On Estimating the Cost of Accessing Records in Blocked Database Organizations‡

Arvola Chan and Bahram Niamir
Laboratory for Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA†

The estimation of the cost of processing a query using a particular access path under a given physical organization has important applications in integrated database environments. When records in a file are stored in fixed-length physical blocks in secondary storage, and mechanisms are available whereby a query can be resolved without the accessing of all of the records, an important measure of the cost of using a particular access path is the number of blocks that have to be accessed in referencing the records of interest. In this paper, a general formula is derived for the expected number of blocks on which a random sample of r records from a file containing n records (which may be of arbitrary lengths, and which may extend across block boundaries) will reside. The specialization of this formula to the case of fixed-length records is discussed. An approximation to this formula which is highly accurate for a wide range of parameters and which can be computed very efficiently is also provided.

INTRODUCTION

The estimation of the cost of processing a query can serve two important purposes in a database system. Firstly, when the physical organization of the database is to be designed or tuned, it is essential that the expected cost of processing a given query under any proposed organization can be evaluated. Secondly, the provision of a query cost estimation facility can aid casual users in determining if the cost of processing a query is commensurate with the value of the information the query is expected to provide.

For present-day large databases that cannot fit into primary memory, the records in a physical file are usually grouped into fixed-length physical blocks, which constitute the units of transfer between primary and secondary memory. Under such an organization, the major cost component to the processing of a query is the number of blocks that have to be accessed in retrieving the set of records that are of interest. (Auxiliary access structures (e.g. secondary indices) are often maintained to directly identify the records of interest.) In this paper, we consider the problem of estimating the number of block accesses to retrieve the necessary records to resolve a query, given that r distinct records have to be accessed. We assume that the addresses of these r records are available and that we have direct access to blocks on which these r records reside.

Our analysis of block retrievals is based on the following model of the file and the query made to a file. (1) Each record has equal probability of being selected; any subset of r records in the file are equally likely to be selected. This is not an unreasonable assumption for a file that is not sorted on some field(s). Even if the file is sorted or clustered based on values in some field(s), the set of records which satisfy a query which does not involve the sorted field(s) will be effectively unclustered. Since it is common to have records clustered on a primary key field, the file is effectively unclustered for queries which involve only secondary key fields. (For an analysis on estimating access costs in sorted files, see Ref. 1.) (2) A record is not retrieved more than once. Once a record is selected from a block, the probability that another record is selected from the same block decreases, i.e. selection is without replacement. (3) A block is not accessed more than once. Once a block is accessed, all records selected from this block are accessible and hence it is not necessary to access the block again.

Previous studies[2–10] have considered such an estimation problem. However, they have all assumed that each block contains an integral number of records. When records are long, though, it may be desirable to split a record across block boundaries. (This is in fact done in practice.[11]) The purpose of this paper is to derive a general block access formula that is valid even when records (possibly of unequal lengths) may cross block boundaries and when the number of blocks may not be integral, and to report an approximation to the derived formula for the case of fixed length records which is highly accurate for a wide range of parameters and which can be computed very efficiently. (When counting the number of records in a block, we will call a record which is not completely resident in the block and overlaps onto the next block a partial record. A record that is completely resident in the block is called a total record.)

For the rest of this paper, we will use the following notations. Let

n = number of records in the file
b_i, $i = 1, \ldots, m$ = number of records that totally or partially reside in block i
b = number of records that totally or partially reside in a block when all blocks contain the same number of total and partial records
r = number of records to be retrieved

† This research was supported by the Advanced Research Projects Agency of the Department of Defence under ARPA order No. N00014-76-C0944.

‡ Authors' present addresses: Arvola Chan, Computer Corporation of America, 575, Technology Square, Cambridge, MA 02139; Bahram Niamir, Laboratory for Computer Science, Massachusetts, Institute of Technology, 545, Technology Square, Cambridge, MA 02139.

CCC-0010-4620/82/0025-0368$03.50

$A(n, \{b_i\}_{i=1}^m, r)$ = expected number of block accesses to retrieve r records from a file of n records where records are stored in blocks each with b_i total or partial records

$A(n, b, r)$ = expected number of block accesses to retrieve r records in a file of n records with each block containing b records

C_k^j = number of combinations of j objects taken k at a time

RELATED RESULTS

For the case where all blocks contain the same integral number of records, the computation of $A(n, b, r)$ has been considered by various authors.[2-10] Cardenas,[2] Rothnie and Lozano,[4] and Waters[7,8] have suggested the use of an approximation that is a lower bound to $A(n, b, r)$. Schmid,[5] and Siler[6] have derived similar formulations that involve complicated recurrence relations, whose computation for moderate values of the parameters is very costly and also inaccurate because of significant round-off errors. Using a probabilistic argument, Yao,[9] Waters,[8] and Pezarro[3] have independently obtained the following closed form expression for the block access formula

$$A(n, b, r) = \frac{n}{b}\left[1 - \frac{C_r^{n-b}}{C_r^n}\right] \tag{1}$$

whose computation cost is of order $\min(r, b)$. In addition, Yue and Wong[10] have derived the following recurrence relation (with computation cost of order r) for the block access function

$$A(n, b, 0) = 0$$

$$A(n, b, r+1) = \frac{n-r-t}{n-r} A(n, b, r) + \frac{n}{n-r}$$

We have also derived the block access function (1) by solving the above recurrence relation.[12,13]

THE GENERAL EXPECTED BLOCK ACCESS FORMULA

In this section, we derive the general block access formula and then consider its specialization to the case of fixed-length records.

Proposition 1. Consider a file with n records stored on m fixed-length blocks. The records may be of unequal lengths and may extend across block boundaries. If b_i is the blocking factor of the ith block (i.e. the number of records that totally or partially reside in the block), then the expected number of blocks in which a random sample of r records reside is

$$A(n, \{b_i\}_{i=1}^m, r) = \sum_{i=1}^{m}\left[1 - \frac{C_{b_i}^{n-r}}{C_{b_i}^n}\right]$$

Proof. The ith block in the file contains b_i (total or partial) records and will have to be accessed unless none of these b_i records is among those r records being accessed. The probability that the first record accessed does not come from the ith block is $(n - b_i)/n$. The probability that the second record accessed does not come from the ith block given that the first record accessed does not come from the ith block is

$$\left[\frac{n-b_i}{n}\right]\left[\frac{n-b_i-1}{n-1}\right]$$

In general the probability that none of the r records is selected from the ith block is

$$\prod_{j=0}^{r-1}\left[\frac{n-b_i-j}{n-j}\right] = \frac{C_r^{n-b_i}}{C_r^n} = \frac{(n-r)!\,(n-b_i)!}{n!\,(n-r-b_i)!} = \frac{C_{b_i}^{n-r}}{C_{b_i}^n}$$

The probability that any of the r records is selected from the ith block is $1 - (C_{b_i}^{n-r}/C_{b_i}^n)$. Hence, the expected number of blocks accessed is the sum of the probability of each being accessed

$$\sum_{i=1}^{m}\left[1 - \frac{C_{b_i}^{n-r}}{C_{b_i}^n}\right]$$

In the rest of this section, we will specialize the block access expression to cases where the records of the file are of the same length and obtain formulas that are more applicable for block access estimation. The most general of these cases is where records overlap block boundaries and the last block in the file is not full. See Fig. 1 for an example of such a file. Since b is not integral, all blocks containing fragments of a record that spans more than one block need to be accessed when such a record is to be retrieved.

Let S be the block size and L be the record length. Also let b, the blocking factor, be a rational number

$$b = \frac{S}{L} = k + \frac{p}{q}$$

where $k = \lfloor S/L \rfloor$
$p = \text{mod}(S, L)/\gcd(S, L)$
$q = L/\gcd(S, L)$
p and q are relatively prime

There are $\lfloor n/b \rfloor$ blocks that are completely full and one block that is partially full in the file. Call a record that resides fully in a block a total record, and a record that crosses a block boundary a partial record. It is easy to see that, except for the last block, the number of records that totally or partially reside in each block is either $\lfloor b \rfloor$ or $\lfloor b \rfloor + 1$. We will call blocks that contain $\lfloor b \rfloor$ total and partial records *type I* blocks, and blocks that contain $\lfloor b \rfloor + 1$ total and partial records *type II* blocks. The last block is only partially full and has less than or equal to $\lfloor b \rfloor$ records. We will call it a *type III* block. Furthermore, we will call each q blocks which constitute the smallest number of consecutive blocks that contain an integral number of consecutive records, a *complete cycle*. (Thus, a file consists of zero or more complete cycles and possibly one incomplete cycle. See Fig. 1.) Lemmas 1, 2, and 3 below, whose proofs are provided in the Appendix, provide expressions for computing the number of type I and type II blocks in a file, as well as the number of records that reside in the type III block.

Lemma 1. If b is nonintegral ($p \geq 1$, $q \geq 2$), then there are $q - p + 1$ type I blocks and $p - 1$ type II blocks in a complete cycle. If b is integral ($p = 0$, $q = 1$), then there is one type I block and no type II block in a complete cycle.

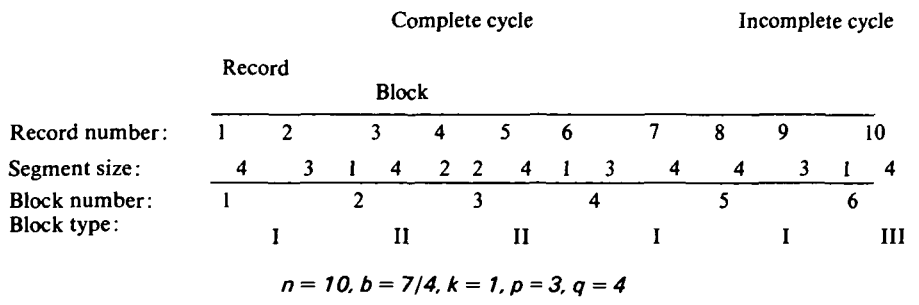

Figure 1. A file with six blocks.

Lemma 2. In an incomplete cycle with a integral blocks (of types I and II), there are $a - \lfloor ap/q \rfloor$ type I blocks and $\lfloor ap/q \rfloor$ type II blocks.

Lemma 3. The number of records that reside in the type III block in a file is $n - \lfloor b\lfloor n/b \rfloor \rfloor$.

With the above lemmas, we can specialize the general expected block access formula (Proposition 1) to cases where records are of fixed length.

Proposition 2. The expected number of block accesses in referencing r records in a file of n records that are stored in blocks of capacity of b records, where b is nonintegral and n is not an integral multiple of b is

$$A(n, b, r) = \left(\left\lfloor \frac{n}{b} \right\rfloor - \alpha\right)\left[1 - \frac{C^{n-r}_{\lfloor b \rfloor}}{C^{n}_{\lfloor b \rfloor}}\right] + \alpha\left[1 - \frac{C^{n-r}_{\lfloor b \rfloor + 1}}{C^{n}_{\lfloor b \rfloor + 1}}\right] + \left[1 - \frac{C^{n-r}_{b'}}{C^{n}_{b'}}\right] \quad (2)$$

where $b = k + p/q$, k integer, $1 \le p < q$, $\gcd(p, q) = 1$

$$\alpha = \left\lfloor \frac{p\lfloor n/b \rfloor}{q} \right\rfloor = \left\lfloor \frac{n}{bq} \right\rfloor$$

$$b' = n - \left\lfloor b\left\lfloor \frac{n}{b} \right\rfloor \right\rfloor$$

Proof. Since each complete cycle is q blocks, there are $\lfloor \lfloor n/b \rfloor / q \rfloor = \lfloor n/bq \rfloor$ complete cycles in the file and $a = \lfloor n/b \rfloor - q\lfloor n/bq \rfloor$ integral blocks in the last incomplete cycle. See problem 1.2.4.35 in Ref. 14 for a proof of the identity $\lfloor \lfloor x \rfloor / n \rfloor = \lfloor x/n \rfloor$ for positive integer n. Hence α, the number of type II blocks in the file is

α = (Number of cycles) × (Number of type II blocks in cycle) + (Number of type II blocks in the last incomplete cycle),

$$= \left\lfloor \frac{n}{bq} \right\rfloor (p - 1) + \left\lfloor \frac{p(\lfloor n/b \rfloor - q\lfloor n/bq \rfloor)}{q} \right\rfloor$$

by Lemmas 1 and 2,

$$= \left\lfloor \frac{p\lfloor n/b \rfloor}{q} \right\rfloor \quad \left\lfloor \frac{n}{bq} \right\rfloor$$

By Lemma 3 the number of records that reside in the type III block is

$$b' = n - \left\lfloor b\left\lfloor \frac{n}{b} \right\rfloor \right\rfloor$$

Corollary 1. The expected number of block accesses in referencing r records in a file of n records that are stored in blocks of capacity of b records, where b is nonintegral and n is an integral multiple of b is

$$\frac{n}{b}\left[\frac{q - p + 1}{q}\left[1 - \frac{C^{n-r}_{\lfloor b \rfloor}}{C^{n}_{\lfloor b \rfloor}}\right] + \frac{p - 1}{q}\left[1 - \frac{C^{n-r}_{\lfloor b \rfloor + 1}}{C^{n}_{\lfloor b \rfloor + 1}}\right]\right]$$

Proof. $n/b = nq/(kq + p)$ is an integer. But q and $kq + p$ are relatively prime (since q and p are relatively prime); therefore n is divisible by $kq + p$ or by bq. Therefore in Eqn (2)

$$\alpha = \left\lfloor \frac{p\lfloor n/b \rfloor}{q} \right\rfloor - \left\lfloor \frac{n}{bq} \right\rfloor = \left\lfloor p\frac{n}{bq} \right\rfloor - \frac{n}{bq} = \frac{n}{b}\left\lfloor \frac{p - 1}{q} \right\rfloor$$

and

$$b' = n - \left\lfloor b\left\lfloor \frac{n}{b} \right\rfloor \right\rfloor = n - \left\lfloor b\left(\frac{n}{b}\right) \right\rfloor = 0$$

Corollary 2. The expected number of block accesses in referencing r records in a file of n records that are stored in blocks of capacity b records, where b is integral and n is not an integral multiple of b is

$$\left\lfloor \frac{n}{b} \right\rfloor \left[1 - \frac{C^{n-r}_{b}}{C^{n}_{b}}\right] + \left[1 - \frac{C^{n-r}_{b'}}{C^{n}_{b'}}\right]$$

where $b' = \mathrm{mod}(n, b)$.

Proof. This follows from the facts that there are $\lfloor n/b \rfloor$ cycles; and by Lemma 1 each cycle consists of a type I block and no type II blocks ($\alpha = 0$). The incomplete block consists of one type III block with b' records where

$$b' = n - \left\lfloor b\left\lfloor \frac{n}{b} \right\rfloor \right\rfloor = n - b\left\lfloor \frac{n}{b} \right\rfloor = \mathrm{mod}(n, b)$$

Corollary 3. The expected number of block accesses in referencing r records in a file of n records that are stored in blocks of capacity of b records, where b is integral and n is an integral multiple of b is

$$\frac{n}{b}\left[1 - \frac{C^{n-r}_{b}}{C^{n}_{b}}\right]$$

Proof. This follows from Corollary 2 where $\lfloor n/b \rfloor = n/b$, and

$$b' = n - b\left\lfloor \frac{n}{b} \right\rfloor = 0$$

COMPUTATION OF $A(n, b, r)$

When estimating the number of block access made to a file by $A(n, b, r)$, we need to evaluate the expression C_b^{n-r}/C_b^n which is the probability that any particular block is not hit.

$$\frac{C_b^{n-r}}{C_b^n} = \frac{(n-r)!\,(n-b)!}{n!\,(n-r-b)!} = \prod_{i=0}^{r-1}\left[1 - \frac{b}{n-i}\right] \tag{3}$$

Expression (3) is symmetric in r and b, and therefore requires on the order of $\min(r, b)$ multiplications and divisions for its computation. This may be too costly for practical purposes.

One approximation to Expression (3) (used in Refs 2, 4, 7, 15) is

$$\left(1 - \frac{b}{n}\right)^r \tag{4}$$

We observe that this approximation is good when the number of records selected is a small fraction of the records in the file. If $n \gg r$, then $n - i|_{0 \le i \le r-1} \cong n$, and $\prod_{i=0}^{r-1}[1 - b/n - i] \cong (1 - b/n)^r$. In fact, if we assume $n \to \infty$ and $b \to \infty$ such that b/n approaches a finite number, then

$$\lim_{n,\, b\to\infty} \frac{C_b^{n-r}}{C_b^n} = \lim_{n,\, b\to\infty} \prod_{i=0}^{r-1}\left[1 - \frac{b/n}{1 - i/n}\right] = \left(1 - \frac{b}{n}\right)^r$$

Intuitively, when $n \gg r$, the probability that a block gets hit (b/n) is the same whether records that were previously hit are replaced or not, and hence the probability that any particular block not being hit by any of the r records is $(1 - b/n)^r$. Note that Expression (4) is an upper bound to Expression (3) since selection is with replacement.

Another approximation to Expression (3) used by Waters[8] is

$$\left(1 - \frac{r}{n}\right)^b \tag{5}$$

This approximation is good if $n \gg b$. Again, the argument follows along the same lines as the previous approximation. The probability that a record gets hit is r/n, and the probability that any group of b records (a block) not getting hit is $(1 - r/n)^b$. Again note that in the above approximation, the probability that a record gets hit is obtained by assuming previous hit records are replaced. Hence Expression (5) is also an upper bound to Expression (3).

Morris[16] and Severance and Duhne[15] use $\exp(-rb/n)$ to approximate Expression (3). To see this we note

$$\log\left(1 - \frac{b}{n}\right) = -\frac{b}{n} - \frac{b^2}{2n^2} - \frac{b^3}{3n^3} - \cdots \quad \left|\frac{b}{n}\right| < 1,$$

and

$$\left(1 - \frac{b}{n}\right)^r = \exp\left(-\frac{rb}{n} - \frac{rb^2}{2n^2} - \cdots\right), \quad \left|\frac{b}{n}\right| < 1.$$

If $n \to \infty$, and $r \to \infty$ such that r/n approaches a finite number, then

$$\lim_{n,\, r\to\infty}\left(1 - \frac{b}{n}\right)^r = \exp\left(-\frac{rb}{n}\right) \tag{6}$$

We also see that Eqn (6) is an upper bound to Expression (4) and hence an upper bound to Eqn (3).

Table 1 compares the three approximations (4), (5) and (6) to C_b^{n-r}/C_b^n.

All of the above approximations are special case approximations in that their accuracy depends on the existence of certain relationships among n, b and r. However, it would be useful to have an approximation that is valid for all values of n, b and r. The Stirling approximation to the factorial function provides us with such an approximation

$$m! \sim (2\pi)^{1/2}\, m^{m+1/2} \exp(-m + 1/12m)$$

$$\frac{C_b^{n-r}}{C_b^n} \cong \exp\Big[(n - r + \tfrac{1}{2})\log(n-r) + (n - b + \tfrac{1}{2})\log(n-b) - (n + \tfrac{1}{2})\log n - (n - r - b + \tfrac{1}{2})\log(n - r - b) + \frac{1}{12}\left(\frac{1}{n-r} + \frac{1}{n-b} - \frac{1}{n} - \frac{1}{n-r-b}\right)\Big] \tag{7}$$

Another approximation that we have found useful is to approximate the b factor product form of Eqn (3) by setting i to the average of the values over which it ranges.

$$\frac{C_b^{n-r}}{C_b^n} \cong \left[1 - \frac{r}{n - (b-1)/2}\right]^b \tag{8}$$

This approximation is almost as accurate as the one using Stirling's approximation and is much easier to compute.

Table 1 compares this approximation to all of the previous approximations. Column 4 gives the ratio of C_b^{n-r} to C_b^n. Each of the remaining columns is the difference between C_b^{n-r}/C_b^n and the indicated approximation, divided by C_b^{n-r}/C_b^n. Considering the accuracy (8) and its ease of calculation (which is a constant number of multiplications when put in the exponentiation-logarithmic form), we suggest using Eqn (8) for the computation of C_b^{n-r}/C_b^n.

Table 1

n	b	r	C_b^{-r}/C_b^r	Ref. 11	Ref. 14	Ref. 16	Ref. 13	Ref. 3
10	2	2	0.622	−0.0285	−0.0285	−0.0773	4.46e−6	−0.00168
10	2	4	0.333	−0.228	−0.0800	−0.347	2.27e−5	−0.00554
10	5	2	0.222	−0.125	−0.474	−0.655	7.75e−5	−0.0678
10	5	4	0.0238	−1.62	−2.26	−4.68	0.00223	−0.312
100	2	2	0.960	−2.20e−4	−2.20e−4	−6.61e−4	8.80e−4	−1.10e−6
100	2	10	0.809	−0.00986	−0.00112	−0.0119	5.54e−6	−5.59e−6
100	2	50	0.247	−0.471	−0.0102	−0.486	4.42e−6	−7.76e−5
100	5	2	0.902	−5.53e−4	−0.00210	−0.00312	−1.14e−6	−2.21e−5
100	5	10	0.583	−0.0256	−0.0115	−0.0390	−2.25e−6	−1.12e−4
100	5	50	0.0281	−1.73	−0.110	−1.91	−9.92e−7	−0.00165
100	10	2	0.809	−0.00112	−0.00986	−0.0119	5.54e−6	−1.19e−4
100	10	10	0.330	−0.0550	−0.0550	−0.113	3.31e−6	−0.00112
100	10	50	5.59e−4	−7.68	−0.645	−10.35	4.42e−6	−0.0155
100	20	2	0.638	−0.00253	−0.0457	−0.0500	4.47e−6	−0.00187
100	20	10	0.0951	−0.128	−0.278	−0.422	1.10e−6	−0.0108
100	20	50	8.87e−8	−161.31	−9.84	−515.31	3.32e−6	−0.180
1000	2	2	0.996	−2.20e−6	−2.20e−6	−6.60e−6	2.22e−5	0.0
1000	2	10	0.980	−9.90e−5	−1.10e−5	−1.11e−4	−1.15e−4	0.0
1000	2	50	0.902	−0.00254	−5.52e−5	−0.00264	2.21e−5	−2.29e−8
1000	2	250	0.562	−0.0780	−3.33e−4	−0.0786	−7.75e−5	−2.22e−7
1000	10	2	0.980	−1.10e−5	−9.90e−5	−1.11e−4	−1.15e−4	−1.17e−7
1000	10	10	0.903	−4.45e−4	−4.45e−4	−9.96e−4	−6.69e−5	−9.92e−7
1000	10	50	0.597	−0.0128	−0.00238	−0.0154	−1.11e−4	−4.46e−6
1000	10	250	0.0554	−0.461	−0.0152	−0.479	−1.14e−4	−3.32e−5
1000	50	2	0.902	−5.52e−5	−0.00254	−0.00264	2.21e−5	−2.22e−5
1000	50	10	0.597	−0.00238	−0.0128	−0.0154	−1.11e−4	−1.11e−4
1000	50	50	0.0719	−0.0690	−0.0690	−0.140	4.47e−7	−6.60e−4
1000	50	250	3.37e−7	−6.28	−0.529	−9.06	−3.35e−5	−0.00443
1000	100	2	0.809	−1.11e−4	−0.0106	−0.0108	4.44e−5	−1.19e−4
1000	100	10	0.346	−0.00504	−0.0550	−0.0603	−1.10e−4	−9.98e−4
1000	100	50	0.00447	−0.151	−0.322	−0.505	−1.13e−5	−0.00528
1000	100	250	5.53e−14	−66.90	−4.98	−258.37	−2.24e−5	−0.0396
10000	2	2	0.999	−2.29e−8	−2.29e−8	−5.59e−8	5.57e−4	0.0
10000	2	10	0.998	−9.90e−7	−1.10e−7	−1.11e−6	9.92e−4	−1.14e−8
10000	2	50	0.990	−2.24e−5	−5.50e−7	−2.25e−5	7.71e−4	−1.14e−8
10000	2	250	0.950	−6.63e−4	−2.25e−6	−6.63e−4	1.14e−4	−1.14e−8
10000	2	1250	0.765	−0.0171	−1.14e−5	−0.0172	5.50e−4	−1.14e−8
10000	5	2	0.999	−5.59e−8	−2.20e−7	−2.29e−7	−2.23e−5	−5.59e−8
10000	5	10	0.995	−2.22e−6	−1.10e−6	−3.35e−6	−1.12e−4	−4.44e−8
10000	5	50	0.975	−6.61e−5	−5.50e−6	−6.67e−5	−6.65e−4	0.0
10000	5	250	0.881	−0.00158	−2.25e−5	−0.00161	−6.63e−4	−2.29e−8
10000	5	1250	0.512	−0.0435	−1.14e−4	−0.0437	−8.80e−4	−4.44e−8
10000	10	2	0.998	−1.10e−7	−9.90e−7	−1.11e−6	9.92e−4	−1.14e−8
10000	10	10	0.990	−4.45e−6	−4.45e−6	−9.85e−6	−2.24e−4	−7.74e−8
10000	10	50	0.951	−1.12e−4	−2.22e−5	−1.14e−4	−3.34e−4	−5.59e−8
10000	10	250	0.776	−0.00317	−1.11e−4	−0.00329	6.61e−4	−5.59e−8
10000	10	1250	0.262	−0.0890	−6.64e−4	−0.0897	−1.19e−5	−1.17e−7
10000	50	2	0.990	−5.50e−7	−2.24e−5	−2.25e−5	7.71e−4	−2.28e−7
10000	50	10	0.951	−2.22e−5	−1.12e−4	−1.14e−4	−3.34e−4	−4.40e−7
10000	50	50	0.777	−6.61e−4	−6.61e−4	−0.00124	−0.00124	−6.52e−7
10000	50	250	0.281	−0.0160	−0.00315	−0.0192	−4.48e−4	−2.29e−6
10000	50	1250	0.00123	−0.534	−0.0177	−0.559	2.20e−4	−1.16e−5
10000	100	2	0.980	−1.10e−6	−9.99e−5	−1.10e−4	−5.57e−4	−4.49e−7
10000	100	10	0.904	−4.45e−5	−4.49e−4	−5.54e−4	3.37e−5	−1.15e−6
10000	100	50	0.604	−0.00124	−0.00250	−0.00376	−0.00180	−4.49e−6
10000	100	250	0.0785	−0.0324	−0.0128	−0.0455	−6.61e−4	−2.22e−5
10000	100	1250	1.14e−6	−1.36	−0.0738	−1.52	4.44e−4	−1.13e−4
10000	500	2	0.902	−5.52e−6	−0.00258	−0.00259	9.92e−4	−2.27e−5
10000	500	10	0.598	−2.23e−4	−0.0130	−0.0132	−4.47e−4	−1.11e−4
10000	500	50	0.0764	−0.00649	−0.0670	−0.0737	−8.82e−4	−5.56e−4
10000	500	250	2.22e−6	−0.181	−0.392	−0.632	−7.73e−4	−0.00292
10000	500	1250	1.15e−30	−88.79	−5.34	−451.18	−8.81e−4	−0.0174

REFERENCES

1. B. T. Berkowitz, Cost models for the datacomputer, SB thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, (1978).
2. A. F. Cardenas, Analysis and performance of inverted database structures. *Communications of the ACM* **18** (No. 5) (May 1975).
3. M. T. Pezarro, A note on estimating hit ratios for direct-access storage devices. *Computer Journal* **19** (No. 3), 271–272, (1976).
4. J. B. Rothnie and T. Lozano, Attribute based file organization in a paged memory environment. *Communications of the ACM* **17** (No. 2) (February 1974).
5. H. A. Schmid and P. A. Bernstein, A multi-level architecture for relational database systems, *Proceedings of the International Conference on Very Large Data Bases* (September 1975).
6. K. F. Siler, A stochastic evaluation model for database organizations in data retrieval systems. *Communications of the ACM* **19** (No. 2) (February 1976).
7. S. J. Waters, Estimating magnetic disc seeks. *Computer Journal* **–8** (No. 1) (February 1975).
8. S. J. Waters, Hit ratios, *Computer Journal* **19** (No. 1) (February 1976).
9. S. B. Yao, Approximating block accesses in database organizations. *Communications of the ACM* **20** (No. 4) (April 1977).
10. P. C. Yue and C. K. Wong, Storage cost considerations in secondary index selection. *International Journal of Computer and Information Science* **4** (No. 4) (1975).
11. Datacomputer Version 1 User Manual, Computer Corporation of America, 575 Technology Square, Cambridge, Massachusetts 02139 (1975).
12. A. Y. Chan, Index selection in a self-adaptive relational database system, TR-166, Laboratory for Computer Science, Massachusetts Institute of Technology (September 1976).
13. B. Niamir, Attribute partitioning in a self-adaptive relation database system, TR-192, Laboratory for Computer Science, Massachusetts Institute of Technology (1978).
14. D. E. Knuth, *Fundamental Algorithms, The Art of Computer Programming*, Vol. 1, 2nd Edn. Addison-Wesley, Reading, Massachusetts (1973).
15. D. Severance and R. Duhne, A practitioner's guide to addressing algorithms. *Communications of the ACM* **19** (No. 6) (June 1976).
16. R. Morris, Scatter storage techniques. *Communications of the ACM* **11** (No. 1) (January 1968).

Received September 1981

APPENDIX

Proof of lemma 1

Let the size of a block be S bytes and let the length of a record be L bytes. We will define a segment to be a unit of storage equal to gcd(S, L) bytes. A record consists of $L/\gcd(S, L) = q$ segments and a block consists of $S/\gcd(S, L) = kq + p$ segments. Before we prove this lemma, we will first show that each of the q blocks in a complete cycle begins with a partial record (i.e. a record that is spilled over from the previous consecutive block) which is 0, or 1, . . ., or $q - 1$ segments long. Furthermore, no two blocks in the same cycle begin with partial records of the same length. Assume that this is not true and there exists blocks i and j ($i \neq j$) such that they both begin with a partial record of s ($0 \leq s \leq q - 1$) segments, and both are in the same cycle. Consider blocks $i + \lambda$ and $j + \lambda$, where λ = an integer such that $\lambda p + s$ is divisible by q. Blocks $i + \lambda$ and $j + \lambda$ both begin with a partial record of $\mod(\lambda(kq + p) + s, q) = 0$ segments, i.e. they both begin with a total record. They are also less than a cycle apart since blocks i and j were within the same cycle. But a cycle is by definition the number of blocks between two such blocks. Hence blocks $i + \lambda$ and $j + \lambda$ cannot be less than a cycle apart, a contradiction. Thus for each $s = 0$, . . ., $q - 1$, there is one and only one block in the cycle which begins with a partial record of s segments.

For a block to contain $[b] + 1$ (total or partial) records (i.e. for the block to be of type II), it must start with a partial record of $s = 1, \ldots, p - 1$ segments. To see this, assume the first partial record of the block is of s segments ($1 \leq s \leq p - 1$). Then there are $kq + p - s$ segments remaining in the block. Since a record is of q segments, there will be $k = [b]$ total records and a partial record of $p - s$ segments in the remainder of the block: a total of $[b] + 1$ records. (See Fig. 1.) Conversely, if the first record of the block is of $p \leq s \leq q$ segments, then there are $(k - 1)\, q + q + p - s$ segments remaining in the block and there are $k = [b]$ total and partial records in the remainder of the block. Hence the block will be of type I.

Since one and only one block each with $s = 1, \ldots, p - 1$ segments occur in a complete cycle, we can conclude that there are $p - 1$ type II blocks and hence $q - p + 1$ type I blocks in a cycle (for $p \geq 1$).

If b is integral ($p = 0, q = 1$), then each block contains an integral number of records and a cycle consists of one block with b records.

Proof of lemma 2

To prove this lemma, we need some preliminary results. Define

$$D(i,j) = \begin{cases} 1 \text{ if } i \text{ is divisible by } j \\ 0 \text{ otherwise} \end{cases}$$
$$= 1 - \left[\left\lceil \frac{i}{j} \right\rceil - \left\lfloor \frac{i}{j} \right\rfloor\right]$$

For integers i and j we have

$$\left\lceil \frac{i}{j} \right\rceil = \left\lfloor \frac{i+j-1}{j} \right\rfloor \qquad j \geq 0$$

(See problem 1.2.4.48 in Ref. 13.)
Thus

$$D(i,j) = 1 + \left\lfloor \frac{i}{j} \right\rfloor - \left\lceil \frac{i}{j} \right\rceil \qquad j \geq 0$$
$$= 1 + \left\lfloor \frac{i}{j} \right\rfloor - \left\lfloor \frac{i+j-1}{j} \right\rfloor \qquad j \geq 0$$
$$= \left\lfloor \frac{i}{j} \right\rfloor - \left\lfloor \frac{i-1}{j} \right\rfloor \qquad j \leq 0$$

As stated earlier, a type II block must begin with a partial record of size $s = 1, \ldots, p-1$ segments. Assume there are a blocks in the last incomplete cycle of the file ($0 < a < q$ since there are q blocks to a complete cycle). There will be a type II block that begins with a partial record of s segments in the last incomplete cycle if

$$\sum_{i=0}^{a-1} D(i(kq+p)+s, q) = 1 \qquad 1 \le s \le p-1$$

Therefore, the number of type II blocks in the partial cycle of a blocks is

$$\sum_{s=1}^{p-1}\sum_{i=0}^{a-1} D(i(kq+p)+s, q)$$

We have

$$D(i(kq+p)+s, q) = D(ip+s, q) = \left\lfloor \frac{ip+s}{q} \right\rfloor - \left\lfloor \frac{ip+s-1}{q} \right\rfloor$$

Hence, the number of type II blocks is

$$\sum_{s=1}^{p-1}\sum_{i=0}^{a-1}\left[\left\lfloor \frac{ip+s}{q} \right\rfloor - \left\lfloor \frac{ip+s-1}{q} \right\rfloor\right]$$

$$= \sum_{i=0}^{a-1}\sum_{s=1}^{p-1}\left[\left\lfloor \frac{ip+s}{q} \right\rfloor - \left\lfloor \frac{ip+s-1}{q} \right\rfloor\right]$$

$$= \sum_{i=0}^{a-1}\left[\left\lfloor \frac{ip+p-1}{q} \right\rfloor - \left\lfloor \frac{ip}{q} \right\rfloor\right]$$

$$= \sum_{i=0}^{a-1}\left[\left\lfloor \frac{ip+p-1}{q} \right\rfloor - \left\lfloor \frac{ip}{q} \right\rfloor + \left\lfloor \frac{(i+1)p}{q} \right\rfloor - \left\lfloor \frac{(i+1)p}{q} \right\rfloor\right]$$

Since p and q are relatively prime and $i+1 \le a < q$, $(i+l)p$ is not divisible by q. Therefore,

$$\left\lfloor \frac{ip+p-1}{q} \right\rfloor - \left\lfloor \frac{(i+1)p}{q} \right\rfloor = -D((i+1)p, q) = 0$$

Thus the number of type II blocks is

$$\sum_{i=0}^{a-1}\left[\left\lfloor \frac{(i+1)p}{q} \right\rfloor - \left\lfloor \frac{ip}{q} \right\rfloor\right]$$

By the telescoping property, the above summation yields $\lfloor ap/q \rfloor$. It follows that the number of type I blocks in an incomplete cycle is $a - \lfloor ap/q \rfloor$.

Proof of lemma 3

The number of integral blocks in the file is $\lfloor n/b \rfloor$. Hence the number of records that completely reside in these blocks is $\lfloor b\lfloor n/b \rfloor \rfloor$. Therefore, the number of records that totally or partially reside in the partially filled block is $n - \lfloor b\lfloor n/b \rfloor \rfloor$.