計算機アーキテクチャ 79－2
（１９８９．１１．２０）

# マルチプロセッサシステム“砂丘”の共有メモリアーキテクチャについて

井上　正人・井上　倫夫・小林　康浩
鳥取大学工学部

本報告では，全てのプロセッサが大容量共有メモリを，アクセス競合による待機時間を回避して利用するためのアーキテクチャとして，共有メモリの階層化およびメインメモリのマルチリード・ワンライトメモリ方式について述べる．

具体的には，メインメモリのリードアクセス用バスとライトアクセス用バスを分離し，マルチリード・ワンライトメモリ方式を採用することによって，各プロセッサの稼働率を落とさずに接続できる台数を多くできること，そのときメモリアクセスに占めるリード動作の割合が0.7～0.8であるとき最も能率がよいこと，さらに共有メモリの階層化について，メインメモリのアクセスの割合を 0.8くらいに保てば，アクセス競合による性能低下を起こさずに稼働できるプロセッサの台数を最大にできることなどを示した．

# Shared Memory Architecture of The Multimicroprocessor System “SAKYU”

Masato INOUE · Michio INOUE · Yasuhiro KOBAYASHI
Department of Electrical Engineering, Faculty of Engineering, Tottori University

This paper proposes a memory architecture which is necessary for scaling up a tightly coupled multiple microprocessor system and is useful for implementing highly parallel processing.

The proposal consists of (1) introduction of a concept of hierarchy into memory organization, (2) furnishing of shared memories with two ports, (3) equipment of two kind of shared memories; system memories for storaging prime data, and main memories for offering common working areas, (4) adoption of multiple access for read operations and once access for write operations, (5) construction of exclusive read buses and exclusive write buses, (6) use of two-way interleaved main memories, (7) provision of the omega network connecting to processor units through exclusive write buses.

Usefulness of the above measures is discussed with theoretical investigations.

## 1. はじめに

市販の1チップマイクロプロセッサ（以下μPと記す）を複数台結合して，演算能力の向上を図るマルチμPシステムが各種提案されている[1]，[2]．筆者らは複数のμPを接続する方法として，資源共有型（共有バス方式）によるマルチμPシステムの開発を行っている[5]～[10]．普通このようなシステムでは，接続可能なμPの台数に物理的制限があり，処理能力の飛躍的な向上は困難である．しかし，既存のハードウェア技術を活用して，システムの処理能力を数十倍に向上させることは比較的容易である[6]，[7]．また，各処理ユニット（PU）の機能を均質にすることによって，種々の処理アルゴリズムにも簡単に対応できるような利便性のよいシステムの構築が可能である．

マルチμPシステムで能率よく並列処理を行うには，ジョブを複数の並列タスクに分割して複数台のPUで均質に処理することが望まれる．ところが，一つのジョブ全体で見れば，必ずしも全ての部分を並列処理（並列タスクへの分割）できるわけではなく，単一のPUで処理をしなければならない部分もある．このため，システムの処理速度はPUの構成台数に比例するとは限らない．これはシステムの構成規模を決定する上で重要な因子であり，基本設計時に十分考慮する必要がある．

また資源共有型システムでは，アクセス競合による各PUの待機時間の増加を抑制できなければ，複数のPUを使用したにもかかわらず，システムの処理能力の向上が期待できない．共有メモリ方式では，共有バスのアクセス競合を緩和，または回避するための有効な対策をたてることが必要不可欠である．

筆者らは，α－16に関する一連の研究成果[5]～[7]をもとに，現在16ビットμP（MC68000）を用いたマルチμPシステムの開発を行っている[8]～[10]．

本報告では，現在開発中の並列計算機“砂丘”のシステム構成，大容量共有メモリの制御方法としてのマルチリード・ワンライトメモリ方式の採用，メモリの利用度別階層化等を示す．そして，それら共有メモリの制御方法を用いた場合のシステムの稼働率について述べる．

## 2. システム構成

図1は，当研究室で製作を進めている並列計算機“砂丘”のシステム構成図である．

本システムは，最大64台のPUが接続可能な密結合型マルチμPシステムである．その構成は，バスコントロールユニット（BCU）と最大4台のPUを接続できる共有バスを1グループとし，その4グループを1ブロックとする．この4ブロックをマトリックススイッチ（MTX）を介してシステムエリア，メインメモリユニット（MMU），オメガネットワークに接続した．物理メモリの構成は，ローカルメモリ（PUに付随），メインメモリユニット（MMU），システムメモリ（システムエリア内）の3階層にした．ここで，ローカルメモリはプログラム格納エリアとして，メインメモリは特定のプロセッサごとの共通のワークエリアとして，システムメモリは全プロセッサで共同利用するデータエリアとして，それぞれ機能分化している．また，BCUは同一のバスに接続されるPUの共有資源へのアクセスを調停する．MTXは4×4のバススイッチである．4入力は全てBCUに接続する．4出力は，1出力がシステムエリアのアクセスに，別の3出力がメインメモリのアクセスに使用される．メインメモリのアクセス（3出力）は，2出力が2インターリーブ構成のリードアクセスパスとして使用され，別の1出力がライトアクセスパスとしてオメガネットワークに接続される．オメガネットワークは，大別して個別利用モードと一斉放送モードの2機能を持ち，ライトアクセスパスだけに使用される．システムエリアはMTXを2ユニット使用して，システムメモリ，グラフィックメモリ，割り込みコントローラ等が接続される．

## 3. 共有メモリの制御方法

一般に，並列計算機“砂丘”のような資源共有型の密結合システムにおいては，バス結合の制御が頻繁に行われる．したがって，各ＰＵから共有資源へのアクセス要求を効率よく制御し，アクセス競合による待機時間の累積をどのように抑制するかが，本システムを能率よく運用するための重要な課題である．本システムではこの問題を解決するため，以下に述べるような方法の採用を検討した．

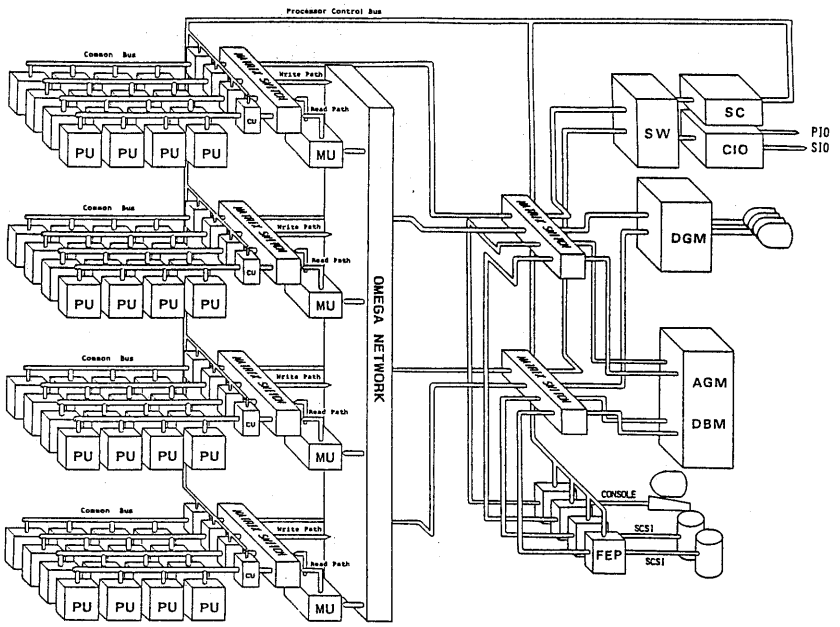
### 3.1 メモリ空間の階層化

本システムでは，アクセス競合を緩和する一手段として，メモリ空間を分割して使用する階層型メモリ構造を採用した．これによって，各ＰＵはローカルメモリ，メインメモリ，システムメモリの３種類のメモリ空間をアクセスできる．この中でメインメモリとシステムメモリは全てのμＰがアクセス可能な共有メモリである．このように，共有メモリを複数使用し，かつこれらを有効に使用することによって，アクセスが一つの共有メモリに集中する可能性を小さくした．例えば，入力データ，演算結果，表示データ等は，頻繁にアクセスされるものではないので，これらのデータ類はメインメモリではなくシステムメモリに保管することにし，メインメモリの負担を軽減することにした．

### 3.2 マルチリード・ワンライトメモリ方式の採用

共有メモリのうち，メインメモリは各ＰＵブロックごとに１ユニットずつ配置した．そして，このメモリの運用をプログラムの処理アルゴリズムに応じて，ある程度変更できるようにした．

今回さらに本システムを能率よく運用するため




PU :Processing Unit
CU :bus Control Unit
MU :Memory Unit
SW :System Work area
SC :System Controller
CIO:Communication Input/Output
DGM:Digital Graphic Memory
AGM:Analog Graphic Memory
DBM:Data Buffer Memory
FEP:Front End Processor

図１　システム構成図

に，メインメモリにマルチリード・ワンライトメモリ方式を採用した．

(1) リードアクセスパス

ＰＵのメモリアクセスにおいて，一般にはリードアクセスの方がライトアクセスより多く行われる．したがって，複数のＰＵをブロック化し，限られたＰＵで一つのメインメモリを共有させ，それぞれのブロックにおけるリードアクセスをそのメモリユニットに限定すれば，アクセス競合はそのブロック内でだけ考慮すればよい．図２にリードアクセスパスの様子を示す．リードアクセス用の共有バスは２インターリーブ方式にした．

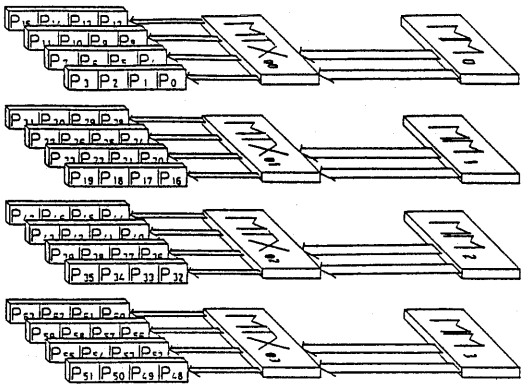


図２　リードアクセスパス

(2) ライトアクセスパス

本システムでは，４ユニットのメインメモリを分散配置し，４つのＰＵブロックをオメガネットワークで接続する．図３にライトアクセスパスの様子を示す．これは，メインメモリの運用を能率よくするためであり，いろいろな処理アルゴリズムに対応して，その利用形態をダイナミックに変更することが可能となる．ただし，全てのμＰを使用する並列動作時は，ライト動作でのアクセス競合に注意しなければならない．

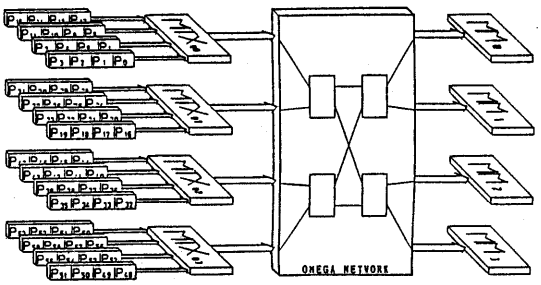



図３　ライトアクセスパス

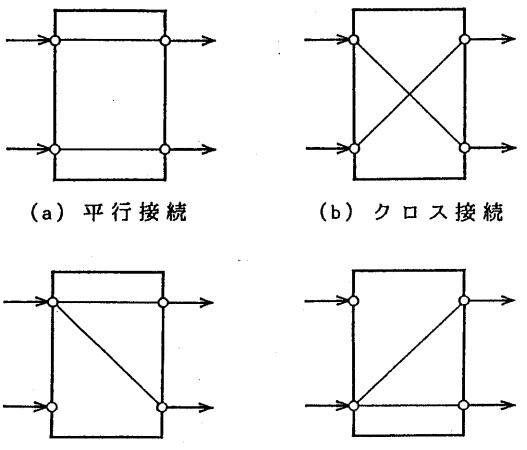
### ３．３ インターリーブ方式の採用

１つのメインメモリユニットは，８Ｍバイトのメモリ容量を持ち，偶数アドレスと奇数アドレスそれぞれ４Ｍバイトずつの２ブロックに分けられている．これをＰＵブロックに接続するのに２インターリーブ方式とした．

これにより，同じＰＵブロック内の２台のＰＵが隣合うアドレスをリードアクセスする場合，メインメモリ上ではアクセス競合が生じないことになる．４バイト単位以上のデータを取り扱う処理系においては有効な手段である．

### ３．４ オメガネットワークの使用

ライトアクセスパスにオメガネットワークを使



(a) 平行接続　(b) クロス接続

(c) 上ブロードキャスト接続　(d) 下ブロードキャスト接続

図４　バススイッチの接続モード

用したことにより，次の３種類の基本的利用形態が実現できた．ここで，オメガネットワークに使用されるバススイッチの接続モードを図４に示す．オメガネットワークは，これらの接続モードをジョブまたはタスクレベルで固定することも，ライトサイクルごとに自由に切り替えることも可能である．

(1) 一斉放送モード

バススイッチの接続モードに，上下ブロードキャスト接続を使用し，ライトアクセスごとに変更するようにすれば，特定のＰＵのライトアクセスを，４つのメインメモリユニットに同時に実行できる．これによって，全てのメモリユニットが同じ内容を保持できる．

(2) パイプライン接続モード

バススイッチの接続モードに，平行接続，クロス接続両モードを併用すれば，任意の隣接したメモリユニットにライトアクセスを実行できる．これによって，タスクのパイプライン処理が可能となる．

(3) 個別接続モード

バススイッチの接続モードを，平行接続に固定すれば，それぞれのＰＵブロックのライトアクセスを，リードアクセスで使用するのと同じメモリユニットに対し実行できる．つまり，１つのＰＵブロックが使用できるメインメモリユニットが１つだけ決定されて，各ＰＵブロックが４つのメモリユニットをそれぞれ独立して使用できる．

### ３．５ 共有メモリの多重ポート化

本システムでは，メインメモリとシステムメモリの２種類のメモリを共有メモリとして使用する．これらのメモリは，バスの切り替え等による無駄時間を少しでも省くために多重ポート化する．

メインメモリには，ライトアクセス用バスをアドレスの最下位ビットによって，偶数または奇数アドレスブロックに分けて接続する．このときのアクセスタイミングは，メインメモリの制御部により時分割制御される．

システムメモリには，共有バスを，32台ずつの２グループに分けて接続する．各ポートは，メインメモリと同様に時分割制御される．

ところで，この方法はＰＵ１台当りのアクセス間隔は長くなるが，１つのポートが使用されている間に，次のポートのリクエストを受け付けられるので，すぐに次のポートのアクセスを実行できる．それにより，共有バス上での遅延時間がラップされるという利点がある．また，競合の対象となるＰＵの台数を半分にできるので，各μＰの稼働率を低下させないで接続可能台数を増加できる．しかし，どちらか一方のバスが未使用の時でもポートを切り替えるという無駄が生じるので，両方のポートを均等にアクセスするように，各ＰＵに処理させるタスクを分配することが重要である．

## ４．検討

並列計算機の基本目的は，個々のμＰの処理能力が低くても，その多数を組織化することによって，システム全体の処理能力をμＰの台数に比例して向上させることにある．しかし，処理すべき課題の分割方法や共有資源の利用方法等に問題があり，理想通りにいかないのが現実である．

これまで，共有資源の利用時に発生するアクセス競合の回避策や低減措置について述べてきた．ここでは，本システムの共有メモリの制御方法によるシステム全体の稼働率について，予測と検討を述べる．

### ４．１ μＰの稼働率

共有メモリをμＰがアクセスするとき，同期をとるために挿入するウェイトによって，μＰの持つ能力がどれくらい低下するのかの指標として稼働率を定義する．共有メモリのアクセス回数ｋ回のプログラム実行時間をＴ(ｋ)[ｓ]とすると，一度も共有メモリをアクセスしないプログラムの実行時間はＴ(０)[ｓ]となり，稼働率Ｐは次のよう

に定義できる.

$$P = \frac{T(0)}{T(k)} \times 100 \quad [\%] \tag{1}$$

ここで，共有メモリをアクセスする時にμＰへ挿入される平均ウェイトをｔｗ［ｓ］とすると，

$$T(k) = T(0) + k \cdot t_W \quad [s] \tag{2}$$

となり(１)式は

$$P = \frac{100}{1 + \frac{t_W}{Th}} \quad [\%]$$

$$Th = \frac{T(0)}{k} \quad [s] \tag{3}$$

となる．ここで，Ｔｈ［ｓ］はμＰが共有メモリをアクセスする平均間隔に相当し，以後これをシンクタイムと呼ぶことにする．(３)式は，システムのハードウェア特性（ｔｗ）とソフトウェア特性（Ｔｈ）との関数として表されていて，ｔｗを短く，Ｔｈを長くすることがシステムの稼働率を上げる条件であることを示している．

## ４．２ 並列動作時の稼働率

アクセス競合が頻発し，待機状態に陥るμＰが増加するのは，各μＰが同時に処理時間の等しいタスクを処理しているときである．筆者らはこの点に着目して，システムの処理性能を陽的に解析できることを示した．そこで，システムのハードウェア特性（共有資源のアクセス時間，サイクル時間等）およびソフトウェア特性（タスクの平均シンクタイム）とを用いて，並列動作時の各μＰの平均稼働率を次のように表す．

$$P(n,Th) = \frac{100}{1 + \frac{tac + m_n \cdot t_S}{Th}} \quad [\%] \tag{4}$$

ここで，

$$m_n = \begin{cases} 0 & (n \leq i_B \cdot n_0) \\ \frac{n}{i_B} - n_0 & (n > i_B \cdot n_0) \end{cases} \tag{5}$$

ただし，

ｎ　：同時に動作しているμＰの数

Ｔｈ　：タスクの平均シンクタイム［ｓ］

$t_S$　：共有資源のサイクルタイム［ｓ］

$t_{ac}$　：各μＰが共有資源をアクセスするときの平均アクセスタイム［ｓ］

$$t_{ac} = t_B + t_A + \frac{1}{2} t_S \tag{6}$$

$t_A$　：共有資源のアクセスタイム［ｓ］

$t_B$　：ＢＣＵ，ＭＴＸ等での遅延時間［ｓ］

$i_B$　：インターリーブ数

$n_0$　：アクセス競合を起こさずに動作できるμＰの最大台数

$$n_0 = 1 + \frac{Th}{t_S} \tag{7}$$

## ４．３ マルチリード・ワンライトメモリ方式による稼働率

メインメモリのリードアクセス用バスは，１ユニット当り２インターリーブ構成としたので，メインメモリ全体では８インターリーブ方式と同等とみなせる．

マルチリード・ワンライトメモリ方式ではμＰの接続可能台数および稼働率が，アクセスのリード・ライト比（ソフトウェア）に依存する．メインメモリへのアクセス全体を１としたときのリードアクセスの割合をｒとする．このとき(７)式を拡張すると，リード動作においてアクセス競合を起こさず動作できるμＰの最大台数$n_R$は，

$$n_R = 8 \left(1 + \frac{Th/r}{t_S}\right) / r \tag{8}$$

で表せる．同様にライト動作での$n_W$は，

$$n_W = \left(1 + \frac{Th/(1-r)}{t_S}\right) / (1-r) \tag{9}$$

で表せる．

(８)(９)式の関係を図５に示す．この図は，メインメモリのアクセスにおいて，シンクタイムが５［μｓ］のとき，リードアクセスの割合が 0.7～0.8 付近でμＰを 128台以上並列稼働できること

を表している．また，メインメモリへのアクセスに対するリード動作の割合が最適な状態にあるとき，アクセス競合を起こさずに並列稼働できるμＰの台数が最大になることがわかる．

稼働率を表す式は(４)(５)式に(８)(９)式を用いて，

$$P(n) = \frac{100}{1 + \frac{t_{ac} + m \cdot t_s}{Th / r} + \frac{t_{ac}' + m' \cdot t_s}{Th / (1-r)}} \qquad (10)$$

と表せる．

(10)式において，リード動作の割合ｒが 0.8のときのμＰの台数と稼働率の関係を図６に示す．破線は共有メモリを８インターリーブ方式で具体化した場合の稼働率を示す．この図より，マルチリード・ワンライトメモリ方式は，接続するμＰの台数が同じ場合，８インターリーブ方式に比べてシンクタイムをより短くできることがわかる．

### ４．４ メモリの階層化による稼働率

共有メモリであるメインメモリとシステムメモリの階層化について，それぞれのメモリでアクセス競合を起こさずに稼働できるμＰの台数を知ることにより，稼働率を求めることができる．共有メモリへのアクセスを１とし，そのうちメインメモリへのアクセスの割合をｒとすると，メインメモリでアクセス競合を起こさず稼働できるμＰの最大台数 $n_M$ は，

$$n_M = 8 \left( 1 + \frac{Th/r}{t_s} \right) / r \qquad (11)$$

で表せる．同様にシステムメモリでの $n_S$ は，

$$n_S = 2 \left( 1 + \frac{Th/(1-r)}{t_s} \right) / (1-r) \qquad (12)$$

で表せる．

(11)(12)式の関係を図７に示す．この図から，共有メモリのアクセスにおいて，シンクタイムが５[μｓ]のとき，メインメモリのアクセスの割合が0.83付近で，約 120台のμＰを接続できること，アクセスの割合が 0.8以上のとき64台のμＰがアクセス競合を起こさずに稼働できることがわかる．また，共有メモリのアクセスの割合が最適な状態

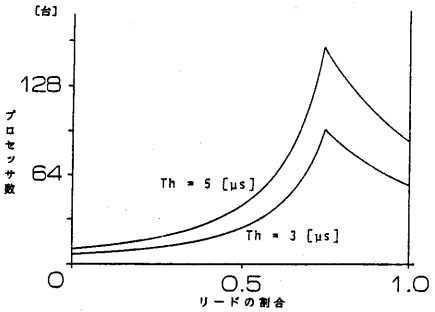



図５ メインメモリ利用時のリード・ライト比と競合を起こさずに稼働できるμＰ数

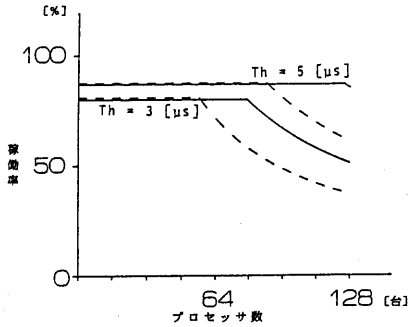



図６ メインメモリ利用時のμＰ数と稼働率

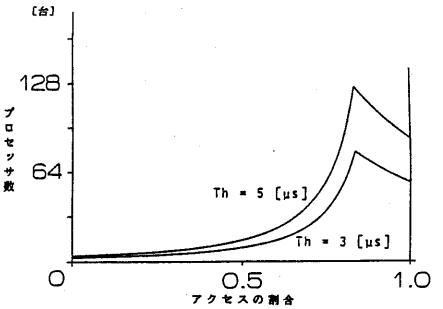



図７ 共有メモリ利用時の各メモリのアクセス比と競合を起こさずに稼働できるμＰ数

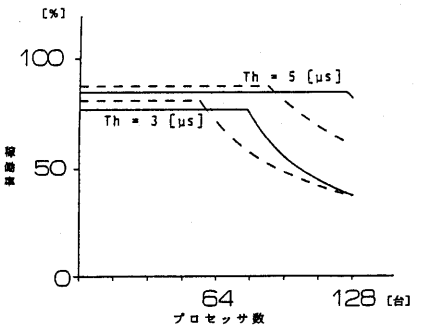



図８ 共有メモリ利用時のμＰ数と稼働率

にある場合，接続できるμＰの台数を最大にできることがわかる．

稼働率を表す式は，(10)式と同様に(11)(12)式を用いると，

$$P(n) = \frac{100}{1 + \frac{tac + m \cdot ts}{Th / r} + \frac{tac' + m' \cdot ts'}{Th / (1-r)}} \qquad (13)$$

で表せる．

(13)式において，メインメモリへのアクセスの割合ｒが 0.8のときの，μＰの台数と稼働率の関係を図８に示す．破線は共有メモリを階層化しなかった場合の稼働率である．この図より，接続するμＰの台数が同じ場合，共有メモリの階層化によってシンクタイムをより短くできることがわかる．

## 5．おわりに

以上，資源共有型マルチμＰシステム“砂丘”の共有メモリアーキテクチャについて述べた．具体的には，複数のμＰが共有資源を効率よく利用するために，メモリの階層化，マルチリード・ワンライトメモリ方式の採用等が効果的であることを，その稼働率を求めることで示した．

また，共有メモリのうちのメインメモリを分散配置したことによって，次のようなシステムの処理形態を実現できる．

(1) 一斉放送モードによって，全メモリユニットの同一内容保持ができる．
(2) ライトパスを隣接するメモリユニットに接続して，タスクレベルでのパイプライン処理ができる．
(3) 各ＰＵグループ別にメモリユニットを利用して，グループ別処理ができる．
(4) 上記 (1)，(2)，(3)をアクセスサイクルごとにダイナミックに変更できる．

今後の課題は，本システムを使用して，各種の具体的な問題に対する評価を行うことにある．

## 参考文献


[１] 白川　他：“並列計算機ＰＡＸ－１２８”通信学論，Vol.J67-D，No.8，pp.853-860，Aug.　(1984)

[２] 出口　他：“コンピュータグラフィックシステムＬＩＮＫＳ－１における画像生成の高速化手法”　情報処理論文誌，Vol.25，No.6，pp.944-952，Nov.　(1984)

[３] Rodrigue. G. : “Parallel Computation” Academic Press　(1982)

[４] Paker. Y. : “Multi-microprocessor” Academic Press　(1983)

[５] 井上・小林：“マイクロプロセッサを用いた並列処理システムα－16”シミュレーション　第２回研究会資料，pp.19-24，March　(1982)

[６] 井上・小林：“マルチマイクロプロセッサシステムα－16のアーキテクチャ”情報処理論文誌，Vol.25，No.4，pp.632-639，July　(1984)

[７] 井上・小林：“α－16マルチマイクロプロセッサシステムの性能評価”情報処理論文誌，Vol.25，No.4，pp.640-646，July　(1984)

[８] 井上・小林：“資源共有型マルチマイクロプロセッサシステムにおけるアクセス競合の調停について”　電子情報通信，回路とシステム研究会資料，CAS84-206，pp.9-16　(1984)

[９] 山根　他：“並列計算機“砂丘”のハードウェアアーキテクチャ”　電子情報通信，コンピュータシステム研究会資料，CPSY87-3，pp.15-22，June　(1987)

[10] 加納　他：“マルチマイクロプロセッサシステムの大容量共有メモリの一構成法”情報処理学会　計算機アーキテクチャ研究会資料，CA-66-2，pp.1-8，July　(1987)