# 高速並列処理ワークステーション（ＴＯＰ－１）
## ― アーキテクチャ ―

7N-2

大庭信之、小原盛幹、清水茂則、中田武男、森脇淳、若林真一

日本アイ・ビー・エム株式会社東京基礎研究所

## １．はじめに

高速並列処理ワークステーション（ＴＯＰ－１）は、汎用目的のマルチプロセッサ・システムである（以下ＴＯＰ－１と略記）。本稿では、ＴＯＰ－１がどのようなアーキテクチャ上の方針に基づいて研究開発されているかについて、主たるいくつかの特徴について述べ、次いで、実際のハードウェア構成について概説する。

## ２．アーキテクチャ上の特徴

### ２．１　完全共有メモリ

ＴＯＰ－１のメモリシステムは、すべての処理ノードから完全に均質に共有された共有メモリのみから構成され、ローカルメモリは含まれていない。従来、共有メモリ型マルチプロセッサ・システムは、共有メモリとローカルメモリの組合せで実現され、各処理ノードに固有なデータ及びコードをローカルメモリに割当てることによって、共有メモリ型マルチプロセッサ・システムの欠点であるバス及びメモリ競合の軽減を図るというのが一般的であった。しかしローカルメモリを用いたそのような構成は、汎用的な使用を前提とした場合、静的なメモリ割当て、さらには、動的なメモリ再配置等かなり複雑な問題をメモリ管理に引き起こし、ソフトウェアの生産性を著しく低下させることも事実である。一方、ＴＯＰ－１で採用している完全共有メモリシステムは、ソフトウェアに対して、単純できれいなメモリモデルを提供することができる反面、バス及びメモリ競合に起因するシステム性能低下の問題を改善するには、ハードウェア上かなりの工夫が必要となる。

### ２．２　スヌープ・キャッシュ

バス及びメモリ競合低減の為にＴＯＰ－１で採用している第一の機構は、各処理ノード毎に搭載されているスヌープ・キャッシュである。各処理ノードにキャッシュを搭載することの意義は、１つには、もちろん、共有メモリとプロセッサとの速度差を高速のキャッシュ・メモリによって吸収することにあるが、マルチプロセッサ・システムの場合、むしろ、各プロセッサの共有バスにたいするアクセス要求の低減を図り、バス競合に起因する性能低下の問題を改善することにあると言える。マルチ・キャッシュのデータ一致性を実現する手法としてスヌープ・キャッシュと呼ばれる方式が広く用いられているが、その実現方式としては、いくつかの変形が提案、実現されている。ＴＯＰ－１では、バス使用頻度の低減を第一の目標とし、共有メモリのアクセス速度等も考慮の結果、ＴＯＰ－１プロトコルと呼ぶ独自のプロトコルに基づくスヌープ・キャッシュ方式を採用している。プロトコルの詳細、また、キャッシュ・サイズ、ライン・サイズ、アソシアティビティ等の設計諸元に関する検討は、別稿［１］で論ずる。

### ２．３　高速共有バス

バス競合を低減させる為のより直接的な方策は、共有バスの転送能力を増大させることである。ＴＯＰ－１の共有バスは、高いデータ転送能力を実現する為に、アドレス境界で二重化されており、それぞれ６４ビットのデータ幅を持つ。二重化されたそれぞれのバスは、独立に動作可能であり、合計で８５ＭＢ／秒の実効データ転送能力を提供する。前項のスヌープ・キャッシュの効果と併せて、この二重化共有バスがシステム性能の向上に大きく寄与することは、性能評価シミュレーションの結果からも明らかである［２］。
実効的な性能向上の為には、共有バスの転送帯域幅を増大することにも増して、高速かつ公平なバス調停方式を実現することも重要である。ＴＯＰ－１では独自のバス調停方式によって、ランドロビンとよく似た公平なバス調停を１バス・サイクル（５０ｎｓあるいは６２．５ｎｓ）で実現している［３］。各処理ノードには、その使用目的に従って、バス使用優先権のレンジを与えることができ、さらに、そのレンジの変更をソフトウェアによって動的に行うことが可能である。

### ２．５　プロセッサ間非同期通信

ＴＯＰ－１は、共有メモリ型のマルチプロセッサ・システムであるので、プロセッサ間の同期は共有メモリ上の共有変数を通して行うのが基本である。しかし、一般に、マルチプロセッサ・システムでは、この他に、要求駆動的な非同期通信の手段を実現することが必須である。ＴＯＰ－１では、このプロセッサ間非同期通信を実現する為に、１対１、１対多の非同期通信を行うメッセージ・パシングの機構を採用している［４］。

## ３．　システム構成

---

High-Performance Multiprocessor Workstation (TOP-1), -- Architecture -- Nobuyuki Ohba, Moriyoshi Ohara, Shigenori Shimizu, Takeo Nakada, Atsushi Moriwaki, Shin'ichi Wakabayashi
IBM Research, Tokyo Research Laboratory

ＴＯＰ－１のハードウェア構成が図１に示される。標準構成では、１０枚のプロセッサ・カード、２枚の共有メモリ・カード、１枚のハードディスク制御カード、及び１枚のマイクロチャネル・インタフェース・カードの合計１４枚の同じ大きさのカードより構成される。プロセッサ・カードは、１つの処理ノードを構成し、固定小数点演算用にインテル８０３８６、また浮動小数点演算用にＷｅｉｔｅｋ１１６７を有している。また、二重化共有バスに対応してキャッシュ・システムもアドレス境界で二重化されており、それぞれ６４Ｋバイト、合計１２８Ｋバイトのキャッシュ・サイズを有する。各共有メモリ・カードは、１６Ｍバイトの容量を持つ。ＥＣＣ補正はカード内で行われ、共有バスを通して１８７．５ｎｓのサイクル・タイムで使用される。ハードディスク制御カードは、１枚のプロセッサ・カードと対になって使用され、それぞれ３００Ｍバイトの容量を持つ４台のハードディスク装置の制御及び、管理を行う。ＴＯＰ－１の筐体中に実装されている４台のハードディスク装置は、この制御カードによって４台並列に制御され、さらに、４Ｍバイトのディスク・バッファを通して非常に高いアクセス速度及び、転送能力を提供する。ハードディスクと共有メモリの間のデータ転送は、対になって使用されるプロセッサ・カードによってスヌープ・キャッシュを通して行なわれる。連続したメモリ領域に対するブロックＩ／Ｏについて、キャッシュ・システムの性能を最適に保つ為に、特別なスヌープ・プロトコルが用意されている。この特別なプロトコルは、他の処理ノードで使用されているＴＯＰ－１プロトコルと一貫して動作し、データの一致性を保証すると同時に、効率の良いブロックＩ／Ｏを実現する［５］。

マイクロチャネル・インタフェース・カードは、ＩＢＭ ＰＳ／５５モデル５５７０あるいは、ＰＳ／２モデル８０をＴＯＰ－１の共有バスに接続するのに使用される。プロセッサ・カードと同一のスヌープ・キャッシュを通して共有バスに直接的に接続されるので、高い転送能力が実現される。このような構成によって、ＰＳ／５５モデル５５７０あるいは、ＰＳ／２モデル８０の持つ標準的なＩ／Ｏ装置のすべてが、ＴＯＰ－１のＩ／Ｏとして使用可能である。また、現行のシステムでは、ＰＳ／５５モデル５５７０はＴＯＰ－１のシステム・コンソールとしても使用されている。

## ４．　まとめ

我々は、試作に先立って、ＴＯＰ－１のハードウェア・アーキテクチャによってどの程度の性能が達成されるかについて、ソフトウェア・シミュレータによって評価を行っているが、実用的な実際の応用プログラムに対してどの程度の性能を達成できるかについては、今後の評価を待つところである。ＴＯＰ－１には、キャッシュ性能、バス性能等の実動作時測定を可能とする特別なハードウェアが実装されている。この性能評価用ハードウェアによって、アーキテクチャを再評価し、今後の研究の助けとしたい。また、ＯＳとして、※ＵＮＩＸをベースとしたマルチプロセッサＯＳを現在試作中であるが、これについての報告は、応用プログラム等と併せて、別の機会に譲ることにする。

参考文献

［１］清水　他、'高速並列処理ワークステーション（ＴＯＰ－１）－スヌープ・キャッシュー'，３７回全国大会（１９８８）
［２］大庭　他、'－性能評価－'，同上
［３］若林　他、'－バス・システム－'，同上
［４］小原　他、'－並列処理同期化機構－'，同上
［５］中田　他、'－入出力システム－'，同上

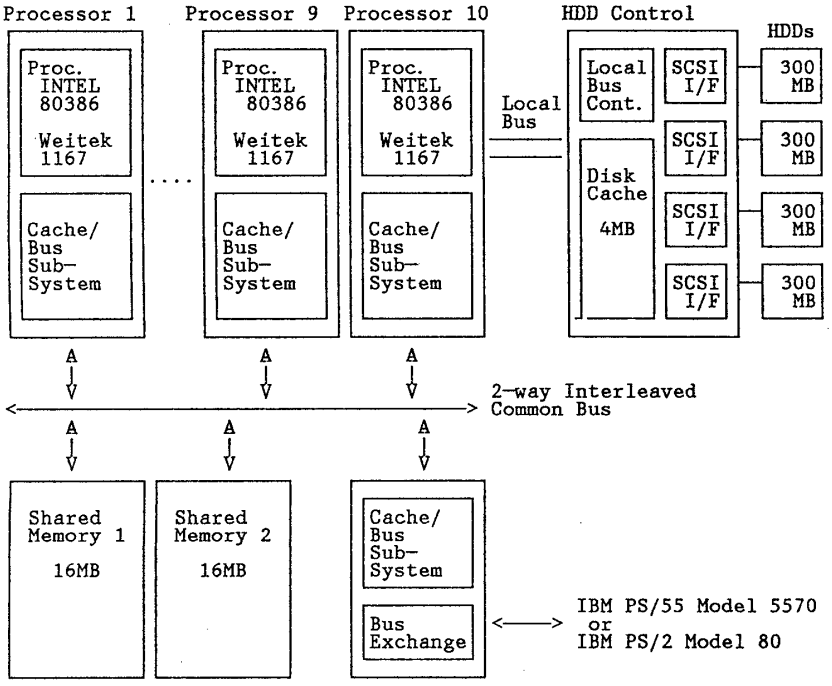


図１．　ＴＯＰ－１のハードウェア構成

＊　ＵＮＩＸは米国ＡＴ＆Ｔ社ベル研究所で開発したソフトウェアの名称です。