Optimizing Random Retrievals from CLV format Optical Disks

Daniel Alexander Ford

Department of Computer Science
University of Waterloo,
Waterloo, Ontario, N2L 3G1

Stavros Christodoulakis†

Department of Electronic and Computer Engineering,
Technical University of Crete,
P.O. Box 133, Chania, Greece 73100

ABSTRACT

One technique often employed to improve retrieval performance from storage devices is to reduce seek costs by to clustering frequently accessed data together in locations on the storage device that are physically close. For magnetic disks determining the best position on the disk to place frequently accessed data is straightforward, for optical disks with their many different recording formats the solution is much more difficult. We develop a detailed model for the placement of data on Constant Linear Velocity (CLV) format optical disks that includes distribution of storage capacity across the disks surface (which is variable for CLV format optical disks), the seek performance of the disk drive, delays due to rotational latency, and the distribution of accesses over the data set. We derive closed form expressions which determine the position of frequently accessed data that will minimize the expected cost of random accesses to the data set.

1. Introduction

An important goal of physical database design is to obtain excellent retrieval performance from the storage system or device on which a database resides. An accurate measure of retrieval performance is the expected time delay required to access the records qualifying for a query. For both magnetic and optical disks this delay is dominated by the time needed to reposition the device's disk physically head during the retrieval process. A typical seek time for a magnetic disk is 30 milliseconds (or better), but for optical disks, the value tends to be approximately 400 milliseconds. And for some optical disks, such as *Compact Disk Read Only Memory* (CD ROM), seek times can be as much as one second [Laub 86]. For good retrieval performance from optical disks, it is obviously of critical importance to minimize the expected delays that result from their slow seek performance.

A technique often employed to improve retrieval performance by reducing seek costs is to cluster frequently accessed data together in locations on the storage device that are physically close, such as in the same or adjacent tracks on a disk. The ISAM file organization uses this technique as does the UNIX fast file system [Mckusick 84]. This physical grouping reduces both the expected number of seeks that the disk head will be required to execute as well as the distance it will travel.

We can extend the idea of positioning data to improve retrieval performance to encompass the entire arrangement of sectors on a disk, the goal of the placement procedure would then be to find a global arrangement of all of the disk sectors that will minimize the expected cost of a single disk access.

This *optimal sector placement problem* is an important one for optical disks. Their large storage capacities and low cost make them ideal for large database systems, but their slow seek performance is a drawback. Any technique that can mitigate the impact on retrieval costs of the slow seek performance of optical disks will be of great benefit.

This is particularly true for optical disks which employ the *Constant Linear Velocity* (CLV) recording format. First, CLV format optical disks typically have the slowest access times of all optical disks; extra delay is incurred because of the need to adjust the rotation rate of the disk to match the position of the disk head. And second, data on CD ROM's and CLV format *Write Once Read Many* (WORM) optical disks are never modified or moved from position to position (unlike magnetic disks where the placement of data is modified frequently and is usually transparent to users). An example application would be determination the positions on the disk of frequently accessed indices, or of files that receive

† On leave from the Department of Computer Science, University of Waterloo, Canada.

particularly heavy amounts of retrieval traffic.

The optimal sector arrangement problem has been investigated before for magnetic disks, but the results of those investigations, have very limited application to optical disks. The differences in the physical characteristics of magnetic and optical disks are significant enough to invalidate many important underlying assumptions used in determining placement solutions for magnetic disks. For instance, all previous investigations for magnetic disks [Grossman 73] [Yue and Wong 73] [Wong 80] [Wong 83], implicitly assumed that the distribution of storage capacity over the disk surface was uniform (CAV format). For instance, with the CLV format, the distribution of storage space varies across the disk surface; the tracks nearest the centre of the disk have fewer sectors than those nearest the outer edge.

A skewed storage distribution produces variations in both the amount of clustering possible and in the rotational delay encountered during accesses from different positions on the disk. These variations can be exploited to improve access performance. The cost expression and solution that we derive reflects the impact of these variations and finds the optimal position on a CLV format disk.

It is interesting to note that a uniform storage capacity distribution, such as that produced by the CAV format, eliminates the possibility of trading-off positional performance improvements entirely. All tracks on such disks have the same capacity and rotational delay. This uniformity makes the distance between disk sectors the only factor in determining the expected retrieval cost and simplifies the problem considerably. The Organ-Pipe permutation [Hardy 34], which minimizes this distance in order of the frequency of sector accesses, is the resulting optimal solution.

In the next section, we develop a model for our analysis that encompasses virtually all aspects of the placement problem. In section 3, we develop proofs restricting the form of the optimal solution. In section 4, we analyze the positional performance tradeoff and develop a closed form expression for the optimal solution. Using this expression in section 5, we examine the roles played by the model parameters in determining an optimal sector placement. In section 6, we extend our analysis to include a slightly more general seek model. In section 7, we show how to extend our model to allow for general sector access probability distributions. In section 8, we validate our model and analysis by comparing the solutions they predict with measurements made from a CLV format disk drive. In the final section, we summarize our results.

2. The Placement Model

To avoid some of the inherent complications of employing a discrete model of a disk with a smoothly varying storage capacity, we develop a continuous model for our analysis. In moving from the discrete to the continuous domain the problem changes from one of placing sectors on a discrete disk to one of positioning probability masses on a continuous disk. The immense capacities of optical disks allow this approach; a typical disk can easily be organized into more than 40000 tracks and more than 1 million separate sectors, and often many more.

Storage Capacity Distribution

As mentioned above, we develop a continuous model for our analysis. This model matches the smoothly varying storage capacity found on a CLV format optical disk and eliminates many of the problems that arise from a discrete model.

We adopt a representation that models CLV format disks and their storage capacity distributions in relative terms, and then develop our analysis for this relative representation. This approach allows any disk to be described by just two parameters, the capacity of the innermost position/track of the disk relative to the capacity of the middle position/track, and the slope of the change of storage capacity across the disk, also relative to the capacity of the middle position/track. The relative capacity of the middle position in the continuous model is by definition one unit.

A position/track on the disk in our relative model is represented by a number between 0 and 1. Position 0 corresponds to the innermost track on the disk and position 1 corresponds to the outermost track. The middle track on the disk is represented by position 0.5. A relative model for a disk is illustrated in Figure 1.

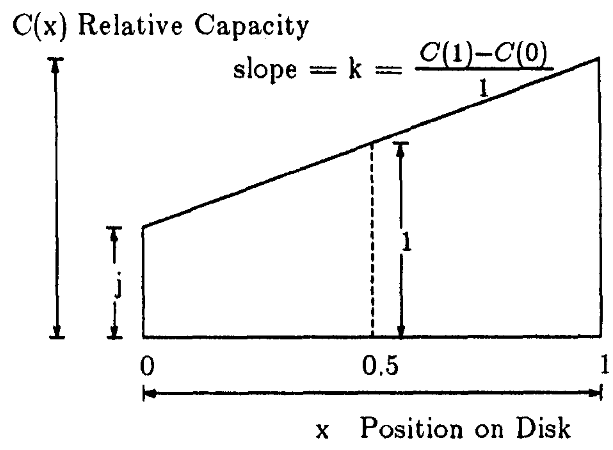

Figure 1: Model of Distribution of Storage Capacity

Letting k be the relative slope of the change in the storage capacity across the disk and j be the relative capacity of the innermost track, we define the relative capacity of the position x to be:

$$C(x) = k\,x + j,\ 0 \le x \le 1,\ 0 \le j \le 1,\ 0 \le k \le 2$$

Since the capacity of the innermost track can never be less than zero, the value for j also cannot be less than zero.

Random Access Probability Distribution

Our probability model assumes that sector access requests are independent of each other. This is consistent with the models used previously for CAV format magnetic disks [Grossman 73] [Yue and Wong 73] [Wong 80] [Wong83]. In addition, to simplify our analysis, we restrict the access probabilities of the point masses to one of two relative values, P_1 and P_2 ($P_1 > P_2$).

The proportion of point masses with relative access probability P_1 is r; the proportion with value P_2, is $1-r$. Figure 2 illustrates a two value probability distribution.

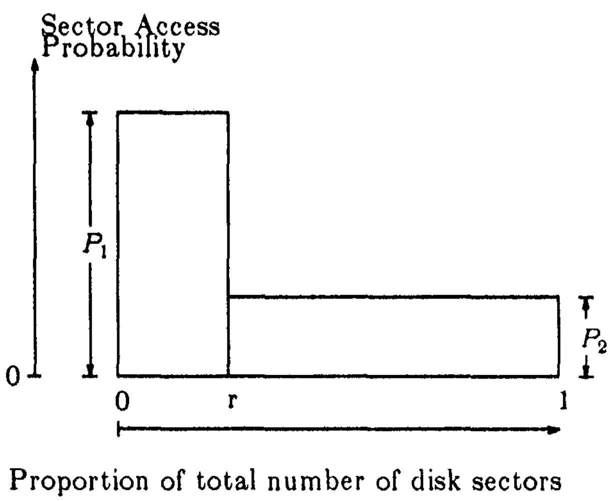

Figure 2: Two Value Rel. Access Prob. Distribution

The absolute access probability value of a probability mass which is composed of points having relative probability density value P_1 is $\dfrac{P_1}{P_1 r + P_2(1-r)}\dfrac{1}{\mu}$, and for P_2 is $\dfrac{P_2}{P_1 r + P_2(1-r)}\dfrac{1}{\mu}$, where μ is the number of probability points (proportional to the area occupied by these points). This number tends to infinity in the continuous model. We define $W = P_1 r + P_2(1-r)$, the normalizing factor.

Seek Cost Function

The cost in time units of seeking to a position t on a disk as:

$$Sc(t) = \begin{cases} a\,t + b & \text{if } t > Q \\ d\,t + c & \textit{otherwise} \end{cases}$$

We do not model proximal window accesses, which have a more complex seek cost function [Ford 91]. A proximal window access is an access performed by tilting the objective lens in the optics of the disk head. This type of access is usually faster than seek accesses which require the disk head to move before data retrieval begins. The amount of data that is present in a window is very much smaller than the capacity of a optical disk. Typically, a proximal window will allow access to at most 0.1% of the disk capacity; as such, it is reasonable to assume that successive random access will not fall within the same proximal window, and that the cost function we described above provides an accurate measure of the expected seek cost.

Note that in the studies of optimal placement for magnetic disks, the cost model used was only $cost = a \cdot t$ (i.e., cost was proportional only to the distance, $b = 0$, $Q = 0$).

Rotational Delay Function

Rotational delay on CLV format disks varies as a function of the capacity of a position on the disk; the more sectors in a track, the longer it takes to read them. Since in our model storage capacities are expressed relative to the capacity of the middle track, we must also specify our rotational delay parameter in relative terms.

We define h to be the time required for the entire middle position of the disk to be read. From this, the expected rotational delay (latency) of accesses from a position y is:

$$Rd(y) = \tfrac{1}{2}\, h\, C(y)$$

The Expected Random Access Retrieval Cost

The objective of our analysis of the optimal placement problem will be to determine an arrangement of probability masses that minimizes the expected random access retrieval cost (seek and rotational delay). The overall expected retrieval cost is computed by summing the cost in time units of successive accesses to each possible pair of initial and destination positions on the disk. The cost of moving from one position to another is the sum of the value of the seek cost function for moving the disk head the distance between the two positions, and the value of the rotational delay function at the destination position. For each pair of positions, the cost must be computed in both directions since the rotational delay cost will vary with the destination. We assume successive accesses are independent, so each term in this sum will be weighted by the product of the probability of accessing the initial position and the probability of accessing the destination position. Our approach below is to first develop a discrete expression for the expected random access retrieval cost and then extend it to the continuous domain.

The discrete cost function in our model is:

$$Cost = \sum_{i=1}^{T}\sum_{j=1}^{T} Pm(i)Pm(j)(Sc(|(i-j)/T|)+Rd(j)) \tag{1}$$

Where i is an initial position (track) on the disk, j is a destination position (track), and $Pm(i)$ is the probability mass assigned to position i. $Sc(|(i-j)/T|)$ is the seek cost between the positions i and j, and $Rd(j)$ is the expected rotational delay at position j.

We can separate out the seek and rotational delay cost components:

$$Cost = \sum_{i=1}^{T}\sum_{j=1}^{T} Pm(i)Pm(j)\,Sc(|(i-j)/T|) + \sum_{i=1}^{T}\sum_{j=1}^{T} Pm(i)Pm(j)\,Rd(j) \tag{2}$$

We expand the function $Sc(|(i-j)/T|)$ (and change the limits on the summations). In the discrete case, we use $\overline{Q}$ to represent the number of tracks, to the left or right of the current track, which are in the current knee span; a and b are the slope and intercept of long seeks on the disk, respectively, and d and c are the slope and intercept of short seeks on the disk, respectively.

$$\begin{aligned} Cost = &\sum_{i=1}^{T-\overline{Q}-1}\sum_{j=i+\overline{Q}+1}^{T} Pm(i)Pm(j)(\frac{a}{T}(j-i)+b) \\ &+ \sum_{i=\overline{Q}+2}^{T}\sum_{j=1}^{i-\overline{Q}-1} Pm(i)Pm(j)(\frac{a}{T}(i-j)+b) \\ &+ \sum_{i=1}^{T-\overline{Q}-1}\sum_{j=i}^{i+\overline{Q}} Pm(i)Pm(j)(\frac{d}{T}(j-i)+c) \\ &+ \sum_{i=T-\overline{Q}}^{T}\sum_{j=i}^{T} Pm(i)Pm(j)(\frac{d}{T}(j-i)+c) \\ &+ \sum_{i=\overline{Q}+2}^{T}\sum_{j=i-\overline{Q}}^{i} Pm(i)Pm(j)(\frac{d}{T}(i-j)+c) \\ &+ \sum_{i=1}^{\overline{Q}+1}\sum_{j=1}^{i} Pm(i)Pm(j)(\frac{d}{T}(i-j)+c) \\ &+ \sum_{i=1}^{T} Pm(i) \sum_{j=1}^{T} Pm(j)Rd(j) \end{aligned} \tag{3}$$

After some simple but tedious manipulations, we have:

$$\begin{aligned} Cost = &\sum_{i=1}^{T}\sum_{j=i}^{T} Pm(i)Pm(j)\frac{a}{T}(j-i) \\ &+ \sum_{i=1}^{T}\sum_{j=1}^{i} Pm(i)Pm(j)\frac{a}{T}(i-j) \\ &+ b + \sum_{i=1}^{T} Pm(i)Pm(i)b \\ &- \sum_{i=1}^{T-\overline{Q}-1}\sum_{j=i}^{i+\overline{Q}} Pm(i)Pm(j)(\frac{(a-d)}{T}(j-i)+(b-c)) \\ &- \sum_{i=T-\overline{Q}}^{T}\sum_{j=i}^{T} Pm(i)Pm(j)(\frac{(a-d)}{T}(j-i)+(b-c)) \\ &- \sum_{i=\overline{Q}+2}^{T}\sum_{j=i-\overline{Q}}^{i} Pm(i)Pm(j)(\frac{(a-d)}{T}(i-j)+(b-c)) \\ &- \sum_{i=1}^{\overline{Q}+1}\sum_{j=1}^{i} Pm(i)Pm(j)(\frac{(a-d)}{T}(i-j)+(b-c)) \\ &+ \sum_{j=1}^{T} Pm(j)Rd(j) \end{aligned} \tag{4}$$

Expected Cost for Small Knee Span Sizes

We develop our expression for the expected cost by assuming that the knee span size is small in comparison to the number of tracks on the disk. This is the case for conventional disk drives; the typical number of tracks on an optical disk platter will range from 20000 to 40000, or more, depending on the diameter of the disk. A typical value for the knee span would be from 40 to 100 tracks.

Our expression for the expected (discrete) random access retrieval cost simplifies to the following:

$$\begin{aligned} Cost = &\sum_{i=1}^{T}\sum_{j=i}^{T} Pm(i)Pm(j)\frac{a}{T}(j-i) \\ &+ \sum_{i=1}^{T}\sum_{j=1}^{i} Pm(i)Pm(j)\frac{a}{T}(i-j) \\ &+ b + \sum_{i=1}^{T} Pm(i)Pm(i)b + \sum_{j=1}^{T} Pm(j)Rd(j) \end{aligned} \tag{5}$$

For $T \to \infty$, the first two terms of the discrete cost expression become:

$$\begin{aligned} &\lim_{T\to\infty} \sum_{i=1}^{T}\sum_{j=i}^{T} Pm(i)Pm(j)\frac{a}{T}(j-i) \\ &= \int_{x=0}^{1}\int_{y=x}^{1} Pm(x)Pm(y)\,a(y-x)\,dy\,dx \end{aligned} \tag{6}$$

$$\lim_{T\to\infty} \sum_{i=1}^{T} \sum_{j=1}^{i} Pm(i)\,Pm(j)\,\frac{a}{T}(j-i) \qquad (7)$$

$$= \int_{x=0}^{1} \int_{y=0}^{x} Pm(x)\,Pm(y)\,a\,(x-y)\,dy\,dx$$

When we move to the continuous domain the clustering term vanishes because the probability of successive accesses to the same position becomes smaller as the number of positions increases.

Expressed mathematically:

$$\lim_{T\to\infty} \sum_{i=1}^{T} Pm(i)\,Pm(i)\,b \qquad (8)$$

$$\leq \lim_{T\to\infty} \sum_{i=1}^{T} (P_1 C(1)\frac{1}{T})^2 b = \lim_{T\to\infty} \frac{b}{T} P_1 C(1) = 0$$

And, the rotational delay term becomes:

$$\lim_{T\to\infty} \sum_{j=1}^{T} Pm(j)\,Rd(j) = \int_{x=0}^{1} Pm(x)\,Rd(x)dx \quad (9)$$

Thus, for $T \to \infty$, $\frac{Q}{T} \to 0$, the cost expression becomes:

$$Cost = \int_{x=0}^{1} \int_{y=x}^{1} Pm(x)\,Pm(y)\,a(y-x)\,dy\,dx \quad (10)$$

$$+ \int_{x=0}^{1} \int_{y=0}^{x} Pm(x)\,Pm(y)\,a\,(x-y)\,dy\,dx$$

$$+ b + \int_{x=0}^{1} Pm(x)\,Rd(x)dx$$

3. Proof of Consecutivity and Unimodality

We now prove some results about the form an optimal sector placement will take. Specifically, we show that a position on the continuous disk must contain only probability masses of one of the two subsets, either P_1 and P_2. And, we show that the form of any solution must be *unimodal*, meaning that in an optimal placement, the probability masses from the P_1 group must be placed as close together as possible.

Theorem 1 {Consecutivity}: In an optimal arrangement, there cannot exist two different positions x and y on the disk, such that there are two probability elements α_1 and α_3 at position x and two elements α_2 and α_4 at position y, such that $\alpha_1 < \alpha_2 < \alpha_3$.

Proof: Assume that there are two such positions. Consider the change in the expected retrieval cost if we exchange the α_2 mass at position y with the α_3 mass at position x. If the expected retrieval cost increases, any increase in probability at y with equivalent decrease in probability at x would also increase the expected retrieval cost. Thus, decreasing the probability at y and correspondingly increasing that at x will decrease the cost. Thus, the opposite operation of exchanging the α_1 mass element at position x with the α_2 mass element at position y would decrease the expected retrieval cost. This would violate our assumption that the initial arrangement was optimal.

The implication of Theorem 1 {Consecutivity} is that the sectors or (probability masses) that are assigned to a particular position/track on the disk must have a particular relationship. Namely, that their probability values (their probabilities of being accessed) must form a consecutive subsequence of all of the probability values are sorted in order. So, for instance, the track that contains the sector with the highest probability of being accessed should also contain the sector with the second highest probability of being accessed, etc. This result is essentially the same result as obtained in [Wong 80] [Wong 83] (called the greedy partition scheme) for CAV (magnetic) format disks. This proof however is much simpler.

We will show that the optimal placement is unimodal. This means that given two positions with probability elements of value P_1, there cannot exist a position between them with probability elements of value P_2.

Theorem 2 {Unimodality}: The optimal arrangement of two probability masses is unimodal.

Proof: Found in [Ford 91].

4. Analysis

We showed in the previous section that our optimal arrangement of probability masses must be unimodal (i.e., all the P_1 probability mass elements will be as close to each other as is possible). Thus, the form of an optimal solution in our model will be a specification of the position of the P_1 probability mass on the disk. We shall specify this position as the location on the disk of the mid-point between the left and right boundaries of the P_1 group. We denote this point as m ($m = \frac{Xl + Xr}{2}$). The relevant variables are illustrated in Figure 3. The goal of our analysis is then to develop an expression for m which can be computed from the parameters of the model.

Before tackling the main problem, we first derive some preliminary expressions. To compute Xl and Xr from m we need to know the "width" in relative disk units of the high probability group. We know that the proportion of the area occupied by the high

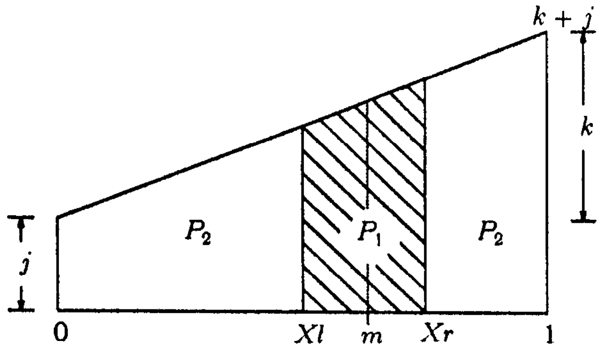

Figure 3: Optimal Placement Parameters

probability group must be equal to r.

If w is the width of the group, then from the formula for the area of a trapezoid, we have:

$$r = \frac{Area\ of\ P_1}{Total\ Area} = \frac{C(m)\,w}{1} = \frac{(k\,m\,+\,j)\,w}{1} \tag{11}$$

Therefore:

$$w = \frac{r}{k\,m\,+\,j} = \frac{r}{C(m)} \tag{12}$$

From this we derive the simple expressions for Xl and Xr.

$$Xl(r,m) = m - \tfrac{1}{2}\,\frac{r}{C(m)} \tag{13}$$

$$Xr(r,m) = m + \tfrac{1}{2}\,\frac{r}{C(m)}$$

Given these bounds, we can specify our probability assignment function:

$$Pm(x) = \begin{cases} \dfrac{P_1}{P_1 r + P_2(1-r)}\,C(x), & \text{if } Xl \le x \le Xr \\[2ex] \dfrac{P_2}{P_1 r + P_2(1-r)}\,C(x), & otherwise \end{cases} \tag{14}$$

From the model, our expected retrieval cost (Equation 10) is:

$$Cost = b + 2\int_{x=0}^{1}\int_{y=x}^{1} Pm(x)\,Pm(y)\,a\,(y-x)\,dy\,dx$$

$$+ \int_{x=0}^{1} Pm(x)Rd(x)\,dx$$

Where b is the intercept of the long seek cost, $Pm(x)$ is the probability of accessing location x on the disk, a is the slope of the long seek cost function, and $Rd(x)$ is the rotational delay cost at location x.

Recall that $W = (P_1 r + P_2(1-r))$, the seek component after substitution is:

$$= b + \frac{2P_2^2}{W^2}\int_{x=0}^{Xl(r,m)}\int_{y=x}^{Xl(r,m)} C(x)C(y)a(y-x)dy\,dx \tag{15}$$

$$+ \frac{2\,P_2 P_1}{W^2}\int_{x=0}^{Xl(r,m)}\int_{y=Xl(r,m)}^{Xr(r,m)} C(x)\,C(y)\,a\,(y-x)\,dy\,dx$$

$$+ \frac{2\,P_2^2}{W^2}\int_{x=0}^{Xl(r,m)}\int_{y=Xr(r,m)}^{1} C(x)\,C(y)\,a\,(y-x)\,dy\,dx$$

$$+ \frac{2\,P_1^2}{W^2}\int_{x=Xl(r,m)}^{Xr(r,m)}\int_{y=x}^{Xr(r,m)} C(x)\,C(y)\,a\,(y-x)\,dy\,dx$$

$$+ \frac{2\,P_1 P_2}{W^2}\int_{x=Xl(r,m)}^{Xr(r,m)}\int_{y=Xr(r,m)}^{1} C(x)\,C(y)\,a\,(y-x)\,dy\,dx$$

$$+ \frac{2\,P_2^2}{W^2}\int_{x=Xr(r,m)}^{1}\int_{y=x}^{1} C(x)\,C(y)\,a\,(y-x)\,dy\,dx$$

Where P_1 and P_2 are the weights of the two probability masses, r is the proportion which are P_1, m is the location of the centre of the P_1 mass, $Xr(r,m)$ and $Xl(r,m)$ are the right and left most extents of the P_1 mass, and $C(x)$ is the capacity of position x. Where h is the time to read the middle position of the disk, the rotational delay component is:

$$\tfrac{1}{2}h\int_{x=0}^{1} Pm(x)\,C(x)\,dx \tag{16}$$

$$= \frac{\tfrac{1}{2}h}{W}P_2\int_{x=0}^{Xl(r,m)} C(x)^2\,dx + \frac{\tfrac{1}{2}h}{W}P_1\int_{x=Xl(r,m)}^{Xr(r,m)} C(x)^2\,dx$$

$$+ \frac{\tfrac{1}{2}h}{W}P_2\int_{x=Xr(r,m)}^{1} C(x)^2\,dx$$

The expression for the expected random access retrieval cost has been evaluated using the Maple symbolic mathematics package [Char 85] and is given in [Ford 91]. For brevity, we show only a part of the resulting expression below:

$$Cost(k,j,a,b,P_1,P_2,r,h,m) = \tag{17}$$

$$\frac{(60P_2^2 j^7 h\ +\ \cdots -300P_2^2 j^6 khmr)}{120\,(km+j)^5\,(P_1 r+P_2(1-r))^2}$$

To derive the optimal value for m we take the derivative of the cost expression with respect to m, again using the Maple symbolic mathematics package. The complete derivative is in [Ford 91].

$$\frac{\partial\, Cost(k,j,a,b,P_1,P_2,r,h,m)}{\partial\, m} = \tag{18}$$

$$\frac{r\,(48aP_2^2 j^7 + \cdots + 48k^6 m^6 j a P_2^2)}{(km+j)^6 (P_1 r + P_2 - P_2 r)^2}$$

Because P_1 and P_2 are relative values, we select $P_2 = 1$ to simplify the derivative. Using Maple again, we determine the roots of the derivative and finally arrive at the desired expression for the optimal location of m (shown in its entirety):

$$m_{optimal} = -\frac{j}{k} + \frac{\left(\sqrt{3}\,C_1 + \sqrt{48 j^4 a^2 + C_2 k + C_3 k^2 + C_4 k^3 + C_5 k^4}\right)^{1/2}}{2\sqrt{2}\;3^{1/4} k \sqrt{a}} \tag{19}$$

$$C_1 = a\,(4j^2 + 2k^2 + 4jk) + hk^2(-P_1 r + r - 1)$$

$$C_2 = 96 j^3 a^2$$

$$C_3 = 32 P_1 r^2 a^2 - 24 j^2 P_1 a r h + 24 j^2 a r h - 24 j^2 a h + 96 j^2 a^2 - 48 r^2 a^2$$

$$C_4 = -24 j a h - 24 j a P_1 r h + 24 j a r h + 48 j a^2$$

$$C_5 = 12a^2 + 3P_1^2 r^2 h^2 - 6P_1 r^2 h^2 + 6 P_1 r h^2 - 12 a P_1 r h - 12 a h + 3h^2 + 12 a r h + 3 r^2 h^2 - 6 r h^2$$

5. Impact of Model Parameters on Optimal Data Placement

An expression for the optimal position of the group of the most frequently accessed sectors was derived in the previous section. In the discussion below, we examine and explain the variation in the optimal solution as the values of the parameters of the placement problem are changed.

Impact of Rotational Delay

The graph in Figure 4 plots the optimal position of m as a function of the rotational delay. The solid line in the graph plots m while the top and bottom lines plot the positions of the boundaries of the P_1 group, Xr and Xl. The graph shows that the rotational delay plays a significant role in determining the optimal position. As the retrieval delays due to rotational delay become more significant in relation to other sources of delay, the optimal position shifts towards the lower capacity tracks which have a lower rotational latency (and the spread between Xl and Xr increases).

To determine a typical value for P_1, the relative access frequency of the most frequently accessed disk sectors, we use the well known rule-of-thumb that states that 80% of all disk accesses will be directed to 20% of the disk sectors. This means that our value for r is 0.20 and that the area in the P_1 group in the distribution must be 80% of the total distribution area.

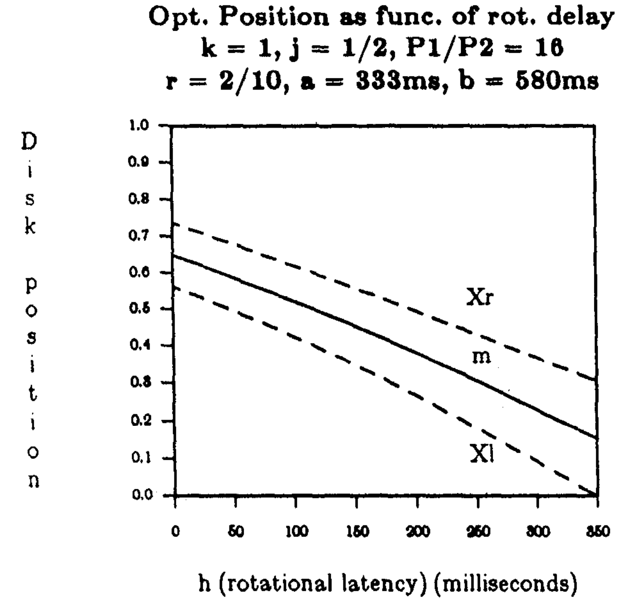

Figure 4: Optimal Position as function of rot. delay

$$\frac{r \cdot P_1}{r \cdot P_1 + (1 - r) \cdot P_2} = 0.8 \tag{20}$$

With $r = 0.2$ and $P_2 = 1$, we find $P_1 = 16$.

Impact of Seek Cost

The graph in Figure 5 plots the optimal position as a function of the seek cost parameter a (slope of seek cost). It shows that as the value of the seek cost function becomes more dependent upon distance (greater value for the slope a), the optimal position moves away from the inner edge of the disk. A similar behaviour would be observed, if for a constant slope a, the rotational delay would decrease (so that the seek cost would become more important). The limiting position as a increases depends upon the exact distribution of storage across the disk, and is essentially the location of the centre of gravity of the "probability mass".

The Impact of the Storage Distribution

The next parameters we examine are those which determine the distribution of storage capacity on the disk. For two different values of the rotational delay parameter h, the graph in Figure 6 plots the optimal position as a function of k, the relative slope of the capacity distribution function.

When the rotational delay is significant with respect to the slope of the seek cost function, the optimal position shifts towards the inner edge of the disk as the distribution becomes more skewed to take advantage of the reduced rotational delay at that position. When the rotational delay is insignificant with respect to the seek cost slope (e.g., $h = 10$ milliseconds), the optimal position shifts towards the

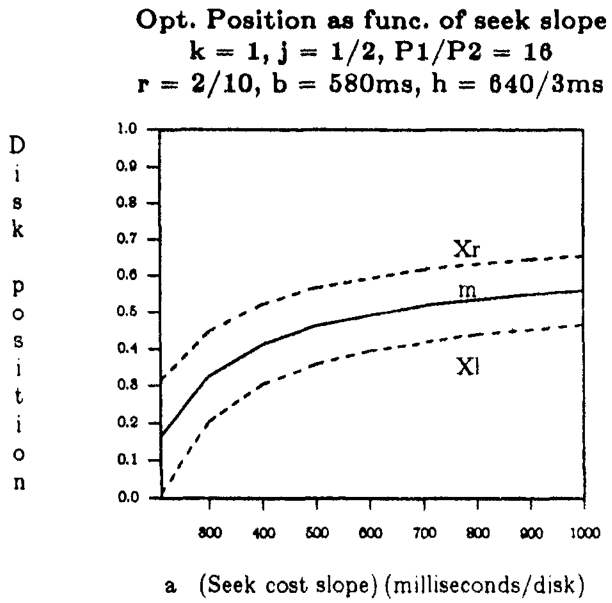

Figure 5: Optimal Position as function of seek slope

outer edge of the disk.

Note that for $k = 0$ (uniform distribution) the optimal position for both values of h are at the centre of the disk ($m = 0.5$) as we would expect since that is the optimal position for CAV format disks which have a uniform distribution of storage capacity [Grossman 73] [Yue and Wong 73] [Wong 80] [Wong 83].

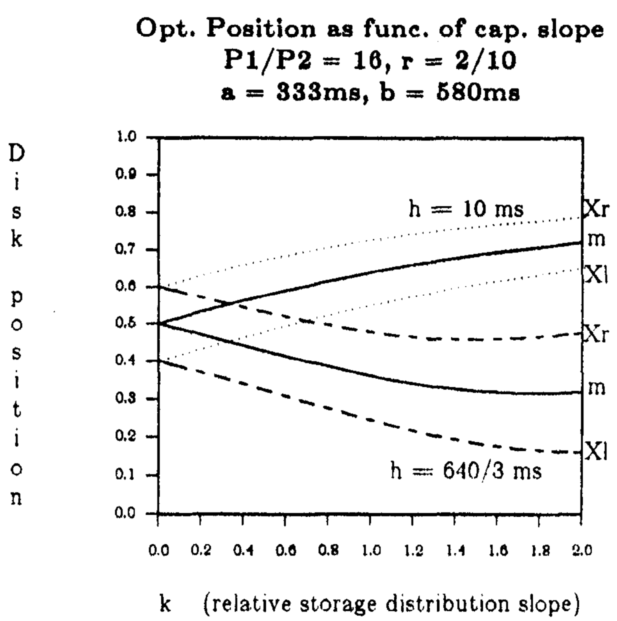

Figure 6: Optimal Position as func. of capacity slope

The Impact of the Data Access Probabilities

The access probability distribution also plays a role in determining the optimal solution. As the distribution becomes more skewed (P_1/P_2 increases, keeping r constant) the impact of the rotational delay component for the high probability masses (P_1) becomes the more significant component of the cost. As a result, the optimal location of the high probability mass moves towards the inner tracks to take advantage of the reduced rotational delay. The plot in Figure 7 shows how the optimal solution moves towards the lower capacity tracks as the P_1/P_2 ratio increases. The proportion r, will also affect the optimal position of the P_1 group. For brevity we omit a discussion of its impact here.

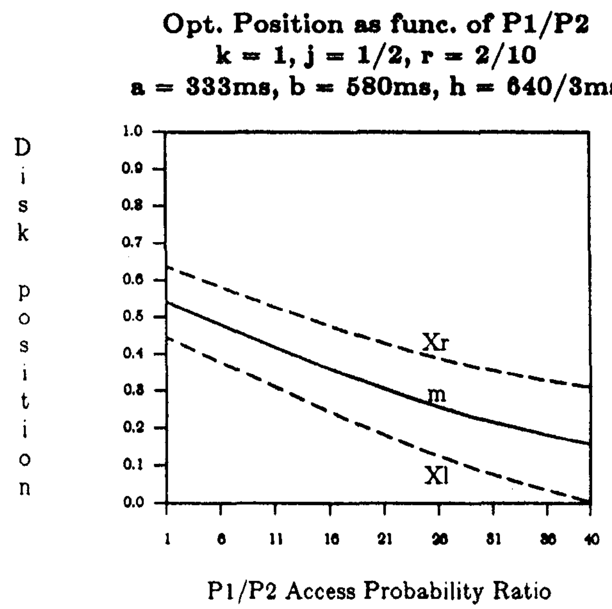

Figure 7: Optimal Position as func. of P1/P2

6. Sector Placement and Extensions to more general distributions

Having developed an optimal solution for our continuous two probability value model, we now describe its application to the discrete sector placement problem for general access probabilities.

Given a CLV format optical disk and a set of disk sectors with access probabilities not limited to just two values, our problem is to determine the track to which each sector should be assigned so that the total expected random access retrieval cost will be minimized. To determine this track assignment using our solution for the continuous model we approximate the general sector access probability distribution and the discrete disk using an equivalent two value probability mass model and a continuous disk. This approximation will divide the set of discrete sectors into two subsets, one corresponding to the P_1 probability mass and one to the P_2 mass. The dividing point (the value of r, the proportion of sectors which are represented by the P_1 mass) between the two subsets is chosen as described below to obtain a good placement. The values used for P_1 and P_2 will be the average access probability for the sectors in their corresponding subsets.

The optimal positions of the P_1 and P_2 probability masses in the continuous model will determine the approximately optimal placements of their corresponding discrete sector subsets. This placement can be refined by recursively applying the approximation again to the P_1 sector subset, treating the portion of the disk it occupies as a complete (but smaller) disk. This process of recursive subdivision can continue until the placement problem becomes trivial, such as when all sectors have the same access probability, or when the number of tracks on the disk is very small. An example of three levels of this recursive subdivison process are illustrated in Figure 8.

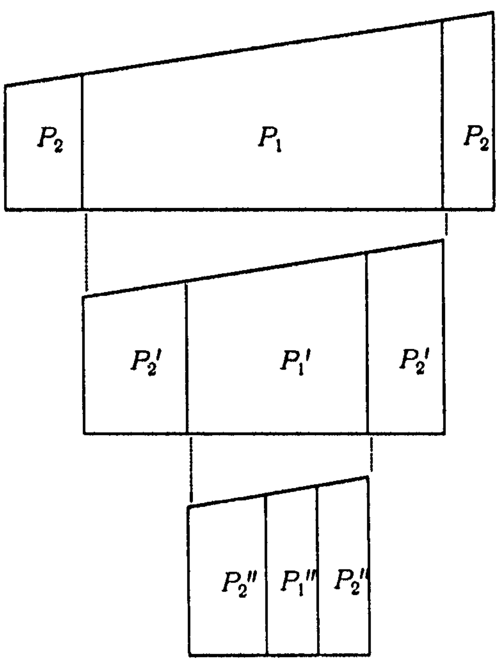

Figure 8: Recursive Subdivision

At the end of the subdivision process, the entire set of disk sectors will be divided into a series of subsets, each of which will be associated with a set of disk tracks. If the result of the subdivision is such that each sector subset corresponds to exactly one track, the placement problem is complete. It is more likely, however, that each sector subset will be associated with several tracks, which are also likely to be divided into two different groups, corresponding to the two P_2 masses produced at each level of the recursion. The problem then is to determine the exact assignment of each sector subset to its corresponding group of tracks on the disk. We employ heuristics to aid us in this task.

From theorems 1 {Consecutivity} and 2 {Unimodality}, it is known that an optimal sector arrangement must be unimodal, and that the set of sectors in each track must have access probabilities which are consecutive in the total distribution (i.e., if a track has three sectors with access probabilities 0.1, 0.2, and 0.4, then there cannot be another sector in a different track with access probability 0.3).

If there is only one group of adjacent tracks for a sector subset, these constraints are sufficient to determine a unique placement for the subset. If there are two groups of tracks, we use a heuristic that alternates the assignment of sectors from one group to the other. For the sectors corresponding to the P_1 mass at the lowest level of the recursion, the sectors are placed in alternate tracks from the outside into the centre of the group. For a disk with a uniform distribution of storage capacity these heuristics produce the Organ-Pipe permutation which alternates the assignment from one side of the middle track to the other, and is optimal.

We can use these heuristics at each level of the subdivision process to produce a placement. The expected access cost for the placement at each level will allow us to monitor the progress of the algorithm and can be used as a stopping criterion, for instance, stopping if the cost begins to increase. The exact details of the subdivision algorithm, which are not complicated, are given in [Ford 91], we omit them here for brevity.

7. Validation

We validate our model and analysis by comparing the predicted value for the optimal position of the frequently accessed data, with actual measurements from a CD ROM optical disk. We measured the average time for accesses on a Hitachi CD-1503S CD ROM drive from a file of 160000 disk sectors (320 megabytes) placed at the innermost position of the disk. Of the 160000 sectors, 38000 (24%) belonged to the group of frequently accessed sectors (P1). The relative access probability of the frequently accessed group was 16:1. Each measurement was obtained by first placing the group of frequently accessed sectors at a position on the disk and then measuring the time needed to complete each of 2500 accesses, the average of those times was then computed.

The sectors chosen for access were selected at random according to the relative access probabilities of the two sector groups. The graph in Figure 9 plots the average access time as a function of the position of the centre of the group of frequently accessed sectors.

We compute the optimal position of m, the centre of the group of frequently accessed sectors, using the equation derived previously in our analysis and the performance parameters of the disk drive. The optimal position of m is computed to be at relative

Measured cost as func. of P1 pos.

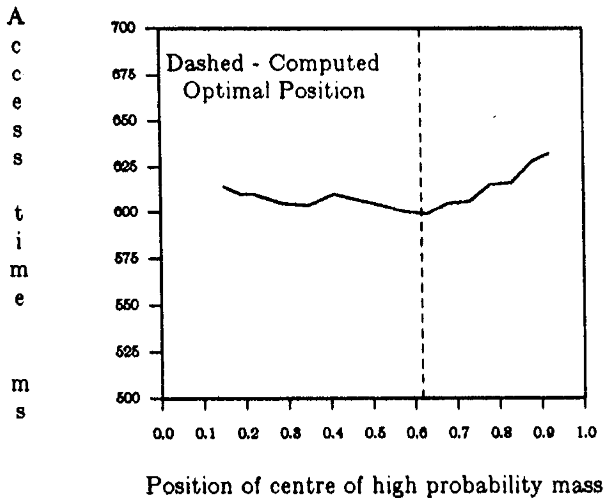

Figure 9: Measured cost for position of P1 group

position $m = 0.62$; this is indicated by the dashed line on the graph in Figure 9.

As we can see that m, the computed optimal position of the centre of the frequently access sector group, corresponds to the position of the group which had the lowest average random retrieval cost.

8. Summary

We have presented a model for studying the problem of optimal placement of data with known access probabilities on the recording surface of CLV optical disks. The model takes into account the non-uniform distribution of storage capacity on the disk and the dependency of the rotational delay on the track location, as well as a parameterized seek cost function. We have shown that the optimal placement satisfies a unimodality property for the placement of high probabilities, and we have derived an analytic solution to the optimal data placement problem. We have shown that the optimal data placement may be drastically different than the optimal data placement on magnetic disks.

The data access probabilities were described by a parametrized, two valued, probability distribution. This problem formulation allowed us to derive optimal locations for the high probability data items. Since in many real environments, precise knowledge about the access probabilities of data items may not be known, this problem formulation will be adequate for these environments. As a special case, indices are often considered to be frequently accessed items (in comparison to data values). In this context our results suggest an optimal position for indices given the device characteristics.

In environments where more detailed knowledge of the access frequencies of the data items may be available, our method can be extended using a recursive approximation of the access probabilities. We have outlined such an algorithm.

We have also validated our model and analysis against measurements made from CLV format optical disks and showed that the positions that they predicted as being the optimal locations for frequently accessed data, corresponded to the positions with the lowest measured average random retrieval cost.

9. References

[Laub 86]
Laub, L., "What is CD ROM", *The New Papyrus*, Microsoft Press, 1986, pp. 47-71

[Mckusick 84]
Mckusick, M.K., Joy, W.N., Leffler, S.J., Fabry, R.S., "A Fast File System for UNIX", *ACM Transactions on Computer Systems*, Vol. 2, No. 3, (August 1984), 181-197.

[Grossman 73]
Grossman, D.D., Silverman, H.F., "Placement of Records on a Secondary Storage Device to Minimize Access Time", *Journal of the Association for Computing Machinery*, Vol. 20, No. 3, July 1973, pp. 429-438.

[Yue and Wong 73]
Yue, P.C., Wong, C.K., "On the Optimality of the Probability Ranking Scheme in Storage Applications", *Journal of the Association for Computing Machinery*, Vol. 20, No. 4, October 1973, pp. 624-633.

[Wong 80]
Wong, C.K., "Minimizing Expected Head Movement in One-Dimensional and Two-Dimensional Mass Storage Systems", *Computing Surveys*, Vol. 12, No. 2, June 1980, pp. 167-178.

[Wong 83]
Wong, C.K., *Algorithmic Studies in Mass Storage Systems*, Springer-Verlag, Berlin, 1983

[Hardy 34]
Hardy, G.H., Littlewood, J.E., Pólya, G., *Inequalities*, Cambridge University Press, Cambridge, England, 1934.

[Ford 91]
Ford, D.A., *Performance Optimizations for Optical Disc Architectures*, Ph.D. Thesis, *Technical Report CS-91-08*, Department of Computer Science, University of Waterloo, March, 1991.

[Char 85]
Char, B.W., Geddes, K.O., Gonnet, G.H., Watt, S.M., *MAPLE Reference Manual*, 4'th edition, University of Waterloo, Waterloo, Ontario, June 1985.