# 光ディスクを用いたVirtual Hard Diskシステムの開発*

川崎製鉄技報
24 (1992) 1, 45-51

# Development of Virtual Hard Disk System Using Optical Disk

沢田 要
Kaname Sawada
新事業本部 システム・エレクトロニクス事業部開発部 主査(課長)

高橋 泰隆
Yoshitaka Takahashi
新事業本部 システム・エレクトロニクス事業部開発部 主査(掛長)

橋詰 要
Kaname Hashizume
新事業本部 システム・エレクトロニクス事業部開発部 主査(掛長)

百瀬 惇
Atsushi Momose
新事業本部 システム・エレクトロニクス事業部開発部 主査(部長補)

## 要旨

光ディスクをベースとして，ハードディスクのエミュレーション機能を有し，かつ当社独自のキャッシュアルゴリズムによりハードディスク並の高速性を実現したVHD (virtual hard disk) システムを開発した。ハードディスク・エミュレーションにより，ワークステーションやパーソナル・コンピュータの主要機種において，専用ソフトウェアを必要とせずハードディスクとして動作した。また，ワークステーションやパーソナル・コンピュータを用いたベンチマーク・テストでは，当初の目標どおり，光ディスクを大幅に上回りハードディスクに近い高速性を得た。また，キャッシュの有効性はアプリケーションに依存するため，3方式のキャッシュ・アルゴリズムを開発した。

## Synopsis:

A new storage device named VHD (virtual hard disk) system has been developed by Kawasaki Steel Corp. The VHD system is based on the technique of the optical disk drive, and has two major advantages over it. The VHD appears as a hard disk drive to the operating system. The second advantage is that the transfer rate of the VHD is nearly equal to the hard disk drive by proprietary cache algorithm. The VHD can eliminate the need for a special device driver, and achieve plug compatibility with the standard hard disk drive for many personal computers and workstations. The VHD can give good results as were expected from the bench mark test, that is, the VHD has a higher transfer rate than that of the optical disk drive and almost catches up with the hard disk drive. In addition the VHD has three cache modes to keep high performance for many applications.

## 1 緒　言

光ディスクは，リムーバビリティと大容量性の大きな特長から，従来にない画期的なメモリ・デバイスとして登場した。最初に再生専用型，次いで追記型が製品化された。再生専用型としては光学式ビデオ・ディスクやコンパクト・ディスク，追記型としてはドキュメントの蓄積・検索を主用途とする光電子ファイルなどが代表的な製品としてあげられる。

さらに，コンピュータ用外部記憶装置を主なターゲットとして3年前から書換型が登場した。コンピュータ用外部記憶装置としては，現在ハードディスクを中心としてフロッピーディスクや磁気テープなどが使用されている。しかし，光ディスクが，コンピュータ用外部記憶装置として広く普及するためには，光ディスクの優れた特長を生かし，かつ既存のメモリ・デバイス，特にハードディスクの機能を包含し得る技術や製品が必要になる。

本論文は，上記の観点から開発に取り組んだVHD (virtual hard disk) システムについて述べる。

## 2 VHDシステムの概要

### 2.1 開発のねらい

光ディスクの主な特長として次の4点があげられる。

(1) リムーバブル・メディアである。
(2) 記憶容量が大きい（記憶密度が高い）。
(3) 非接触記録方式のため，ハードディスクで問題となるクラッシュがない。

* 平成3年11月20日原稿受付

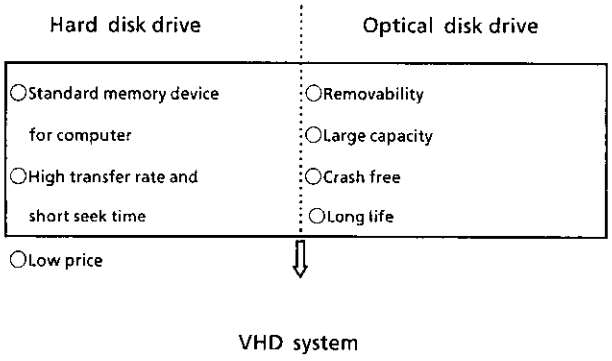


Fig. 1 Concept of VHD system

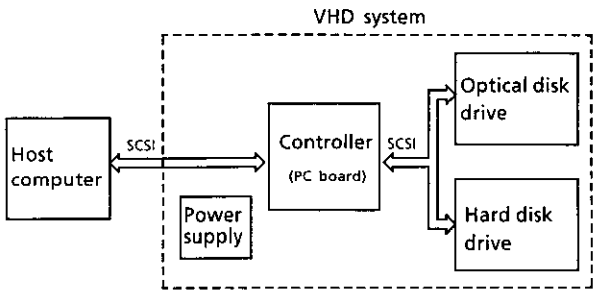


Fig. 2 Configuration of VHD system

(4) データの保存性が高く，10年程度のライフが期待できる。

一方，ハードディスクは以下の点が高く評価されている。

(1) 現在最も代表的なコンピュータ用外部記憶装置として，多くのソフト/ハードの下で動作する。

(2) 高速データ転送かつ高速シーク。

(3) 500 M Byte クラス以下では低価格である。

VHD システムは，光ディスクをベースとしていることから，上記の光ディスクの特長を有しており，かつ VHD システム内部でハードディスクのエミュレーションを行うことにより，ハードディスクの実績・資産を生かしている。また当社独自のキャッシュアルゴリズムにより，ハードディスクのもう一つの特長である高速性を実現している。このようにハードディスクと光ディスクのそれぞれの長所を合わせ持つ「リムーバブルなハードディスク」の実現を VHD システムの開発のねらいとしている (**Fig. 1**)。

### 2.2 基本構成

VHD システムの基本構成を **Fig. 2** に示す。本システムは，片面 325 M Byte，両面 650 M Byte の記憶容量を有する 5.25 インチ光ディスク装置と，40 M Byte あるいは 325 M Byte の 3.5 インチハードディスク装置と，これらを用いてハードディスクのエミュレーションとデータ・キャッシュを実現する基本制御部（プリント基板化）などから構成される。本システムの外観を **Photo 1** に示す。

VHD システムの外部インターフェースとして，ハードディスク装置で最も普及している SCSI (small computer system interface)[1] を採用している。光ディスクには，ハードディスクのエミュレーションのために書換可能な光磁気型を用いている。

基本制御部のブロック図を **Fig. 3** に示す。基本制御部は，データ転送のためのハードロジックと，ハードディスク・エミュレーションとデータ・キャッシュを実現するファームウェアとから構成されている。

基本制御部には SCSI ポートが二つあり，1 ポートは VHD システムの外部インターフェイスポートとして HOST コンピュータに接続され，残る 1 ポートには光ディスク装置とハードディスク装置がデイジーチェイン接続されている。基本制御部は，HOST コンピュータに対してはターゲットとして動作し，一方，光ディスク装置やハードディスク装置に対してはイニシエータとして動作する。

二つの SCSI ポート間，すなわち HOST コンピュータと VHD システムに内蔵されているハードディスク装置や光ディスク装置間で高速にデータを転送するための制御回路を設けており，これによってバーストで最大 4 M Byte/s のデータ転送速度を実現している。HOST コンピュータが VHD システムにアクセスしていない間，データ・キャッシュのためにバックグラウンド処理としてハードディスク装置と光ディスク装置との間でデータ転送が行われるが，転送はすべて 32 K Byte のデータ・バッファ RAM を介して行われ

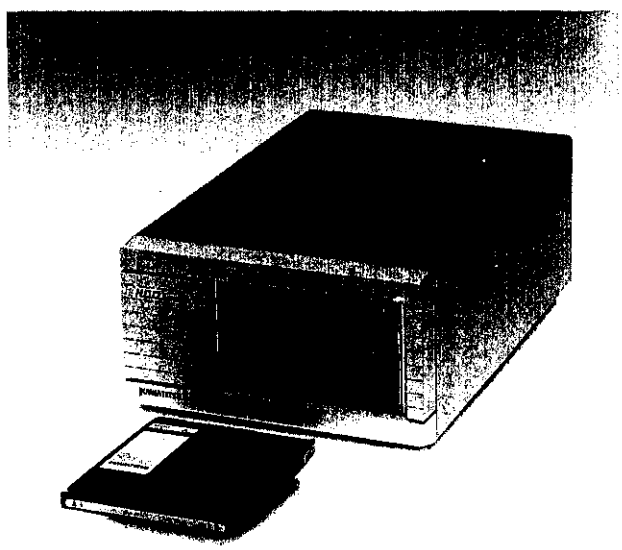

Photo 1 Appearance of VHD system

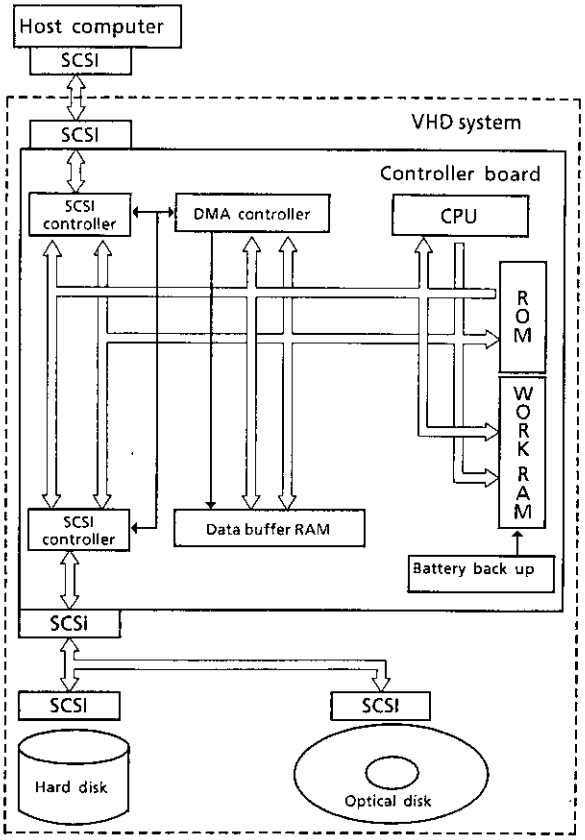


Fig. 3 Block diagram of controller

る。

基本制御部では 16 ビット CPU を用いている。ファームウェアの規模は，後述するキャッシュの各方式によって異なるが，実行型ファイル換算で約 90 K Byte，ステップ数は約 18 000 である。プログラム言語はアセンブラとC言語で，ファームウェアのオーバヘッドが無視できる処理はC言語，無視できない処理は極力アセンブラ化している。

CPU に直結している 32 K Byte のワーク RAM は，CPU の汎用ワークエリアとして用いられるだけでなく，キャッシュ管理情報もストアされている。このキャッシュ管理情報をもとに，光ディスク装置とハードディスク装置との間でバックグラウンド処理が行われる。バックグラウンド処理が終了する前に電源が切られることによりキャッシュ管理情報が消失し，HOST コンピュータにとって終結したはずの WRITE 処理が実行されなくなるのを防ぐために，ワーク RAM に対して電源バックアップを行っている。

## 3 ハードディスク・エミュレーション

コンピュータにメモリやプリンタなどの各種デバイスを接続し動作させるためには，デバイスドライバとよばれるソフトウェア・インターフェースが必要になる。コンピュータにはハードディスク装置用のデバイスドライバが標準で組み込まれているが，光ディスク装置用のデバイスドライバはサポートされていない場合が多い。そのため，光ディスク装置をコンピュータに接続するためには，**Fig. 4** (a) に示すようにユーザーサイドで独自にデバイスドライバを作成し，OS に組み込む必要があった。

VHD システムでは，ハードディスク装置用デバイスドライバの

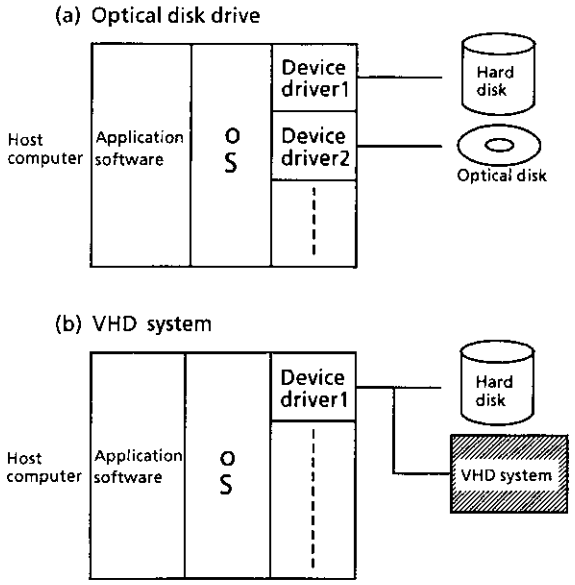


Fig. 4 Configuration of software

Table 1 List of compatible computers

| | |
|---|---|
| Workstation<br>Middle range computer | · Universe (Kawasaki Steel)<br>· SUN SPARC (Sun microsystems)<br>· LUNA (Omron) |
| Personal computer | · PC98 (NEC)<br>· PC/AT (IBM)<br>· PS/55 (IBM Japan)<br>· Mac (Apple)<br>· FMR (Fujitsu) |

下で動作するように，装置内の基本制御部でハードディスク装置のエミュレーションを行っている。従来のソフトウェア・インターフェースであるデバイスドライバに対し，VHD システムの方式は「ハードウェア・デバイスドライバ」と表現することもできる。これにより，Fig. 4 (b) に示すように特別のデバイスドライバを必要とせず，プラグ・インで VHD システムをコンピュータの外部記憶装置として使用できる。

**Table 1** にハードディスク・エミュレーションを行った HOST コンピュータの一覧を示す*。ワークステーションやパーソナルコンピュータの主要製品について，既設ハードディスク装置の代替あるいは追加メモリとしてプラグ・インで動作が可能である。

## 4 キャッシュ機能

光ディスクの本格的な普及を阻害する要因の一つとして，光ディスクのスループットがハードディスクと比べて低いことがあげられる。その要因として，

(1) 記録再生ヘッドが重いことなどによるシークの時間の長さ

(2) エラー訂正前のデータ・エラーレイトが悪く，WRITE 後の

* Universe は Charles River Data Systems, Inc. が登録商標を有するコンピュータであり，川崎製鉄㈱が日本国内において独占販売している。

ベリファイ処理が不可欠

(3) 記録再生ヘッドとディスクの機械特性から，ディスク回転数を十分に上げられない

などがあげられ，加えて光ディスク装置が光磁気方式の場合，オーバーライトができないため，ライトサイクルの前にイレーズサイクルが必要になる。

後述するキャッシュ方式は，光ディスクのスループットの低さを補う目的で当社が独自に考案したもので，キャッシュメモリとしてハードディスク装置を用いている。

また，キャッシュの有効性はアプリケーションに依存するため，今回3種類のキャッシュ方式を開発した。ランダムアクセスに有利な第1方式，アクセス頻度の多いエリアが比較的限定できる場合に有利な第2方式およびキャッシュの有効性が限定されず，大量データのバックアップ時にも高速性が損なわれない第3の方式がある。以下，それぞれのキャッシュ方式について詳しく述べる。

### 4.1 ランダムアクセス方式（第1方式）

低価格性をねらって 40 M Byte のハードディスクを用いている。

WRITE 時のデータフローを **Fig. 5** に示す。HOST コンピュータから転送されるデータは，いったんハードディスクに記録され，その後ハードディスク装置から光ディスク装置に転送される。後者の転送は，HOST コンピュータが VHD システムに起動をかけていない間，すなわちバックグラウンド処理として実行される。バックグラウンド処理実行中に HOST コンピュータから新たな起動がかかった場合は，実行中のバックグラウンド処理を中断し，HOST コンピュータから起動されたシーケンスを実行した後，再び残ったバックグラウンド処理を再開する。なお，ハードディスクが未転送データで埋まった時にさらに HOST コンピュータからデータが転送されてきた場合は，直接光ディスクへ記録する。

本方式および後述する第2，第3方式においてデータのアドレス管理体系の破たんを防ぐため，VHD システム内に処理すべき WRITE データが残っている場合，すなわち WRITE 時のバックグラウンド処理実行中は，光ディスクを交換できないようにプロテクトする。なお READ 時は，バックグラウンド処理中でも光ディスクの交換は可能である。

WRITE 時のデータ配列方法を **Fig. 6** に示す。Fig. 6 (上) はハードディスク装置や光ディスク装置などのダイレクトアクセスデバイスのアクセス方法を示したもので，指定されたアドレスに応じてシークが発生する。矢印はシーク動作を，図中の数字は記録するブロックの順序を表す。ランダムアクセスが多い場合，WRITE に要する時間に対して，シークに要する時間は無視できない。VHD システムのデータ配列方法は，この問題点に着目して考案したもので，Fig. 6 (下) にその概要を示す。HOST コンピュータから転送された WRITE データは，指定されたアドレスにかかわらず，ハードディスクにシーケンシャルに記録していく。HOST コンピュータからの指定アドレスとハードディスク上にシーケンシャルに記録されたデータとの対応を示す情報を，基本制御部内のワーク RAM にストアする。バックグラウンド時，ワーク RAM にストアされているアドレス情報に基づいて，ハードディスク上のデータを光ディスクに転送する。このデータ配列方法により，従来のダイレクトアクセスデバイスで問題になっていたランダムアクセス時のシーク動作を除くことができ，WRITE 時のスループットを向上できる。

次に READ 時のデータフローを **Fig. 7** に示す。HOST コンピュータから READ 指定されたデータがハードディスク上に存在している場合はハードディスク装置から (hit)，存在しない場合は光

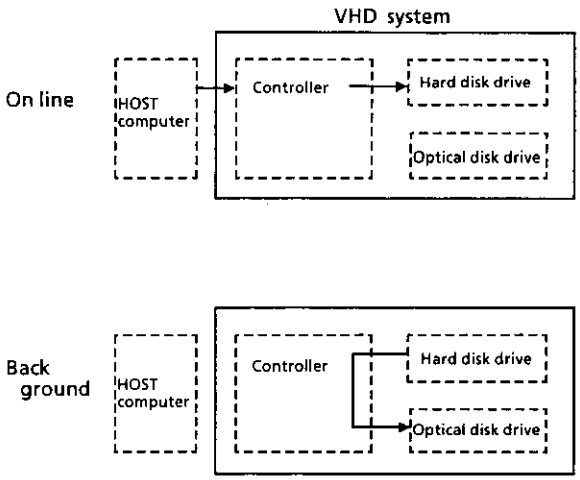


Fig. 5 Dataflow in write mode

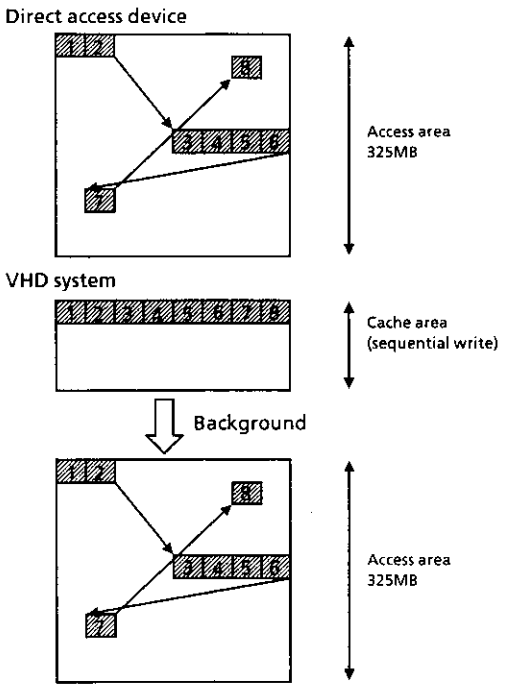


Fig. 6 Flow of background sequence

ディスク装置から転送される (fail)。hit 率を高めるために，アクセス頻度が高いと予想される固定アドレス領域や，シーケンシャル・アクセスを予想したプリフェッチ（先読み）がバックグラウンド時に実行される。

ハードディスクのエリア配分を **Fig. 8** に示す。ハードディスクの領域を，"hot" なデータのアクセスに備えた LRU (least recently used) エリアと，シーケンシャル・アクセスを予想したプリフェッチ・エリアと，OS の管理情報用エリアとしての固定エリアの以上3エリアに分割している。

LRU エリアは，前述の WRITE キャッシュのために使用されるエリアで，時間的に少し前に WRITE されたデータへの READ 起動に対して，このエリアにストアされているデータが，HOST コ

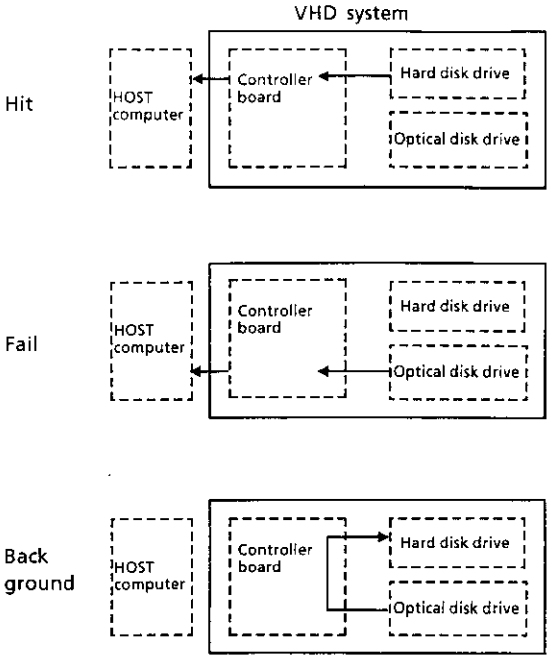


Fig. 7 Dataflow in read mode

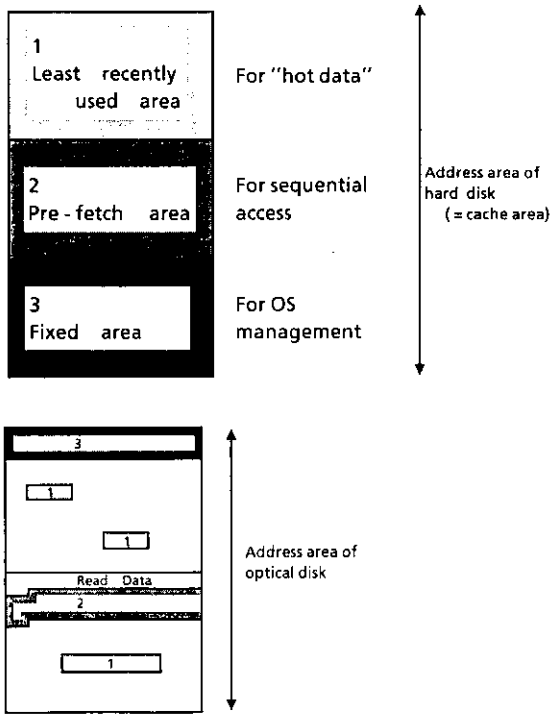


Fig. 8 Structure of cache area assignment

ンピュータに転送される。図中にも示すように，LRUエリアのデータはランダムアクセスに対応し，通常全アドレス空間に点在する。

HOSTコンピュータから外部記憶装置へのアクセスは，シーケンシャルに行われる場合がある。あるデータへアクセスされた後，続くアドレスがアクセスされる場合，バックグラウンド処理で行われるプリフェッチが有効になる。

MS-DOS（マイクロソフト）では，ファイルの生成に伴って

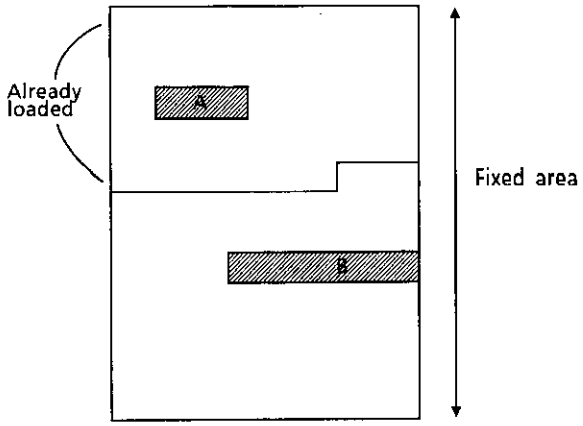


Fig. 9 Structure of fixed area

FAT (file allocation table) やディレクトリが生成，更新される。MS-DOSのディスクは，アドレスの先頭からFAT領域，ディレクトリ領域，データ領域の順に配列されているため，アドレスの先頭領域へのアクセスが多い。このようにOSの管理領域をターゲットとしたキャッシュ・エリアが固定エリアである。次に固定エリアのバックグラウンド処理方法を **Fig. 9** に示す。光ディスクがVHDシステムにセットされた時，直ちに光ディスクの先頭エリアのデータがハードディスク装置へ転送される。Fig. 9は，バックグラウンド処理の途中の，ハードディスク上の固定エリアが完全に埋まっていない時点を示す。すでにハードディスクに転送されたエリアにHOSTコンピュータからWRITEする場合はハードディスクに上書きし(A)，まだ転送されていないエリアにWRITEする場合は光ディスクに直接WRITEする(B)。

### 4.2 自動マッピング方式（第2方式）

ランダムアクセス方式と同様に，低価格性をねらって40 M Byteのハードディスク装置を用いているが，キャッシュ方式は異なる。ここではランダムアクセス方式との違いを中心に述べる。

光ディスクの全エリアを分割し，それぞれのエリアごとにHOSTコンピュータからのアクセス頻度をカウントする。そして光ディスクのアドレス空間内のアクセス頻度の多いエリアを，光ディスクと同様に分割したハードディスクのエリアに自動的にマッピングする方式である。

自動マッピング方式の概要を **Fig. 10** に示す。ハードディスクを8 M Byteずつ A～Eの5エリアに分割する。同じく光ディスクも8 M Byteずつ1～39（ただし第39エリアの容量は他のエリアと比べて少ない）に分割する。そしてHOSTコンピュータからアクセスされるごとに，光ディスク上の1～39のそれぞれのエリアのアクセス頻度カウンタをインクリメントする。

バックグラウンド処理では，アクセス頻度カウンタの値を比較し，上位5エリアを降順にハードディスク上のA～Eの5エリアに割り当て，割り当てられたエリアの光ディスクのデータをハードディスク上にコピーする。なお光ディスクをVHDシステムにセットした時は，デフォルト状態として光ディスクの先頭エリアの1～5がハードディスク上に割り当てられる。これは，システムディスクとして使用されるとき，OSの管理情報へのアクセスが多いと予想されるためで，ランダムアクセス方式において固定エリアを設けたのと同様のねらいである。

このように自動マッピング方式は，光ディスクの全アドレス空間

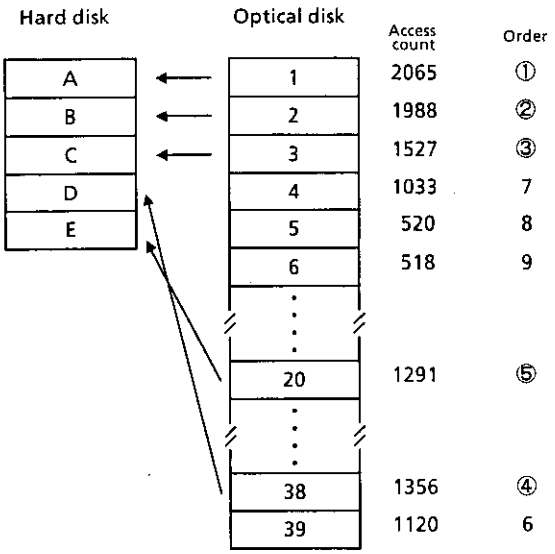


Fig. 10 Method of cache area auto mapping

にわたってランダムにアクセスされるような場合はあまり有効ではないが，アクセスされるエリアがある程度集中しているような場合に有効である。

### 4.3 大容量キャッシュ方式（第3方式）

第1方式や第2方式と異なり，光ディスクと同じ記憶容量の325 M Byte のハードディスクをキャッシュ・メモリとして活用した方式である。

この方式は，原理的にはランダムアクセス方式の機能の一つである固定エリアをベースとしたもので，キャッシュ・メモリのアドレス空間を光ディスクのアドレス空間まで拡張したものである。この方式では，光ディスクがVHD システムにセットされた直後は，キャッシュがfailする可能性があるが，バックグラウンド処理で光ディスクからハードディスクにデータがすべて転送された後は，キャッシュは100% hit する。

また本方式のVHD システムは，テンポラリなミラーディスク装置とみなすこともできる。HOST コンピュータから記録されたデータは，光ディスクが交換されるまでVHD システム内で自動的に2重化されているため，データの信頼性は高い。

### 4.4 性能評価

次にデータ・キャッシュの評価結果について述べる。

市販パーソナルコンピュータでの基本パフォーマンス・テスト結果を **Fig. 11** に示す。横軸に示したセクタ数を1単位として，ランダムシークを行いながら256回 WRITE/READ した時の処理時間を縦軸にプロットした。VHD システムのほかに，比較のためにハードディスク装置や光ディスク装置の測定も行った。WRITE 時の測定結果として，処理サイクル数とシーク時間の違いから光ディスク装置と比べて大きな優位性が得られただけでなく，Fig. 6 で示したようにシーク処理を除けることから，シーク時間の無視できない短いセクタ数の場合では，ハードディスク装置をも上回るデータが得られた。ただし，ハードディスク装置に対する優位性は，セクタ数が大きくなるにしたがって低下し，今回の結果では10セクタでクロスしている。一方，READ 時のパフォーマンスは，キャッシュがhitすればハードディスク装置に，failすれば光ディスク装置

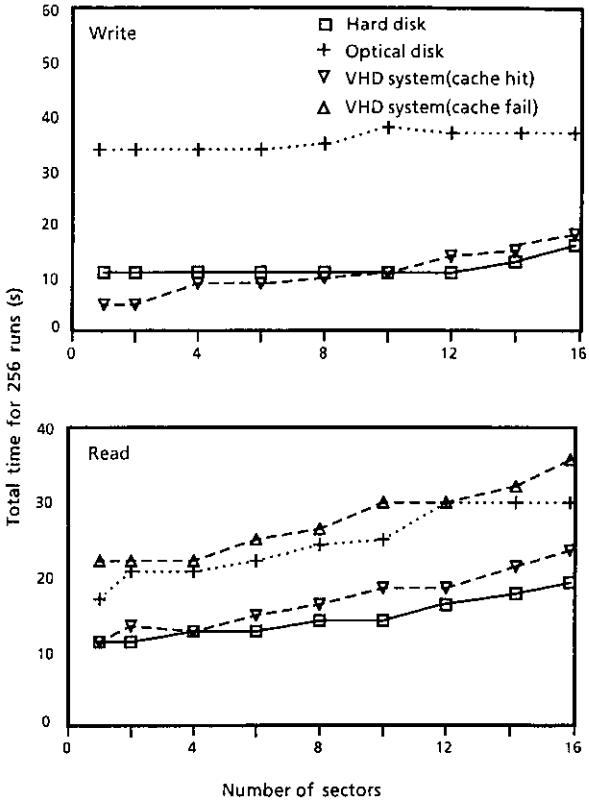


Fig. 11 Result of basic performance test

Table 2 Results of bench mark test (s)

| Item | Hard disk | VHD system | Optical disk |
|---|---|---|---|
| MS-DOS boot | 3.9 | 4.6 | 7.4 |
| 8 Files (500 KB) write | 27 | 22 | 53 |
| 8 Files (500 KB) read | 23 | 21 | 31 |
| Word processor 'Ichitarou dash' (Just system) | | | |
| Boot | 8.6 | 10.3 | 16.0 |
| Read (50 KB) | 11.3 | 11.6 | 15.7 |
| Save (50 KB) | 16.5 | 19.5 | 22.6 |
| Lotus 1-2-3 (Lotus) | | | |
| Boot | 6.5 | 7.0 | 11.6 |
| Read (100 KB) | 4.7 | 5.8 | 7.9 |
| Save (100 KB) | 4.2 | 6.2 | 15.8 |

に近付く。Fig. 11 でハードディスク装置や光ディスク装置と比べて若干データが落ちているのは，基本制御部内のファームウェアのオーバーヘッドに因る。

市販パーソナルコンピュータでのベンチマーク・テスト結果を **Table 2** に示す。同じくハードディスク装置や光ディスク装置についても測定しているが，VHD システムは光ディスク装置のパフォーマンスを上回り，ハードディスク装置に近いデータを得ている。

次にVHD システムを社外のユーザーA社で，市販パーソナルコ

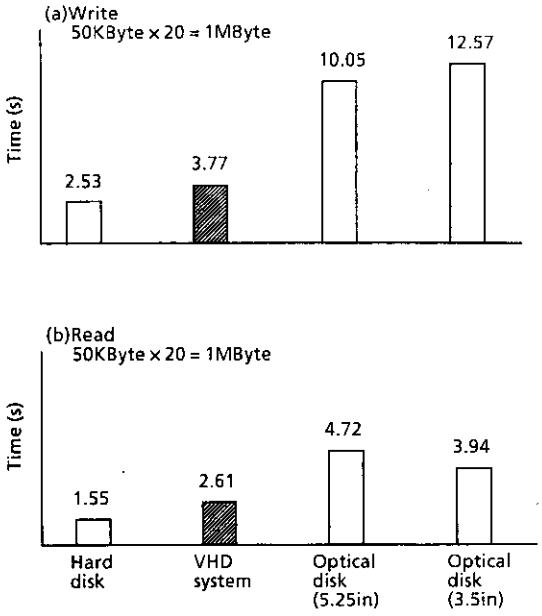


Fig. 12 Results of basic performance test

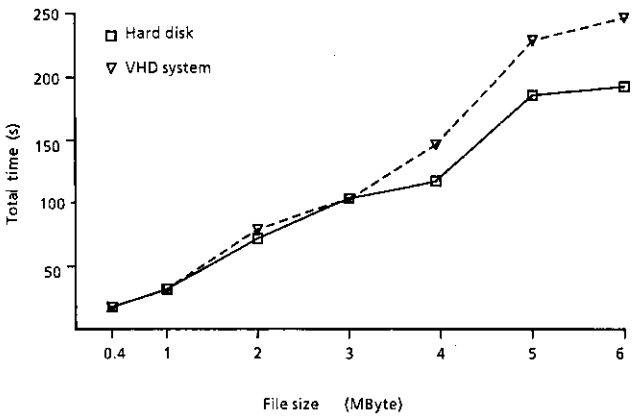


Fig. 13 Results of bench mark test in the case of CAD

ンピュータを用いて基本パフォーマンスの評価をいただいた結果を **Fig. 12** に示す。当社での評価と同様，光ディスク装置と比べ高速で，かつハードディスク装置に近い結果を得た。

同じく社外のユーザーA社で実業務に使用いただいた時の評価結果を **Fig. 13** に示す。A社では製品のデザイン設計に市販ワークステーション上で動作するメカニカルCADを使用しており，ワークステーション用ハードディスクと比較・評価いただいた。なお Fig. 13 に示す処理時間には，メモリアクセス以外の処理時間も含まれている。ファイルサイズが小さい場合はハードディスク装置との差はないが，4 M Byte 以上で差がでてくる。

なお上述の Fig. 11～13 および Table 2 は第1方式を用いた場合の結果である。

## 5 結　言

当社では，光ディスクの特長を生かし，さらにコンピュータ用外部メモリとしての幅広い普及をめざして VHD システムを開発した。今回の開発成果を以下にまとめる。

（1） ハードディスク用デバイスドライバの下で動作させることにより，既存ハードディスクとの互換性を実現した。対象コンピュータとしてワークステーション，パーソナルコンピュータの主要機種を網羅した。

（2） 当社独自のキャッシュ方式による高速化アルゴリズムを開発し，光ディスクを大幅に上回りハードディスクに近い高速性を得た。

（3） 幅広いアプリケーションで高速性を発揮するため，3種のキャッシュアルゴリズムを開発した。

今後は，ハードディスク・エミュレーションの対象 HOST コンピュータおよび内蔵する光ディスク装置の適用機種を拡げることに取り組み，さらにメモリの階層化アーキテクチュア技術を発展させて行きたい。

**参 考 文 献**

1) ANSI X3. 131-1986 "Small Computer System Interface"