5K－6

光バスクラスタ計算機Euphoriaの開発 (2) メモリアクセス機構

下山朋彦　濱口一正　福井俊之　柴山茂樹
キヤノン（株）情報メディア研究所

1. はじめに

従来のワークステーションに相当するノードを，光波長多重回線を用いて接続することにより，ノード間でのメモリの共有を可能とした「光バスクラスタ計算機」のプロトタイプ“Euphoria”の開発・試作を行った[1]．本報告ではEuphoriaのメモリアクセス機構について述べる．

2. Euphoriaのメモリシステム

Euphoriaではネットワークに，光波長多重通信を採用している．そのことにより従来のLAN等に比べてより高速，大容量の通信を実現することが可能となる．このネットワークの性能を活かすためには，従来とは異なったアプローチが必要となる．

一般的にネットワークの進歩に伴い，分散共有メモリ型のプログラミングモデル持つクラスタ構成のコンピュータシステムが出現してきている．そのメモリの共有の方法には，大きく分けてハードウェアによる方法と，ソフトウェアによる方法の2通りが考えられる．

分散共有メモリシステムをソフトウェアによって実現した場合，既存のアーキテクチャと親和性は良いが，データ転送速度に対するそれらのプロトコル処理時間等の負荷を考えると，ネットワークの転送速度を活かし切れない懸念がある．

そこで，Euphoriaでは，ハードウェアによって分散共有メモリを実現することにした．各ノードに付属したOPC（OPtical-connection Control unit）が，相互にパケットを通信してこの仕事を行う．あるノードから外部ノードのメモリに対するアクセスが行われると，OPCがこれを検出しメモリアクセス要求の通信パケットを作成してリモートメモリアクセスが行われる．

より詳細に述べると，OPCは内部バスを監視しており，ノード外のアドレスに対するアクセスを検出する．検出するとOPCは，光バスアービタに対してアクセス対象デバイス（メモリなど）を持つノードとの回線接続要求を出す．回線がつながるとOPCは，相手ノードのOPCに対してアクセス要求パケットを送る．相手側のOPCはパケットで示されたメモリアクセスを代行し，アクセス結果を返送する．アクセス結果を受けとったOPCは，アクセスを行ったデバイス（MPUなど）に対してアクセス結果を伝える．この様子を図1に示す．競合するアクセスが同時に行われた場合は，コンセントレータにおいて処理を直列化し，矛盾の発生を防ぐ．

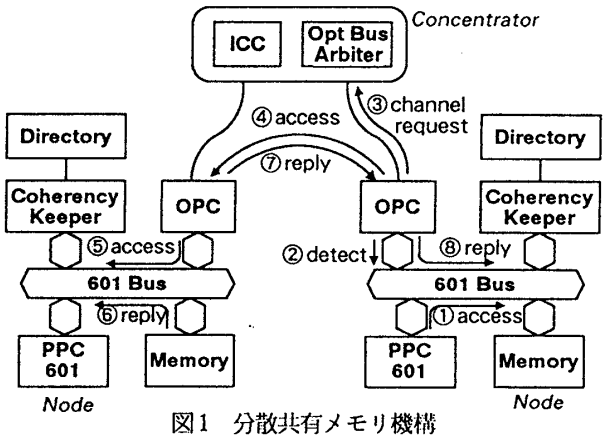

図1　分散共有メモリ機構

3. キャッシュ管理システム

Euphoriaではメモリアクセスコストの低減のために，キャッシュシステムを採用している．これは特に，参照コストの高い外部ノードのメモリのアクセス時間の低減に効果的である．

複数のMPU（ノード）がキャッシングを行うためキャッシュの一貫性保持動作が必要となるが，これはMPU内蔵のスヌープ機構と，外付けのディレクトリ方式のキャッシュ一貫性保持機構を組み合わせることで対処している．

Design and Implementation of an Optical Bus Cluster Computer “Euphoria”
(2) Memory-access Mechanism
T. SHIMOYAMA, K. HAMAGUCHI, T. FUKUI and S. SHIBAYAMA
Media Technology Laboratory, Canon Inc.

キャッシュ機構は，各ノードとコンセントレータに機能を分散している．一貫性保持に必要な情報の各ノードへの通知をコンセントレータで行い，処理の高速化を図る．

3.1 ノード内のキャッシュ管理装置

各ノードはキャッシュ管理装置として，CK (Coherency Keeper) とディレクトリを持つ．

CKは内部バス上に発行されるアクセスを監視し，必要に応じて光バスアービタを通じてメンテナンスパケットを送出する．またパケットを受けとったノードでは，CKがそのパケットを解釈し，内部バス上に一貫性保持のためのアクセスを発行する．

ディレクトリは自ノードのメモリブロックをキャッシングしているノードをキャッシュブロック (32B) 単位で記録する．CKは外部ノードからのアクセスが生じた際に，アクセスを行ったノードをディレクトリの対応箇所に登録する (図2)．一貫性保持のためのパケットの送付はディレクトリに記録されたノードに対してのみ行う．

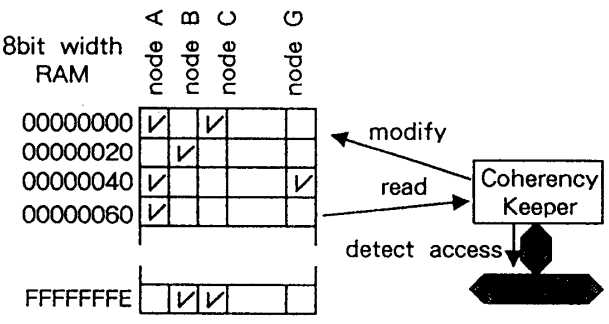

図2 CKとディレクトリ

3.2 光バスアービタ内のキャッシュ管理装置

コンセントレータ内の光バスアービタは，回線接続時や各ノードから要求があった場合に一貫性保持のための動作を行う．

光バスアービタは回線接続要求パケットを受けとると，ノード間の回線を接続する前に，当該メモリブロックに対応するディレクトリエントリをメモリを持っているノードから取り寄せ，ディレクトリエントリに登録されているノードに対してそのアクセスにより必要となる一貫性保持操作を指示し，その後回線を接続する．

コンセントレータにはICC (Internode Caching information Cache) と呼ぶ，ディレクトリ情報のためのキャッシュを備える．目的のディレクトリ情報がICC に存在した場合，光バスアービタは当該メモリブロックの存在するノードからディレクトリ情報を取り寄せる必要がなく，性能の向上が望める．

3.3 キャッシュシステムの動作

自ノードのメモリ参照時には，内部バス上にアクセスが発行されるとCKがこれを検出し，アクセス先メモリブロックに対応するディレクトリを検索する．検索の結果，他のノードがメモリブロックをキャッシングしていることが分かった場合，CKは光バスアービタにそれらのノードに対する一貫性保持のためのパケットの送付を依頼する．光バスアービタは必要なノードに一貫性保持のためのパケットを送付する．光バスアービタからのアクノレッジを受けとると，CKはディレクトリを更新し，アクセスを完了させる．

他ノードのメモリ参照時には，内部バス上にアクセスが発行されるとOPCがこれを検出し，光バスアービタに回線要求パケットを出力する．光バスアービタはICCからアクセス先メモリブロックに対応するエントリを読み出し，当該メモリブロックを持つノードに必要となる一貫性保持操作を指示した後に回線を接続する．アクセス元のOPCは接続された回線を通じてリモートアクセスを実行する．メモリ側のノードのCKはアクセスを検出し，当該メモリブロックに対応するディレクトリエントリを更新し，アクセスを完了させる．

4. おわりに

光バスクラスタ計算機“Euphoria”を設計・試作し，分散共有メモリ機構，ディレクトリ方式のキャッシュ機構を実現した．

今後，メモリシステムの評価としては，メモリアクセスパターンの調査などを行う予定である．

参考文献

[1] 福井 他，「光バスクラスタ計算機Euphoriaの開発 (1) 概要」，
第49回情報処理学会全国大会5K − 05，1994.