエリアにおける耐障害性にすぐれたシステム構築の一手法として、計算機および共有メモリ間の光伝送結合を検討した。また、光ファイバを用いることにより、伝播遅延時間が無視できないことを示し、共有メモリに適したシステム構成とデータ転送の高速化について考察を行なった。

参考文献

- 1) 齊藤他：FIFO 付きブロードキャストメモリを用いた共有メモリの実現とその評価、信学論 (D-I), J75-D-1, 12, pp. 1125-1131(1992-12).