論　文

高速応答ビデオサーバ向きの2層ディスクアレー方式

西村　一敏[†]　丸山　充[†]　阪本　秀樹[†]　鈴木　偉元[†]
中野　博隆[†]

A Double-Layered Disk Array for Quick-Response Video Servers

Kazutoshi NISHIMURA[†], Mitsuru MARUYAMA[†], Hideki SAKAMOTO[†],
Hideharu SUZUKI[†], and Hirotaka NAKANO[†]

あらまし　読取り制御に循環タイムスロット方式を用いるビデオサーバを対象に，2層ディスクアレー新方式とこれに適合したスキップサーチを提案すると共に，性能についての最悪値理論解析を行って以下のことを解明している．今日のバンデッドレコーディング磁気ディスク装置を用いる場合のタイムスロット時間の最悪値は，最大アクセス時間と平均データ転送速度というディスク性能で近似できる．単一ディスク従来方式に対するディスクアレー従来方式の読取り多重度の改善倍率は，秒単位実時間の約0.7～1.1倍となる正規化最大待ち時間と1の和が限界値である．新方式では，更に，最大待ち時間で正規化した先頭映像の許容切捨て時間と1の和の倍率まで改善できる．パリティディスクを付加した高信頼2層ディスクアレー方式では，切捨て時間の制約がなければ，サブアレーの並列ディスク台数に数台という最適値が存在する．更に，ビデオサーバの試作により，提案方式の機能を検証すると共に，最大待ち時間1秒の高速応答と許容切捨て時間3秒を条件に，1.5 Mbit/s映像の100多重の読取り性能を実現している．

キーワード　ビデオオンデマンド，ビデオサーバ，多重読取り，ディスクアレー

1. まえがき

近年注目されているビデオオンデマンドでは，蓄積センタ内の中核装置として，磁気ディスク装置（DK）への映像情報の蓄積と多重読取りを行うビデオサーバが使用される．ビデオサーバには，経済性とサービス性の観点から，相反関係にある高多重読取りと高速応答が同時に要求される．映像情報の要求から読取り開始までの待ち時間を小さく制約した高速応答の条件では，循環タイムスロット方式[1]～[3]がスキャン方式[4],[5]よりも大きな許容多重度を実現できるので優れている[4],[6]．

循環タイムスロット方式を採用する各種ビデオサーバの性能を解明し比較することは，蓄積センタを設計する上で極めて重要である．しかし，従来の理論解析は固定データ転送速度のDKについてしかなされておらず[2]～[4]，データ転送速度がシリンダ位置に依存する今日のバンデッドレコーディングDKについての報告，および最大待ち時間とDK総数を同一条件とした比較についての報告は見当たらない．

本論文では位相シフト時分割多重アクセス方式[7]を拡張した2層ディスクアレー方式とこれに適合したスキップサーチを新たに提案し，これらを試作検証する．また，理論解析により，バンデッドレコーディングDKの読取り処理時間の最悪値を解明し，単一ディスク[1]～[3]，ディスクアレー[1],[4],[5],[8]の両従来方式と新方式について，最悪値設計により保証できる許容多重度を最大待ち時間やDK総数を同一にして比較する．

2. で読取り処理時間の最悪値を明らかにした後，**3.** で従来方式の性能を解明する．**4.1** で2層ディスクアレー新方式の性能を従来方式と比較し，**4.2** でパリティディスクを付加した高信頼2層ディスクアレー方式の最適構成を明らかにする．**4.3** でスキップサーチを提案すると共に，**4.4** で新技術について試作検証する．

2. タイムスロット時間

循環タイムスロット方式の多重読取り原理を図1に示す．符号化した映像情報をセグメントと呼ぶ単位に

[†] NTTヒューマンインタフェース研究所，武蔵野市
NTT Human Interface Laboratories, 3-9-11 Midori-cho, Musashino-shi, 180 Japan

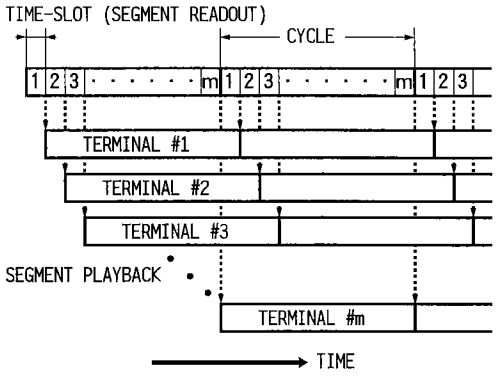

図1　循環タイムスロット方式の多重読取り
Fig.1　Multiple-readout using cyclic time-slots.

分割してDKに蓄積しておく．DKからは1回の処理で1セグメントを読み取る．セグメントにアクセスしてバッファメモリへ転送している時間をタイムスロットと呼ぶ．1セグメントを端末で復号して再生している間に，DKからは他の複数の端末のために順にセグメントが読み取られる．また，その再生が終了する前には，後続のセグメントが読み取られて補給される．このように，循環タイムスロット方式では，m 台の端末のためにセグメントの読取りを実行する m 個のタイムスロットが周期的に繰り返される．

許容多重度を決定するタイムスロット時間は，DKの制御オーバヘッド時間，シーク時間，回転待ち時間，およびセグメントのデータ転送時間の和である．固定データ転送速度のDKの場合には，シーク時間と回転待ち時間が最大となる場合に，タイムスロット時間が最悪となる．しかし，表1に例を示す今日のDKでは，記憶容量を増大できるバンデッドレコーディングを採用している．すなわち，内周から外周に向けてデータ転送速度が増大するので，最悪タイムスロット時間は従来のように単純には求まらない．

最悪タイムスロット時間は，データ転送速度が最小となる最内周シリンダ内のセグメントと，あるシリンダ内のセグメントを交互に読み取る場合に生じる．シリンダ数から1を減じた値を c として，最内周の隣のシリンダ番号を $1/c$，最外周のシリンダ番号を1に正規化して任意のシリンダ番号を x で表すと，最内周シリンダとシリンダ x 間のシーク時間 $Seek$ は，

$$Seek(x) = a + b\sqrt{x} \qquad (1/c \leqq x \leqq 1) \qquad (1)$$

で近似できる[9]．但し，a と b は最小および最大シーク時間から求まる定数である．

表1　3.5インチ磁気ディスク装置の仕様例
Table 1　Example of 3.5-inch disk-drives.

		I	II	III
フォーマット容量〔MB〕		1,050	4,300	4,512
シリンダ数		3,898	3,832	4,416
バンド数		29	16	10
制御オーバヘッド〔μs〕		~25	~50	~280
シーク時間〔ms〕	最小	1	0.8	0.5
	最大	22	20	16.5
回転数〔rpm〕		5,400	7,200	7,200
データ転送速度〔Mbit/s〕	最小	37	35.4	76.7
	最大	62	68.3	100.6
信頼度 MTBF〔万時間〕		80	80	100

データ転送速度は，最小値 $D_{min.}$ から最大値 $D_{max.}$ まで x に比例して変化すると仮定する．回転待ち時間は，最悪値を議論しているので最大値 L とする．制御オーバヘッド時間は，他の時間と比較して十分に小さいので無視する．このとき，最内周シリンダ内およびシリンダ x 内のデータ量 S のセグメントを交互に読み取る場合の1セグメント読取り処理当りに平均化したタイムスロット時間 T は，

$$T(x) = Seek(x) + L + \frac{S}{2D_{min.}}\left(1 + \frac{1}{1+(R-1)x}\right) \qquad (2)$$

で与えられる．但し，R は

$$R \equiv D_{max.}/D_{min.} \qquad (3)$$

で定義される速度比である．ここで，付録1.より

$$\theta = MIN[\pi/6, \tan^{-1}\sqrt[3]{R-1}] \qquad (4)$$

として，

$$S \leqq bD_{min.}/(\sqrt{R-1}\sin\theta\cos^3\theta) \qquad (5)$$

の条件が満足されれば，式(2)は単調に増加する．但し，$MIN[y,z]$ は y と z のうちで小さい方を与える関数である．上記により，平均化タイムスロット時間の最悪値は，

$$A \equiv a + b + L \qquad (6)$$

$$D \equiv 2/(1/D_{min.} + 1/D_{max.}) \qquad (7)$$

で定義される最大アクセス時間 A および逆数で平均したデータ転送速度 D で表現できる．

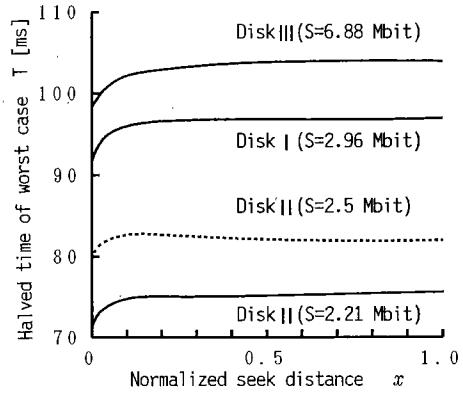

図2 2シリンダ交互読取りの処理時間
Fig. 2 Operation time of reciprocal readout from two cylinders.

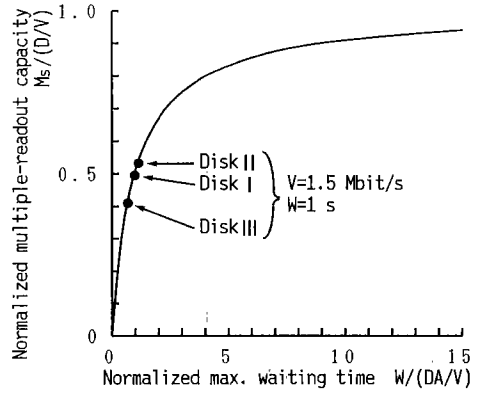

図3 単一ディスク方式の多重読取り性能
Fig. 3 Multiple-readout performance of a single disk-drive.

式(2)で求めた平均化タイムスロット時間 T を図2に示す．表1のDKについて，式(5)の等号が成立する条件とした．縦軸は T，横軸は正規化シーク距離 x を表す．T を単調増加させるセグメントデータ量 S の上限は，DK個別のデータ転送速度の大きさに比例して大きくなり，2.21~6.88 Mbitであった．段階的に変化する真のデータ転送速度に基づく T は，図2に例示した近似曲線と両端で一致し，途中で複数回交差する折れ線状となる．この T と近似理論値との間の誤差は，付録2.の方法で大きく評価しても，DKのI, II, IIIについて，各々1.1, 2.9, 1.6%と十分に小さい．なお，S が大きくなって式(5)が成立しなくなると，図2に破線で示すように T はN字型の曲線へと変化する．

3. 従来方式の性能

3.1 単一ディスク

セグメントのデータ量 S が2.の式(5)を満足する範囲において，許容多重度の最悪値 M_s は，

$$M_s = (S/V)/(S/D + A) \tag{8}$$

で与えられる[2]．但し，V は映像情報のビットレートである．M_s は厳密には式(8)の右辺を超えない最大の整数であるが，本論文では簡略化して実数として扱う．M_s 台の端末から同時に映像情報の要求があり，最後に処理された先頭セグメントの読取りが完了するまでの最大待ち時間 W は，

$$W = S/V \tag{9}$$

で与えられ，周期時間に等しい．式(9)を使い式(8)の S を消去して式(10)が得られる．

$$M_s = \frac{D}{V} \cdot \frac{VW/(DA)}{1 + VW/(DA)} \tag{10}$$

式(10)で求めた単一ディスク方式の多重読取り性能を図3に示す．縦軸は D/V で正規化した許容多重度 M_s，横軸は DA/V で正規化した最大待ち時間 W を表す．図3中の点は，V =1.5 Mbit/s, W = 1sの条件で，表1のDKで実現できる M_s (DKのI, II, IIIについて各々15, 16, 23) を示す．式(5)の前提条件が成立する S の最大値で与えられる W は，DKのI, II, IIIについて各々2.0, 1.5, 4.6秒である．これよりも W が大きくなると M_s に誤差を生じ，許容多重度の真の理論限界である $D_{min.}/V$ に漸近する．

従来の $D_{min.}$ を用いた最悪値設計による許容多重度は，DKのI, II, IIIについて各々13, 14, 22である．D を用いた方が，大きくて15%高性能となる．この差は，R, S が大きく，A が小さいほど大きくなる．

3.2 ディスクアレー

ディスクアレー方式は，セグメントを更に細分して通常のDK複数台に分散蓄積し，これらを並列に読み取ってDK台数倍のデータ転送速度を実現する．DK総数を s とすれば，1タイムスロットでDK1台から読み取るデータ量は，単一ディスクの場合の $1/s$ と小さくなる．従って，式(5)が単一ディスクで成立していれば，ディスクアレーでも成立する．以上より，ディスクアレーの許容多重度 M_a は式(10)の D を sD に

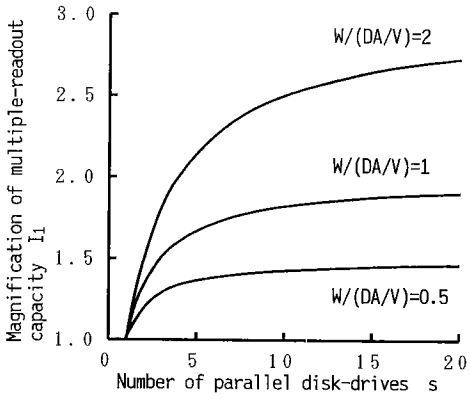

図4 ディスクアレー方式の効果
Fig. 4 Effect of disk array.

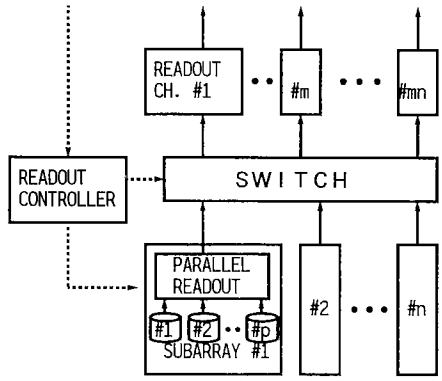

図5 2層ディスクアレー方式のビデオサーバ
Fig. 5 Video server with double-layered disk array.

置き換えて，

$$M_a = \frac{D}{V} \cdot \frac{VW/(DA)}{1+VW/(sDA)} = \frac{sD/V}{1+sDA/(VW)} \qquad (11)$$

で与えられる．厳密には右辺を超えない整数であるが，M_s と同様，実数として扱う．式(11)と式(10)の比で，ディスクアレーの単一ディスクに対する許容多重度の改善倍率 I_1 を定義する．

$$I_1 \equiv \frac{M_a}{M_s} = \frac{1+VW/(DA)}{1+VW/(sDA)} \qquad (12)$$

式(12)で求めたディスクアレー方式の効果を図4に示す．縦軸は I_1，横軸はDK総数 s を表す．パラメータとして正規化最大待ち時間を0.5，1，2と変化させた．正規化最大待ち時間が小さいほど，I_1 も小さくなり，かつ s の増大に伴って早く飽和傾向になることがわかる．これは，最大待ち時間が小さいほどセグメントデータ量も小さく，データ転送速度の増大によるタイムスロット時間減少の飽和が早いためである．また，いずれの場合にも I_1 は正規化最大待ち時間と1の和で飽和する．映像ビットレートを1.5 Mbit/sとすると，正規化最大待ち時間が2（DKのI，II，IIIについて各々実時間で2.0，1.8，2.9秒）の場合でも，改善は3倍以下である．

4. 新 方 式

4.1 2層ディスクアレー

新たに提案する2層ディスクアレー方式の構成を図5に示す．DK p 台のディスクアレーである許容多重度 m のサブアレー n 台が，スイッチを介して mn 個の読取りチャネルと接続される．読取り制御部は，サブアレーの多重読取りとスイッチの切換えを制御する．映像情報のセグメントは，先頭からサブアレー1，2，…n，1，2，…の順に蓄積する．

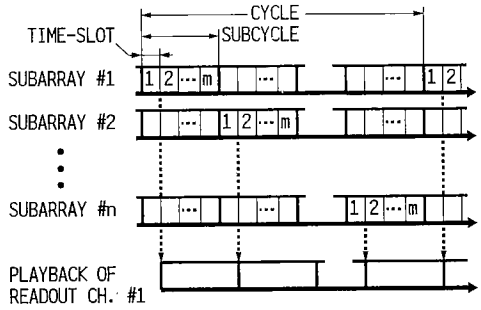

図6 2層ディスクアレー方式の多重読取り動作
Fig. 6 Multiple-readout from double-layered disk array.

本方式における多重読取り動作の原理を図6に示す．サブアレーのタイムスロット循環周期を小周期（Subcycle）と呼ぶ．この小周期ごとに，セグメントの蓄積順にサブアレーを切り換えて読取りを行う．n 台のサブアレーから並行して m 多重の読取りを実行するので，全体では $m \times n$ の許容多重度が得られる．許容多重度以内であれば，新規の要求を受け付けた時点から小周期時間の範囲内でサブアレー1から n の順に調べれば，必ず空きタイムスロットが見つかる．そこから読取りを開始すれば先頭部分のセグメントを切り捨てることになるが，最大待ち時間は小周期時間に短縮される．

$n=1$ の場合はディスクアレー方式であり，$p=1$

の場合は位相シフト時分割多重アクセス方式[7]である．これらと2層ディスクアレー方式の優劣を要求条件と許容多重度の観点から明らかにする．また，2層ディスクアレーはディスクアレーのグループ分けでも実現できる．実現規模は，インタフェースボードの実装量・転送速度に制約されるが，サブアレーを装置として分離すれば拡大する．なお，映画ビデオ20タイトルについて調査したところ，製作会社の表示映像から本編開始までの時間が9秒以上であった．また，約4〜7分の音楽ビデオ20タイトルの場合は，題名・歌手名または導入映像の表示から音楽開始までの時間が6秒以上であった．一般的にも同様に，映像情報の先頭部分の工夫によって，数秒程度の先頭部切捨ては利用者の許容範囲にできる．

式(11)の s を p に置き換えて得られるサブアレーの許容多重度 m が，整数となる場合について考察する．DK総数 s には

$$s = pn \tag{13}$$

の関係があるので，2層ディスクアレーの許容多重度 M_d は，

$$\begin{aligned} M_d &= mn \\ &= \frac{pD/V}{1+pDA/(VW)} \cdot n \\ &= \frac{sD/V}{1+pDA/(VW)} \end{aligned} \tag{14}$$

で与えられる．先頭映像の最大切捨て時間 F は，

$$\begin{aligned} F &= (n-1)W \\ &= (s/p-1)W \end{aligned} \tag{15}$$

となる．式(14)より，$p=1$ かつ W が許容される範囲の最大値で M_d は最大になり，このとき n に等しい s に比例することがわかる．この場合，サブアレーは単一のDKとなり，位相シフト時分割多重アクセス方式[7]そのものとなる．

以下では，許容切捨て時間 F_p が W の整数倍で与えられ，かつ $F_p < F(p=1)$ の場合，すなわち

$$s > F_p/W + 1 \tag{16}$$

の場合について考察する．このとき p より n が優先して制約され，式(15)を F_p に等しいと置いて，

$$n = F_p/W + 1 \tag{17}$$

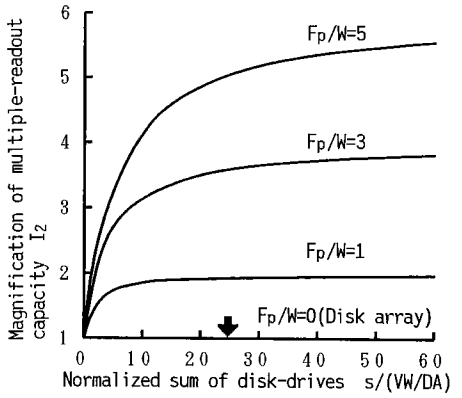

図7 2層ディスクアレー方式の効果
Fig. 7 Effect of double-layered disk array.

を得る．$F_p = 0$ のときは $n=1$ となり，ディスクアレー方式そのものとなる．式(14)と式(11)の比で，2層ディスクアレーのディスクアレーに対する許容多重度の改善倍率 I_2 を定義すると，

$$\begin{aligned} I_2 &\equiv \frac{M_d}{M_a} \\ &= \frac{1+sDA/(VW)}{1+(F_p/W+1)^{-1}sDA/(VW)} \end{aligned} \tag{18}$$

となる．但し，式(13)で p を消去すると共に，式(17)で n を置き換えた．

式(18)で求めた2層ディスクアレーの効果を図7に示す．縦軸は I_2，横軸は正規化最大待ち時間で正規化したDK総数 s を表す．パラメータとして最大待ち時間 W で正規化した許容切捨て時間 F_p を1，3，5と変化させた．正規化許容切捨て時間が大きいほど，I_2 も大きくなり，かつ s の増大に伴う飽和傾向が緩和されることがわかる．これは，s の増大に連れて，式(11)で与えられるディスクアレーの許容多重度が飽和するのに対し，2層ディスクアレーの許容多重度は切捨て時間を増大すれば比例傾向になるからである．また，いずれの場合にも I_2 は正規化許容切捨て時間と1の和で飽和する．

4.2 高信頼2層ディスクアレー

長時間の映像情報を蓄積する場合には，DK台数が多くなるのでDK故障に備えた対策が必要となる．そこで，サブアレーごとにパリティDK1台を追加した高信頼2層ディスクアレー方式を提案する[10]．Oyangらも同様の構成を提案しているが，スキャン方式でのバッファメモリ削減をねらいとしているので，セグメントの蓄積配置と読取り方法が異なる[11]．パリティ

DKには，DK間にまたがって演算したパリティデータを蓄積する．サブアレー内の1台のDKが故障しても，そのデータは残りのDKのデータから復元できる．このサブアレーは，RAID (Redundant Arrays of Inexpensive Disks) のレベル3または4に相当するが，レベル5でも同等機能を実現できる．DK故障が発生した場合やデータ復元中には，サブアレーの性能が10%程度低下するので，許容多重度を変更して対処する．

サブアレー内のデータDKの台数を p とすると，DK総数 s，許容多重度 M_r および最大切捨て時間 F は，各々式(13)，式(14)，式(15)と同様に，

$$s = (p+1)n \tag{19}$$

$$M_r = \frac{pW/A}{p + VW/(DA)} \cdot \frac{s}{p+1} \tag{20}$$

$$\begin{aligned} F &= (n-1)W \\ &= \{s/(p+1) - 1\}W \end{aligned} \tag{21}$$

で与えられる．高信頼2層ディスクアレーでは，有効なデータDKの台数がDK総数よりも小さくなるので，評価にはこれを加味すべきである．そこで，許容多重度と，有効蓄積容量を同一にした蓄積コストとの比で，評価関数 E を定義する．

$$\begin{aligned} E(p) &\equiv M_r/\{s(p+1)/p\} \\ &= \frac{W}{A} \cdot \frac{p^2}{(p+1)^2\{p + VW/(DA)\}} \end{aligned} \tag{22}$$

サブアレーに予備DKを装備する構成もあるが，DKは故障時に取り替えることを想定して，E の中には取替え部品のコスト（運用経費）は考慮していない．

ここでまず，F が制約されない場合について考察する．E の W に関する偏導関数は常に正なので，W が許容範囲の最大値のときに E が最大となる．従って，W は許容最大値とする．E の p に関する偏導関数は p が

$$P_{opt.} \equiv \left\{1 + \sqrt{1 + 8VW/(DA)}\right\}/2 \tag{23}$$

で定義される $P_{opt.}$ のときに0になり，このとき E は最大となる．すなわち，サブアレー内のデータDK台数 p には，最適値 $p_{opt.}$ が存在する．$P_{opt.}$ が整数の場合は $p_{opt.} = P_{opt.}$ である．そうでなければ $P_{opt.}$ の切下げと切上げを行った二つの整数のうち，式(22)を大きくする方の整数が $p_{opt.}$ である．

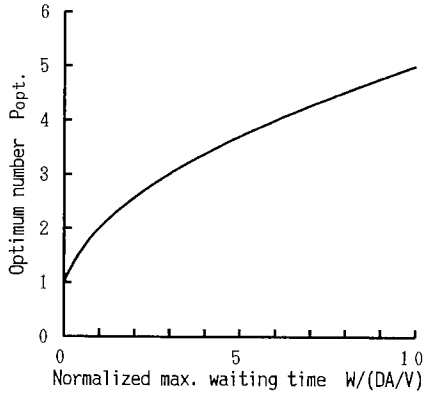

図8　高信頼2層ディスクアレー方式における並列データDKの最適数
Fig. 8　Optimum number of parallel data disk-drives for redundant double-layered disk array.

式(23)で求めた $P_{opt.}$ を図8に示す．縦軸は $P_{opt.}$，横軸は正規化最大待ち時間を表す．正規化最大待ち時間を0から大きくするに連れて，$P_{opt.}$ は1から次第に大きくなる．これは下記の理由による．最大待ち時間（周期時間）が十分に小さくてサブアレーの許容多重度が $p=1$ で1であれば，$p=2$ でも1となり，E は式(22)より各々1/4および2/9となる．すなわち，$p=1$ が最適となる．最大待ち時間を大きくすると，図4で示したように並列DK増大による許容多重度改善倍率が大きくなり，かつDK総数に対する有効DK台数の割合も大きくなるので，$P_{opt.}$ が増大する．最大待ち時間が数秒程度の高速応答条件では，サブアレーの並列DKは数台が最適である．

次に，許容切捨て時間 F_p が W の整数倍として与えられ，かつ $F_p < F(p = p_{opt.})$ の場合，すなわち

$$s > (p_{opt.} + 1)(F_p/W + 1) \tag{24}$$

の場合について考察する．このとき p より n が優先して制約され，式(21)を F_p に等しいとおいて，

$$n = F_p/W + 1 \tag{25}$$

$$p = sW/(W + F_p) - 1 \tag{26}$$

を得る．

以上の高信頼2層ディスクアレー方式の設計フローを図9に示す．まず，設計条件として映像ビットレー

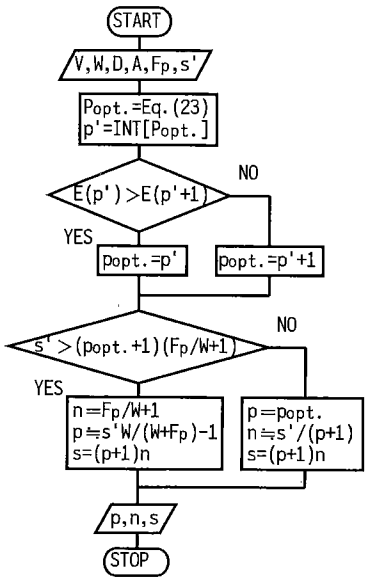

図9 高信頼2層ディスクアレー方式の設計フロー
Fig. 9 Design flow for redundant double-layered disk array.

ト V，最大待ち時間 W，DK の逆数平均データ転送速度 D，最大アクセス時間 A，許容切捨て時間 F_p，および仮の DK 総数 s' を与える．次に，評価関数 E を最大にする $P_{opt.}$ を式 (23) で求め，これに最も近い方の二つの整数の中で E をより大きくする方を最適データ DK 台数 $p_{opt.}$ とする．図 9 中の関数 $INT[y]$ は，yを超えない整数を与える．次に，式 (24) の条件について調べ，成立する場合にはサブアレー台数 n を優先して式 (25) で決定すると共に，式 (26) で得られる値に近い値にデータ DK 台数 p を決定する．反対の場合には，p を優先して $p_{opt.}$ に決定すると共に，式 (19) で得られる値に近い値に n を決定する．最後に，DK 総数 s が $p+1$ と n の積として確定する．

4.3 スキップサーチ

ビデオサーバには，許容多重度を低減させないビジュアルサーチが要求される．筆者らが既に提案した専用データ方式は蓄積データ量を増加させる [8]．蓄積データ量の増加がなく，2 層ディスクアレーに適合したビジュアルサーチの新方式として，スキップサーチを提案する [10], [12]．Chen らが提案している方法は，これをディスクアレー向きに変形したものとなっている [13]．

スキップサーチの動作原理をサブアレーが 4 台の場合を例に図 10 に示す．映像情報はセグメントに細分され，先頭からサブアレー 1，2，3，4 を繰り返す順に蓄積されている．1 セグメントの再生時間は，説明の便宜上 1 秒だと仮定する．セグメントの読取りに使用するサブアレーは小周期ごとに 1，2，3，4 の順に繰り返して切り換えられ，通常再生の場合は先頭のセグメントから順に読み取られる．

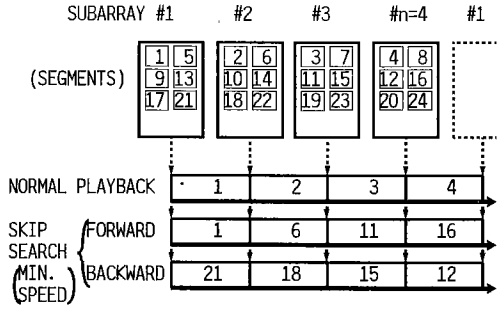

図10 2層ディスクアレー方式のスキップサーチ
Fig. 10 Skip search for double-layered disk array.

順方向のスキップサーチの場合は，次に使用するサブアレーに蓄積されている後続セグメントを 1 個（一般には J 個）飛ばして読み取る．端末では，1 秒間の通常再生の後，4 秒（一般には $J \times n$ 秒）分をスキップする動作が繰り返される．一方，逆方向のスキップサーチの場合には，次に使用するサブアレーに蓄積されている先行セグメントを読み取る（一般には $J-1$ 個飛ばして読み取る）．端末では，1 秒間の通常再生の後，2 秒（一般には $J \times n - 2$ 秒）分をスキップする動作が繰り返される．J を指定することにより，$J \times n + 1$ 倍速の順方向スキップサーチ，および $J \times n - 1$ 倍速の逆方向スキップサーチを実現できる．

4.4 試作ビデオサーバ

高信頼 2 層ディスクアレー方式およびスキップサーチを採用したビデオサーバを試作し，これらの機能を検証した．試作ビデオサーバの構成を図 11 に，その外観を図 12 に示す．映像音声情報は MPEG (Moving Picture Coding Experts Group) −1 で符号化し，そのビットレートは 1.5 Mbit/s である．

最大待ち時間は 1 秒，従ってセグメントを 15 フレームの GOP (Group of Pictures) 二つ分とした．また，DK 機種は表 1 の I，DK 総数は 28 台，許容切捨て時間は 3 秒とした．式 (23) より $P_{opt.} = 2.0$ となり，直ちに並列データ DK の最適台数として 2 が得られる．式 (24) の関係が成立するので許容切捨て時間が優先

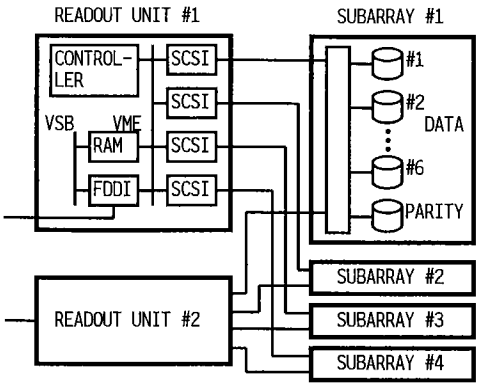

図11　試作ビデオサーバの構成
Fig. 11　Configuration of experimental video server.

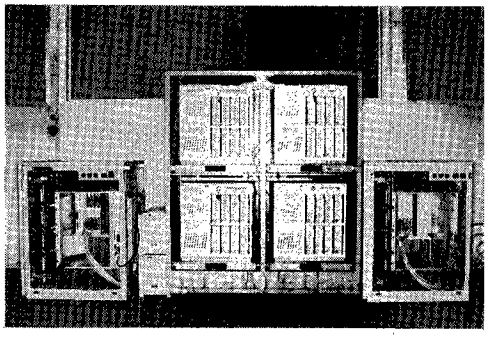

図12　試作ビデオサーバの外観
Fig. 12　Experimental video server.

されることになり，式(25)よりサブアレーは4台（並列データDKは6台）の構成となる．式(11)と式(20)で与えられるサブアレーおよび高信頼2層ディスクアレーの許容多重度の最悪理論値は，各々25および100である．単一ディスクの15多重に対してディスクアレーでは1.9倍の29多重となり，それよりも更に3.4倍の改善効果がある．

端末との通信インタフェースには，TCP/IP (Transmission Control Protocol/Internet Protocol) を高速処理するFDDI (Fiber Distributed Data Interface) ボード[14]を採用した．1枚の通信能力は1.5 Mbit/sでは50チャネルなので，読取りユニット2組に各々2ホストインタフェースを有するサブアレー4台を接続する構成とした．図5に示した複数の読取りチャネルは，バッファメモリ（RAM）のアドレス空間に分散している．また，スイッチは，サブアレーの2ホストインタフェース，および読取りユニットのVME (Versa Module Europe) バスに分散している．バッファメモリから高速FDDIボードへのデータ転送には，VSB (VME Subsystem Bus) を採用している．

特殊再生として，ポーズ，コマ送り，スロー，ジャンプ，順逆両方向の早送りおよびスキップサーチを実現した．低速系の再生は，読取りおよび通信の休止と再開によるフロー制御[2]で実現した．ジャンプは，次の小周期で使用するDKに蓄積されているうちで，ジャンプ先に最も近いセグメントから読み取ることで実現した．従って，読取りの開始時と同様，最大待ち時間1秒を保証する代わりに，最大3秒分のジャンプ誤差を有する．早送りは専用データ方式[8]で15倍速を実現した．すなわち，15フレームごとに1フレームを抽出し，これをフレーム内符号化して作成した順方向と逆方向の専用データをセグメントに分割して蓄積した．早送りが要求されると，ジャンプと同様に，次の小周期で使用するDKに蓄積されている要求時点以降の最も近い専用データセグメントから読取りを開始する．スキップサーチは，4.3の方法で，順方向97倍，逆方向99倍までの可変速度を実現した．

片方の読取りユニットに端末9台を接続し，残りの負荷はワークステーションで代行して，100多重の読取り試験を行った．端末から映像情報を要求して表示を開始するまでの応答時間は，100多重の負荷をかけた状態での平均で1.1秒であった．このうち，ビデオサーバが要求を受信して映像情報を送信開始するまでの時間は，0.54秒と高速であった．また，このときスキップサーチなどの特殊再生を実行しても，他端末での再生に影響のないことを確認した．以上により，新方式の有効性と解析の妥当性が検証できた．

5. むすび

ビデオオンデマンドの経済性とサービス性の向上をねらいとして，循環タイムスロット方式に分類されるビデオサーバを対象に，新方式の提案，その試作検証，および最悪値理論解析による従来方式との性能比較を行った．提案した2層ディスクアレー新方式は，m 多重読取りのディスクアレー方式サブアレー n 台により，$m \times n$ の許容多重度を可能にする．かつサブアレーにパリティディスクを付加することにより，大量映像蓄積時の信頼性を向上できる．提案したスキップサーチは，通常再生と同様の読取り動作により，許容多重度の劣化なしでビジュアルサーチを可能にする．試作ビデオサーバでは，映像ビットレート1.5 Mbit/s,

先頭映像の切捨て時間3秒以下の条件で，100多重の読取りと待ち時間1秒以下の高速応答を達成した．

理論解析では，今日のバンデッドレコーディングDKを使用する場合について，以下のことを解明した．読取り処理時間の最悪値は，逆数で平均したデータ転送速度と最大アクセス時間で近似できる．映像ビットレート1.5 Mbit/s，最大待ち時間1秒の条件では，単一ディスク方式の許容多重度は20程度である．これに対するディスクアレー方式の改善倍率は，秒単位実時間の約0.7~1.1倍の正規化最大待ち時間と1の和が限界値である．新方式では，更に，最大待ち時間で正規化した許容切捨て時間と1の和の倍率まで改善可能である．高信頼方式のサブアレー内データDK台数は，切捨て時間の制約がなければ数台程度が最適である．

バンデッドレコーディングDKに関する近似が成立するのは，現状では読取りデータ量が数Mbit以下の場合である．この成立範囲は，今後のデータ転送速度の向上に伴って拡大する．性能解析の内容は，固定データ転送速度の従来型DKについても成立する．今後の課題は，スイッチを中心とする拡張性に優れた構成法の実現である．

謝辞 本研究の機会を与えて頂いた吉利誠メディア応用システム研究部前部長，ならびに試作ビデオサーバの負荷試験に御協力頂いた関係各位に深謝します．

文　献

[1] K. Nishimura, T. Mori, Y. Ishibashi, and N. Sakurai, "System architecture for digital video-on-demand services," Proc. IEEE Singapore ICIP '92, pp.602–606, Sept. 1992.

[2] 西村一敏，森　達男，石橋　豊，"多重読取り可能なビデオオンデマンドシステム，" テレビ誌，vol.48, no.3, pp.287–294, March 1994.

[3] P.V. Rangan and H.M. Vin, "Designing file systems for digital video and audio," Proc. ACM SOSP '91, Oper. Sys. Rev., vol.25, no.5, pp.81–94, Oct. 1991.

[4] 梶谷浩一，"動画サーバのためのディスクアレー管理法についての考察，" 信学論 (D-I), vol.J77-D-I, no.1, pp.66–76, Jan. 1994.

[5] F.A. Tobagi, J. Pang, R. Baird, and M. Gang, "Streaming RAID^{TM} - A disk array management system for video files," Proc. ACM Multimedia 93, pp.393–400, Aug. 1993.

[6] 鈴木偉元，阪本秀樹，西村一敏，"動画情報のディスクアクセス方式の評価，" 平7前期情処学全大，2G-9.

[7] 阪本秀樹，西村一敏，中野博隆，"ビデオ情報の大規模多重アクセス方式，" 信学論 (D-II), vol.J78-D-II, no.1, pp.76–85, Jan. 1995.

[8] T. Mori, K. Nishimura, H. Nakano, and Y. Ishibashi, "Video-on-demand system using optical mass storage system," Jpn. J. Appl. Phys., vol.32, Part 1, no.11B, pp.5433-5438, Nov. 1993.

[9] D. Bitton and J. Gray, "Disk shadowing," Proc. VLDB '88, pp.331–338, Aug. 1988.

[10] 西村一敏，阪本秀樹，鈴木偉元，森　達男，"ディジタル動画情報の高多重読取り方式，" テレビ学技報，vol.18, no.20, pp.1–6, March 1994.

[11] Y. Oyang, M. Lee, C. Wen, and C. Cheng, "Design of multimedia storage systems for on-demand playback," Proc. IEEE Data Engineering '95, pp.457–465, March 1995.

[12] 鈴木偉元，西村一敏，阪本秀樹，森　達男，"多重読取り特殊再生方法，" 特開平7-226909 (1993-12出願)．

[13] M. Chen, D. Kandlur, and P.S. Yu, "Support for fully interactive playout in a disk-array-based video server," Proc. ACM Multimedia 94, pp.391–398, Oct. 1994.

[14] M. Maruyama, O. Nakano, K. Nishimura, and H. Nakano, "Communication processing techniques for multimedia systems," Proc. IEEE Singapore ICCS '94, pp.974–980, Nov. 1994.

付　録

1. 式(2)が単調増加する条件

式(2)をxについて微分することにより，

$$\frac{b}{2\sqrt{x}} = \frac{S(R-1)}{2D_{min.}\{1+(R-1)x\}^2} \tag{A·1}$$

を満足する点xでTは極大，極小，または変曲点となることがわかる．ここで，

$$(R-1)x = \tan^2\theta$$

$$\left(\tan^{-1}\sqrt{(R-1)/c} \leqq \theta \leqq tan^{-1}\sqrt{R-1}\right) \tag{A·2}$$

で置き換えた式(A·1)の両辺の逆数をとって，

$$\frac{\sin\theta}{b\sqrt{R-1}\cos\theta} = \frac{D_{min.}}{S(R-1)\cos^4\theta} \tag{A·3}$$

$$\therefore\ \sin\theta\cos^3\theta = bD_{min.}/(S\sqrt{R-1}) \tag{A·4}$$

を得る．

式(A·4)の左辺は，凸型の曲線であり，$\theta=\pi/6$で最大値をとる．従って，式(A·4)の右辺が十分に大きければ，式(A·4)すなわち式(A·1)は成立しないので，Tは単調に増加する．右辺が小さくなってθの1点でだけ式(A·4)が成立する場合は，その点でTは変曲点となるが，単調に増加することに変わりはない．右辺が更に小さくなると式(A·4)はθの2点で成立することになり，最初の点および次の点でTは各々極大および極小となる．従って，

$$\sin\theta\cos^3\theta \leq bD_{min.}/(S\sqrt{R-1})$$

$$\left(\theta = MIN[\pi/6, \tan^{-1}\sqrt{R-1}]\right) \qquad \text{(A·5)}$$

の条件が満足されれば，T は単調増加関数となる．

2. データ転送速度の近似による式 (2) の誤差

データ転送速度の近似値と真値の差の絶対値 d を

$$d = (D_{max.} - D_{min.})/(B-1) \qquad \text{(A·6)}$$

と見積もる．但し，B はバンド数である．データ転送速度の誤差がシリンダ位置 x によらず一定だとすれば，平均化タイムスロット時間 T に及ぼす影響は，データ転送速度が最小となる $x = 1/c$ の場合に最大となる．そこで，$x = 1/c$ でのデータ転送速度が $D_{min.}$ よりも d だけ小さかったとして，T の誤差を大きく見積もる．このときの平均化タイムスロット時間 T' は最小シーク時間を a' として，

$$T' = a' + L + \frac{S}{2}\left(\frac{1}{D_{min.}} + \frac{1}{D_{min.}-d}\right) \qquad \text{(A·7)}$$

で与えられる．従って，過大評価した誤差 ε は，

$$\varepsilon = \left\{T' - T(1/c)\right\}/T(1/c) \qquad \text{(A·8)}$$

となる．

（平成7年6月7日受付，8年3月14日再受付）

西村　一敏　（正員）

昭48熊本大・工・電子卒．同年日本電信電話公社（現NTT）に入社．以来，磁気テープ記憶装置，光ディスク装置，マスストレージシステム，およびビデオオンデマンドシステムの研究実用化に従事．現在，NTTヒューマンインタフェース研究所研究グループリーダ．応用物理学会会員．

丸山　充　（正員）

昭60電通大大学院修士課程了．同年日本電信電話（株）入社．以来，メッセージ通信処理システムの研究，高精細画像情報提供システムの研究，ビデオオンデマンドシステムにおける高速通信処理技術の研究に従事．現在，NTTヒューマンインタフェース研究所メディア応用システム研究部主任研究員，電通大非常勤講師．ソフトウェア科学会，IEEE各会員．

阪本　秀樹　（正員）

昭59阪大・工・通信卒．昭61同大大学院修士課程了．同年日本電信電話（株）入社．以来，HDTV高速ビデオテックスシステム，ビデオオンデマンドシステムの研究開発に従事．現在NTTヒューマンインタフェース研究所主任研究員．情報処理学会，IEEE，ACM各会員．

鈴木　偉元　（正員）

平1千葉大・工・機械卒．平3同大大学院修士課程了．同年日本電信電話（株）入社．以来，高精細画像情報提供システム，ビデオオンデマンドの研究開発に従事．現在NTTヒューマンインタフェース研究所メディア応用システム研究部研究主任．

中野　博隆　（正員）

昭47東大・工・電気卒．昭52同大大学院博士課程了．同年日本電信電話公社（現NTT）に入社．以来，ビデオテックスシステム，ビデオオンデマンドシステムなど画像通信システムの研究開発に従事．現在NTTヒューマンインタフェース研究所メディア応用システム研究部部長．工博．