1987

# MODELING AND EVALUATION OF A HYBRID OPTICAL AND MAGNETIC DISK STORAGE ARCHITECTURE

George Diehr
*University of Washington*

Bernie Han
*University of Washington*

# MODELING AND EVALUATION OF A HYBRID OPTICAL AND MAGNETIC DISK STORAGE ARCHITECTURE

George Diehr
Bernie Han
Department of Management Science
University of Washington

## ABSTRACT

A hybrid storage system combining optical disks and magnetic disks is proposed and evaluated via mathematical models. Unlike most current applications of optical disk technology, which consider static databases or deferred update, this research considers environments with a moderate level of near real-time updates. An example of such an environment is databases for administrative decision support systems (DSS).

The proposed hybrid storage system uses a write-once, read-many optical disk device (ODD) for the database and a conventional magnetic disk (MD) for storage of a differential file. Periodically, the differential file is used to "refresh" the ODD file by writing updated records to free space on the ODD. When available free space on the ODD is exhausted, the file is written to new ODD media -- a "reorganization" operation.

Models of storage cost are developed to determine the optimum refresh cycle time, $t^*$, and optimum reorganization cycle time, $T^*$. Parameters of the model include data file volatility, file size, device costs, and costs for refresh and reorganization. Numerical results indicate that the hybrid system is attractive for a broad range of database environments.

## 1. INTRODUCTION

Most current applications of optical disk storage devices are for static or relatively static databases. Databases such as encyclopedias and bibliographies are now available on CD-ROMs (Malloy 1986; Desmarais 1985; Lowe, Lynch and Brownrigg 1985). The write-once devices are being used in environments where additions to the databases are common but record changes are few (Shaffer, Shelin and Thomas 1983; Ammon, Calabria and Thomas 1985). Both additions and changes are applied in a batch. These applications reflect, of course, characteristics of the CD-ROM and WORM devices. CD-ROMs are not writeable; they are literally stamped out. The WORM can be written but not rewritten. In addition, the space overhead for a write discourages writing of small units -- that is, it encourages collection of updates into batches so that large units may be written. Characteristics of optical devices are described in Section 2.

In contrast to most current applications of optical storage devices, our focus is on environments which have a low to moderate level of updates which must be available for online access within a short time after they occur. The objective is to explore ODD technology for such applications. Increasingly prevalent examples of such environments are decision support databases (Sprague and Carlson 1982). A qualitative analysis of DSS databases and the congruence of their requirements with features and characteristics of optical storage systems is presented in Section 3.

Section 4 describes a proposed storage architecture which utilizes a differential file approach. The differential file is stored on a conventional magnetic disk (MD) and the "base" file is stored on a WORM optical disk device (ODD) with distributed free space for subsequent updates. Unlike a pure optical storage system, the hybrid system supports online updates by writing them to the MD's diffe-

rential file. Periodically, the differential file is copied into free space on the ODD to refresh it's base file. When free space fills, it is merged with the base and differential files to generate a new base file on new ODD media.

Section 4 also presents a cost model which accounts for the MD and ODD device and media costs, cost of refresh processing, and cost of reorganization. The cost is a function of the refresh time period, t, and reorganization time period, T. Section 5 develops an iterative algorithm for determining optimal values for t and T.

Results for a range of file characteristics and several cost assumptions are given in Section 6. Section 6 also presents closed form results derived through model approximations. These results suggest that, at least from a storage cost standpoint, the ODD/MD hybrid represents a very cost effective approach for a wide variety of database applications.

Section 7 includes a summary and suggestions for future work.

## 2. CHARACTERISTICS OF OPTICAL DISK DEVICES

Figure 1 illustrates the data organization on an ODD. Tracks are organized into groups called "bands." Access is in two steps: a coarse access to the band followed by a fine access to a track within the band. This two stage access is required due to the very high inter-track density. Average access time is significantly greater than for magnetic disks.

The write-once, read-many optical disk represents an important addition to the hierarchy of computer storage devices. Figure 2 summarizes characteristics of both primary and secondary storage technologies in terms of storage cost and access performance. As seen, the ODD fills a large gap between earlier mass storage systems and conventional magnetic disks. This figure does not, of course, tell the whole story. The ODD cannot be rewritten and while rewriteable ODDs are in the labs they appear to be several years from commercial availability. Furthermore, when available they are not expected to replace the WORM devices to lower density (Vitter 1985; Maier 1982).

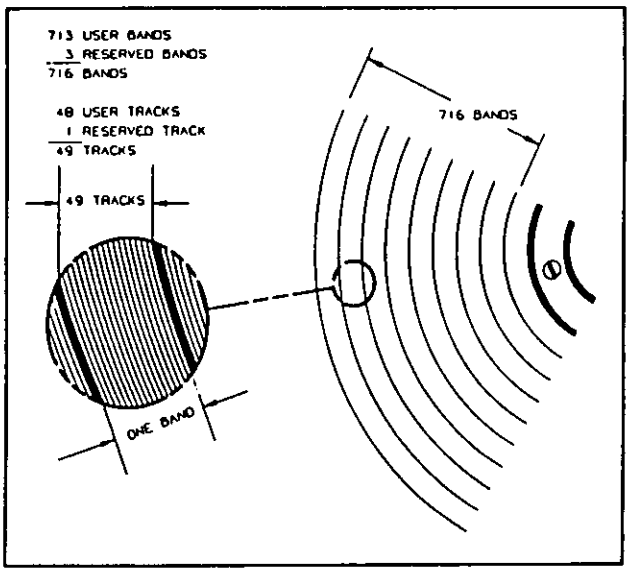


**Figure 1.** ODD Data Organization

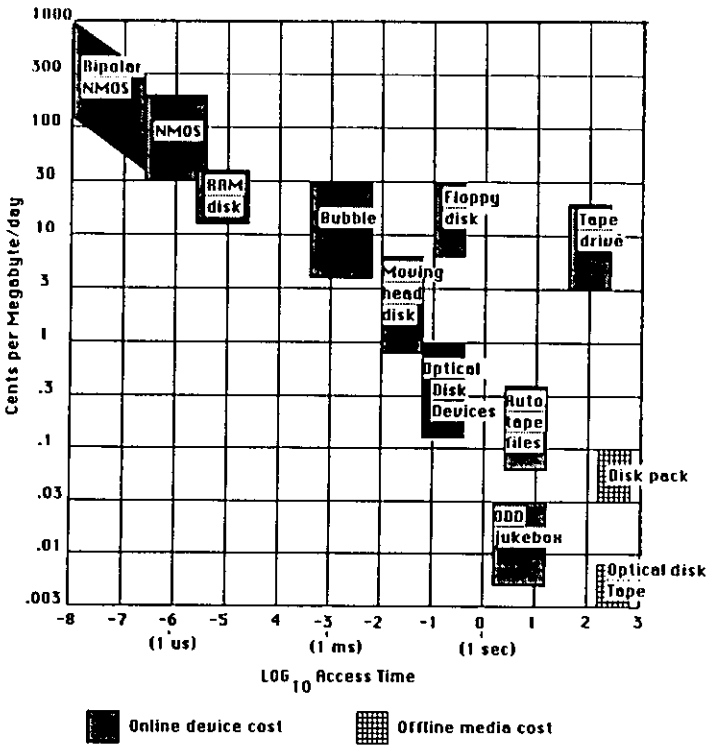


**Figure 2.** Random Access Time and Storage Cost for Main Memory and Secondary Storage Systems (Adapted from Pohm 1984)

An important limitation of both the WORM and rewriteable ODDs is the high overhead for error correction code associated with each block written. This encourages writing large blocks. Furthermore, the random access time is on the order of five times that of the magnetic device. This discourages application in environments with a high volume of reads to small units.

In addition to its very low cost, the ODD has several other advantages over the MD. One is data transfer rate which ranges from roughly comparable to that of MDs to twice as fast. Thus, its performance may even exceed that of the MD if the size of the data unit transferred is large. For example, using access times of 30 and 150 milliseconds and transfer rates of one and two megabytes per second for MD and ODD respectively, the two devices have identical random retrieval times to 240 kilobyte units.

Table 1 summarizes characteristics of the ODD and MD. Figure 3 characterizes database applications in the two dimensions of static to highly volatile and random to sequential access. Within this region, areas are classified in terms of the most cost effective storage system: conventional magnetic disk, pure optical disk (CD-ROM or WORM) device, or the hybrid storage system. One of the objectives of this research is to quantify these boundaries, particularly the boundaries between the hybrid ODD/MD and pure MD storage systems.

**Table 1.** Characteristics of Optical and Magnetic Disk Devices

| | ODDs | MDs |
|---|---|---|
| Capacity | 1 - 4 GB | 100 MB - 1 GB |
| Cost | $ 5/MB | $50/MB |
| Band Access | Yes | No |
| Data Integrity | High | Moderate |
| Data Transfer | 2 - 3 MB/sec | .8 - 2 MB/sec |
| Removability | Yes | Depends |
| Life Span | 10 years | 2 - 3 years |
| Random Access | 150-200 ms | 25-45 ms |
| Read/Write | Write once Read many | Read/write many |
| Update in Place | No | Yes |

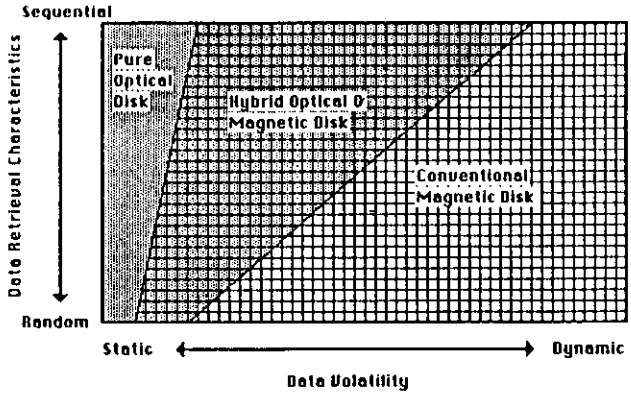


**Figure 3.** Most Cost Effective Storage System as a Function of Data Volatility and Retrieval Characteristics

## 3. OPTICAL DISK TECHNOLOGY AND DECISION SUPPORT DATABASES

Most current applications of ODDs are limited to bibliographic databases (Desmarais 1985), textual databases (Lowe, Lynch and Brownrigg 1985), office documents (Christodoulakis 1985), and image databases (Shafer, Shelin and Thomas 1983; Ammon, Calabria and Thomas 1985). In some cases, such databases are literally stamped out onto CD-ROMs. In other applications, WORM devices are used but the updates are collected over some time before being written to the WORM. In many of these cases the document (or object) is large so that the overhead of small writes is not incurred. These databases can be characterized by the large size of the stored unit, retrieval and update of large units, and a low level of updates (most of which are additions as opposed to changes).

At another extreme are traditional administrative databases used for high volume transaction processing ("production system database" or PSDB). Such databases are poorly suited for WORM devices due to both high volatility and the relatively small size of the stored units, e.g., tens to hundreds of bytes. This leaves a broad range of applications in which the update rate is low to moderate and most of the retrievals involve database scans as opposed to random accesses. One example of such applications is the management decision support database (DSDB).

A DSDB provides data for tactical and strategic level ad hoc decision making. We can characterize

a DSDB and contrast it with a PSDB in three dimensions: data characteristics, access characteristics, and response time requirements. These and other characteristics of transaction processing and decision analysis are summarized in Table 2.

**Table 2.** Comparative Characteristics of Decision Support and Operational Support Environments

| | Decision Support Environment | Operation Support Environment |
|---|---|---|
| Data Sources | Extractions from PSDB historical data, external sources | Production database (PSDB) |
| Currency - time lag from event to update | Minutes to days | Real time to hours |
| Query Type | Ad Hoc, hard to format | Use access routine |
| Data Volume Retrieved | High, many records | Single or a few records |
| Decision Type | Unstructured, semi-structured | Routine, programmable |
| Data Use | Trend analysis, forecasting, prediction | Daily operations |
| Decision Time | Minutes to days | Real time to minutes |

*Data Characteristics.* There are three sources for DSDB data: extracts from the PSDB, internal historical data, and external sources (Sprague and Carlson 1982, p. 241). Historical data is (almost by definition) static. Data from external sources tends to have low volatility -- new data will be added but older data is relatively static. Data from the PSDB will have a significant range of volatility.

Historical data is particularly problematic; it is usually considered to be too voluminous to maintain online on conventional magnetic disks unless it is needed by the PSDB (and it usually is not). If the DSDB is limited to conventional devices, the alternatives to online storage of massive amounts of historical data are aggregation (Sprague and Carlson 1982, p. 240), offline storage on magnetic tape, or use of mass storage devices such as the IBM 3850. All of these have drawbacks.

Consider, for example, an aggregation of product sales records. Should the aggregation be over geographical region, time, salesperson, product type, customer, or some combination of these attributes of a sale? A seemingly innocuous aggregation such as summarizing sales by month but maintaining full detail of product, customer, salesperson, etc., would frustrate any subsequent attempt to determine if occasional short-term spurts in demand for specific products are random phenomenon or follow some predictable cycle. As stated by Sprague and Carlson, "Often an aggregate value will cause the decision maker to want to examine the detailed data that were used to create the aggregate...." (1982, p. 240).

Use of offline tape storage adds significant cost and time overhead to data access. Extracting data for analysis may require special tape mount jobs, merging data with incompatible formats, and the need to write code in a procedural language. Often, the end result is simply to forego the analysis or resort to manual extraction of relevant data from hardcopy reports. A further problem with magnetic tape is its deterioration over time. While definitive information is not available, lifespan estimates for ODDs are upwards of ten years.

The IBM 3850 variety of mass storage device has one major barrier for DSDB storage: very high fixed cost. Only the largest organizations are able to justify the $2 million initial cost of this class of device. Furthermore, applications which require random accesses across several physical tape cartridges (e.g., a join which crosses cartridges) are not practical.

Another data characteristic is "currency" -- what is the time lag from event to data update. In the PSDB, this is usually short (from hours down to real time). By contrast, decision analysis seldom requires real time data. An update delay of hours to even days is often acceptable. "Few decision support systems have a requirement for real time data. In fact, some DSS users prefer not to use real time data in order to keep the problem description (i.e., the extracted database) focused on a static time frame" (Sprague and Carlson 1982, p. 250). For example, many budget analysts prefer to use snapshots of accounting data (the "month-end close").

The optical device represents an almost ideal combination of magnetic disk and mass tape

characteristics for the DSDB environment. Online storage cost per byte is comparable to the mass tape device (see Figure 2) but fixed cost is comparable to the magnetic disk. While the optical device is inefficient for online applications requiring frequent and small unit updates, a batched update is acceptable in the DSS environment. Furthermore, the hybrid ODD/MD architecture we propose can support real time update as necessary.

*Access characteristics.* A PSDB is characterized by high volume of random access to small units of data -- a single record or small group of records. A DSDB accesses larger units, often requiring a full file scan to extract a relevant subset of records and data items or to create summaries.

Most PSDB access is limited to a set of predefined paths. The overhead of index management with its concomitant small writes is cost effective. By contrast, since DSS analyses tend to be ad hoc, the accesses to a DSDB are also ad hoc. Thus, the use of indexes and other access paths are usually not recommended and the need for small writes is reduced.

Furthermore, a DSS application is primarily read-only. If output is created (e.g., a file extract), it may be written to a magnetic device if it has a short life or written to an optical device if it will be used over an extended period of time. Production applications involve a high volume of update making rewriteable and efficiency of small writes mandatory (i.e., conventional MDs).

Therefore, the characteristics of the optical devices are well suited to the access characteristics of a DSDB environment -- minimal update, acceptable random access time, and very high data transfer rate.

*Response requirements.* Production environments require rapid response times. Long response times reduce operator throughput, cause delays to the client, and can produce long queues of transactions. In the decision support environment, the pace is more leisurely. The time span from problem recognition to decision is typically hours to days. A DSS application usually has three phases: 1) formulation of the query, extraction, or report specifications, 2) database access time, and 3) analysis. The time involved in formulation and analysis is minutes to hours (or even days). Therefore, database access times which run to several seconds or minutes usually have a minimal impact on the overall time.

Our *qualitative* conclusion is that the optical device meshes well with the requirements placed on a DSDB. The strengths and advantages of the device are congruent with DSDB requirements -- low fixed and very low variable costs, rapid sequential access, acceptable random access times, and long life. The deficiencies of the device, such as high overhead of small writes and inability to update in place, are either not of concern in the DSS environment or can be ameliorated by the hybrid ODD/MD architecture.

We recognize, however, that while there are general features of DSDBs and PSDBs, individual files and databases in both environments can exhibit atypical characteristics. For example, a specific DSDB file may require both frequent update and high level of currency making the optical device less attractive. Conversely, there will be PSDB files which are relatively static or for which deferred update is acceptable. Therefore, our objective in the remaining sections is to develop quantitative models which will allow us to determine, on a file by file basis, the preferred storage device.

Before ending this section, we comment briefly on the issue of a separate database for decision support. One might argue that historical and external data be maintained on a separate DSDB but that requirements for data from the production database be simply accessed directly from that database. This is certainly feasible and is recommended where the volume of DSS applications is very low. In general, we and other writers [e.g., Sprague 1983, p. 109) advocate extracting data from the PSDB into an independent DSDB. The advantages include 1) the ability to restructure the DSDB for its specific environment and 2) insulation of the production and decision support systems from each other in terms of file access demands.

A restructure includes all relational operations such as projection of a subset of data items, selection of records, aggregations (e.g., SQL "Group By"), and joins. A DSDB may significantly benefit from non-BCNF relations and other redundancy; since its updates are obtained from other sources or databases, update integrity is not a problem. The DSDB may also remove or add access paths, recode data items into "friendlier" representations, and physically reorder records.

Extractions from the PSDB into a DSDB represent "materialized views" or "snapshots." Clearly, it is very expensive to maintain such views by complete re-extraction and rewrite at desired intervals (e.g., daily, weekly, etc.). Furthermore, complete rewrite would make the use of WORMs quite expensive since the space occupied by earlier versions of a snapshot could not be recovered. Several recent papers have presented algorithms to reduce the cost of updating snapshots (Blakeley, Larson and Tompa 1986; Lindsay et al. 1986). These methods can be adapted to our proposed hybrid storage system.

To conclude this section, Figure 4 summarizes transaction processing, decision support, and library information environments in terms of two dimensions: update intensity and degree of random versus sequential access. This graph is partitioned into areas where each of the three storage systems, pure MD, pure ODD, and hybrid ODD/MD, is most cost effective. An objective of this research is to more precisely characterize these areas.

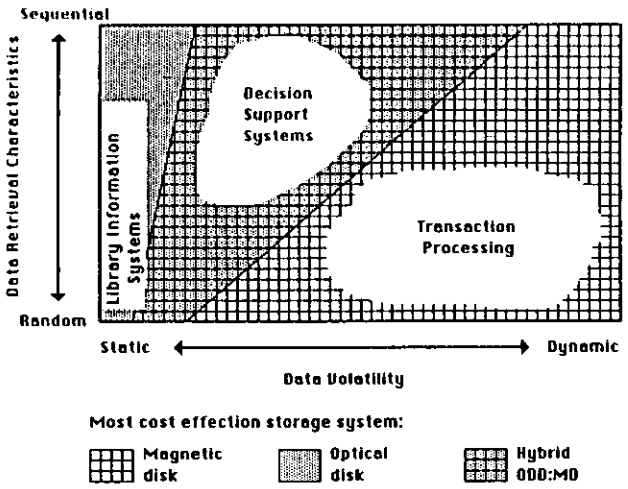


**Figure 4.** Characterization of Operational Processing Systems, Library Systems, and Decision Support as a Function of Data Volatility and Access Characteristics

## 4. A HYBRID ODD/MD ARCHITECTURE AND COST MODEL

The objectives and constraints on the hybrid storage system architecture include:

1. Updates should be available shortly after they are recorded on the operational database. The permissible time lag will be a function of the applications using the hybrid system. We would expect reasonable limits to range from seconds to several hours.

2. Small writes are to be avoided. Thus, updates must be batched into units of reasonable size (several thousand bytes) before being written to the ODD.

3. An update of a record on the ODD should be in physical proximity to its base record to minimize random accesses.

### 4.1 ODD/MD Architecture

Figure 5 illustrates the architecture of the proposed storage system and shows the data flows for retrieval, update, refresh, and reorganization. Specifics of the design and assumptions follow the figure.

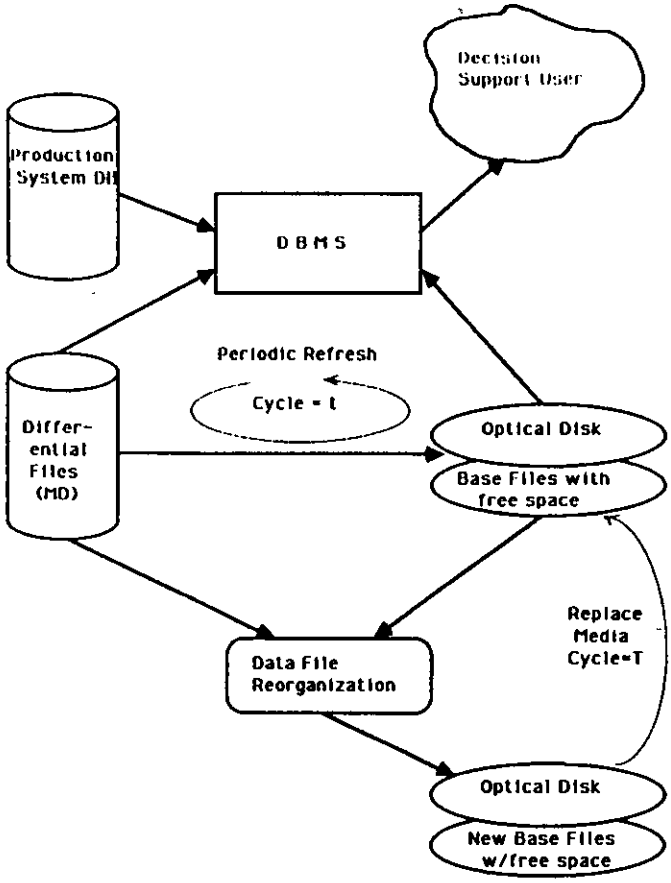


**Figure 5.** Architecture and File Maintenance, Hybrid Optical/Magnetic Disk Storage System

1. A differential file is created on the MD for each base ODD data file. Records in the differential file are kept in the same order as records in the base file so that the current records may be retrieved by a parallel scan of the two files. The differential file uses the same access method as the ODD base file, for example, a VSAM-like scheme.

2. There are several possible methods for transmitting updates to the differential file. Records might be read continuously or in batch from the audit trail or one of the methods for updating materialized views used.

3. The base file includes distributed free space within each data track. The free space is used to write updated versions of records within the track. This keeps updates in close proximity to base records.

4. Record updates are uniformly distributed over the file and occur with constant frequency.

5. At each time interval t, the records in the differential file are written to the free space in the base file tracks (a "refresh").

6. At interval T, the latest version of each record from the free space, base file, and differential file is written to a new WORM media (a "reorganization").

In the next section, a cost model is developed which is used to solve for the optimum intervals for refresh, $t^*$, and reorganization, $T^*$. This model is also used to compare costs of the hybrid and conventional magnetic disk systems.

### 4.2 Cost Model

The model includes costs related to storage and to operations of refresh and reorganization. The parameters include:

| | |
|---|---|
| $C_{OD}$ | Cost of optical disk device per megabyte/day |
| $C_{MD}$ | Cost of magnetic disk per megabyte/day |
| $C_{med}$ | Cost of WORM media per megabyte |
| D | Size of file in megabytes |
| v | Volatility -- fraction of file modified/day |
| $S_r$ | Fixed setup cost for file refresh |
| $S_R$ | Fixed setup cost for file reorganization |
| $U_r$ | (Update) refresh cost/megabyte |
| $U_R$ | (Update) reorganization cost/megabyte |

While the proposed hybrid architecture supports record addition, deletion and modification, to simplify the initial development of a cost function, we consider only updates which change existing records. No record additions or deletions occur.

#### 4.2.1 Storage cost

Under the assumption of dynamic space allocation on the *MD*, the storage space required is given by the size of the differential file at each point in time. If multiple updates to the same record were each stored, the size of the differential file at time x would be D(1 + vx). We assume, however, that only the latest version of a record's update is stored in the differential file so that its size at time x is $D(1-e^{-vx})$. (Thus, the differential file size asymptotically approaches D as vt goes to infinity.) Therefore, with refresh interval, t, the average size of the differential file is:

$$1/t \int_0^t D(1 - e^{-vx})dx = D/t(t - 1/v + e^{-vt}/v).$$

Storage space required on the *ODD* is the base file plus the free space. Each refresh requires $D(1-e^{vt})$ megabytes. Assuming an integral number of refreshes per reorganization (i.e., T/t an integer), the total ODD free space is:

$$D(1 - e^{vt}) (T/t - 1).$$

The *-1* is due to the fact that the final refresh is incorporated into the reorganization. For example, if t=T, then no refresh is done and no free space is required. Total MD plus ODD storage cost is given by:

$$C_{STO} = C_{OD} [D + D(1 - e^{-vt}) (T/t - 1) ] + C_{MD} D/t [t - 1/v + e^{-vt}/v]$$

#### 4.2.2 Refresh cost

The refresh stores updated records from the MD differential file into the distributed free space on the ODD. The base and differential are in the same order so that writing records to the free space is a sequential process. The cost for each refresh is assumed to involve a setup cost, $S_r$, plus a cost based on the size of the differential file. The average daily cost for refreshes is:

$$C_{ref} = [S_r + U_r D (1 - e^{-vt})] (T/t - 1)/T$$

### 4.2.3 Reorganization cost

Reorganization creates a new file on new WORM media by writing the latest version of each record which may be in the base file, free space, or in the differential file. Given adequate internal memory (i.e., space for one ODD track plus differential file updates for that track), the process is also sequential. The average daily reorganization cost is estimated as a setup cost, $S_r$, a charge for the space required on new optical media, plus a processing cost which depends on the sizes of the base file, occupied free space, and differential file:

$$C_{Reorg} = \{S_R + DC_{med} + U_R [D + D(1 - e^{-vt})(T/t - 1) + D(1 - e^{-vt})]\}/T$$

Simplifying gives:

$$C_{Reorg} = \{S_R + DC_{med} + U_R [D + D(1 - e^{-vt})T/t]\}/T$$

Total cost is given by:

$$F(t,T) = C_{STO} + C_{ref} + C_{Reorg} \quad (1)$$

## 5. SOLUTION ALGORITHM

There is no closed form expression for the minimum value, $F^*$, of F(t,T). However, it is possible to show that for T in the interval [0,T-], F(t,T) is convex for $0 \leq t \leq T$. While F(t,T) may not be convex for T > T-, we can also show that F(t,T) exceeds $F^*$ for T > T- and $0 \leq t$. Thus, a solution approach is to determine the minimum value, $F^*(\cdot,T)$, of F over t for each value of T in the range [0,T-]. The minimum of the $F^*(\cdot,T)$ over T is the global minimum.

Unfortunately, there is no closed form expression for the value of t which minimizes F(t,T) for fixed T. Therefore, an iterative algorithm is used which makes successive linear approximations to $(1-e^{-vt})$. The algorithm is:

1. For each integer value of T in the range 1 to T-
   a. Approximate $(1 - e^{-vt})$ by vt. This approximation of F(t,T) is denoted f(t,T).
   b. Repeat until successive values of $f(t^*,T)$ are sufficiently close:
      1. Determine $t^*$ such that $f(t^*,T)$ is minimized
      2. Reapproximate $(1 - e^{-vt})$ by a + bvt, selecting a and b so that a + bvt is tangent to $(1 - e^{-vt})$ at $t^*$.
2. Select $T^*$ such that $f(t^*,T^*)$ is minimized

## 6. NUMERICAL RESULTS AND ANALYSIS

The model was run with a variety of cost, file size, and volatility assumptions to gain insight into the following two questions:

1. What is the impact of varying file size, volatility, and costs on the optimal refresh and reorganization intervals?
2. What are the characteristics in terms of file size and volatility which favor use of the proposed hybrid storage system over a conventional magnetic disk?

Sections 6.1 and 6.2 address these questions by numerical solution of the model. Section 6.3 uses approximations to the model to obtain closed form results.

### 6.1 Effects of File Size, Volatility, and Costs on $t^*$ and $T^*$

Optimal refresh and reorganization intervals were determined for several different values of the cost parameters as functions of file size, D, and volatility, v. The objective was to gain insight into the impact of these parameters on $t^*$ and $T^*$. The costs used in all runs were:

$C_{OD}$ = \$.0042 = Cost of an optical disk device per megabyte per day.

$C_{MD}$ = \$.042 = Cost of magnetic disk per megabyte per day -- based on \$50,000 device with one Gigabyte capacity.

$C_{med}$ = \$.025 = Cost of WORM media per megabyte. Based on a cost of \$100 for a four Gigabyte media.

$S_r$ = \$10 = Fixed setup cost for file refresh

$S_R$ = \$15 = Fixed setup cost for file reorganization

$U_r$ = \$.05 = Refresh cost per megabyte. This is based on estimates of processing time and amortization of processor hardware costs.

$U_R$ = \$.10 or .30 = Reorganization cost per megabyte. Two different values were used to explore the impact of this cost.

Figure 6 shows the relationship between file size and optimal refresh and reorganization intervals for volatility of 1%. Due to the fixed costs of refresh and reorganization, increasing file size results in increasing the frequency of both of these updates. A file of up to about eleven megabytes will have the same refresh and reorganization intervals--that is, reorganization only. Files in the range eleven to fifty megabytes have $t^* = T^*/2$ -- one refresh.

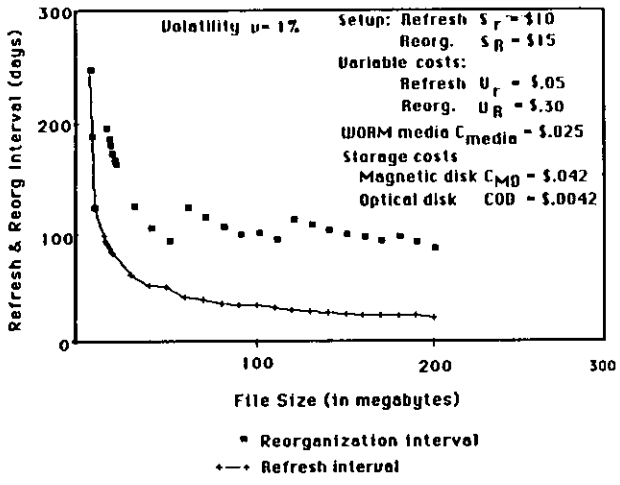


**Figure 6.** Optimum Refresh and Reorganization Intervals Versus File Size

Figure 7 shows the impact of different volatilities on $t^*$ and $T^*$ for file size of 100 megabytes. As expected, increased volatility results in more frequent reorganization. However, the refresh interval is roughly constant for volatilities above 1%. The plots are discontinued at a volatility of 3.8% -- the point at which the pure MD system has lower cost.

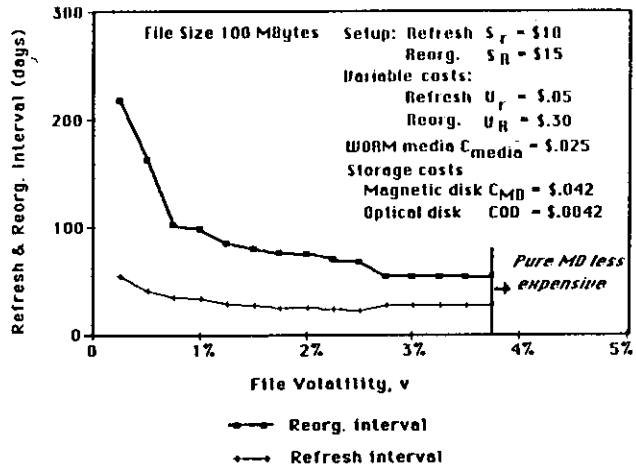


**Figure 7.** Optimum Reorganization and Refresh Intervals Versus File Volatility

### 6.2 Comparative Costs: Hybrid versus Pure MD Storage Systems

An important objective of this study is to compare the economics of the proposed hybrid ODD/MD storage system with a conventional magnetic disk. Our intuition, and "in the limit" arguments, tell us that the hybrid system will have lower cost for low volatility files and the pure MD system will be the optimum choice for high volatility. In addition, since we have assumed a fixed overhead cost for refresh and reorganization, we also expect that a very small file will have lower cost if stored on a MD even if its volatility is rather low.

To quantify these notions, the model was run for varying file sizes and for two values of variable reorganization cost ($U_R$), \$.10 and \$.30. Results are presented in Figure 8 as two isocost lines -- points of equal cost for the hybrid and pure MD storage systems. Points below an isocost line favor the hybrid system. As expected, larger file sizes and lower volatility favor the hybrid system. For example, with $U_R$ = \$.30, a file of ten megabytes must have volatility under 1% to favor the hybrid system. By contrast, a file of 200 megabytes can have volatility up to about 4.1% before the pure MD is favored. Of course, with lower reorganization costs (e.g., $U_R$ = \$.10), the isocost line is higher.

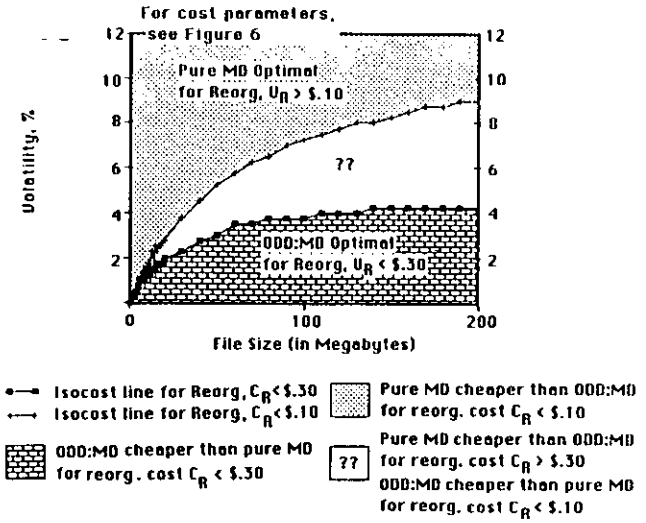


**Figure 8.** Isocost Lines for Pure MD and ODD:MD Systems as a Function of File Volatility and File Size for Two Reorganization Costs

Figure 9 uses isocost lines to compare storage system costs as a function of volatility and the ratio of hybrid to pure MD storage costs for two files sizes, 50 and 150 megabytes. As expected, higher MD/hybrid cost ratios increase the volatility limits at which the hybrid system is preferred. It shows that the hybrid system remains viable at costs of up to one-half conventional magnetic disk cost if volatility is very low. This figure again demonstrates that, other things equal, the hybrid system's attractiveness is enhanced by larger files.

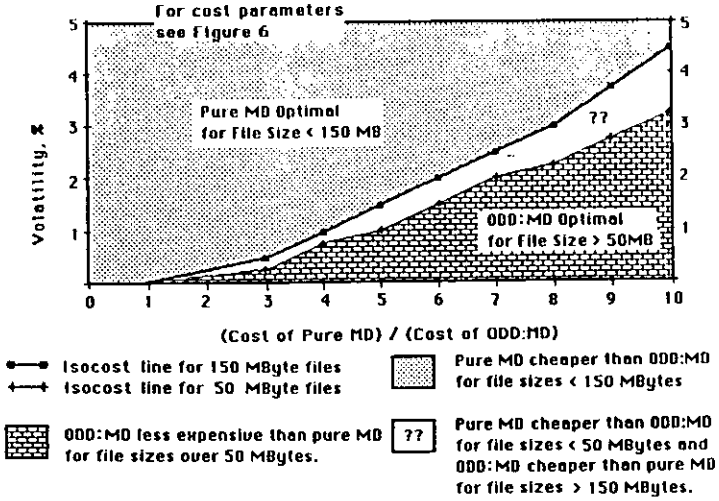


**Figure 9.** Isocost Lines for Pure MD and ODD:MD Systems as a Function of Storage System Cost Ratio and Volatility, for Two File Sizes

Clearly, the appropriate storage system depends on factors in addition to those captured in our cost model. As shown in Figure 3, a critical factor will be the access characteristics -- relative degree of random versus sequential. In an environment of a high volume of random accesses to small units of data, the hybrid system could become a bottleneck.

### 6.3 Model Approximation and Analysis

The approximation, $(1 - e^{-vt}) = vt$, allows closed form solutions for $t^*$ and $T^*$. Numerical solutions to the model demonstrate that this approximation has only a minor effect on the total cost. Since the primary objective of our work is to broadly characterize environments in which the proposed hybrid storage system is cost effective, this level of error is acceptable.

The approximated model can can be written in the form:

$$f(t,T) = K_1 + K_2 t + K_3 t/T + K_5/T + K_6 T$$

where: $K_1 = (C_{OD} D + U_r Dv + U_R Dv)$
$K_2 = (C_{MD}/2 - C_{OD})Dv$
$K_3 = -U_r Dv$
$K_4 = S_r$
$K_5 = S_R + C_{med} D + U_R D - S_r$
$K_6 = C_{OD} Dv$

Solving for $t^*$ and $T^*$ gives:

$$t^* = [K_4 T^*/(T^* K_2 + K_3)]^{0.5}$$

$$T^* = [(K_3 t^* + K_5)/K_6]^{0.5}$$

With reasonable assumptions for cost parameters and volatility, the Hessian is positive definite thus assuring the solution is a global minimum.

Furthermore, except for virtually static files (i.e., v close to zero), typical refresh and reorganization time intervals are such that $|K_3 t^*| << K_5$ and $|K_3| << T^* K_2$. Thus, $t^*$ and $T^*$ may be further approximated by:

$$t\# = [K_4/K_2]^{0.5} = \{S_r/[(C_{MD}/2 - C_{OD})Dv]\}^{0.5}$$

$$T\# = [K_5/K_6]^{0.5} = [(S_R - S_r + C_{med} D + U_R D)/(C_{OD} Dv)]^{0.5}$$

This allows the following observations: 1) Solutions for t# and T# are independent. 2) The approximation for t# shows that as the cost of the optical disk approaches one-half the cost of the magnetic device, the refresh interval increases -- that is, only reorganizations are performed. 3) The approximation T# shows that if the fixed reorganization and refresh costs are equal, the reorganization interval is independent of file size.

## 7. SUMMARY AND CONCLUSIONS

Optical storage technology has made it economical to store massive databases online. However, the pure optical systems present several problems. Their write-once limitation and high overhead for small writes discourages their use in applications with moderate levels of online updates. An example of such applications are management decision support databases in which the database size and access characteristics make the (pure) ODD an attractive candidate but where update needs make it inappropriate. This research proposed a marriage of optical and magnetic storage technologies which exploits the advantages of each to address such applications.

A cost model for the proposed hybrid storage system was developed which determines optimal refresh and reorganization policies. Numerical solutions of the model evaluated the impact of file size and volatility on refresh and reorganization intervals. The model was used to make cost comparisons between conventional magnetic disk and the hybrid storage system. The comparisons allow determination of the more cost effective system as a function of application characteristics such as file size and volatility, and other factors such as the ratio of MD to ODD device costs and the cost of reorganization.

This preliminary evaluation indicates that the proposed hybrid system is practical and cost effective for a broad range of applications. In particular, database environments with large files, low to moderate levels of update, and predominance of sequential file scans are prime candidates for the hybrid system.

We plan to extend the model to incorporate file growth and access characteristics. These extensions will allow us to better identify appropriate environments for the hybrid system.

## ACKNOWLEDGEMENTS

The authors wish to thank the anonymous referees for excellent comments that have improved the quality of this paper.

## ENDNOTES

[1] Note that no time stamping is required. Any record in the differential file is the latest version. If a record is not in the differential file, then the physically last version of a record in the free space is the latest version. Otherwise, the base copy is the current version.

[2] This can alternatively be considered as (1) a seven-year amortization at 10.7% interest rate of a four gigabyte device with cost of $20,000 plus reasonable allowance for operating costs (i.e., 10% of purchase cost per year for maintenance), or (2) the full maintenance lease cost of the device based on typical industry monthly lease rates of purchase cost divided by 40.

[3] This can alternatively be considered as (1) a seven-year amortization at 10.7% interest rate of a four gigabyte device with cost of $20,000 plus reasonable allowance for operating costs (i.e., 10% of purchase cost per year for maintenance), or (2) the full maintenance lease cost of the device based on typical industry monthly lease rates of purchase cost divided by 40.

## REFERENCES

Ammon, G. J.; Calabria, J. A.; and Thomas, D. T. "A High-Speed, Large-Capacity, 'Jukebox' Optical Disk System." *IEEE Computer*, Vol. 18, No. 7, May 1985, pp. 36-45.

Blakeley, J. A.; Larson, P.; and Tompa, F. W. "Efficiently Updating Materialized Views." *ACM-SIGMOD 1986*, pp. 61-71.

Christodoulakis, S. "Issues in the Architecture of a Document Archiver Using Optical Disk Technology." *ACM-SIGMOD 1985*, pp. 34-50.

Desmarais, N. "Laser Libraries." *BYTE*, Vol. 11, No. 5, May 1985, pp. 235-246.

House, W. C. (Ed.). *Decision Support Systems*. Petrocelli Books, Inc., 1983.

Lindsay, B.; Haas, L.; Mohan, C.; Pirahesh, H.; and Wilms, P. "A Snapshot Differential Refresh Algorithm." *ACM-SIGMOD 1986*, pp. 53-60.

Lowe, J. B.; Lynch, C. A.; and Brownrigg, E. B. "Publishing Bibliographic Data on Optical Disks: A Prototypical Applications and Its Implications." *SPIE*, Vol. 529, Optical Mass Data Storage, 1985, pp. 227-236.

Malloy, R. "A Roundup of Optical Disk Drives." *BYTE*, Vol. 11, No. 5, May 1985, pp. 215-224.

Maier, D. "Using Write-Once Memory for Database Storage." *Proceedings of the First Annual ACM Symposium on PODS*, March 1982, pp. 239-246.

Pohm, A. V. "High-Speed Memory Systems." *IEEE Computer*, Vol. 17, No. 10, October 1984, pp. 162-171.

Severance, D. B., and Lohman, G. M. "Differential Files: Their Applications to the Maintenance of Large Databases." *ACM TODS*, Vol. 1, No. 3, September 1976, pp. 256-267.

Shaffer, J. B.; Shelin, J. W.; and Thomas, D. T. "Database and File Management Approach for Large Optical Disk Systems." In E. V. Labudde (ed.), *Optical Disk Systems and Applications*, Proceedings SPIE 421, 1983, pp. 20-30.

Sprague, R. H. "A Framework for the Development of Decision Support Systems." In W. C. House (ed.), *Decision Support Systems*, Petrocelli Books, Inc., 1983, pp. 85-123.

Sprague, R. H., and Carlson, E. D. *Building Effective Decision Support Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1982.

Vitter, J. S. "An Efficient I/O Interface for Optical Disks." *ACM TODS*, Vol. 10, No. 2, June 1985, pp. 129-162.