

# 光ディスク間最適配置機能をもつ階層ストレージシステム

Hierarchical Storage Management System with Function for Optimizing Data Storage Locations among Optical Discs

神山 忠信  
T. Kamiyama

菅野 雅之  
M. Kanno

高木 志郎  
S. Takagi

特集  
II

情報のマルチメディア化に伴い、今後さらにニーズが拡大する高性能・高信頼の大容量ドキュメントストレージシステム向けに、半導体メモリ、磁気ディスク、光ディスクと光ディスクオートチェンジャからなる階層記憶制御機構を標準 PCI (Peripheral Component Interconnect) ボード 1 枚に収めた MicroHSM<sub>TM</sub>を開発し、当社ドキュメントファイリングシステムの WindowsNT<sup>®</sup>(注1)モデルの大容量ストレージシステムとして製品化した。従来の階層記憶機構がデータのアクセス頻度に基づき、半導体メモリから光ディスクオートチェンジャまでの格納位置を垂直方向に最適化するだけであるのに対し、MicroHSM<sub>TM</sub>は複数の光ディスク間の水平方向の格納位置の最適化を行い、オートチェンジャの光ディスク交換回数を削減し、アクセス性能とシステムの信頼性を大幅に改善した。

We have developed MicroHSM<sub>TM</sub>, a large-capacity storage system for the WindowsNT<sup>®</sup> version of the Toshiba document management system. MicroHSM<sub>TM</sub> is a hierarchical storage management (HSM) system consisting of a semiconductor memory, magnetic disc, optical discs and a jukebox with a specially developed function for optimizing data storage locations.

The conventional HSM system optimizes storage locations only vertically, while MicroHSM<sub>TM</sub> optimizes them horizontally as well. This two-dimensional strategy reduces the frequency of disc changes in the jukebox, thus improving access time and system reliability.

## 1 まえがき

情報化社会の進展によってオフィス情報は増大の一途をたどっているが、近年のマルチメディア技術の深耕により、オフィスで扱われる情報にイメージ情報や映像情報が加わり、今後さらに幾何級数的に増大していくものと予想されている。それに伴い、オフィス情報を効率よく格納し、高速なアクセス性能と高い信頼性をもつ廉価な超大容量ストレージシステムへのニーズが、近年急速に高まっている。

今回、当社がこれまで TOSFile<sub>TM</sub>などで培ってきたオフィス情報管理技術のノウハウを生かしながら、マルチメディア時代の新しい要請にこたえた新しいコンセプトのストレージシステムを開発した(図1)。ここでは、新たに開発した光ディスク間最適配置機能と、それを搭載する階層ストレージシステム MicroHSM<sub>TM</sub>について述べる。



図1. 階層ストレージシステム MicroHSM<sub>TM</sub> PCI シングルボードのコントローラで高性能の階層ストレージ制御機構を実現した。  
MicroHSM<sub>TM</sub> hierarchical storage management system

## 2 MicroHSM<sub>TM</sub>のコンセプト

### 2.1 Two Dimensional Relocation Strategy

一般に情報にはライフサイクルがあり、生成直後はアクセス頻度が高く、時間の経過に従ってアクセス頻度が下がっていく。オフィスには新旧織り交ぜて大量の情報(ドキュメント)が保有されているが、そのすべてに均一にアクセス

(注1) WindowsNTは、Microsoft社の商標。

が分布していることはなく、比較的新しい少数の情報にアクセスが集中している。全体の20%の情報にアクセスの80%が集中していると言われるが(パレートの法則)、通常のオフィス業務のなかではこれを上回るアクセスの集中が起こっていると考えられる。

したがって、情報をストレージする場合に複数種のメディアを用意し、少量だがアクセス頻度の高い情報はビット

単価が高いが高アクセス性能のメディアに格納し、アクセス頻度は低いが大量の情報は低アクセス性能だがビット単価の安いメディアに格納する。そして情報を個々のライフサイクルに沿ってメディアを選択して格納すれば、全体としては性能を確保しながら廉価で大容量のストレージシステムを構築できる。

MicroHSM<sub>TM</sub>は半導体メモリ、磁気ディスク、および光ディスクと光ディスクオートチェンジャからなる階層的なストレージシステムを構成し、個々の格納情報へのアクセス頻度を監視してライフサイクルを判定し、格納位置を最適化する機構をもつ。MicroHSM<sub>TM</sub>ではこの最適化原理をもった、各データ格納メモリ間の階層間垂直方向の最適配置機構 (Vertical Relocation) を搭載するだけでなく、オートチェンジャに格納される複数の光ディスク間の階層内水平方向の最適配置機構 (Horizontal Relocation) を搭載した (図2)。この新しい階層ストレージ制御方式を “Two Dimensional Relocation Strategy” と呼んでいる。

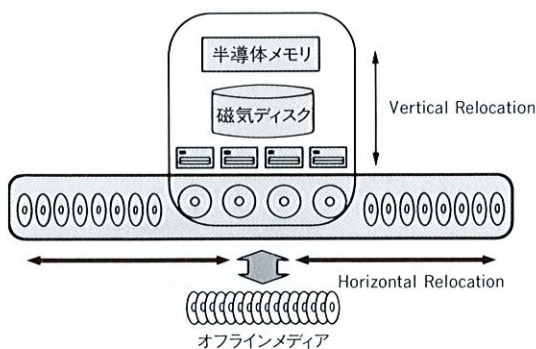


図2. MicroHSM<sub>TM</sub>の階層ストレージ制御方式 Two Dimensional Relocation Strategy 情報のライフサイクルに基づく階層間垂直方向の最適配置機構と、階層内水平方向の最適配置機構を搭載している。 Two-dimensional relocation strategy

水平方向の最適配置機構が加味されたことにより、アクセス頻度の高い情報とアクセス頻度の低い情報は、光ディスクを分けて格納されることになる。これによりアクセス頻度が高い情報へのアクセス光ディスクが限定されるため、オートチェンジャのディスク交換回数を大幅に減らすことができる。また、格納スペースを確保するためオートチェンジャ内の光ディスクの一部を排出し新規の光ディスクを投入する場合でも、アクセス頻度が低い情報が格納されている光ディスクから順次選択していけば、情報のオフライン化による影響が最小限に押さえられる。

## 2.2 MicroHSM<sub>TM</sub>の特長

MicroHSM<sub>TM</sub>は、上述の階層ストレージ制御方式を搭載するストレージシステムとして製品化された。MicroHSM<sub>TM</sub>は次の特長をもつ。

- (1) 半導体メモリ、磁気ディスク、光ディスクと光ディスクオートチェンジャにより階層型ストレージシステムを構成し、“Two Dimensional Relocation Strategy”により高性能のデータアクセスを実現した。
- (2) 旧来の階層ストレージシステムが、ファイル単位で配置制御されているのに対し、MicroHSM<sub>TM</sub>はブロック単位のきめ細かい最適化が施される。これにより、例えばデータベースシステムなどの巨大ファイルに対しても記録レコードごとの最適化が可能となる。
- (3) これらの制御機構を高性能 CPU や新規開発の高速データ転送 LSI を搭載した標準 PCI ボード 1 枚に実装した (MicroHSM<sub>TM</sub>コントローラ)。これにより、従来のソフトウェアによる階層記憶管理システムに比較し、ストレージサーバの CPU やシステムバスの負荷が軽減され、高いスループットを実現した。また、階層記憶制御や最適配置制御などはすべてこのボード上で処理され、各種の OS (Operating System) やファイルシステムに容易にこの階層ストレージ機構が搭載できるようになった。
- (4) 複数枚の光ディスクをグルーピングし、大容量仮想ディスクとして使用することができる。オートチェンジャ全体を一つの大容量な仮想ディスクとして使用することも可能で、最大 2 T (テラ) バイトのボリューム構成ができる。
- (5) オートチェンジャから排出された光ディスクメディアのオフラインメディア管理が可能である。これによりオートチェンジャの光ディスク収納枚数を超える容量のストレージシステムの構築が可能となる。
- (6) 光磁気ディスク (MO)、コンパクトディスク (CD-ROM, CD-R)、DVD (注2) (DVD-ROM, DVD-R, DVD-RAM) など標準フォーマットの各種光ディスクメディアに対応する。異種の光ディスクメディアが混在する統合的なストレージシステムの構築が可能である。
- (7) MicroHSM<sub>TM</sub>コントローラや光ディスクオートチェンジャは二重化が可能で、システムの可用性とデータの保全性を大幅に向上させることができる。
- (8) 光ディスク管理や論理ドライブの割当てなどの仮想ディスク制御のための運用管理ソフトウェアを提供する。操作性の高い GUI (Graphical User Interface) により、大容量ストレージシステムの操作環境を大幅に改善した。

## 3 MicroHSM<sub>TM</sub>ストレージシステム

### 3.1 システム構成

図3に MicroHSM<sub>TM</sub>ストレージシステムの構成を示す。階

(注2) DVD: Digital Versatile Disc。

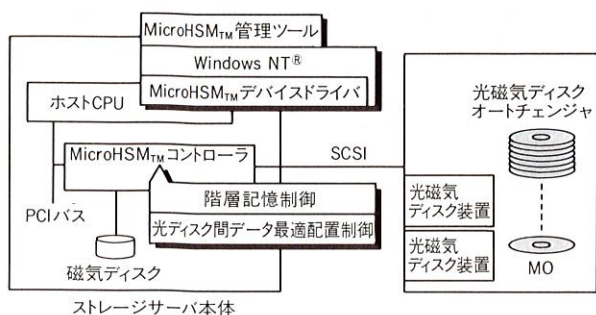


図3. MicroHSM<sub>TM</sub>ストレージシステムの構成 MicroHSM<sub>TM</sub>コントローラをPCIバスを装備するパソコンやワークステーションに実装することで、容易に高性能な階層ストレージシステムが構築できる。  
Configuration of MicroHSM<sub>TM</sub> storage system

層ストレージ制御機構を搭載する MicroHSM<sub>TM</sub>コントローラは、PCIバスを実装する汎用（はん）用のパソコンあるいはワークステーションに接続される。この MicroHSM<sub>TM</sub>コントローラには、高性能 CPU、メモリ、2ポートの SCSI (Small Computer System Interface) 制御 LSI、各メモリ階層間を高速にデータ転送する新規開発の高速データ転送制御 LSI が搭載され、高性能の階層ストレージ制御機構を実現している。MicroHSM<sub>TM</sub>コントローラを実装するパソコンあるいはワークステーションに対しては、MicroHSM<sub>TM</sub>デバイスドライバが用意される。現在、WindowsNT<sup>®</sup> 版がリリースされている。また、光ディスクオートチェンジャについては、最新の5型4倍密(2.3Gバイト)光磁気ディスクを32枚あるいは76枚収容する機種を採用している。将来はCDおよびDVDを搭載するオートチェンジャもラインアップする予定である。

MicroHSM<sub>TM</sub>ストレージシステムの仕様を表1に示す。

表1. MicroHSM<sub>TM</sub> (MO版) 基本仕様  
Specifications of MicroHSM<sub>TM</sub> MO version

MicroHSM <sub>TM</sub> コントローラ	
プロセッサ	Pentium <sup>TM</sup> (注3)
メインメモリ (標準)	8Mバイト, 最大16Mバイト
キャッシュメモリ	16Kバイト
不揮発性メモリ	8Kバイト
専用LSI	高速データ転送LSI
ホストインタフェース	PCIバス
SCSIインタフェース	2チャンネル標準
ボードサイズ	PCIボード 1枚
ソフトウェア	
MicroHSM <sub>TM</sub> デバイスドライバ	WindowsNT <sup>®</sup> 3.51以上
階層記憶制御機構	
最適配置制御機構	
運用管理アプリケーション, ほか	
光ディスクオートチェンジャ	
光磁気ディスク装置	5型4倍密(2.3Gバイト)MO2台内蔵
メディア収容枚数	32枚/76枚

### 3.2 ソフトウェア構成

MicroHSM<sub>TM</sub>のソフトウェア構成を図4に示す。Micro-

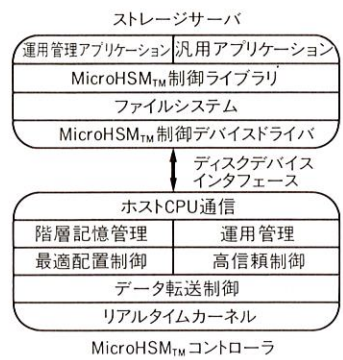


図4. ソフトウェア構成  
MicroHSM<sub>TM</sub>ストレージシステムにおけるストレージサーバ本体とMicroHSM<sub>TM</sub>コントローラ内部の主なソフトウェア構成を示す。  
Software architecture

HSM<sub>TM</sub>は実装されるストレージサーバのOSに独立したディスクデバイスインタフェースを提供している。ストレージサーバ側は、運用管理アプリケーション、制御ライブラリ、ファイルシステム下のデバイスドライバから構成される。

MicroHSM<sub>TM</sub>コントローラ側は、通信部、階層記憶制御部、運用管理部、最適配置制御部、高信頼制御部、データ転送制御部から構成される。

### 4 最適配置制御方式

MicroHSM<sub>TM</sub>は、複数枚の光ディスクから構成され“Two Dimensional Relocation Strategy”を実行して高性能で大容量の仮想ディスクを構築する MicroHSM<sub>TM</sub>フォーマットと、通常フォーマットで光ディスク片面を一つの仮想ディスクと見せる互換フォーマットの2種のフォーマットをサポートする。格納する情報のライフサイクルに基づく最適配置アルゴリズムは、最適化処理のためにシステムにかかる負荷の大きさと判定精度とのトレードオフから、種々のシミュレーションを通して決められた。

図5および図6は、“Two Dimensional Relocation Strategy”の効果を示すシミュレーション結果である。光ディスク40面が搭載されるオートチェンジャストレージシステムに対し、アクセス全体の80%が20%の局所データに集中し、かつアクセスの20%が書き込み処理で、時間経過とともにアクセス頻度の高い情報が交替していくという現実に近いモデルのアクセスパターンを発生させ、シミュレーションを実行した。

図5は従来方式とMicroHSM<sub>TM</sub>方式により、各光ディスク面上の高アクセス頻度データの配置状態を求めたもので、MicroHSM<sub>TM</sub>方式では、高アクセス頻度のデータが特定メディアに集中化されることがわかる。図6はこの結果生ずるアクセス性能差を示しており、MicroHSM<sub>TM</sub>方式では従来方式に比較してアクセスに要する平均時間が約1/2まで改善されることが示されている。

(注3) Pentiumは、Intel社の商標。

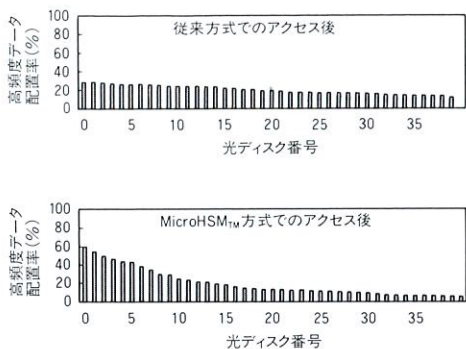


図5. 高頻度データの配置状態(シミュレーション結果) 従来方式では高頻度データが各光ディスクに分散するが、MicroHSM™では特定メディアに集中する。

Distribution of frequently accessed data

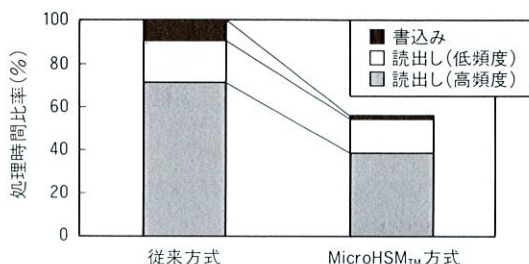


図6. MicroHSM™の性能(シミュレーション結果) MicroHSM™は従来方式と比較してデータアクセスに要する時間が約1/2になる。書き込みおよび高頻度データの読出し性能が向上した。

Relative performance of MicroHSM™

## 5 運用管理アプリケーション

光ディスクの片面ないし複数枚の光ディスクから構成される仮想ディスク一つ一つをMicroHSM™ではユニットと呼ぶ。運用管理アプリケーションは、光ディスクやこのユニットに対する次のような種々の操作をGUIベースで行うことができる。図7にそのGUIを示す。

- (1) ユニットの作成/削除(複数枚の新規光ディスクからユニットサイズやユニット名を設定してユニット作成、または作成したユニットを削除し、新規ディスクとする)
- (2) ユニットのオープン/クローズ(ユニットをファイルシステムに組み込み、アプリケーションから通常の磁気ディスクとして利用可能にする、またはクローズすることによりユーザ空間からユニットを切り離す)
- (3) ユニット/光ディスクの収納/返却(光ディスクをオートチェンジャに収納し、フォーマットを表示、または選択したユニット/光ディスクをオートチェンジャから返却)
- (4) ユニットのフラッシュ(磁気ディスク上にキャッシュされている更新データを光ディスクに反映)

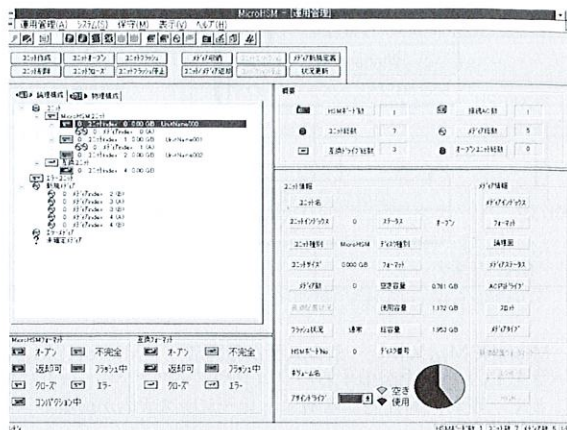


図7. MicroHSM™運用管理アプリケーション GUI 光ディスク管理や仮想ディスク制御を操作性のよいGUIで提供する。

Administration GUI of MicroHSM™

- (5) エラーログ表示(MicroHSM™内でのエラーを表示、ファイルとして保存)
- (6) システム構成表示(オートチェンジャのスロット数、光ディスクドライブ数、磁気ディスク容量、ソフトウェアのバージョン、キャッシュのウォータマーク値の設定、キャッシュのアクセス回数などを表示/設定)
- (7) 障害回復(トラブル発生時のデータ修復)

## 6 あとがき

旧来の階層ストレージシステムの性能を大幅に改善する“Two Dimensional Relocation Strategy”技術を開発し、MicroHSM™として製品化した。今後はアクセス性能のさらなる向上に加え、DVDなどのメディアバリエーションを整え、超大容量で高性能、かつ廉価なシステムを実現し、マルチメディア時代の要請にこたえるストレージシステムとして、種々の展開を図っていく。



神山 忠信 Tadanobu Kamiyama

マルチメディア技術研究所開発第四部グループ長。  
ドキュメントシステムアーキテクチャの研究開発に従事。  
Multimedia Engineering Lab.



菅野 雅之 Masayuki Kanno

青梅工場コンピュータ・マルチメディア設計部グループ長。  
コンピュータハードウェアの開発設計に従事。  
Ome Works



高木 志郎 Shiro Takagi

マルチメディア技術研究所開発第四部グループ長。  
ドキュメントシステムアーキテクチャの研究開発に従事。  
Multimedia Engineering Lab.