計算機システムの解析と制御 13－1
（1981. 6. 12）

ポリプロセッサシステムEPOSの オペレーティングシステムとその性能評価について

田中哲男　石田勝世　上田隆司　山崎　勇
（東芝総合研究所）

1．まえがき

ポリプロセッサシステムEPOS*のオペレーティングシステムと，その性能測定結果について報告する。EPOSはマイクロプロセッサPULCE**を中心としたコンピュータモジュールで構成された計算機複合体である。EPOSオペレーティングシステムはこのハードウェアの特徴を生かし，拡張性と適応性に秀れたシステムをつくりあげた。ここで拡張性とは，業務量が増大したときにコンピュータモジュールを増設することにより，システムの能力を拡張することをいう。適応性とは，個々の応用に対する適応と，システムに荷せられる負荷のパターンの動的な変動に，システムの内部構成を変化させて対応することとをいう。

この２つを実現するために，我々はマイクロプログラムによるコンピュータモジュールの専用機械化をおこない機能分散システムをつくりあげた。システム資源の管理という面から言えば管理機能を階層的に分散させ，計算機複合体における管理の困難性を解決した。専用機械個々に対する負荷量の変動に動的に対応するために，ダイナミックマイクロプログラミング技術をもちいて，専用機械としての役割を動的に変化させた。

EPOSオペレーティングシステムは，利用者にはタイムシェアリングシステムとしての環境を提供する。性能測定をおこない，端末数が増大したとき，コンピュータモジュールの増設により応答速度の改善が得られることを明らかにした。また特別な応用として，並列的な漢字変換を実装し，単位時間あたりの変換量がコンピュータモジュール台数に比例して増大することを確認した。

2．ハードウェアの構成

あとの議論の準備としてハードウェアの構成を説明する。

EPOSは8台のコンピュータモジュール（CM）と2台の入出力モジュール（IOM）とから構成される。CMおよびIOMはシステムバス（S-BUS）を介して結合されている。CM，IOMは総計32台まで増設できる。S-BUSは最大4本である。S-BUSの本数は，マイクロプログラムレベルからは見えない。

各コンピュータモジュールは，ハードウェア的には同一であって，LSI-PULCEを中心に構成されている。PULCEは7000ゲート以上の論理回路を集積したSOS-LSIであり、外部から32ビット長のマイクロ命令を供給されて動作する。コンピュータモジュールはそれぞれ個有のマイクロプログラムメモリ（MPM），ローカルメモリ（LM）をもつ。大部分の仕事はコンピュータモジュール内部で閉じておこなわれる。

コンピュータモジュール，入出力モジュールは，相互にS-BUSを経由して割込み信号を出すことができ，他方のローカルメモリもアクセスするこ

本研究は，通産省工技院大型プロジェクト「パターン情報処理システムの研究開発」の一環としておこなわれた。

＊Experimental Polyprocessor System　＊＊PIPS Universal Computing Element

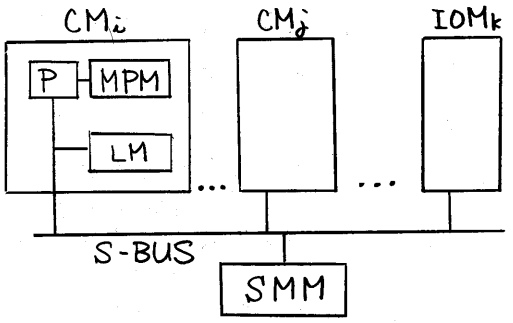

図1　EPOSの構成

とができる。このローカルメモリアクセス機能は入出力モジュールによるデータ転送にだけ用いられている。S-BUSには、共有メモリモジュール(SMM)が接続され、コンピュータモジュール、入出力モジュールによってアクセスされる。共有メモリモジュールはコンピュータモジュール、入出力モジュール間での通信用バッファとして用いられている。

3. 拡張性

共有メモリを主体としたマルチプロセッサシステムの問題点は、次の2点に要約される。第1点は、すべての情報が共有メモリ上に在るために、アクセスの衝突を緩和することが重要となる。共有バス方式、クロスポイントスイッチのいづれを採用しても、速度をあげるために高価となり、接続台数も制限をうけ、拡張性阻害の要因になる。第2には、すべてのプロセッサが共有するシステム資源の管理アルゴリズムを対等に実行する。このために相互排他を組み入れた複雑なアルゴリズムになってしまう。

ローカルメモリをもったEPOSのような複合体では、第1の問題点は、ほぼ解消されている。EPOSは、共有メモリをもっているが、それは主としてプロセス間通信の際のメッセージバッファとして用いられるものでありアクセス頻度は小さい。

マルチプロセッサに比べて、EPOSの場合により大きく問題になるのは、コンピュータモジュール間の情報伝達の遅れの問題である。あるコンピュータモジュールは、他のコンピュータモジュールの各資源が、今現在どのように使われているかを正確に知ることはできない。たとえ何らかの通信手段によって通知を受けたとしても、それは必ず過去の状態を表現したものである。したがって複合体を構成するすべての資源について、一元的に管理することは、実際上不可能である。

この問題を解決するためにEPOSでは、管理機能の階層的分散という考え方に基づいてシステムを構成した。すなわち、資源の管理に関してコンピュータモジュールを分類し、下位のレベルのコンピュータモジュールは、ある範囲の資源だけを排他的に管理する。その直接上位のコンピュータモジュールは、下位のモジュールを、モジュールの単位で管理する。この方式に依れば、各コンピュータモジュールは、自分が直接担当する管理対象についての情報だけを保持すればよく、他の同一レベルのコンピュータモジュールと、共有情報を持つ必要はない。情報伝達

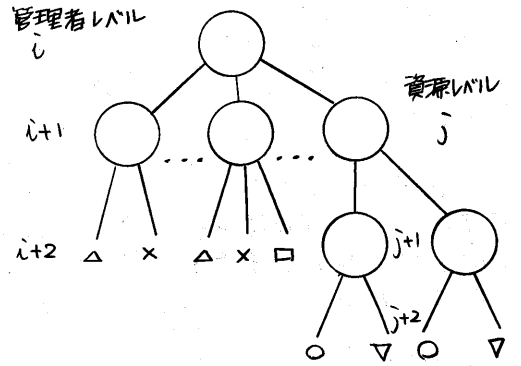

図2　管理機能の階層的分散

の時間遅れの問題を解決し、マルチプロセッサに関するオ2の問題も解決したことになる。

さらにこの方式は、システムの拡張を容易にする。ある構成要素を付加することは、その上位の構成要素にとって管理すべき対象が1つ増えたという変更をもたらすだけである。同位の他の構成要素には、新しい構成要素の付加は、何ら影響を与えない。

4、適応性

EPOSにおける適応性は2つの面から考えることができる。

オ1は適応性の深さである。適応性の深さとは、どれだけ精密に、ある応用に専用化できるか、ということである。EPOSでは、個々のコンピュータモジュールを、マイクロプログラムをもちいて特定の機能を効率良く実現するよう専用機械化する。利用者のPascalプログラムを実行する専用機械としてPascalマシン[2]が提供される。利用者のAPLプログラムを実行する専用機械としてAPLマシン[3]が提供される。またその他のオペレーティングシステムの機能を実現するために、ファイルシステム専用機械[4]，端末制御専用機械等が用意されている。

オ2の適応性は、システムに対する負荷パターンの変動に動的に対応することである。システムの利用者の大部分がAPLをもちいているときにPascalマシンが何台も存在しているのは不都合である。EPOSではダイナミックマイクロプログラミングをもちいて、動的に専用機械としての役割を変えさせている。

オ2の適応性は、システムの運用の場面でも要求される。計算機合体では、構成要素の個々のものが障害を起したとき、あるいは定期保守の場合にも、残りの構成要素で運用を継続していくことが望まれる。EPOSでは、このためにも専用機械の集まりとしてのシステム構成を変化させて動作を継続する。

5、オペレーティングシステムの構造

EPOSのオペレーティングシステムの構造を図3，4に示す。図3は、管理機能の階層構造である。図4は、オペレーティングシステムの各構成要素のハードウェア上での配置図である。

JCP(Job Control Process)は、資源管理の階層の最上位に位置するものである。JCPは端末から，APL，Pascal等のサブシステム選択コマンドを受け取り，それぞれの実行を担当する問題処理専用コンピュータモジュール(PCM)を選択し，その処理を依頼する。このときJCPは、各PCMの内部の資源の状態には関知しない。選択の根拠とするのは、それまでに，JCP自身が，どのようなサブシステムをもちいる仕事を，何個ずつ，各PCMに依頼しているかという記録だけである。

PMP(問題管理プロセス)は、各問題処理専用コンピュータモジュール上で動作するプロセスである。PMPは，自分が存在するコンピュータモジュール上の，マイクロプログラムメモリ，ローカルメモリを管理する，オ2のレベルの管理者である。JCPから割りあてられた仕事に、これらの資源を割りつけて，動作させる。

FAP(File Access Process)[4]，TCP(端末制御プロセス)もオ2のレベルの管理者である。FAPは，ファイルシステム，入出力モジュール，ディスク装置を管理し，他のプロセスからの要求に応答する。ファイルシステムの物理的な構成のしかたは，他のプログラムからは隠され，抽象的なファイルシステムの構成とアクセス手段

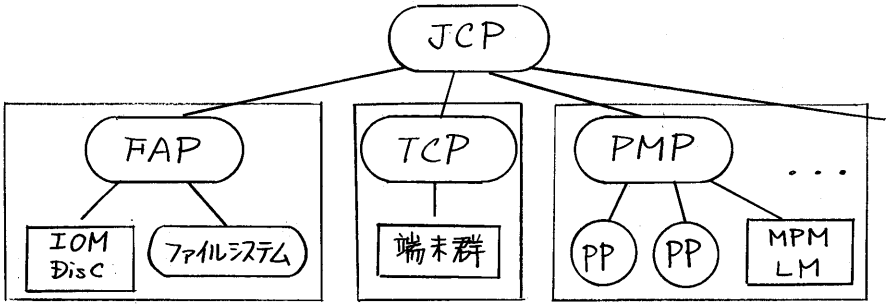

図3　管理機能の階層構造
　　はソフト的資源
　　はハード的資源

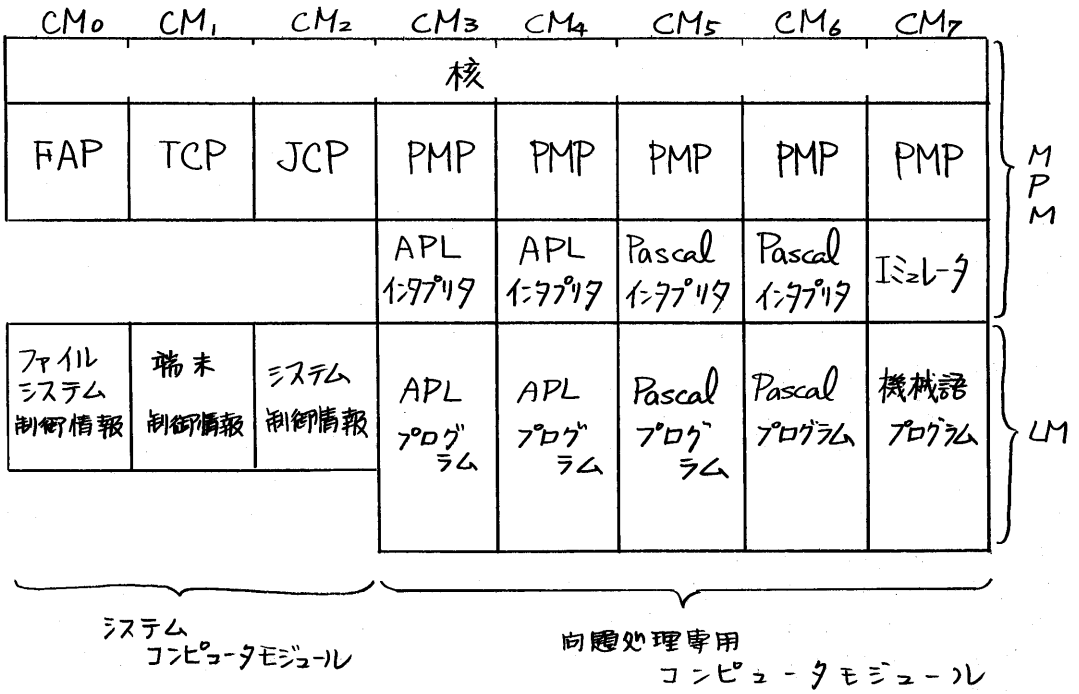

図4　オペレーティングシステムの
ハードウェア上での配置.

だけが外部に提示される。TCPは端末の入出力を制御する。

第3のレベルの管理者は「核」である（第4図）。各コンピュータモジュール上では、上記の管理者としてのプロセス、および利用者の仕事の実行としてのプロセスが動作する。核はコンピュータモジュールのタイムスライスを管理し、これらのプロセスに割りつけて動作させる。核はマイクロプログラムで記述されており、各コンピュータモジュールに同一のものが、コピーされて置かれる。核はこの他に、各プロセス間の依頼・応答情報を伝達するための、プロセス間通信機能も提供する。プロセス間通信機能は、現在共有メモリ上のメッセージバッファと、SBUSによる割込み機能をもちいて実装されている。これを共有メモリを持たずに、通信路だけで実装することも可能である。

問題処理専用コンピュータモジュール（PCM）は、ある時点でどのような仕事を割りあてられているかによって専用機械としての役割が定まる。APLをもちいる仕事だけが割りあてられていれば、そのPCMは、APLマシンとして動作する。APL, Pascalの両方の仕事を割りあてられていれば、マイクロプログラムを入れ換えて、動的に役割を変える。マイクロプログラムの入れ換えは、当然性能の低下をもたらす。JCPはこの点を配慮して、仕事を割りあてなければならない（第6節参照）。

EPOSの最小構成は、コンピュータモジュール台数が2台の場合である。1台はシステム用としてFAP, TCP, JCPの3つのプロセスが動作する。他の1台は、問題処理用である。拡張は負荷の量に応じてシステム用コンピュータモジュールを増設するか、あるいは、問題処理用コンピュータモジュールを増設することによっておこなわれる。

6. JCPによる仕事の割りあてとシステム再構成

JCPは問題処理用コンピュータモジュールの負荷量（割りあてる仕事の数）が均等化されるよう，また，各問題処理用コンピュータモジュールでのマイクロプログラムの入れ換えの可能性が少なくなるよう仕事の割りあてをおこなう。この割りあては、利用者があるサブシステムを選択したとき，およびサブシステムの使用を終了したときにおこなわれる。後者の場合にはすでに割りあててあるコンピュータモジュールから仕事を取りあげ，負荷の小さいコンピュータモジュールに再度割りあてる。

JCPによる問題処理用コンピュータモジュール選択の方法は以下のとおりである。

各コンピュータモジュールの負荷をL_iとする。

(1) 全問題処理用コンピュータモジュールのうちで，最小の負荷をもつコンピュータモジュールをjとする。

(2) 各問題処理用コンピュータモジュールについて，「その中で一番多く動作しているサブシステムの種類」が，「これから割りつけようとする仕事が必要とするサブシステムの種類」と一致するコンピュータモジュールのうちで，負荷が最小であるコンピュータモジュールをkとする。

このとき，

$L_j + a \leq L_k$ならば、コンピュータモジュールjを，

$L_j + a > L_k$ならば，コンピュータモジュールkを選択する。

システムの再構成は、各コンピュー

タモジュールが果している機能が何であるかによって詳細は異なるが，大局的にはほぼ同じである。

例えば，コンピュータモジュールiで動作しているFAPを，コンピュータモジュールjに移動させる場合を考えてみる。

FAPは現在オープンされているファイルに関する情報を，ローカルメモリ上に保持している。これを移動先のFAPが引き継がなければならない。またこの移動の処理中にも，ファイルシステムに対するアクセス要求が到着する可能性がある。これが放棄されてはならない。

これを解決するために，図5に示すように，FAPに2つの内部状態NORMとURGとを設けた。それぞれの状態で他からの要求を受信するメッセージセマフォFAPREQとFAPURGが用意されている。

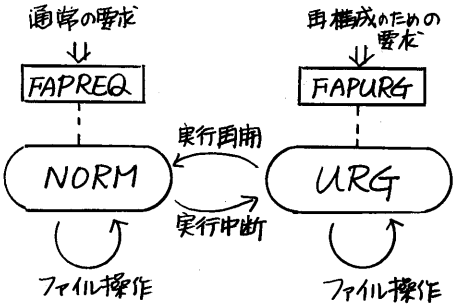

図5　FAPの状態遷移
xx→：要求xxに対する遷移

FAPは実行中断，実行再開の要求以外の要求に対しては，NORM，URGいづれの状態にあっても全く同じ動作をおこなう。実行中断要求に対しては，FAPは，内部情報をディスクに保存し，URG状態に遷移する。実行再開要求では，内部情報を回復し，中断前からの処理を継続する。

JCPによるFAP移動手続きは次のとおりである。

(1) JCPはFAPREQに実行中断のメッセージを送信する。

(2) JCPは，ディスクファイルからFAPのマイクロプログラムを読み出し，コンピュータモジュールjにロードする。

(3) JCPは，FAPURGに実行再開の要求を送信する。このメッセージは，新しいFAPによって受信され，以前からの動作を継続する。

7. 性能測定

EPOSは54年度までに製作を終え，55年度には，総合調整・テストをおこない，性能評価をおこなった。ここでは，TSSとしての応答速度，ダイナミックマイクロプログラミングが性能に及ぼす効果，特別な応用例として並列な漢字変換をおこなったときの変換速度の3つの測定結果について報告する。

測定のための道具としては，端末シミュレータ[5]とソフトウェアモニタ[6]を用意した。端末シミュレータは，最大32台の端末の動作を，個々の端末毎にあらかじめ定められたコマンド系列に沿って模擬する。端末シミュレータはコマンド毎に入力時刻，それに対する応答時刻，応答文字列を記録する。端末シミュレータに設定するコマンド系列は，TSSのテキストエディタによって作成変更することができる。

端末シミュレータからのコマンド系列を実行しつつ，ソフトウェアモニタは，一定周期(6.5秒)毎に動作し，オペレーティングシステムの内部情報，たとえば，FAPに対する要求長などのデータを収集し，磁気テープに出力する。

TSSの応答速度.

図6は，TSSの応答速度の測定結果である。測定にもちいたコマンド系列は、全ての端末について同一のものである。エディタをもちいてPascalソースプログラムを修正し，Pascalサブシステムを起動して、翻訳・実行するというサイクルを繰り返す。思考時間を平均20秒とし，端末台数および問題処理用コンピュータモジュール台数を変化させて測定した。

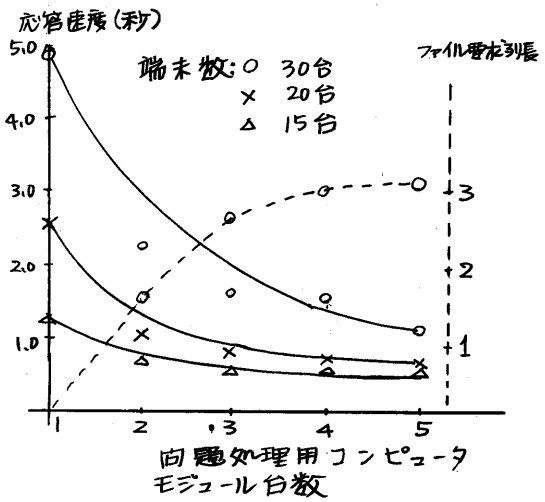

図6　TSSの応答速度とコンピュータモジュール数.
(思考時間：平均20秒．破線は，ファイルでの要求列長)

EPOSのシステム構成から容易に想像されるように，問題処理用コンピュータモジュール台数の増加は，コンピュータモジュールあたりの仕事量を減少させ，応答速度の向上に寄与する。しかしこれは一方では，システムに唯一つしかないファイル専用機械に対する要求頻度を増大させることになり，応答速度の向上に限界を課すことになる。この測定の範囲内では，ファイルシステムに対する要求列長（サービスを受けているものを除く）は最大の場合でも0.3程度であるが，これ以上の要求が発生する場合には，ファイルシステム専用機械を増設する必要があろう。

ダイナミックマイクロプログラミングの効果の測定

JCPによる仕事の割りあてが正しく動作することは、TSSの応答速度の測定の際の、割りあて表を観察することによって確認された。マイクロプログラムの動的な入れ換えが、システムの性能に及ぼす影響を測定するためには、特別な環境設定が必要となる。すなわち，各問題処理用コンピュータの役割が，固定的に定められているようなシステム構成を設定し，それと，EPOS本来の姿であるダイナミックマイクロプログラミングがおこなわれているシステム構成の場合とで，応答速度を比較する。

与えるコマンド系列は，APL，Pascal，テキストエディタをほぼ均等に利用するが，他に比べてAPLを利用するコマンドが多いものを用意した。

表1に4つのシステム構成と，それぞれの場合の応答時間の平均値を示した。

システム構成II，III，IVは、システムのオペレータが，負荷の分布を予測し，固定的に役割を定めた場合に相当する。測定の結果によれば，コマンド系列にとって最適に近いと考えられるシステム構成であるIIの場合が，一番良い性能を示している。これには及ばないものの、全面的にダイナミックマイクロプログラミングがおこなわれているIの構成が，それに近い性能を示しているといえる。

システム構成					測定結果（応答時間の相対値）
No.	可変CM数	APL固定CM数	Pascal固定CM数	エディタ固定CM数	
Ⅰ	5	0	0	0	1
Ⅱ	0	3	1	1	0.87
Ⅲ	0	1	3	1	1.23
Ⅳ	0	1	1	3	1.34

表1　ダイナミックマイクロプログラミングの効果の測定

並列かな漢字変換の性能

パターン情報処理システム統合システムプロトタイプの応用例の1つとして，日本語文章処理による特許情報検索システムが開発された[7]。この中でMPOSは，手書き文字認識サブシステムで読み込まれた，かな文（特許公開公報）を受け取り，漢字かな混り文に変換する役割りを受け持った。JCPは，変換要求を受け取ると，その時点で存在している，問題処理用コンピュータモジュール台数分だけ変換プロセスを生成し，各PMPに処理を依頼する。変換プロセスは並列に動作し，原文の順序どおりに変換結果を並べるために相互に同期をとりながら変換処理をおこなう。

測定は，4つの文例について，問題処理用コンピュータモジュール数を変化させておこなった。その結果を図7に示す。

変換に要する時間は，原文の文例によってかなりのバラツキを示すが，5台で並列に処理したとき，1台の場合に比べて，3.4～4.7倍の処理能力向上が認められる。各変換プロセスは，原文バッファから順次文節をとり出し，原文の順序を保存しつつ，変換文バッファに格納する。従って，変換に長時間を要する文節があると，それが全体の処理時間に影響を与える可能性がある。

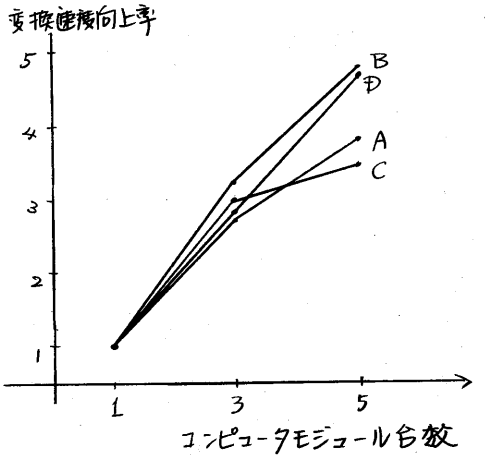

図7．並列化によるかな漢字変換速度の向上率．
（A，B，C，Dは測定に使用した文例）

8、あとがき.

ポリプロセッサシステムEPOSのオペレーティングシステムの構成，その性能測定結果について報告した。EPOSは，発達したLSI技術の適用例を示すものであり，そのオペレーティングシステムは，複合計算機システムのOSとして，管理機能の階層的分散という考え方のもとに，マイクロプログラム技術を駆使してつくりあげられたものである。

EPOSの開発において不十分であったのは，マイクロプログラム作成環境の整備である。EPOS自身の上でマイクロプログラムを開発する環境をもたなかったことが，開発速度を遅れた。

本研究にあたり，多大な御支援と御助言をいただいた，通産省工業技術院および電子技術総合研究所の関係各位に感謝の意を表する。

参考文献

1) 飯塚他 "Development of a high performance universal computing element -PULCE" Proc. AFIPS NCC, vol 47, 1255-1264 (1978)

2) 森本 「ファームウェアを利用したAPLインタプリタの構成法および評価」情報処理学会・記号処理研究会 9-2 (1979)

3) 吉村他 「EPOSにおけるパスカルプロセッサについて」情報処理学会 計算機アーキテクチャ研究会. 35-4 (1979)

4) 田中他 「ポリプロセッサシステムEPOSにおけるファイルプロセッサ」昭和54年電子通信学会統合全国大会, p. 1458 (1979)

5) 西尾他 「TSS応答速度測定用端末シミュレータ装置」昭和56年度電子通信学会 統合全国大会 p. 1439 (1981)

6) 前田他 "A Distributed File System in EPOS" COMPCON 80 pp50-54 (1980)

7) 森他 「特許情報検索システム」大型プロジェクトパターン情報処理システム 研究成果発表会論文集. 通産省工技院(編). 日本産業振興協会(発行) 1980.

8) 前川他 "Experimental Polyprocessor System (EPOS) - Operating System" Proc. 6th Ann. Symp. on Comp. Architecture. pp 196-201 (1979)