記憶階層構成に関する一検討

An Analysis Of Storage Hierarchies

木ノ内康夫　山口博　小橋喜嗣　桜井紀彦
YASUO KINOUCHI　HIROSHI YAMAGUCHI　YOSHITSUGU OBASHI　NORIHIKO SAKURAI
（日本電信電話公社　横須賀電気通信研究所）
YOKOSUKA ELECTRIC COMMUNICATION LABORATORY

1．はじめに

現行の計算機センタシステムにおいては、アクセス頻度、レスポンスタイム等の異なる様々なファイルが要求され、それぞれに適合したファイル装置を組合わせ階層構成をとることにより、全体として効率的、経済的なファイルシステムを実現している。

しかし、近年のMOS・RAMの急激な価格低下や、各ファイル装置技術の発展度合の違いにより、性能、価格条件に相対的な変化が現われてきており、またディスクキャッシュに代表される新しい方式技術の採用等もあってファイル階層構成の見直しが迫られている。

本報告は、これら性能、価格条件の変化や新しい方式技術の採用が既存のファイル階層構成に与える影響に注目し、ファイル装置の導入効果や適用領域の見直しを通じて、近い将来における望ましいファイル階層構成についてオンラインファイルを対象に考察したものであり、階層構成を評価するために有用な評価方法を提案するとともに、これを用いて、各種装置や新方式の適用領域などを明らかにしている。

検討にあたっては、まずファイル系に対してどのような評価尺度や評価手法を適用するかが重要となるが、ここでは、現行のオンラインシステムにおけるファイル系への要求を踏まえ、単位時間・単位容量あたりのファイルアクセス回数（アクセス頻度）を満足しかつ所定のレスポンスタイム以下で応答可能という条件を設定し、その下で装置・方式別に最小コストを算出するという方法を用いた。

通常このような評価では、注目するファイル系を待ち行列モデルで記述しアクセス頻度、レスポンスタイムを拘束条件として最小コストを算出するという方法がとられる。しかし、多様な装置や方式を対象に直接この方法を適用することは、複雑さや、方式、装置間の要因個別の影響が見通しにくいという点で問題がある。

そこで、Lum、藤井等[1],[2]が採用した「トランザクション内で時間的、空間的に装置を使用した割合に応じて、装置個別にコスト寄与分を算出し、使用する全装置について和を求める」という手法に注目し、これに対して新しくレスポンスタイムの拘束を考慮可能とする拡張を施した簡易な算出方法を提案した。

次にファイル階層構成の評価にあたっては、オンラインファイルをおおよそ以下の高速系、中速系とに分けて考察を試みた。

1）高速系・・・・磁気ドラム装置、固定ヘッドディスク装置、半導体ファイル装置、主記憶装置
2）中速系・・・・磁気ディスク装置、ディスクキャッシュ

高速系については、MOS・RAMの低価格化を主対象として、

a）新しく主記憶装置をファイル装置として用いた場合と、既存の磁気ドラム装置、固定ヘッドディスク装置との適用領域の比較
b）MOS・RAMを用いて新しく入出力装置を構成した場合の有効性

を中心に議論している。

中速系については、磁気ディスク装置を対象にアクセス頻度と所要コストの関係を明確にすることに努めた。特に比較的アクセス頻度の高いファイルへの対処策として、磁気ディスク装置のファイル収容率を削減する手法が知られている[7]が、その有効性、適用領域を明らかにした。

また、最近主要な方式技術として注目されているディスクキャッシュについては、ヒット率をパラメータとして、装置コストを考慮した適用領域を明らかにしている。

2．評価の方法

2．1　評価尺度の設定

ファイル系について評価の方法を具体化するのに先だち、始めに、本報告での評価尺度のとらえ方を示す。

一般に、ファイル系に対する性能面での要求は、次の2点に要約される。

（1）注目する階層のファイル装置に対して，システムとしての処理能力から定まる所定の回数のアクセスを処理可能なこと。
（2）ファイル装置へのI／O要求が出されてから、主記憶上でデータを使用可能になるまで、ファイル装置の空き待ち時間を含むレスポンスタイムが、所定の値を満足すること。

ここで、（1）に示した要求は、オンラインシステムにおいて、システム全体の処理能力及びシステムコストに直接影響する重要な特性である。所定の回数のアクセスが処理できないファイル装置が存在すると、その装置がネックとなりシステム内のCPUを含む他の装置がアイドルとなる。結果として、システム全体の処理能力は、ネックとなるファイル装置によって押さえられてしまう。

更にファイル装置への要求を、アクセス頻度［回／MB･S］でみると、おおよそ10^{-5}－10^{-6}回／MB･Sから10^{2}－10^{3}回／MB･Sと広い範囲にわたって変化している。このような広い領域を幾つかの領域に細分し、それぞれの領域に効率的なファイル装置を組み合わせることが、ファイル階層構成全体としての効率化、経済化を図る上で主要な課題となっている。

一方、（2）に示されたレスポンスタイムへの要求は、（1）のアクセス頻度に対する要求と同様に重要ではあるが、その要求内容は異っている。

現行システムでは多数個のタスクにより、複数のファイルアクセスを同時に実行可能なことから、ファイル装置のレスポンスタイムは、その増加が直ちに他の装置のアイドル状態を引き起こすことは少なく、システムとしての処理能力に直結する特性とはなっていない。現状では端末ユーザに対して一定の値以下の応答時間を保証するという、サービス品質の主要な評価パラメータとして位置づけられていると考えられる。

例えば、TSSシステムにおいて特定のファイルへ多数回アクセスする場合や、リアルタイムシステムにおいて高度で複雑なデータベース処理を実行する場合など、応答時間を一定値以下とするため、ファイルへのレスポンスタイムの短縮が設計上主要な課題となるケースが知られている。

以上のような状況を考慮し、本報告では、次に示す考え方を基本として評価を進めた。

（ア）　ファイル装置に対する（1），（2）の2つの要求のうち、（1）のアクセス頻度に関連する要求を最も重要視し、システムコストとの関係を含め可能な限り厳密に評価可能とする。
（イ）　レスポンスタイムについては、サービス品質としてとらえられるパラメータであるため、アクセス頻度ほど高い精度でコストとの関係を求める必要はなく、近似的な手法を適用する。

2．2　ファイル系コスト性能比の評価方法

ファイル系のコスト性能比を評価する方法は、これまで様々な手法が用いられているが、アクセス頻度、レスポンスタイムのあつかいに注目すると、おおよそ次の2つに分けられる。

［手法A］：　システム全体または注目するファイル系を待ち行列モデルとして記述し、アクセス頻度・レスポンスタイムを拘束条件として所要コストを算出する。
［手法B］：　トランザクション処理のなかで、使用される装置毎にコスト寄与分を直接求め、使用される全装置に渡って和を求めることによりコストを算出する。装置毎の寄与分は、使用するファイルが注目する装置内に占める割合や装置を時間的に占有する割合と装置価格との積で求める。この手法では、通常アクセス頻度は装置のコスト寄与分に直接反映できる。しかし、これまでのところレスポンスタイムは考慮されていない。

手法Aの待ち行列を用いる方法は、コスト性能比評価の有力な手法として数多くの研究がなされている。[3],[4],[5]必要な限り精密なモデルを設定することによって、高い精度の解を得ることが可能である。しかし、算出方法が複雑になり易いことや、モデル内の各要因個別の寄与の度合が直感的にとらえにくいという問題がある。

これに対して手法Bの方法は、レスポンスタイムが考慮されないという面はあるものの、注目する装置の使用分だけコストをカウントするという直感的で簡単な手法であるため、算出が容易で、要因個別の影響が見通しやすいというメリットがある。

既にLum等　は、手法Bを用いて多階層にわたるファイルのマイグレーションを簡明に評価しており、また藤井等は

、ＭＳＳ、ＭＴを対象としてファイル装置の適用領域を簡単な手法で明らかにしている。

本報告では，磁気ドラム装置から磁気ディスク装置、ディスクキャッシュに至るまで多様な階層、方式にまたがるファイル系構成を簡明に評価する必要があること、レスポンスタイムについては高い精度を必要としないこと等の点を考慮し、Ｌｕｍ、藤井等による手法Ｂを、レスポンスタイムによる拘束条件を近似的に反映可能なように拡張し、それを評価手法として使用した。

拡張にあたっては、”手法Ｂは手法Ａの近似的な手法である”との観点に立ち、ファイル系の基本的な構成を想定した場合の手法Ａによる解を求め、この解とこれまでの手法Ｂとの比較を通じて、具体的にＬｕｍ、藤井等の手法において拡張すべき要因を明らかにした。

以下 2.2.1 において手法Ａに基づいて解を求め、2.2.2 において拡張手法を示す。

2.2.1 待ち時間を考慮した場合のコスト算出方法

ファイル系の基本構成を図2-1 に示すようにNd 台のデバイスとNc 台のコントローラから構成されるものと仮定する。このとき、ファイル容量をＦ[MB]、ファイル全体へのアクセス回数をＡ [回／S]、許容されるレスポンスタイムの上限を$\hat{T}_R$[S] として、所要ファイルコストをＣとすると、以下の３式で記述される。

$$C = Nc \cdot Cc + Nd \cdot Cd \qquad (2\text{-}1)$$

Cc ：コントローラ価格
Cd ：デバイス価格

$$\hat{T}_R \geqq T_d(A) + T_{wd}(A) + T_{wc}(A) \qquad (2\text{-}2)$$

Td(A)：待ち時間を含まない場合のデバイスのレスポンスタイム
Twd(A)：デバイス空き待ち時間
Twc(A)：コントローラ空き待ち時間

$$F \leqq Nd \cdot D \qquad (2\text{-}3)$$

D： 装置一台当りの容量

ここで、Td(A) がＡによらず一定としTwd(A) がＭ／Ｍ／１モデルで評価可能で、かつTwc(A) はTwd(A)に比べて小さく無視することが可能であると仮定すると、2-2 式は次の様に書ける。

$$\hat{T}_R \geqq T_d + \frac{\rho_d(A)}{1-\rho_d(A)} T_d \qquad (2\text{-}4)$$

ただし $\rho_d(A) = A \cdot T_d / Nd$ (2-5)

$\rho_d(A)$：デバイスの使用率

また、通常のオンラインシステムの設計においては、予想以上の一時的な高負荷に対処するためや、一部の装置が障害になった場合に代替装置を確保すること等のために、2-2 式で示される平均的なレスポンスタイムによる条件とは別に装置種別毎に最大使用率を設けることが多い。これを考慮し、デバイス、コントローラの最大使用率をそれぞれρ_d°,ρ_c°とすると、さらに次式を得る。

$$\rho_d(A) \left[= A \cdot T_d / Nd \right] \leqq \rho_d^\circ \qquad (2\text{-}6)$$

$$\rho_c(A) \left[= A \cdot T_c / Nc \right] \leqq \rho_c^\circ \qquad (2\text{-}7)$$

Tc：コントローラ占有時間

以上の結果を整理すると、図2-1 に示した基本的なファイル構成において所要コストＣを求めることは、2-3、2-4、2-6、2-7 を拘束条件として、2-1 式においてNd、Nc を変化させた時の最小値を求める問題として定式化される。

一般的な議論をしやすくするため、コスト、アクセス回数ともにファイル１MB当たりの値、Cm ＝Ｃ／Ｆ、およびａ＝Ａ／Ｆをとることとし、また拘束条件をわかりやすくするために、Nd、Nc についての条件として整理すると以下のようになる。

$$C_m^* = \min_{Nc,\,Nd} \left[Nc \frac{Cc}{F} + Nd \frac{Cd}{F} \right] \qquad (2\text{-}8)$$

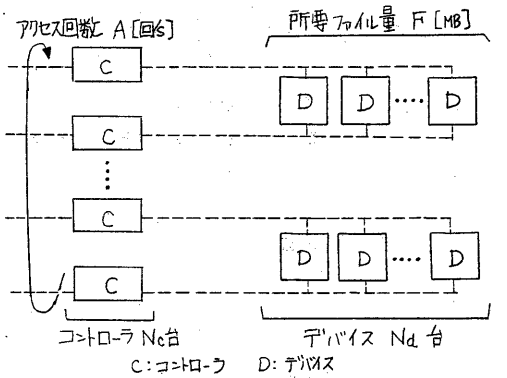


図2-1 ファイル装置の構成

［拘束条件］

(1) $Nd \geqq a \cdot F\, Td \Big/ \left(1 - \frac{Td}{\hat{T}_R}\right)$

(2) $Nd \geqq a \cdot F\, Td / \rho_d^\circ$

(3) $Nd \geqq F/D$

(4) $Nc \geqq a \cdot F\, Tc / \rho_c^\circ$

ここで、拘束条件の形に注目して、解を与える条件を類別しそれぞれについて解を求めると、以下の式を得る。

（ア）拘束条件が (1)および (4)による場合

$$C_m^* = \frac{a \cdot Tc}{\rho_c^\circ} Cc + \frac{a\, Td}{1 - Td/\hat{T}_R} Cd \qquad (2\text{-}9)$$

（イ）拘束条件が (2)および (4)による場合

$$C_m^* = \frac{a \cdot Tc}{\rho_c^\circ} Cc + \frac{a \cdot Td}{\rho_d^\circ} Cd \qquad (2\text{-}10)$$

（ウ）拘束条件が (3)および (4)による場合

$$C_m^* = \frac{a\, Tc}{\rho_c^\circ} Cc + \frac{1}{D} Cd \qquad (2\text{-}11)$$

さらに、（ア）、（イ）、（ウ）の各ケースについてみると、拘束条件の形から、いずれが 2-8式の解になるかは、それぞれの第２項のみを相互に比較した時、最大値をとるか否かによって簡単に決定可能である。これにより 2-8式の解は（ア）、（イ）、（ウ）の各式を以下のようにまとめたものとなる。

$$C_m^* = \frac{a \cdot Tc}{\rho_c^\circ} Cc + \max\left[\frac{a \cdot Td}{1 - Td/\hat{T}_R}, \frac{a \cdot Td}{\rho_d^\circ}, \frac{1}{D}\right] Cd \qquad (2\text{-}12)$$

すなわち、上式第２項での最大値の選択が、（ア）、（イ）、（ウ）いずれの拘束条件が解の決定に関係したかを示している。

2.2.2 評価方法の拡張

前節 2-12 式をベースに、Ｌｕｍ、藤井等の算出式を拡張する。

はじめに 2-12 式を手法Ｂによるコスト算出式としてみると、ａＴｃ 、ａＴｄ はそれぞれ単位時間内に装置を占有している時間を示しており、これを ρ_c 、ρ_d とおき、Ｌｕｍ、藤井等の算出式との相違を点線で囲むと

$$C_m^* = \left[\frac{1}{\rho_c^\circ}\right] \rho_c \cdot Cc + \left[\max\left[\frac{\rho_d}{1 - Td/\hat{T}_R}, \frac{\rho_d}{\rho_d^\circ}\right], \frac{1}{D}\right] Cd \qquad (2\text{-}13)$$

となる。

すなわち、点線で囲んだ項を除いた第１項の $\rho_c Cc$ がコントローラの使用時間に応じたコストを示し、また第２項の Cd/D がファイルを１MB格納しておくためのコストを示している。

更に、点線部の相違の示す意味は次のように考察される。（１）レスポンスタイムを制限することは、装置の使用率上限を変化させることとなり、その使用率上限によって使用時間に応じたコストが補正される。

2-13 式の項 $\rho_d Cd/(1 - Td/\hat{T}_R)$ の分母がそれである。

（２）デバイスのコストは単に空間的にファイルを占有することによる寄与分だけでなく、スループット面で装置の持つ能力のどの程度を使用しているかにも注目する必要がある。アクセス頻度の高いファイルでは、装置内の一部にのみファイルが格納されていたとしても、スループット面で限界近くまで使用されていると当該装置は他に使用できない。このことは、ファイル装置のコスト寄与分を空間的な意味でのファイルの占有率と、時間的な意味での装置占有率との２面からとらえる必要のあることを示しており、 2-13 式での最大値をとる操作は、空間的な占有率、時間的な占有率のいずれかコスト寄与分の大きい方を選択することが、妥当であることを示している。

以上の考察をもとに、本報告では次に示す方法により、アクセス頻度、レスポンスタイムの関数として、コストを算出する。以下この方法によるコストをハード使用コストＨと呼び、次式で定義する。

$$H = \sum_i hi \qquad (2\text{-}14)$$

ここで、ｈｉ は装置でのコスト寄与分であり、装置種別に応じて、次のいずれかにより算出される。

(7) コントローラ等の場合

$$hi = \rho_i\, Ci / \rho_i^\circ \qquad (2\text{-}15)$$

(イ) デバイスの場合

$$h_i = max\left[\frac{\rho_i}{\rho_i^\circ(\hat{T}_R)}C_i, \frac{\rho_i}{\rho_i^\circ}C_i, \frac{C_i}{D}\right]$$

$$= max\left[\frac{\rho_i}{min\left[\rho_i^\circ(\hat{T}_R), \rho_i^\circ\right]}, \frac{1}{D}\right]C_i \quad (2\text{-}16)$$

ρ_i：単位容量のファイルをアクセスするために使用する装置iでの使用時間の割合

$\rho_i^\circ(\hat{T}_R)$：レスポンスタイムの制限により決定される装置使用率

ρ_i°：装置iの最大使用率

C_i：装置iの価格

なお、$\rho_i^\circ(\hat{T}_R)$を厳密に求めることは、手法Aを適用することとなんら差がなくなってしまう。レスポンスタイムに対しては高い精度が要求されていないことから、以下では、主にデバイス待ちを考慮した近似を行なう。待ち行列モデルとしてはM／M／1を使用する。

3．高速系ファイル装置

記憶階層の最上位を構成する高速ファイル系には、一般にジャーナルのバッファ、リカバリ情報等のシステムファイル、あるいはインデックスのようなデータの一部など、特にアクセス頻度、レスポンスタイムに対する要求の厳しいファイルが格納される。

ここでは、高速系ファイル装置として以下のファイル装置を対象に議論を進める。

（1）主記憶装置

近年のMOSメモリ素子の急激な価格低下により、主記憶装置は大容量化する傾向にあり、主記憶装置の一部にファイルを常駐するシステムも出現してきている。

主記憶装置をファイル装置として使用した場合、極めて高いスループットや高速のレスポンスタイムが保証できるが、電源障害時に情報が揮発してしまうことや、疎結合マルチプロセッサ（LCMP）において複数系間での共用ができない等の問題がある。

（2）半導体ファイル装置

MOSメモリを記憶素子として使用し、従来のファイル記憶装置と同様に転送装置配下でI／Oインタフェースを介してアクセスされる装置である。既存のファイル装置に比べアクセスタイム、転送速度は高速である。また、バッテリバックアップ等により不揮発性を保証することも可能である。

（3）磁気ドラム装置（固定ヘッドディスク装置）

従来、ファイル装置で最も高速な装置として位置づけられていたが、コスト、性能ともに改善の傾向に乏しく、適用領域の見直しがせまられている。

（4）固定ヘッド部を持つ可動ヘッドディスク装置

可動ヘッドディスク装置の内部に数シリンダ分の固定ヘッド部を装備した装置である。

3．1　高速系ファイル装置のハード使用コスト

本節では、アクセス頻度a、レスポンスタイム$\hat{T}_R$以下という条件の下でのハード使用コスト評価式を求める。

A．主記憶装置

主記憶装置にファイルを常駐した場合、CPUから直接ファイルにアクセスできるため、転送装置等アクセス経路に関係するコストを考える必要はなく、主記憶上にファイルを常駐するコストのみに着目すればよい。

また主記憶装置の応答は極めて高速（数百ns～数μsのオーダ）であり、本検討のアクセス頻度の範囲（10^3回／MB・S程度以下）では使用率に比例した装置占有コストより単位容量のファイルを収容するコストの方が大きい。従って主記憶装置のハード使用コストは、p_{mem}を単位容量当たりの主記憶装置コストとすると、

$$H = p_{mem} \quad (3\text{-}1)$$

で評価される。

B．半導体ファイル装置、磁気ドラム装置（固定ヘッドディスク装置）

これらの装置は、転送装置配下に接続されたファイル装置であり、アクセスには転送装置、コントローラが介在する。更にCPUが入出力処理を行なうためのオーバヘッドも加算する必要がある。

$$H = h_{cpu} + h_{mem} + h_{CH} + h_c + h_d \quad (3\text{-}2)$$

h_{cpu}：CPU使用コスト　h_{mem}：主記憶入出力バッファコスト

h_{CH}：転送系使用コスト　h_c：コントローラ使用コスト

h_d：記憶装置使用コスト

ここでh_dは各ファイル装置のハード使用コストであり、2-16式に基づいて求められる。磁気ドラム装置、固定ヘッドディスク装置が、半導体ファイル装置に比べて性能面で劣る点は、ρ_iあるいは$\rho_i^\circ(\hat{T}_R)$を通してハード使用コストに反映される。各装置のハード使用コスト算出式を表3-1 に示す。

C．固定ヘッド部

ファイルを固定ヘッド部に格納した場合、アクセスに際しては可動ヘッド部も含めた装置全体を占有する。固定ヘッド部のハード使用コスト h_F は表3-1 に示す通りである。

3．2 高速ファイル系の構成

前節での評価式を用いて、高速ファイル系装置の適用領域等を考察する。

3.2.1 磁気ドラム装置、固定ヘッド部等の適用領域

図3-1 は、一例としてレスポンスタイム $\hat{T}_R$ =30ｍｓの場合をとり上げ、アクセス頻度を横軸にとり、高速ファイル系装置のハード使用コストを示したものである。

この図より、従来はアクセス頻度が10^{-1}～10回／MB･S程度の高速ファイル域に適用領域を持っていた磁気ドラム装置、固定ヘッドディスク装置、固定ヘッド部は、近い将来には適用領域がなくなり、主記憶装置、半導体ファイル装置に代替されることが考えられる。

この主な理由としては、以下の項目が挙げられる。

（ａ）ＭＯＳメモリ素子の価格低下率に比べて磁気ドラム装置、固定ヘッド部は価格低下に乏しく、また性能向上も期待できないことから、相対的に高価な装置になる。

（ｂ）磁気ドラム装置、固定ヘッド部は回転体であるため、半導体ファイル装置に比べると性能は大幅に劣る。このため高速ファイル域、特にアクセス頻度が数回／MB･Sをこえるとファイル収容率を落とさなければならなくなり、例えば単位

表3-1　高速ファイル系ハード使用コストの一例

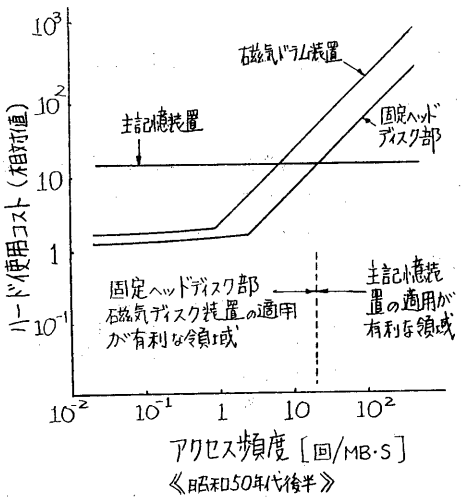
	CPUコスト	転送系コスト	コントローラコスト	記憶装置コスト
主記憶装置	～0	—	—	p_{mem}
半導体ファイル装置	$\frac{P_{CPU}}{\rho^\circ_{CPU}}\frac{DSiacs}{Mips}$	$\frac{P_{CH}}{\rho^\circ_{CH}}\alpha T_{CH}$	$\frac{P_C}{\rho^\circ_C}\alpha T_C$	$\max\left[\frac{\alpha\cdot T_S}{\rho^\circ}, \frac{1}{D}\right]P_S$
磁気ドラム，固定ヘッドディスク装置	〃	〃	〃	$\max\left[\frac{\alpha\cdot T_D}{\rho^\circ}, \frac{1}{D}\right]P_D$
固定ヘッドディスク部	〃	〃	〃	$\max\left[\frac{\alpha\cdot T_F}{\rho^\circ}P_{FHM}, P_F\right]$

ただし　$\rho^\circ = \min\left[\rho^\circ(\hat{T}_R),\ \rho^\circ_d\right]$

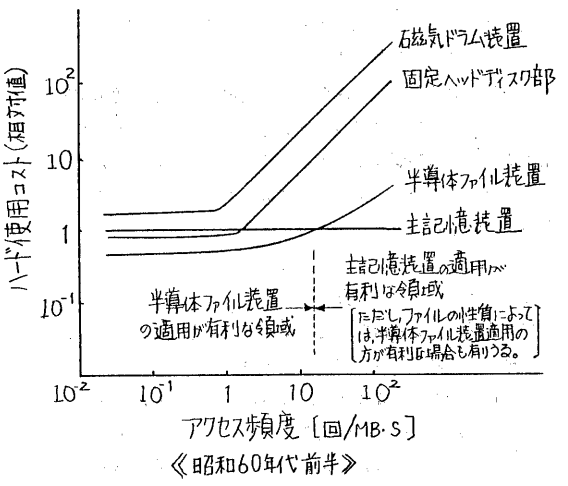
P_{CPU}：CPUのコスト　P_{CH}：転送系装置コスト　P_C：コントローラコスト
P_S　：半導体ファイル装置のコスト　P_{FHM}：可動ヘッド部を含めた固定ヘッドディスクのコスト
P_D　：磁気ドラム装置のコスト　p_{mem}：単位容量当りの主記憶コスト．
p_F　：単位容量当りの固定ヘッドディスクコスト
T_i　：装置iにおける占有時間　D：デバイス容量
$\rho^\circ(\hat{T}_R)$：レスポンスタイムの制限から与えられる装置使用率の上限
ρ°_i　：装置iに定められた最大使用率

容量当たりコストで主記憶装置が数倍程度高くても、主記憶装置を使った方がコスト性能比では優れている。

（ｃ）内部に固定ヘッド部を設けた可動ヘッドディスク装置は、本来高速ファイルを固定ヘッド部に、中速ファイルを可動ヘッド部に格納して、装置を効率良く使おうとすることをねらいとしている。しかし、アクセス頻度が数回／MB･S以上の高速ファイルを固定ヘッド部に格納すると、固定ヘッド部へのアクセスだけで装置使用率が高くなってしまい、可動ヘッド部はほとんど使用することができない。（固定ヘッド部



《昭和50年代後半》



《昭和60年代前半》

図3-1　高速系ファイル記憶装置の性能･コストの評価の一例

に格納するファイルのアクセス頻度と、可動ヘッド部に許される最大使用率の関係の一例を図3-2 に示す。） このため、ファイルのアクセス頻度が大きくなると、可動ヘッド部は無効な装置となり、この分、固定ヘッドディスク部のコストが割高となる。（図3-3 ）

3.2.2 主記憶装置と半導体ファイル装置の適用領域

主記憶装置と半導体ファイル装置は同等のＭＯＳメモリ素子を用いてはいるが、各々以下に示す特徴がある。

（ａ）主記憶装置はＣＰＵの要求から高速性を保証する必要があり、周辺論理回路あるいは素子自身の高速化にコストがかかる。この傾向は大型機ほど顕著である。
（ｂ）一方半導体ファイル装置は、アクセス時にＣＰＵ、チャネル等を使用するため、アクセス頻度が高くなると、これらの装置使用コストが無視できない。

これらの特徴をふまえ、以下で主記憶装置と半導体ファイル装置の適用領域を比較する。

アクセス頻度が数～30回／MB･S程度では、半導体ファイル装置の使用コストは以下のように書ける。

$$H = a \cdot k + p_s \qquad (3\text{-}3)$$

p_s ： 単位容量当りの半導体ファイル装置のコスト

ここで、kは1アクセス当りＣＰＵ、転送装置等を使用するコストである。

主記憶装置のハード使用コストは 3-1 式で表わされるため、半導体ファイル装置のハード使用コストと等しくなる境界は以下の式で求められる。

$$a = \frac{p_{mem}}{k}\left(1 - \frac{p_s}{p_{mem}}\right) \qquad (3\text{-}4)$$

この式で、将来主記憶のコスト p_{mem} が低下した場合、これに相関してＣＰＵ、転送装置等のコストも低下すると予想されることを考慮し、主記憶装置の単位容量当りコストと、半導体ファイル装置の単位容量当りコストの比

$$p_s \ / \ p_{mem}$$

を一定と仮定すると、半導体ファイル装置の適用領域はほとんど変化せず、主記憶装置との適用領域の境界は数～10回／MB･Sとなる。

例えば $p_s / p_{mem} = 0.5$ とした場合、境界は約10回／MB･Sとなる。

なお、主記憶装置と半導体ファイル装置の使い分けにあたっては、上記の他に不揮発性や、系間共用の要否に配慮することが必要である。[20][21]

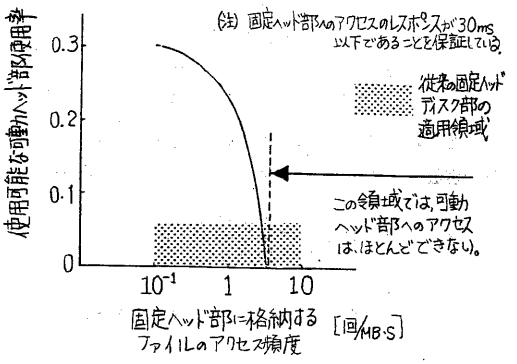


図3-2 固定ヘッド部使用による可動ヘッド部使用率の減少の一例

固定ヘッド部へのアクセスが頻繁になると、固定ヘッド部へのアクセスだけで装置全体の使用率が上がり、可能ヘッド部はほとんど有効に使うことができなくなる。

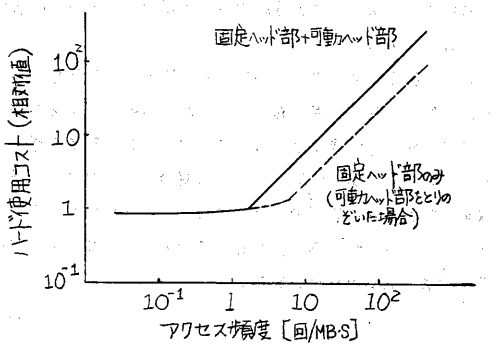


図3-3 固定ヘッドディスク部の性能・コスト

4．中速系ファイル装置

中速系ファイル装置としては、可動ヘッドディスク装置とディスクキャッシュが挙げられる。これらの装置には、総ファイル容量の大部分を占めるデータが格納されるため、これらのファイルの要求条件を満足し、かつコスト性能比にすぐれた構成をとることは、ファイル系設計の上で重要な課題である。

ここでは、中速系ファイル装置について以下の観点から議論を進める。

（１）磁気ディスク装置

ビットコストは10年で1/10と低下の傾向にあるが、反面デバイス容量は増加してきており、また回転待ち時間、シークタイム等アクセスタイムは飽和の傾向にある。主力のファイル装置として、要求されるアクセス頻度、レスポンスタイム

を満足する領域でコスト性能比がどの程度であるかを明らかにする必要がある。

また、磁気ディスク装置のファイル収容率を低下させることによりアクセス頻度の高いファイルを格納する手法が知られているが、これまでは止むを得ない場合の便宜的な手法に留っている。しかし磁気ディスク装置のビットコストの低下は高速ファイル系装置のビットコストの低下に比べて急激であることから比較的高いアクセス頻度のファイルに対して有用な手法となる可能性がある。

（2）ディスクキャッシュ

ディスクキャッシュ方式はディスクコントローラ等にディスク上のデータを一時的に格納しておくキャッシュメモリを設け、アクセス頻度・レスポンスタイムの改善を図り、従来の磁気ディスク装置よりやや高アクセス頻度・高速レスポンスを要求される領域をコスト性能比良くサポートしようとすることをねらいとしている。

既に数社の装置が市販されており、導入時の性能面での改善効果や、アクセス頻度とレスポンスタイムの関係等については数多くの発表がある。しかし、コスト性能比の改善という面でとらえた場合、ヒット率としてどの程度が必要なのか、また代替手法である高速系のファイル装置や、（1）に述べた磁気ディスク装置の収容率を下げる方法と比較してどのような効果があるのか、必ずしも明らかでない。

4．1　中速系ファイル装置のハード使用コスト

4.1.1　磁気ディスク装置

磁気ディスク装置のコスト算出は、基本的には磁気ドラム装置と類似の方法によって可能である。しかし磁気ディスク装置に特徴的な問題として再結合待ち時間の評価がある。磁気ディスク装置の場合、シーク動作時および回転待ち動作時には、デバイスとチャネル、コントローラとの経路をいったん切り離してオフライン動作を行なうのが通常になっており、この時の再結合待ち時間は、一般的にコントローラ使用率の関数となる。従ってデバイス保留時間がコントローラ使用率により変化することとなり、デバイス使用率の算出が複雑となる。ここでは簡単化のため、2章に示したようにコントローラは最大使用率 ρ_c° で使用されるものと仮定した。

この場合、磁気ディスク装置のハード使用コスト h_{DK} は、以下の4-1 式により容易に算出することができる。

なお、CPU、転送系、コントローラについては磁気ドラム装置の場合と同様である。

$$h_{DK} = max\left[\frac{a \cdot T_{DK}}{\rho_{DK}^\circ(\hat{T}_R)} , \frac{1}{D} \right] P_{DK} \qquad (4\text{-}1)$$

ただし $T_{DK} = T_{seek} + \frac{\rho_c^\circ}{1-\rho_c^\circ} T_P + T_{SCT} + \frac{\rho_c^\circ}{1-\rho_c^\circ} T_{rot} + T_{trans}$

$$\rho_{DK}^\circ(\hat{T}_R) = min\left[1 - \frac{T_{DK}}{\hat{T}_R} , \rho_d^\circ \right]$$

T_{seek}：平均シークタイム　　T_P：平均パス使用時間
T_{SCT}：平均回転待時間　　T_{rot}：回転時間
T_{trans}：データ転送時間　　T_{DK}：平均ディスク占有時間

4-1 式には既に、アクセス頻度が高くファイル収容率を下げなければならない場合の効果は近似的に組込まれている。$\frac{a T_{DK}}{\rho^\circ(\hat{T}_R)} > \frac{1}{D}$ となる場合がそれである。

しかし、一般にファイル収容率を下げて使用する時には、装置内部にファイルを分散して配置するのではなく、一箇所に集中し、連続するシリンダに格納することが多い。これにより収容率の低下とともにシークタイムの短縮が可能となるためである。この効果は、ファイル収容率が数十%の範囲では目立たないが、数%程度と極端に低下させた場合は顕著になる。[17]

磁気ディスク装置のファイル収容率を低下させ、高頻度にアクセスされるファイルを格納した場合のコストを正確に評価するため、以下にシークタイム短縮効果を考慮した場合の算出式を求める。

平均シークタイム T_{seek} をファイル収容率 η の関数 $T_{seek}(\eta)$ とすると、収容率 η で格納する場合の磁気ディスク装置占有時間 T_{DK} は次式で与えられる。

$$T_{DK}(\eta) = T_{seek}(\eta) + A_0 \qquad (4\text{-}2)$$

ここで A_0 は η に対して一定の部分であり、

$$A_0 = \frac{\rho_c^\circ}{1-\rho_c^\circ} T_P + T_{SCT} + \frac{\rho_c^\circ}{1-\rho_c^\circ} T_{rot} + T_{trans}$$

とした。4-2 式を 4-1式へ代入するとともに、デバイス中の有効容量が ηD となることを考慮すると、次式を得る。

$$h_{DK}(\eta) = max\left[\frac{a(T_{DK}(\eta)+A_0)}{\rho_{DK}^\circ(\eta)} , \frac{1}{\eta D} \right] P_{DK} \qquad (4\text{-}3)$$

ただし $\rho_{DK}^\circ(\eta) = min\left[1 - \frac{T_{DK}(\eta)+A_0}{\hat{T}_R} , \rho_d^\circ \right]$

ここで η は $0 < \eta < 1$ の範囲で任意に設定でき、コスト $h_{DK}(\eta)$ が最も安くなるように η を選ぶべきであることから

$$h_{DK} = \min_{0<\eta\leq 1}\left[max\left[\frac{a(T_{DK}(\eta)+A_0)}{\rho_{DK}^\circ(\eta)} , \frac{1}{\eta D} \right] P_{DK} \right] \qquad (4\text{-}4)$$

となる。上式に具体的な $T_{seek}(\eta)$ を与えることによっ

て、磁気ディスク装置のハード使用コストが算出可能となる。Tseek（η）は実際の磁気ディスク装置毎に異なるが、一例を図4-1 に示す。

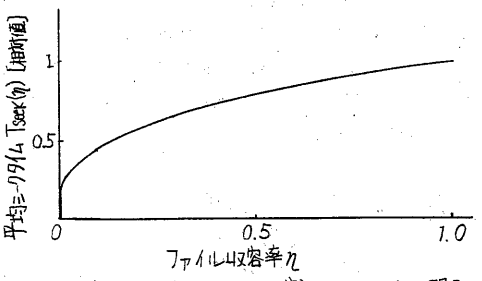


図4-1 磁気ディスク装置におけるファイル収容率と平均シークタイムの関係の一例

4.1.2 ディスクキャッシュ

ディスクキャッシュサブシステムとしては、図4-2 に示すような各々独立に動作可能なディスクコントローラ、デバイス（ディスクユニット）およびキャッシュ装置を想定した。図のトレーン中には、キャッシュの対象となるデバイスと対象とならないデバイスとが混在するものとした。各部分のハード使用コストは以下の考えに基づいている。なお、ディスクキャッシュの制御方式としては、一般的なストアスルー方式を想定した。

（1）キャッシュ装置

キャッシュ装置の使用状況としては次の2つの状態が考えられる。

（a）主記憶へのデータ転送、あるいはステージング等により、キャッシュ装置全体がダイナミックに占有されている状態。

（b）キャッシュ上にデータを保持し、キャッシュ対象ファイルへのアクセスを待っている状態。

（a）のコストは、他のコントローラ等と同様に時間的な占有率に比例するコストを考えれば良い。最大使用率は $\rho^{\circ}_{D/C}$

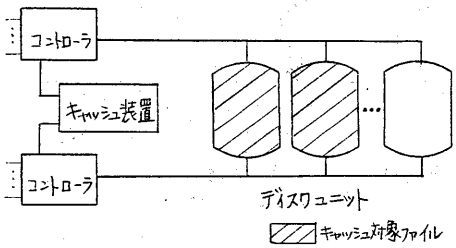


図4-2 ディスクキャッシュ サブシステム構成例

とした。

$$\frac{a\cdot\overline{T}_{D/C}}{\rho^{\circ}_{D/C}}\cdot P_{D/C} \qquad P_{D/C}:\text{キャッシュ装置のコスト}$$

ここで、$\overline{T}_{D/C}$ は、　　各状態での平均レスポンスタイムから計算される期待値である。　一方（b）のコストはキャッシュの空間的な占有率に比例するコストであり、全ファイルでキャッシュ装置全体のコストを分担していると考えられる。

$$\frac{P_{D/C}}{\sum F_i}=\sigma\frac{P_{D/C}}{D_{D/C}}$$

$\sum_i F_i$：キャッシュ対象ファイルの総容量

$D_{D/C}$：ディスクキャッシュ容量

$\sigma \triangleq D_{D/C}/\sum_i F_i$

キャッシュへのアクセス頻度が高く、キャッシュ装置の使用率が高まると、（a）の要因によるコストが支配的であり、またキャッシュ非対象ファイルへのデータ転送等が主で、キャッシュ装置の使用率が低い場合は（b）の要因によるコストが支配的になる。そこで、他の装置と同様に（a）と（b）のどちらかコスト寄与分の大きい方をとることによって、キャッシュのハード使用コストを評価できる。

$$h_{D/C}=\max\left[\frac{a\cdot\overline{T}_{D/C}}{\rho^{\circ}_{D/C}},\ \frac{\sigma}{D_{D/C}}\right]P_{D/C} \qquad (4\text{-}3)$$

（2）ディスクユニット

ディスクユニットのハード使用コストは以下の式で求められる。

$$h_{DKU}=\max\left[\frac{a\cdot\overline{T}_{DK}}{\rho^{\circ}_{DK}},\ \frac{1}{D}\right]P_{DK} \qquad (4\text{-}4)$$

ここで、平均ディスクユニット占有時間 $\overline{T}_{DK}$ は、ミスリード時のステージング動作時間、およびライトアクセス時のディスクユニット占有時間を考慮した期待値として求めた。

更に、最大使用率 ρ°_{DK} は以下のように求めた。

キャッシュ装置空き待ち時間は無視できるものと仮定すると、キャッシュサブシステムに対するレスポンスタイムの制限式は以下のように示せる。

$$\hat{T}_R \geq \overline{T}_{D/C}+\frac{\rho_{DK}}{1-\rho_{DK}}\overline{T}_{DK} \qquad (4\text{-}5)$$

上式より、レスポンスタイムから制限されるディスクユニットの最大使用率 ρ°_{DK} は次の式で求められる。

$$\rho^{\circ}_{DK}=\max\left[1-\frac{\overline{T}_{DK}}{\hat{T}_R-\overline{T}_{D/C}+\overline{T}_{DK}},\ \rho^{\circ}_{a}\right] \qquad (4\text{-}6)$$

4-8 式では、ヒット率の大小により、ディスクユニットの最大使用率が変化することが反映されている。

更に、ファイル収容率を下げてディスクユニットの高速化を図った場合には、4-8 式と同様にしてハード使用コストを求めることができる。

（3）コントローラ

コントローラは、キャッシュ対象ファイル、キャッシュ非対象ファイルのいずれもが共通に使っており、キャッシュ対象ファイルが使っていない時は、キャッシュ非対象ファイルに開放されていると考えられる。このため、コントローラをダイナミックに占有することによるコスト寄与分のみを考えた。

$$h_C = \frac{a\overline{T_C}}{\rho_C^{\circ}} \cdot P_C \qquad (4\text{-}7)$$

$\overline{T_C}$：平均コントローラ占有時間

P_C：コントローラコスト

4．2　中速ファイル系の構成について

前節での結果に基づいて、中速ファイル系記憶装置の適用領域等を考察する。

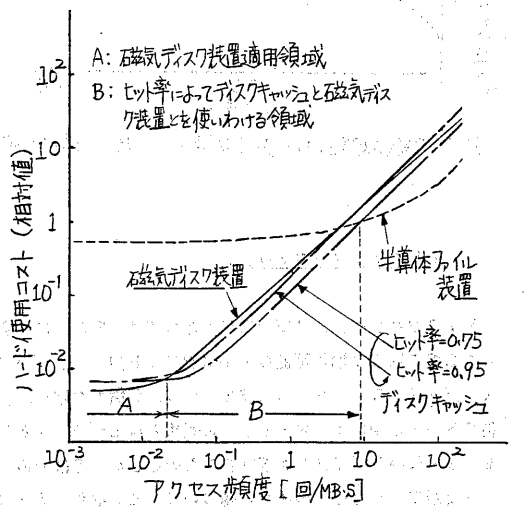
図4-3 に磁気ディスク装置およびディスクキャッシュサブシステムについて、コスト性能比の比較例を示す。中速ファイル系の要求レスポンスタイムの一例として$\hat{T}_R$ =50msの場合をとりあげ、アクセス頻度の関数としてハード使用コストを示した。

また図4-4 には、横軸に要求レスポンスタイム、縦軸にアクセス頻度をとり、各ポイント毎に、各記憶装置のハード使用コストを比較し、コスト最小となる装置を求めることにより、各記憶装置の適用領域を示した。

（1）磁気ディスク装置は、アクセス頻度が 10^{-1}-10^{-2}回/MB·S 以下、要求レスポンスタイムが 50-100 ms以上のファイルを格納するのに適している。また、ディスクキャッシュ導入時のヒット率が 0.7以上期待できず、要求レスポンスタイムが50-100ms程度のファイルに対しては、ファイル収容率の削減等により、アクセス頻度が数回/MB·S までの領域に対して適用可能である。

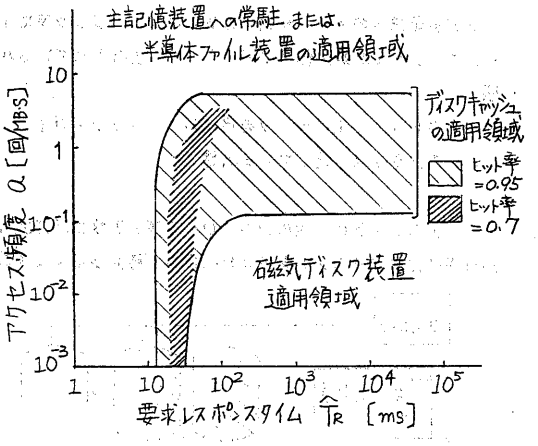
（2）ディスクキャッシュは、0.95程度の高ヒット率が期待できるファイルに対しては、効果が大きい。例えば、要求レスポンスタイム＝50msの場合、アクセス頻度が約 5×10^{-2}-10^{-1} 回/MB·S の範囲で、他装置に比べ最もコスト性能比にすぐれており、磁気ディスク装置に比べて、約1/2 のハード使用コストになっている。しかし、アクセス頻度が10回/MB·S 以上または、レスポンスタイム 10ms 以下の領域は半導体ファイル装置もしくは主記憶装置の有用な領域となり、ディスクキャッシュの適用は経済的ではない。

また、ヒット率が 0.7程度になった場合、磁気ディスク装置との優位差は、ほとんどなく適用領域は急激に狭くなる。これには、ミスリード率が高まるにつれてステージングの頻度が高まり、必要でないデータ転送の割合が増してくることが大きく影響している。



前提条件：要求レスポンスタイム=50ms ；データ転送量4KB
キャッシュ容量比 σ = 1/400 ；リードライト比=3:1

図4-3 中速系ファイル記憶装置 性能・コスト評価の一例



前提条件：データ転送量 4KB ；リードライト比=3:1
キャッシュ容量比 σ=1/400

図 4-4　ファイル記憶装置の適用領域

このように、ディスクキャッシュは、適用するファイルのアクセスローカリティの程度によって導入効果が大きく異なり、特にアクセス頻度が約 10^{-2}-10回/MB·S のファイルについては、ヒット率がどの程度期待できるのかファイルアクセス特性を十分考慮した上で、ディスクキャッシュを導入するのか、磁気ディスク装置でファイル収容率を削減するべきかを決定する必要がある。

5．まとめ

アクセス頻度とともにレスポンスタイムが重要視されるオンラインファイル系の階層構成を評価する手法として「ハード使用コスト」を用いることを提案し、これにより近い将来の階層構成の変化を考察した。

本手法は「要求されたスループットを満足し、かつ要求されたレスポンスタイム以下の応答性を保証するという拘束条件のもとで、アクセスに必要なすべての装置について、使用した分だけのコストの総和を求める。」という簡便な方法であり、従来、Ｌｕｍ，藤井等によって用いられた手法に対してレスポンスタイムが反映可能なよう拡張したものである。この拡張により、大幅に機能、性能、構成の異なる各種方式間の比較評価が可能となった。

ファイル系階層構成の検討にこの手法を適用し、高速ファイル系については、既存の磁気ドラム装置、固定ヘッドディスク装置等に代って主記憶装置、半導体ファイル装置が有用となること、及びその適用条件を明らかにした。

また、磁気ディスク装置を中心とする中速ファイル系については、新方式としてのディスクキャッシュについて、ヒット率をパラメータとして、レスポンスタイム、スループットの両面からコスト／性能を明らかにした。

この結果、ディスクキャッシュの適用領域が磁気ディスク装置と主記憶装置、半導体ファイル装置との中間的な領域となること、またヒット率0.7 程度では必ずしも適用領域が広くないことを示した。また、レスポンスタイムの要求がゆるやかならば、ファイル収容率を大幅に削減するという手法によって、数回/MB·S の高アクセス頻度領域まで磁気ディスク装置の適用領域があることが明らかになった。

以上、本稿では「ハード使用コスト」を用いた評価手法により、オンラインファイル系の構成について各階層でどのような種別の装置、方式が有用かを議論してきた。本手法は、対象とする装置、方式を必要に応じて細分してとらえることにより、性能、コスト面でいずれの箇所がネックになるか、どの程度効率的に動作しているか等を簡単に把握しやすいという特長がある。今後はこの特長を生かし、具体的にどのような装置構成、方式構成が望ましいかを検討してゆきたいと考えている。

謝辞

本研究を進めるにあたって御指導いただいた電電公社横須賀電気通信研究所データ処理研究部橋本統括役、データ処理方式研究室松永室長を始め、数多くの有益な助言をいただいたデータ処理研究部、データ通信研究部ならびに武蔵野電気通信研究所電子装置研究部の皆様に深く感謝いたします。

参考文献

（１）藤井、浅井：階層的ファイル自動管理システムの設計、情報処理学会論文誌、Vol.21 No.6 , pp. 442-453 (1980).
（２）Lum, V. Y. et al. : A Cost Oriented Algorithm for Data Set Allocation in Storage Hierarchies, C. ACM, Vol.18, No.6, pp.318-322 (1975).
（３）Peter. P.S.Chen : Optimal file allocation in multi-level storage system, N.C.C. 1973.
（４）Domenico Ferrari : Computer Systems Performance Evaluation , Prentice-Hall, Englewood Cliffs, N.J.(1978)
（５）E. Gelenbe, I. Mitrani : Analysis and Synthesis of Computer Systems , Academic Press,(1980
（６）山口 他：MT操作自動化方式に関する一検討，昭和58年度信学会総合全国大会 1602
（７）木ノ内、久保田：可動ヘッドディスクの使用容量削減による高速ファイル装置の代替について，昭和57年度信学会全国大会 1530
（８）田尻、木ノ内：ファイル記憶装置の適用領域に関する一考察，昭和56年度信学会全国大会 1548
（９）桜井、小橋、永津：ディスクキャッシュ適用領域に関する一考察，情処学会第26回全国大会 7N-3 .
（10）桜井、宮川：高速半導体ファイル装置構成法に関する一考察，情処学会第25回全国大会 4F-2 .
（11）桜井、宮川：系間共用可能な半導体ファイル装置に関する一考察，情処学会第24回全国大会 6H-5
（12）小橋、木ノ内：データ転送単位の大形化による入出力処理の価格・性能改善について，情処学会第22回全国大会 3J-7
（13）ＩＢＭマニュアルＧＡ32-0062-0 ＩＢＭ3880 STORAGE CONTROLLER MODEL 13
（14）山本 他：ディスク・キャッシュ装置を有する入出力サブシステムの高速化方式の評価，情処学会第26回全国大会 7N-5
（15）木下 他：ディスク・キャッシュを有する入出力サブシステムのシミュレーションによる性能解析，情処学会第26回全国大会 7N-6
（16）金子 他：FACOM 17741Aディスク・キャッシュ機構の方式，情処学会第25回全国大会