

動的配列シリンダディスク上のデータモデル

岡本正昭
大阪ガス

石井博昭
大阪大学

プロセッサ技術のめざましい発展の一方で、2次記憶の機械的制約によるボトルネックがますます顕著になり、この問題の早急な解決が望まれている。

動的配列シリンダディスク(DACYD)は、2次記憶内のデータを物理的に一定な長さで区切り利用頻度順に動的に再配置を行い、ヘッドシーク時間を短縮してコンピュータシステムの性能向上をめざすものである。

ここでは、データをいくつかのアクセス頻度グループにクラス分けしておき、アクセス頻度の変動が少なく再配置後も同一グループ内にとどまるデータは再配置を省略する簡便法を提案する。

この簡便法は、各グループの区間の幅を適切に設定すれば、2、3のグループ化程度で平均シーク距離をかなり短縮できるので、コスト対効果バランスの優れたデータモデルである。

さらに、データ再配置操作の実行間隔の目安を明らかにし実現可能なモデルであることも示す。

Data Model on Dynamically Arranged Cylinder Disk

Masaaki OKAMOTO
Osaka Gas, Ltd.

Hiroaki ISHII
Osaka University

While CPU technology has made remarkable progress, the performance bottleneck between the main memory and the secondary storage becomes more serious and must be solved immediately. Dynamically Arranged Cylinder Disk (DACYD) is aimed to reduce the wasteful head seek time through data rearrangement in the secondary storage.

Here, instead of perfect data rearrangement, we introduce a convenient method which divides data into several access frequency groups within a disk. If the size of each group is adequately decided, this method can decrease the head seek distance effectively even though in the case of a few data groups. Then, the necessary period between data rearrangement is showed to be equivalent to practicable I/O counts; i times of cylinder number.

1. はじめに

高度情報化時代になり、コンピュータシステムは急速に高機能、高性能、大規模化の道を進んでいる。とりわけプロセッサ技術の発達はめざましく、幾何級数的オーダーで性能が向上している。

他方、2次記憶系においては、薄膜・垂直磁化・光技術等を利用して、大容量化の面でかなりの成果が得られているが、高速化に関しては機械的な構造上の制約がありほとんど進歩していない。最近では、キャッシュ・バッファ、半導体ディスク、データベースマシンに見られるように、1次記憶と2次記憶の間に1.5次とも言うべき新しい記憶階層を導入して対応しようする傾向にある。

しかし、情報処理の裾野が広がり、コード情報のみならず文書・画像・音声を含むマルチメディア情報まで取り込もうとすると、それだけでは対応しきれず、大容量の2次記憶系自身の高速化は必須の緊急課題である。

我々が提案する「動的配列シリンダディスク(Dynamically Arranged Cylinder Disk - DACYD)」は、2次記憶内に蓄積されたデータに対するアクセス頻度に関する統計情報を利用して、データを頻度順に動的に再配置することで、機械的ボトルネックの1つであるヘッドシーク時間を短縮して、コンピュータシステムの性能向上をめざすものである。

本報告では、DACYDを適用した場合に予想される再配置作業コストの軽減とデータ管理法の簡素化を目的としてDACYD上でのより実際のモデルの提案を行う。

本モデルにおいて新しく導入したアクセス頻度グループ法によれば、各グループに分類されたシリンダ数の比を適切な割合にすることで、ごくわずかのグループ数でも十分2次記憶に対するアクセス性能の向上が期待できる。

2. 現状の問題点と解決策

従来のディスクでは、基本的に先に発生したデータから順に前づめであるいはキーの値の順序に従って記録されているが、この方法では、データを利用する立場が十分考慮されていない。

すなわち、データの更新が激しくて追加・削除が多数発生すると、ディスク内に虫食い領域や不連続な2次拡張域ができてヘッドシーク距離が増加し、プロセッサとディスク間のアクセスギャップがどんどん拡大してしまう。

そのためユーザーは、定期的なディスク内部のファイルの再編成やディスク間のアクセスバランスをとるためのデータ分散化作業に追いつまられているのが実状である。これは、アプリケーション分野の、華々しい展開に比較し、あまりにも原始的であり、高度情報化の基盤となるデータベースを格納する2次記憶系にふさわしい管理システムが望まれる。

我々は、問題解決のキーワードがデータの流動化にあると考える。従来固定化されていたデータの格納場所を動的に変更することにより、データのアクセス頻

度、検索順序を考慮した管理が可能になり、2次記憶系の高速度アクセスが実現できる様にする。

3. システムの構成

データを流動化するには、3つの基本機能が実現されなければならない。

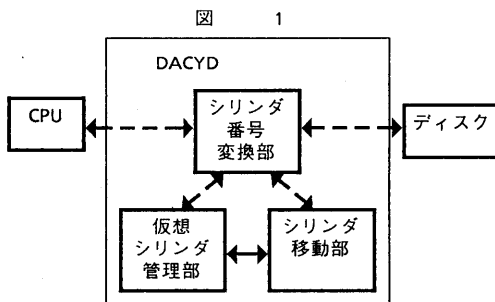
第1点は再配置データの単位で、個々のファイルにとらわれずにシリンダ、トラック、セクタ等の物理的に一定なサイズにする。

第2点は再配置場所で、各データのアクセス頻度、新鮮度、連鎖関係を考慮して決定する。

第3点は再配置時期で、それぞれのディスクに対する総I/O回数が一定回数以上になる毎に、再配置操作をバックグラウンドタスクとして実行する。

これにより、業務システムに影響を与えることなしに、時間経過とともに変動するアクセスパターンに追従可能なデータ構造が2次記憶内に実現できる。

システムの構成は、図1の様にCPUとディスクの間に介在し「シリンダ番号変換部」「仮想シリンダ管理部」「シリンダ移動部」からなる。ここでは、トラック、セクタ等の場合も含めシリンダで表現を統一する。



「シリンダ番号変換部」は、ディスク入出力時にCPUから見たシリンダ番号(仮想シリンダ番号)と実際にデータが記録されているディスク内のシリンダ番号(実シリンダ番号)の間の変換を行う。

「仮想シリンダ管理部」では、仮想シリンダと実シリンダの対応情報、さらにモニタされたデータのアクセス回数、新鮮度、相互結合強度等の情報を保持する。

「シリンダ移動部」は、「仮想シリンダ管理部」に蓄積された情報を利用して、各仮想シリンダの重要度の優先順位を決定し、それに基づき新しい実シリンダ番号に再配置する。

4. DACYDの効果

従来ディスクに記録された情報をDACYDで再配置した場合の概念図を図2に示す。図の横軸はディスク内に記録されたデータのシリンダ位置を、縦軸は各シリンダのアクセス頻度を表す。

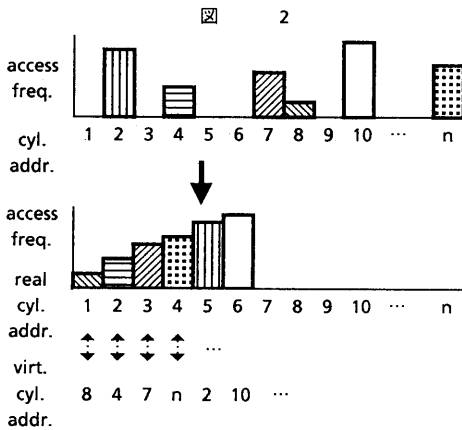


図2からDACYDによるヘッドシーク距離の短縮効果として、シリンダの隙間を詰める圧縮効果と、アクセス頻度順に並べる傾斜効果の2つの因子が相乗的に働くことがわかる。

文献 [1]、[2]によれば、圧縮効果はアクセスされるシリンダの両端間の距離の減少量に比例して平均シーク長の短縮量が決まる。傾斜効果は、三角型のアクセス頻度分布において、最良の山形分布と最悪の谷型分布とで平均シーク長の比が7:12になる。シリンダ幅がnでアクセス頻度が一様分布の場合の平均シーク長n/3に比較して、それぞれ0.7倍、1.2倍の大きさである。

ここで、平均シーク長の短縮効果を調べるための評価式を説明する。[3]

シリンダ幅nのディスクにおいて、ヘッドが任意の位置r (0 ≤ r ≤ n) にアクセスできるものとする。

今rが位置xから位置t (0 ≤ x, t ≤ n) に移動するものとし、x、tにおけるアクセス頻度確率をそれぞれp(x)、p(t)とし、tはxに依存しないとすると、

$$\int_0^n p(r) dr = 1 \quad (1).$$

の条件のもとで、平均シーク長は、

$$L = \int_0^n \int_0^x p(x)p(t)(x-t) dt dx + \int_0^n \int_x^n p(x)p(t)(t-x) dt dx \quad (2).$$

と表される。(x-t)、(t-x)の項をx、tについて微分し、部分積分を実行すると、

$$L = \frac{n}{2} - 2 \int_0^n \left\{ \int_0^x p(r) dr - \frac{1}{2} \right\}^2 dx \quad (3).$$

が得られる。

(3)式の内側の積分は、累積アクセス頻度分布を意味し単調増加関数になっているので、累積アクセス頻度分布が1/2になる位置を中心にアクセス頻度分布のパターンを解析すれば、平均シーク長の変化が定性的に評価可能になることがわかる。そして、アクセス頻度分布が簡単な数式で表現できれば、(3)式から平均シーク長の定量的評価も可能である。

5. 実用データモデル

現実のディスク内のデータの利用傾向は時々刻々変化し、厳密な意味で頻度順に管理するのは困難である。極端なケースでは、過去の統計情報を利用して再配置を行っても、アクセスパターンが大幅に変化すると、折角の労が無駄になるだけでなく逆効果にもなりかねない。

このような実運用システムでは、もう少し大まかな管理でそこそこの効果が得られる方がコスト対効果の面から望ましい。

ここでは、具体的方策のひとつとして、「アクセス頻度グループ」の概念を新しく導入し、その定量的評価を試みる。

アクセス頻度グループは、仮想シリンダをアクセス頻度順に配列した上で、頻度の小さい方から(あるいは大きい方から)順に適当な個数のシリンダ群にクラス分けすることによって構成される。

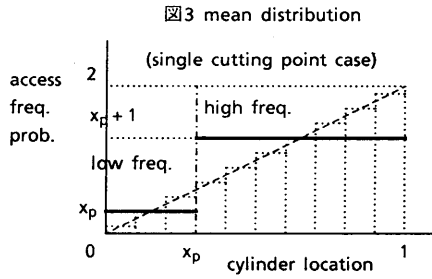
そして、シリンダの再配置を行う際は、アクセス頻度が大幅に変化したためグループ間をまたがって移動する必要のあるシリンダのみ再配置対象とし、アクセス頻度の変化が少なく同じグループ内の移動で済む場合は再配置の対象とせず移動作業を省略する。

必然的にこの簡便法のもとでは、ひとつのアクセス頻度グループ内における各シリンダのアクセス頻度の分布はランダムな状態になる。

さらにグループをディスク内に格納する際は、シリンダの中央に最もアクセス頻度の大きいグループを置き、その両側に次いでアクセス頻度の大きいグループを配置する。以下アクセス頻度が相対的に大きいものが内側になる様に配置し、結果として、ほぼ山型のアクセス頻度分布パターンが形成され自然に最適分布に近づくようにする。

ここで問題は、アクセス頻度グループの数と各アクセス頻度グループに収容されるシリンダの数の決定である。

以下、この問題を、アクセス頻度のシーケンスが一次直線型になっている場合について調べる。



(1) 2グループ分割の効果

最初に2グループに分割した場合の最適な分割位置をアクセス頻度分布図3を利用して検討する。

図3では計算を簡便にするためにシリンダ全体幅を1とする。横軸の x_p を2つのアクセス頻度グループに分割する位置とする。各グループ内のアクセス頻度分布がランダムになるので、個々のシリンダのアクセス頻度をそれが属しているグループのアクセス頻度の平均値に等しいと仮定しても不自然ではない。すると、左側の低頻度グループのアクセス頻度確率は x_p 、右側の高頻度グループのアクセス頻度確率は x_p+1 と見なすことができる。

これらの値を、平均シーク長の評価式(3)に適用すると、

$$L = \frac{1}{2} - 2 \int_0^{x_p} \left\{ \int_0^x x_p dr - \frac{1}{2} \right\}^2 dx - 2 \int_{x_p}^1 \left\{ \int_0^{x_p} x_p dr + \int_{x_p}^x (x_p+1) dr - \frac{1}{2} \right\}^2 dx \quad (4).$$

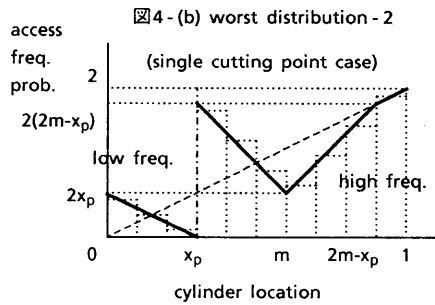
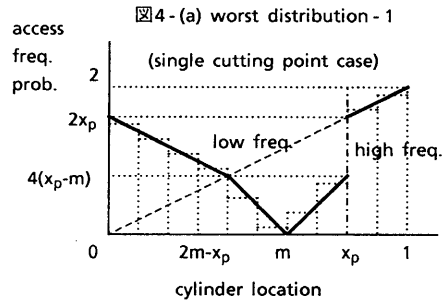
となる。この積分を計算した結果は分割位置 x_p の4次関数

$$L = \frac{1}{3} - \frac{1}{3} x_p (x_p - 1) (2 x_p^2 - 1) \quad (5).$$

になる。

次に2グループ分割において、平均シーク長が最悪になる場合を考える。図4-(a),(b)は、最悪時のアクセス頻度確率の分布を示す。図の横軸の m はアクセス頻度の累積分布が1/2になる位置を表す。(a)は m が低頻度グループに属する場合、(b)は m が高頻度グループに属する場合を表す。平均シーク長が最悪になるのは、 m の付近にそのグループの中で最もアクセス頻度の小さいシリンダが置かれ、 m から遠ざかるにつれてアクセス頻度が大きくなる場合である。

図4-(a)では、最悪のアクセス頻度分布のパターンは、4本の1次直線区間に分けて考えることができる。



すなわち、各区間のアクセス頻度確率を

$$p(x) = \begin{cases} -2(x-x_p) & ; 0 \leq x \leq 2m-x_p \\ -4(x-m) & ; 2m-x_p < x \leq m \\ 4(x-m) & ; m < x \leq x_p \\ 2x & ; x_p < x \leq 1 \end{cases} \quad (6).$$

として

$$\int_0^{2m-x_p} \left\{ -2(x-x_p) \right\} dx + \int_{2m-x_p}^m \left\{ -4(x-m) \right\} dx = \frac{1}{2} \quad (7).$$

の条件を満たす

$$m = \frac{2x_p - (2x_p^2 - 1)^{1/2}}{2} \quad (8).$$

を考慮して、評価式(3)を計算すればよい。形式的には i 番目の区間のアクセス頻度確率を

$$p_i(x) = u_i x + v_i \quad (9).$$

とおき、次の評価式

$$L = \frac{1}{2} - 2 \sum_{i=1}^4 \int_{a_i}^{b_i} \left[\sum_{j=1}^i \int_{c_j}^{d_j} p_j(r) dr - \frac{1}{2} \right]^2 dx \quad (10).$$

から、平均シーク長を分割位置 x_p の関数として求めることができる。 m がアクセス頻度の大きいグループに属している図4-(b)の場合も同様である。

平均シーク長が最良になる場合は図5-(a), (b)の形のアクセス頻度確率分布で考えると、同様に(9), (10)式で平均シーク長が計算できる。

以上求めた分割位置 x_p と平均、最悪、最良アクセス頻度分布時の平均シーク長との関係をグラフにしたのが図6である。

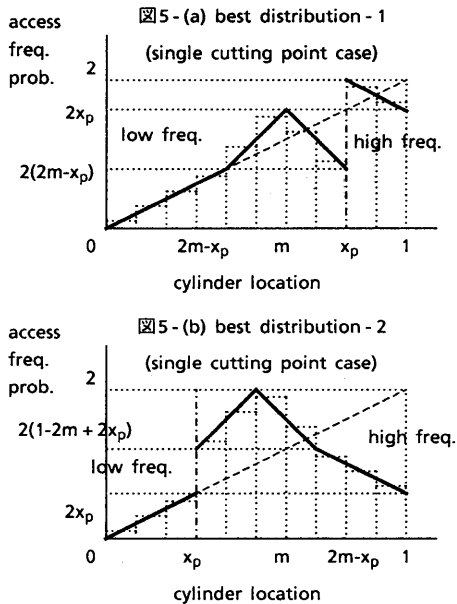


図 6

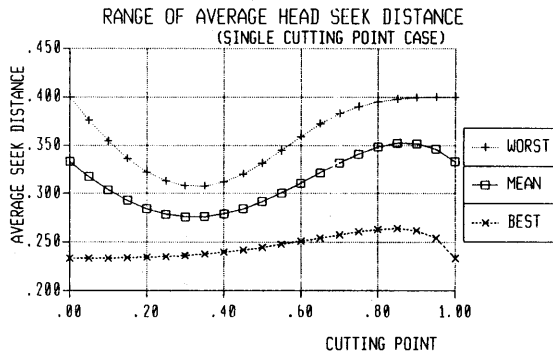


図6から、平均アクセス頻度分布時の平均シーク長が最小値 $L=0.2757$ になる分割位置 $x_p=0.3223$ において、最悪分布時の平均シーク長は 0.3075 とかなり小さくなっており、最良分布時の平均シーク長も 0.2369 であり大きくなっていないことがわかる。

この平均分布時の平均シーク長の最小値 $L=0.2757$ は、一様分布の時の平均シーク長 $1/3$ より約17%小さくなっている。

以上から2グループ分割時には $x_p=0.3223$ が望ましい分割位置と言える。

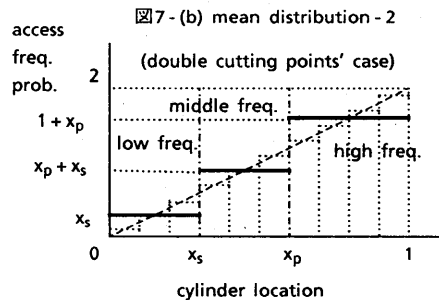
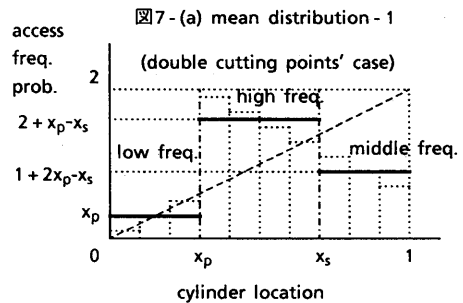
(2) 3グループ分割の効果

3グループ分割の場合も、2グループ分割と同様に考えることができる。この場合の平均的なアクセス頻度分布の状態は図7-(a), (b)のパターンになる。

図7-(a), (b)において、横軸の x_p, x_s は2箇所の分割位置を表す。(a)ではアクセス頻度グループが、小、大、中の順に配置されているので $x_p < x_s$ になっている。(b)ではアクセス頻度グループが、小、中、大の順に配置されているので $x_p > x_s$ になっている。

各グループ内のアクセス頻度分布を2グループの場合と同様にランダムと考えると、個々のシリンダのアクセス頻度をそれが属しているグループのアクセス頻度の平均値に等しいと仮定すると、平均シーク長の評価式(3)は次のような形式で表される。

$$L = \frac{1}{2} - 2 \sum_{i=1}^3 \int_{a_i}^{b_i} \left[\sum_{j=1}^i \int_{c_j}^{d_j} p_j(r) dr - \frac{1}{2} \right]^2 dx \quad (11).$$

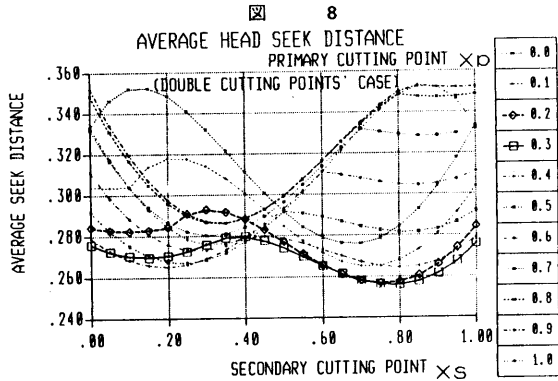


ただし、第*i*区間のアクセス頻度確率は $p_i(x) = u_i \cdot \text{const.}$ である。

(11) 式の計算結果は、2つの分割位置の座標 x_p, x_s の関数となり、グラフにすると図8になる。

図8から、平均的アクセス頻度分布の場合には、 $x_p = 0.2599$ 、 $x_s = 0.7727$ の時に平均シーク長が最小値 $L = 0.2550$ になる。

この値は、一様分布の時の平均シーク長1/3より約23%小さくなっている。



次に、2つの分割位置が $x_p = 0.2599$ 、 $x_s = 0.7727$ の時にアクセス頻度が最悪分布、最良分布の場合の平均シーク長の値を調べる。

3グループの場合の最悪分布も、2グループの場合と同様に考える。ただし、平均分布の平均シーク長が最小になるは $x_p < x_s$ の時だから、図9-(a), (b), (c), (d) の4つのパターンについて平均シーク長を計算すればよい。すなわち、5つの区間にわけて(9), (10)式と同様の計算をして図10のグラフを得る。

グラフから、 $x_p = 0.2599$ 、 $x_s = 0.7727$ の時の最悪分布の平均シーク長は $L = 0.2722$ となり、最悪のケースでもかなり小さい値になっていることがわかる。

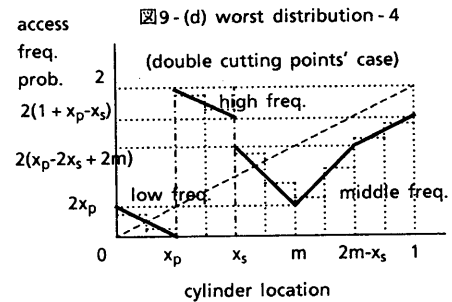
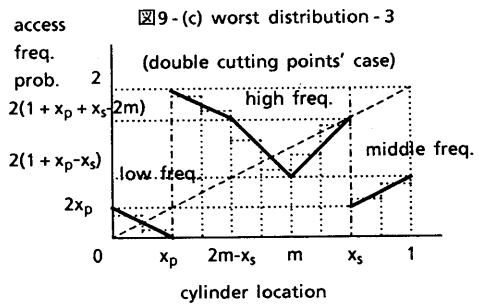
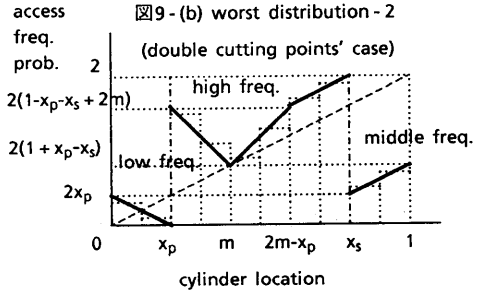
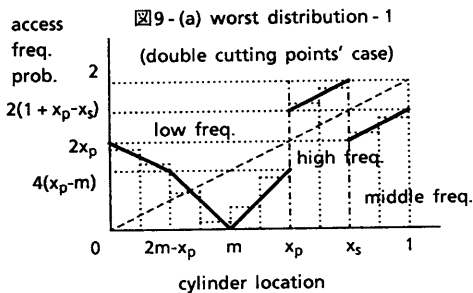
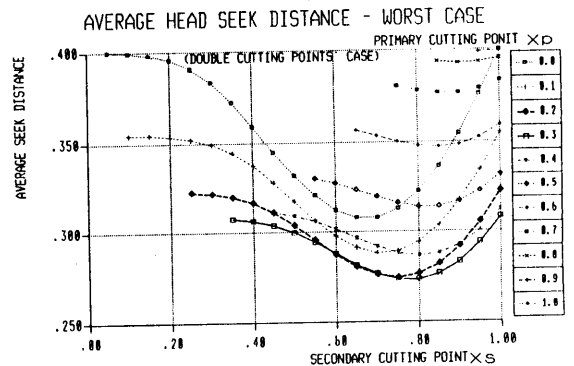
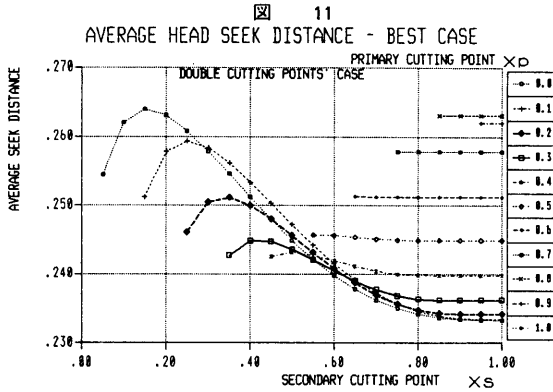


図 10



最良分布の場合も同様にして図11のグラフを得る。
 グラフから、 $x_p=0.2599$ 、 $x_s=0.7727$ の時の最良分布の平均シーク長は $L=0.2359$ となり、最良のケースもあまり大きな値になっていないことがわかる。
 以上から、3グループ分割時の効果的な分割位置として $x_p=0.2599$ 、 $x_s=0.7727$ が得られた。



(3) 2グループ分割のコスト

2グループに分割した場合にデータの再配置に必要な作業コストを検討する。コスト計算方法は、再配置によるデータの移動状況を表すシリンダ配置遷移図12を基に考える。

図12の横軸は再配置前のシリンダ位置を表し、縦軸は再配置後のシリンダ位置を表す。斜線部分は、再配置前後のシリンダ位置が同一グループに属するので、シリンダ移動作業が省略できることを示す。

ここで制約条件として、再配置作業に際し、各グループ内で左寄りに位置するシリンダは再配置先のグ

ループ内でも左寄りに位置し、各グループ内で右寄りに位置するシリンダは再配置先のグループ内でも右寄りに位置するようにするという条件を付加する。この条件は現実の問題に適用するのにそれ程厳しい条件ではない。

すると、再配置前の任意のシリンダ位置 k_i は図12の太い実線に沿って新しいシリンダ位置 m_i に移動するとみなせる。

図12から遷移確率の具体的な値が、

$$q_1 = \frac{x_p}{n} ; 0 < x_a \leq x_p, x_p < x_b \leq n \quad (12).$$

$$q_2 = \frac{n-x_p}{n} ; x_p < x_a \leq n, 0 < x_b \leq x_p \quad (13).$$

となるので、再配置後のシリンダ位置は、

$$m_1 = (k_1 - x_p) \frac{x_p}{n - x_p} ; 0 < m_1 \leq x_p, x_p < k_1 \leq n \quad (14).$$

$$m_2 = x_p + k_2 \frac{n - x_p}{x_p} ; x_p < m_2 \leq n, 0 < k_2 \leq x_p \quad (15).$$

と決定される。

そこで再配置作業コストを、再配置対象データの移動に要するヘッドシーク距離の総和で評価すると、

$$C = \sum_{k_2=1}^{x_p} q_2(m_2 - k_2) + \sum_{k_1=x_p+1}^n q_1(k_1 - m_1) \quad (16).$$

と表すことができる。(16)式の計算結果は、

$$C = n^2 \left\{ -\left(\frac{x_p}{n} - \frac{1}{2}\right)^2 + \frac{1 + \frac{1}{n^2}}{4} \right\} \quad (17).$$

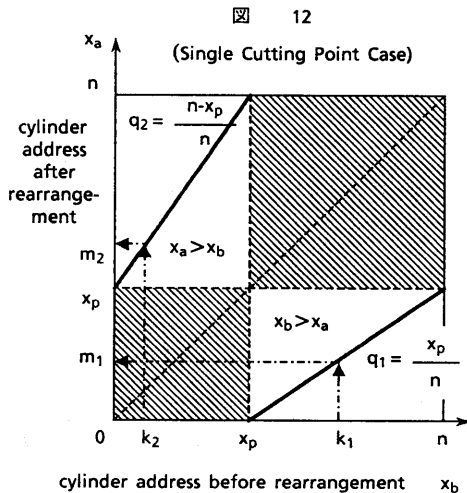
となるが、シリンダ全体幅 n が十分大きい場合には、この式は分割位置 x_p の2次関数

$$C \approx n^2 \left\{ -\left(\frac{x_p}{n} - \frac{1}{2}\right)^2 + \frac{1}{4} \right\} \quad (18).$$

で近似できる。(18)式をグラフに表した図13から、2グループ分割時の最適分割位置 $x_p=0.3223n$ において、再配置コスト $C=0.2184n^2$ を得る。

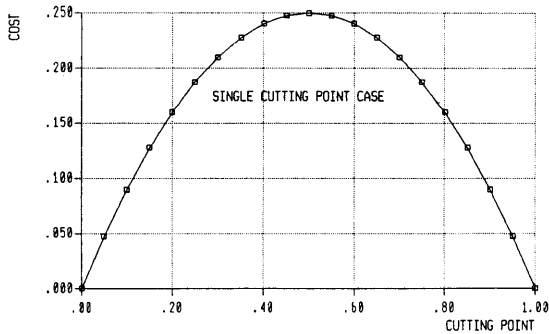
従って、

$$\frac{0.2184n^2}{\frac{n}{3} - 0.2757n} = 3.787n \quad (19).$$



より、総I/O回数が3.787n回以上になる間隔で再配置操作を実行すれば、コストに優る効果が得られる。

図 13
COST OF REARRANGEMENT



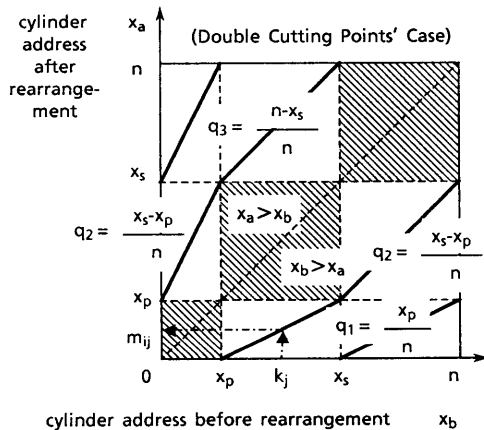
(4) 3グループ分割のコスト

2グループ分割と同様の手順で3グループ分割の場合の再配置作業コストを考えることができる。シリンダ配置遷移図14をもとにして、コスト評価式は

$$C \approx n^2 \left\{ \frac{x_p x_s (x_s - x_p)}{n^3} + \frac{x_s (n - x_s)}{n^2} \right\} \quad (20).$$

となる。(20)式をグラフに表した図15から、3グループ分割時の最適分割位置 $x_p = 0.2599n$, $x_s = 0.7727n$ において、再配置コストが $C = 0.2786n^2$ となることがわかる。

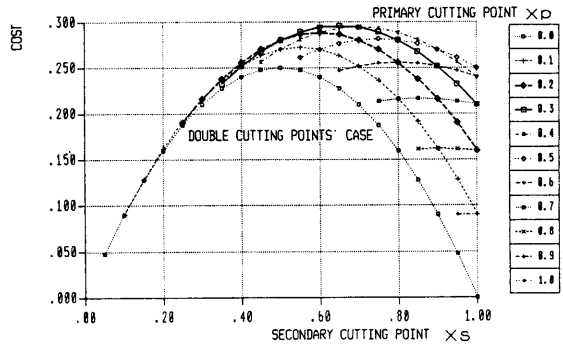
図 14



$$\frac{0.2786n^2}{\frac{n}{3} - 0.2550n} = 3.557n \quad (21).$$

より、総I/O回数が3.557n回以上になる間隔で再配置操作を実行すれば、コストに優る効果が得られる。

図 15
COST OF REARRANGEMENT



6. おわりに

2次記憶の機械的ボトルネック解決の方法として、2次記憶内のデータを流動化し統計情報を利用して再配置する「動的配列シリンダディスク」を紹介した。

さらに、実用的データモデルとして「アクセス頻度グループ」の概念を導入した簡便法を提案し、わずかのグループ数でも各グループの分割位置を適切に設定すれば十分効果があることを示した。そして、再配置作業コストを考慮した場合に有効となる再配置作業間隔を定量的に示した。

ここでは、アクセス頻度のシーケンスが一次直線型になる分布について検討したが、より一般的な分布についても同様の手順で考えることができる。

今後、ヘッドシーク時間だけでなく回転待ち時間も含めて、2次記憶系のパフォーマンスを総合的に検討していきたい。より長期的な展望として、統計情報によるデータ流動化をメインメモリやキャッシュ・バッファを含む階層メモリシステム全体に拡張できれば、データ管理は完全にコンピュータシステム内部の処理にゆだねられ、ユーザーは自らのニーズに基づき業務に専念できるようになるものと期待できる。

<参考文献>

- [1] 岡本: 「動的配列シリンダディスクの構成法」, アーキテクチャシンポジウム, pp.183-189, May, 1988.
- [2] 岡本, 石井: 「動的配列シリンダディスクの効果分析」, 37回全国大会, pp.238-239, Sep., 1988.
- [3] 岡本, 石井: 「動的配列シリンダディスクに関する数理モデル」, OR学会昭63秋季研究発表会, pp.196-197, Sep., 1988.