SIAM J. COMPUT.
Vol. 18, No. 4, pp. 748-758, August 1989

008

SPACE-TIME TRADE-OFFS FOR ORTHOGONAL RANGE QUERIES*

PRAVIN M. VAIDYA†

Abstract. This paper investigates the question of (storage) space-(retrieval) time trade-off for orthogonal range queries on a static data base. Each record in the data base consists of a key that is a d-tuple of integers, and a data value that is an element in a commutative semigroup G. An orthogonal range query is specified by a d-dimensional parallelepiped (box). Two types of response to such a query are considered: one where the output is the semigroup sum of the data values whose keys are located in the query parallelepiped, and the other where the output is a list of all the records whose keys lie in the query parallelepiped. This paper studies two models, the arithmetic model and the tree model and obtains lower bounds on the product of retrieval time and storage space in both models.

Key words. range queries, space-time trade-offs, lower bounds

AMS(MOS) subject classifications. 68Q20, 68Q25, 68U05

1. Introduction. Consider a data base that contains a collection of records, each with a key and a number of data fields. Given a range query, which is specified by a set of constraints on the keys, the data base system is expected to return the set of records, or a function of the set of records whose keys satisfy all the constraints. If the data base is static the collection of records may be preprocessed to achieve a balance between the storage utilized and the time required to answer a query. There is an extensive literature [1], [2], [4], [8], [9], [11], [12] on algorithms for range query, and the space and time requirements have traditionally been used as performance measures for such algorithms. In this paper, we investigate the question of (storage) space-(retrieval) time trade-off for orthogonal range queries on a static data base.

Let G be a commutative semigroup with an addition operation $+$. Let d be a fixed positive integer. Let $N=\{1,2,\cdots,n\}$ and let N^d denote the set of all d-tuples of positive integers less than or equal to n. A record $(k, f(k))$ is a pair of key $k\in N^d$ and datum $f(k)\in G$. The data base consists of n such records. Let $k=(k_1,k_2,\cdots,k_d)$. An orthogonal range query is specified by a $2d$-tuple $(x_{11},x_{12},x_{21},x_{22},\cdots,x_{d1},x_{d2})$ of positive integers satisfying $x_{i1}<x_{i2}$, $1\leqq i\leqq d$. Alternately, the query region for an orthogonal range query is a parallelepiped (box) b, defined by the product $[x_{11},x_{12})\times[x_{21},x_{22})\times\cdots\times[x_{d1},x_{d2})$ of d-semiclosed intervals with positive integer endpoints. A key k is said to be located in a box $b=[x_{11},x_{12})\times[x_{21},x_{22})\times\cdots\times[x_{d1},x_{d2})$ if and only if $x_{i1}\leqq k_i<x_{i2}$, $1\leqq i\leqq d$. We consider two types of response to such a query, one where the output is the sum of the data $f(k)$ whose keys k are located in the query parallelepiped (box) b, and the other where the output is a list of all the records whose keys lie in the query parallelepiped b.

Let $Q(b)$ denote the input $2d$-tuple corresponding to query box$b$, and let K denote the set of keys in the data base. As we shall be studying space-time requirements for orthogonal range query only, we shall assume that the set of query regions is fixed to be the set of boxes.

A space-time trade-off seeks to answer questions such as what is the minimum amount of storage needed to ensure a certain query time? The trade-off between space

* Received by the editors August 18, 1986; accepted for publication (in revised form) November 6, 1988. This research was supported by a Shell Foundation fellowship. A preliminary version of this paper appeared in the Proceedings of the 17th Annual ACM Symposium on Theory of Computing, May 1985 [10].

† Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801. Present address, AT & T Bell Laboratories, Murray Hill, New Jersey 07974.

and time is dependent on the model of data structure and the set of records in the data base. We fix the data structure model and then try to obtain a set of records that makes this trade-off as bad as possible. Thus the lower bounds on space-time products are to be interpreted as worst-case bounds, i.e., there exists a set of n records whose space-time product has the said bounds.

We study two models. In Model A, we work in the general framework defined by Fredman [5]-[7], and consider only data structures and manipulation algorithms that are independent of the choice of the semigroup G. So the set K of keys in the data base together with the set of query regions completely specifies the problem. Given a query box b, the query answering algorithm is expected to return the semigroup sum of the data values whose keys are located in b. Model A is an arithmetic model with unit cost for each arithmetic operation but no cost for memory retrieval. In this model, we show that for orthogonal range query on a static database with n records, there is a space-time trade-off $(\log T)^{d-1}TS \geqq \Omega(n(\log n)^{d-\theta})$, where $\theta = 1$ for $d=2$, and $\theta = 2$ for $d \geqq 3$. Space-time trade-offs for circular range query and interval query in this model are studied by Yao in [13] and [14]. We note that for $d=2$ the results of Yao [14] are considerably stronger for this model; specifically he shows that for a restricted type of range query $T = \Omega(\log n/(\log(S/n) + \log\log n))$ for $d=2$. The complexity of dynamic range queries in this model is discussed by Fredman in [5]-[7].

In Model B (tree model), we study a broad class of tree data structures. In this model, a data structure is a rooted tree, and with each edge in the tree is associated a condition. Given a query, the query answering algorithm starts with the root, and visits a vertex v if and only if the given query satisfies the conjunction of the conditions on the path from the root to v. The output corresponding to a given query is a function of the data associated with the visited vertices. Several standard data structures, such as linked lists, range trees, etc. [1], [2], [4], [8], [9], [11], [12], fit into this model. In Model B, we investigate the orthogonal reporting problem where the response to a query is a list of all the records in the data base whose keys are located in the query box. Since the output size is query dependent, the time required to answer a query is not the correct measure of the overhead involved in producing the desired response to the query. So we define a scaled query time T' that measures the overhead for producing one unit of output. For the orthogonal reporting problem on a static data base with n records, we show that there is a space-time trade-off $(\log T' + \log\log n)^{d-1}T'S \geqq \Omega(n(\log n)^{d-\theta})$, where $\theta = 1$ for $d=2$, and $\theta = 2$ for $d \geqq 3$.

The results in this paper for Model A (arithmetic model) have been significantly strengthened by Chazelle using a different technique [3].

2. An overview.

2.1. Model A. Arithmetic model. In this model [5], [6], [7], a data structure is an infinite array Z of variables $z_0, z_1, z_2, \cdots$, that stores elements from the commutative semigroup G. Given any input query, the query answering algorithm chooses a collection of at most T variables in the array and returns their semigroup sum as the response to the query. Since only arithmetic operations are charged, the query time is T. The data structure is assumed to be independent of the specific semigroup G and so the mapping between elements in G and variables in Z is determined solely by the set K of n keys in the data base. With each variable z_i in Z is associated a subset h_i of K, and the data value $\sum_{k \in h_i} f(k)$ is stored in z_i, where $f(k)$ is the data value associated with k. Let $H \subseteq 2^K$ such that every set in H is associated with some variable in Z and every variable in Z is associated with some set in H. Let R be the set of all possible query boxes, and let $P(K, H, T)$ be the property that for each $b \in R$, $b \cap K$ is expressible as the disjoint union of at most T sets in H. (The lower bounds in this paper are valid even if disjoint union is replaced by union in the definition of $P(K, H, T)$.) The query answering algorithm works correctly if and only if $P(K, H, T)$ is satisfied. The storage space S is defined by

$$S = \max_{K} \min_{H \text{ satisfying } P(K,H,T)} |H|.$$

The following theorem summarizes the results in this model.

THEOREM 1. *In Model* A, *for orthogonal range query on a static database with n records, there is a space-time trade-off* $(\log T)^{d-1} TS \geqq \Omega(n(\log n)^{d-\theta})$, *where* $\theta = 1$ *for* $d = 2$, *and* $\theta = 2$ *for* $d \geqq 3$.

The proof is based on Lemma 1 given below. The lemma asserts that there exists a set K of n keys and a large enough set $B(T, n)$ of query parallelepipeds such that the subsets of K induced by members of $B(T, n)$ satisfy certain intersection conditions. The proof of Lemma 1 is given in § 4.

LEMMA 1. *There is a set K of n keys and a set* $B(T, n)$ *of boxes satisfying the following properties*:

(1) $(\log T)^{d-1} T|B(T, n)| = \Omega(n(\log n)^{d-\theta})$, *where* $\theta = 1$ *for* $d = 2$, *and* $\theta = 2$ *for* $d \geqq 3$.

(2) *For distinct* b_1, b_2, *in* $B(T, n)$, $|b_1 \cap b_2 \cap K| < 1/T \min\{|b_1 \cap K|, |b_2 \cap K|\}$.

Using property (2) in Lemma 1, we show that for any H satisfying $P(K, H, T)$, we must have $|H| \geqq |B(T, n)|$. Then Theorem 1 follows from the lower bound on $|B(T, n)|$ given by property (1) in Lemma 1. Let b_1, b_2, be distinct boxes in $B(T, n)$. As $b_1 \cap K$ is expressible as the union of at most T sets in H, there exists $h_1 \in H$ such that $|h_1| \geqq (1/T)|b_1 \cap K|$ and $h_1 \subseteq (b_1 \cap K)$. Since $|(b_1 \cap K) \cap (b_2 \cap K)| < (1/T)|b_1 \cap K|$, h_1 cannot appear in the decomposition of $b_2 \cap K$ as the union of members of H. So with each b_i in $B(T, n)$ we can associate a distinct h_i in H.

2.2. Model B. Tree model. In the case of the reporting problem the output size is dependent on the given query, and so the arithmetic model is not suitable for investigating this problem. So we study the tree model for data structures. In this model, the data structure is assumed to be a rooted tree. With each vertex v is associated a set of data items and we let $data(v)$ denote the set of data items associated with vertex v. With each edge in the tree is associated a condition. Given an input query in the form of a tuple of numbers, the query answering algorithm first visits the root. A vertex v is visited if and only if it is a son of some vertex u that has already been visited and the input tuple satisfies the condition associated with edge uv. We define $cond(v)$ to be the conjunction of all the conditions on the path from the root to vertex v. Thus for any query box b, on being given the corresponding tuple $Q(b)$ as input, the query answering algorithm visits vertex v if and only if $Q(b)$ satisfies $cond(v)$. The response to a query is a function of the data at the visited vertices.

In the tree model, we investigate the orthogonal reporting problem where the response to a given query is a list of all the records in the data base whose keys lie in the query box. Let $\hat{G}$ be a semigroup consisting of a single element. We shall restrict the universe of records so that the data in each record is the unique element from $\hat{G}$. We shall only consider sets of records which are such that no two records in a set have the same key. Then the set K of keys completely specifies the set of records, and the orthogonal reporting problem is to produce a list of all the keys in K that lie in the given query box. Note that considering this special case does not cause any loss of generality as the lower bounds obtained in the special case trivially extend to the general case. Let r be a fixed constant. We restrict ourselves to trees where every vertex has degree at most r. The condition associated with each edge in the tree is restricted to be a disjunction of at most r binary comparisons. Thus $cond(v)$, the conjunction of the conditions on the path from the root to v, is now a conjunction of disjunctions of comparisons. For each vertex v, $data(v)$ is a set of keys. Vertices may share storage, so $data(v)$ is effectively the set of records accessed via vertex v.

Consider a fixed set K of keys, and a fixed tree for K. For a query box b, let $U(b)$ be the set of all those vertices v such that $Q(b)$ satisfies $cond(v)$. Given a query box b, the query answering algorithm visits all the vertices in $U(b)$, and extracts the set of keys $\bigcup_{v\in U(b)} data(v)$. The set of keys $b \cap K$ is then obtained by explicitly testing for each key in $\bigcup_{v\in U(b)} data(v)$ whether the key is located in b. Thus filtering search [2] is included in this model. Let $T(b)$ be the time required to answer the query corresponding to b. $T(b)$ is lower bounded by $|U(b)|+|\bigcup_{v\in U(b)} data(v)|$. Since the output size is query dependent, the time required to answer a query is not the correct measure of the overhead involved in producing the desired response to the query. With respect to a fixed set K of keys, and a fixed tree for K, we define a scaled query time T' as follows:

$$T' = \max_{1\leqq |b\cap K|\leqq \log_2 n} \frac{T(b)}{|b\cap K|}.$$

For a fixed set of keys, the storage S is defined to be the minimum number of vertices a corresponding tree must have to ensure a scaled query time of T'.

At this point we remark that several common data structures, such as linked lists, range trees, etc. [1], [2], [4], [8], [9], [11], [12], fit into the tree model. Also note that the tree model restricts the manner in which data records are accessed; it does not place a restriction on how the data is stored. As long as there is a fixed tree that defines how the data in the data structure is accessed, and a node in this tree corresponds to a distinct unit of storage in the data structure, the data structure would still fit into the tree model; it would not matter that the data structure itself was not a tree.

THEOREM 2. *In the tree model, for the orthogonal reporting problem on a static data base with n records, there is a space-time trade-off* $(\log T' + \log\log n)^{d-1} T'S \geqq \Omega(n(\log n)^{d-\theta})$, *where $\theta = 1$ for $d = 2$, and $\theta = 2$ for $d \geqq 3$.*

The proof of Theorem 2 is based on Lemma 2 below. A proof of Lemma 2 is given in § 4.

LEMMA 2. *Let c_1 and c_2 be constants dependent on the dimension d. There exists a set of K of n keys that has a subset K' satisfying the following properties*:

(1) $|K'| \geqq c_1 n$.

(2) *With respect to a particular tree for K, let $V(k)$ be the set of all those vertices v that satisfy the conditions* (i) *key $k \in data(v)$; and* (ii) *there is a box b such that $Q(b)$ satisfies $cond(v)$ and $1 \leqq |b \cap K| \leqq \kappa(d, n)$, where $\kappa(d, n) = 1$ for $d = 2$ and $\kappa(d, n) = \log_2 n$ for $d \geqq 3$. Then for each tree for K, for each key $k \in K'$,* $|V(k)| \geqq (\log_2 n)^{d-1}/c_2(\log_2 T' + \log_2 \log_2 n)^{d-1}$.

Let K be a set of n keys and K' be a subset of K such that K and K' satisfy the conditions in Lemma 2. Consider a fixed tree for K. For a key k, let $V(k)$ be as defined in Lemma 2. We must have that

$$\sum_{k\in K'} |V(k)| \leqq \sum_{v\in \bigcup_{k\in K'} V(k)} |data(v)|.$$

For each vertex v in $\bigcup_{k\in K'} v(k)$, $|data(v)| \leqq \kappa(d, n)T'$, since v is visited by a query corresponding to a box containing at most $\kappa(d, n)$ keys. Then from properties (1) and

(2) in Lemma 2 it follows that

$$\left|\bigcup_{k\in K'} V(k)\right| \geqq \frac{c_1 n(\log_2 n)^{d-1}}{c_2\kappa(d,n)T'(\log_2 T'+\log_2\log_2 n)^{d-1}}.$$

Thus the storage S for a set K of n keys satisfying the conditions in Lemma 2 must obey the following constraint:

$$S \geqq \frac{c_1 n(\log_2 n)^{d-1}}{c_2\kappa(d,n)T'(\log_2 T'+\log_2\log_2 n)^{d-1}}.$$

3. Canonical parallelepipeds and almost uniform distributions. We shall utilize a special class of parallelepipeds (boxes) in obtaining the desired space-time trade-offs. Let n be a power of 2 and let $I_l=\{[j2^l+1,(j+1)2^l+1): 0\leqq j<(n/2^l)\}$. I_l is the set of intervals obtained by breaking up $[1, n+1)$ into $n/2^l$ semiclosed intervals of equal size, each interval being closed on the left and open on the right. Let $I=I_0\cup I_1\cup\cdots\cup I_{\log_2 n}$. Then I is defined to be the set of canonical intervals, and I^d is defined to be the set of canonical parallelepipeds, or equivalently canonical boxes.

For a box b, we use $[\pi_{i1}(b), \pi_{i2}(b))$ to denote the interval that is the projection of box b onto to the ith coordinate axis. Equivalently, for $1\leqq i\leqq d$, $\pi_{i1}(b)$ and $\pi_{i2}(b)$ denote the $(2i-1)$st and the $(2i)$th components of the $2d$-tuple $Q(b)$ corresponding to box b. For a box b, *dimensions*(b) is defined to be the d-tuple $((\pi_{12}(b)-\pi_{11}(b)),(\pi_{22}(b)-\pi_{21}(b)),\cdots,(\pi_{d2}(b)-\pi_{d1}(b)))$. We note that since I contains intervals of $\log_2 n+1$ distinct lengths, the total number of choices possible for the dimensions of a canonical box is $(\log_2 n+1)^d$. Let $vol(b)$ denote the volume of a box b, and let $p(x)=2^{\lceil\log_2 x\rceil}$. Let J be the canonical parallelepiped $J_0\times J_1\times\cdots\times J_{d-1}$ where $J_i=[2i(2p(d))^{-1}n+1,(2i+1)(2p(d))^{-1}n+1)$, for $0\leqq i\leqq d-1$. The following lemmas list the properties of canonical boxes that we shall require.

LEMMA 3. *The number of canonical boxes of identical dimensions and of volume* 2^i *is* $n^d/2^i$.

LEMMA 4. *Let* $0\leqq i\leqq(\log_2 n/d^2)$. *Then the number of possibilities for the dimensions of a canonical box of volume* $2^i n^{d-1}$ *is* $\Omega((\log_2 n/d^2)^{d-1})$, *and at most* $(\log_2 n+d)^{d-1}$.

Proof. The number of nonnegative integer solutions to

$$j_1+j_2+\cdots+j_d=i+(d-1)\log_2 n,$$

$$\text{subject to}\quad 0\leqq j_l\leqq\log_2 n,\quad 1\leqq l\leqq d,\quad i\leqq(\log_2 n/d^2)$$

is at least $(\log_2 n/d^2)^{d-1}$ for large enough n, and at most $(\log_2 n+d)^{d-1}$ for any i in the desired range. That gives the required bound on the number of possibilities for the dimensions of a canonical box of volume $2^i n^{d-1}$. □

LEMMA 5. *Let* $b_1, b_2,\cdots, b_\rho$ *be canonical boxes of volume* α *such that* $\bigcap_{i=1}^{\rho} b_i\neq\emptyset$. *Then* $vol(\bigcap_{i=1}^{\rho} b_i)\leqq\alpha 2^{1-\rho^{1/(d-1)}}$.

Proof. For $1\leqq m\leqq d$, let

$$L_m=\{[\pi_{m1}(b_j),\pi_{m2}(b_j)): 1\leqq j\leqq\rho\}.$$

The intervals in L_m can be ordered by containment, and the ratio of the lengths of the largest and the smallest intervals in L_m is $2^{|L_m|-1}$. Let $|L_m^*|=\max_m|L_m|$. By pigeonholing, $|L_m^*|\geqq\rho^{1/(d-1)}$. Then

$$vol\left(\bigcap_{i=1}^{\rho} b_i\right)\leqq\alpha 2^{1-|L_m^*|}\leqq\alpha 2^{1-\rho^{1/(d-1)}}.$$ □

LEMMA 6. *Let $\hat{b}$ be a fixed canonical box of volume α. Then the number of canonical boxes b of volume α which satisfy the condition $vol(b \cap \hat{b}) \geqq 2^{-j} vol(\hat{b})$ is at most $(2j+d+1)^{d-1}$.*

Proof. Since $vol(b \cap \hat{b}) \geqq 2^{-j} vol(\hat{b})$ and $vol(b) = vol(\hat{b})$, for each m, $1 \leqq m \leqq d$, either $[\pi_{m1}(b), \pi_{m2}(b))$ contains $[\pi_{m1}(\hat{b}), \pi_{m2}(\hat{b}))$ or $[\pi_{m1}(\hat{b}), \pi_{m2}(\hat{b}))$ contains $[\pi_{m1}(b), \pi_{m2}(b))$, and

$$2^{-j} \leqq \frac{\pi_{m2}(b) - \pi_{m1}(b)}{\pi_{m2}(\hat{b}) - \pi_{m1}(\hat{b})} \leqq 2^{j}.$$

Thus for $1 \leqq m \leqq d$, there are at most $(2j+1)$ possibilities for the mth interval defining box b, and as the volume of the boxes b is fixed to be α the total number of possibilities for the boxes b is bounded by $(2j+d+1)^{d-1}$. □

Having described canonical boxes, we shall proceed to almost uniform distributions of n keys. The distributions are termed almost uniform because the number of keys in a canonical box does not deviate too far from the volume of the box divided by n^{d-1}. For $d=2$, we can explicitly construct such distributions, and thereby get Theorem 3. For $d \geqq 3$, we have to resort to counting arguments and show that the number of distributions of n keys, which do not satisfy the properties in Theorem 4, is less than the total of n^{dn} possible distributions.

THEOREM 3. *For $d=2$, there is a set K of n keys such that for each canonical box b,*

$$\left\lfloor \frac{vol(b)}{n^{d-1}} \right\rfloor \leqq |b \cap K| \leqq \left\lceil \frac{vol(b)}{n^{d-1}} \right\rceil.$$

Proof. It is adequate to obtain a set K of n keys such that each canonical box of volume n contains exactly one key. We use an inductive construction. Let x_1 and x_2 denote the two attributes of a key. Let K_m denote a set of m keys satisfying the conditions in Theorem 3. We shall obtain K_{2m} from K_m. Let

$$K'_m = \{(2x_1 - 1, x_2) \colon (x_1, x_2) \in K_m\}$$

and

$$K''_m = \{(2x_1, x_2 + m) \colon (x_1, x_2) \in K_m\}.$$

We let $K_{2m} = K'_m \cup K''_m$. A canonical box of volume $2m$ and x_1 dimension equal to 1 contains exactly one key, as each key has a distinct value for x_1. Let $b = [x_{11}, x_{12}) \times [x_{21}, x_{22})$ be a canonical box of volume m corresponding to $n=m$, and let $b_1 = [2x_{11} - 1, 2x_{12} - 1) \times [x_{21}, x_{22})$ and $b_2 = [2x_{11} - 1, 2x_{12} - 1) \times [x_{21} + m, x_{22} + m)$. Then b_1 and b_2 are canonical boxes of volume $2m$ corresponding to $n = 2m$. If $(x_1, x_2) \in b \cap K_m$ then $(2x_1 - 1, x_2) \in b_1 \cap K'_m$ and $(2x_1, x_2 + m) \in b_2 \cap K''_m$, and b_1, b_2 do not contain any other key in K_{2m}. Furthermore, all canonical boxes corresponding to $n=2m$, of volume $2m$ and x_1 dimension at least 2, may be derived in this manner from canonical boxes of volume m corresponding to $n=m$. □

THEOREM 4. *Let σn^{dn} be the number of distinct sets K of n keys, each key in N^d, that satisfy the three properties given below. Then σ tends to 1 as n tends to ∞ and $\sigma = (1 - o(1/n))$.*

(1) *Let $a(n) = 2p(\log_2 n)$. For each canonical box b,*

$$\left\lfloor \frac{vol(b)}{a(n) n^{d-1}} \right\rfloor \leqq |b \cap K| < 6a(n) \left\lceil \frac{vol(b)}{a(n) n^{d-1}} \right\rceil.$$

(2) *Each canonical box of volume n^{d-1} contains at most $\log_2 n$ keys.*

(3) $|J \cap K| \geqq (n/(4(2p(d))^{2d})) = n/4(c_3)^2$.

Proof. The total number of possible key distributions (sets) K is n^{dn}. Let F_1, F_2, and F_3, be the fraction of distributions that do not satisfy properties (1), (2), and (3), in Theorem 4, respectively. We shall show that each of F_1, F_2, and F_3, is $o(1/n)$.

To bound F_1 note that, if there is a canonical box whose volume is not equal to $a(n)n^{d-1}$ and that violates property (1) above, then there is necessarily a canonical box of volume $a(n)n^{d-1}$ that violates property (1) in the same manner. Let F_{11} be the fraction of distributions K such that there is a canonical box of volume $a(n)n^{d-1}$ that does not contain a key in K, and let F_{12} be the fraction of distributions K such that some canonical box of volume $a(n)n^{d-1}$ contains at least $6a(n)$ keys in K. Then $F_1 \leqq F_{11} + F_{12}$. A bound on F_{11} may be obtained by noting that there are at most $n(\log_2 n + d)^{d-1}$ choices for a canonical box of volume $a(n)n^{d-1}$ and all the keys in K must lie outside the chosen canonical box. Thus

$$\begin{aligned} F_{11} &\leqq n(\log_2 n + d)^{d-1}\left(1 - \frac{a(n)}{n}\right)^n \\ &\leqq O(n(\log n)^{d-1} e^{-a(n)}) \\ &= o\left(\frac{1}{n}\right). \end{aligned}$$

An upper bound on F_{12} is obtained by observing that we may choose a canonical box of volume $a(n)n^{d-1}$ in at most $n(\log_2 n + d)^{d-1}$ ways, and we may choose $6a(n)$ keys to lie in the chosen box and then let the remaining keys be located anywhere in $[1, n+1)^d$. Then

$$\begin{aligned} F_{12} &\leqq n(\log_2 n + d)^{d-1}\binom{n}{6a(n)}\left(\frac{a(n)}{n}\right)^{6a(n)} \\ &\leqq n(\log_2 n + d)^{d-1}\frac{(a(n))^{6a(n)}}{6a(n)!} \\ &\leqq n(\log_2 n + d)^{d-1}\left(\frac{e}{6}\right)^{6a(n)} \cdots \quad \text{(using Stirling's approximation)} \\ &= o\left(\frac{1}{n}\right). \end{aligned}$$

Thus

$$F_1 \leqq F_{11} + F_{12} = o(1/n) + o(1/n) = o(1/n).$$

A bound F_2 is obtained as follows. A canonical box of volume n^{d-1} may be chosen in at most $n(\log_2 n + d)^{d-1}$ ways; then $\log_2 n$ keys may be selected to lie in the chosen box, and the remaining keys may lie anywhere in $[1, n+1)^d$. Thus,

$$\begin{aligned} F_2 &\leqq n(\log_2 n + d)^{d-1}\binom{n}{\log_2 n} n^{-\log_2 n} \\ &\leqq \frac{n(\log_2 n + d)^{d-1}}{(\log_2 n)!} \\ &= o\left(\frac{1}{n}\right). \end{aligned}$$

To bound F_3, we choose $(1-4^{-1}c_3^{-2})n$ points to lie outside the canonical box J, and let the remaining points be located anywhere in $[1, n+1)^d$. Furthermore, the volume of J is n^d/c_3. Thus

$$F_3 \leqq \binom{n}{(1-4^{-1}c_3^{-2})n}\left(1-\frac{1}{c_3}\right)^{n(1-4^{-1}c_3^{-2})}.$$

Using Stirling's approximation for factorials and taking logarithms we get

$$\log_e F_3 \leqq \frac{n}{4c_3^2}\left(\log_e(4c_3^2)+(4c_3^2-1)\log_e\left(1-\frac{4c_3-1}{4c_3^2-1}\right)\right)+O(\log n).$$

Then noting that $\log_e(1-x) \leqq -x$ for $0<x\leqq 1$, we get

$$\log_e F_3 \leqq \frac{n}{4c_3^2}(\log_e(4c_3^2)+1-4c_3)+O(\log n)$$

$$\leqq -\frac{7n}{4c_3^2}+O(\log n) \quad \text{as } c_3 \geqq 16.$$

Thus $F_3 = o(1/n)$. □

4. Proofs of lemmas. In this section we give proofs of Lemmas 1 and 2 used to prove Theorems 1 and 2 in §§ 2.1 and 2.2, respectively. For the purposes of this section we shall let the set K of keys be fixed. For $d=2$ let K be a fixed set of n keys as specified by Theorem 3, and for $d\geqq 3$ let K be a fixed set of n keys as specified by Theorem 4. We assume that n is a power of 2. Let $\beta(\alpha)$ denote the set of canonical boxes of volume α, and let β_J denote the set of those canonical boxes that are also subboxes of J (for definition of J see § 3).

Proof of Lemma 1. We shall give a proof for $d\geqq 3$, the proof for $d=2$ is similar. Let $\chi(T)=(64T(2p(d))^{2d})$. Let $B(T, n)$ be the largest set of boxes satisfying the following conditions:

(1) For all $b\in B(T, n)$, $b\in\beta(\chi(T)a(n)n^{d-1})$, and $|b\cap K|\geqq 6Ta(n)$.

(2) For any two boxes b_1 and b_2 in $B(T, n)$, $vol(b_1\cap b_2)\leqq a(n)n^{d-1}$.

By Theorem 4, the number of keys in K located in a canonical box of volume $\chi(T)a(n)n^{d-1}$ is at most $6\chi(T)a(n)$. It is then easily shown that the number of boxes in $\beta(\chi(T)a(n)n^{d-1})$ that have identical dimensions and that contain at least $6Ta(n)$ keys is $\Omega(n/(T\log_2 n))$. Then from Lemmas 4 and 6 in § 3, we can conclude that $|B(T, n)|=\Omega(n/T\log_2 n(\log_2 n/\log_2 T)^{d-1})$. The intersection of any two distinct boxes b_1, b_2 in $B(T, n)$ is a canonical box of volume at most $(a(n)n^{d-1})$ and so by Theorem 4, $|b_1\cap b_2\cap K|<6a(n)<(1/T)\min\{|b_1\cap K|, |b_2\cap K|\}$. □

In Model B, as we restrict ourselves to binary comparisons, the only possible comparisons are those between two components of the input tuple, and those between a component of the input tuple and a constant. We shall focus on canonical boxes that are subboxes of the canonical box J. For each box $b\subseteq J$, the input tuple $Q(b)=(x_{11}, x_{12}, x_{21}, x_{22}, \cdots, x_{d1}, x_{d2})$ is such that $x_{11}<x_{12}<x_{21}<x_{22}<\cdots<x_{d1}<x_{d2}$, and so a comparison between two components of the input tuple has the same outcome for each subbox b of J. Then, in Model B we need to analyze only comparisons between a component of the input tuple and a constant. We note that, if the input tuple satisfies a comparison between x_j, the (j)th component of the tuple, and a constant, when x_j takes on the value z_1 as well as when x_j takes on the value z_2, then the input tuple satisfies the comparison whenever x_j takes on any value between z_1 and z_2. The following lemma follows directly from these observations.

LEMMA 7. *Let b_1 and b_2 be subboxes of J. Let C be a conjunction of binary comparisons, and let both $Q(b_1)$ and $Q(b_2)$ satisfy C. Let b be a subbox of J such that either $\pi_{ij}(b_1) \leqq \pi_{ij}(b) \leqq \pi_{ij}(b_2)$ or $\pi_{ij}(b_2) \leqq \pi_{ij}(b) \leqq \pi_{ij}(b_1)$, for $1 \leqq i \leqq d$, $1 \leqq j \leqq 2$. Then the tuple $Q(b)$ also satisfies C.* □

LEMMA 8. *Let t be an integer greater than 1, let γ be an integer greater than 0, and let m be an integer such that $1 \leqq m \leqq d$. Let $C = \neg(C_1 \vee C_2 \vee \cdots \vee C_{t-1})$, where each of C_i, $1 \leqq i \leqq t-1$, is a conjunction of comparisons. Let $b_1, b_2, \cdots, b_{t+1}$ be distinct canonical boxes satisfying the following conditions:*

(1) *For all i, $1 \leqq i \leqq t+1$, $b_i \in (\beta_J \cap \beta(\alpha))$.*
(2) *Each of the tuples $Q(b_1), Q(b_2), \cdots, Q(b_{t+1})$ satisfies condition C.*
(3) *For all i, $1 \leqq i \leqq t+1$, $\pi_{m2}(b_i) - \pi_{m1}(b_i) = \pi_{m2}(b_1) - \pi_{m1}(b_1)$.*
(4) *For all i, $1 \leqq i \leqq t$, $\pi_{m1}(b_{i+1}) - \pi_{m2}(b_i) \geqq \gamma(\pi_{m2}(b_1) - \pi_{m1}(b_1))$.*

Then there are at least γ boxes b in $\beta_J \cap \beta(\alpha)$ such that $Q(b)$ satisfies condition C.

Proof. Let $\lambda([y_1, y_2), m, b)$ be the box obtained by replacing the mth interval defining box b by the interval $[y_1, y_2)$. Note that I is the set of canonical intervals. For $1 \leqq i \leqq t$, let

$$A_i = \{\lambda([y_1, y_2), m, b_i): [y_1, y_2) \in I, y_2 - y_1 = \pi_{m2}(b_1) - \pi_{m1}(b_1),$$
$$\pi_{m2}(b_i) \leqq y_1 < y_2 \leqq \pi_{m1}(b_{i+1})\}.$$

Then $|A_i| \geqq \gamma$, and each box in A_i is a subbox of J and a canonical box of volume α. There exists an i such that for each box b in A_i, $Q(b)$ satisfies C. Suppose this is not true. Then there exist boxes $\hat{b}_{i_1} \in A_{i_1}$, $\hat{b}_{i_2} \in A_{i_2}$, $1 \leqq i_1 < i_2 \leqq t$, such that both $Q(\hat{b}_{i_1})$ and $Q(\hat{b}_{i_2})$ satisfy C_l, for some l, $1 \leqq l \leqq t-1$. Then by Lemma 7 it follows that $Q(b_{i_2})$ must satisfy C_l and thereby not satisfy C which is a contradiction. □

Proof of Lemma 2. We shall give a proof for $d \geqq 3$, the proof for $d = 2$ may be constructed along similar lines. Fix a tree for the set of keys K. For the purposes of the proof we shall restrict ourselves to canonical boxes that are subboxes of J. Note that as K satisfies the conditions in Theorem 4 in § 3, a canonical box in $\beta(n^{d-1})$ contains at most $\log_2 n$ keys in K, and so the query time T for such a box cannot exceed $T' \log_2 n$. We shall show that if the conditions in Lemma 2 do not hold then there must be a canonical box $b \in \beta(n^{d-1})$ such that $Q(b)$ satisfies $cond(v)$ for more than $T' \log_2 n$ vertices v. Then the query time for b would exceed $T' \log_2 n$, and that would be a contradiction.

Let $|J \cap K| = c_5 n$, by Theorem 4 we know that such a c_5 exists. For each key $k \in J \cap K$, there are at least $c_4(\log_2 n)^{d-1}$ boxes in $\beta_J \cap \beta(n^{d-1})$ that contain k, for some constant c_4 dependent on d. Let $\delta = 2^d(T' \log_2 n + 1)^d (\log_2 n + 1)^{d+2}$, and let $\psi(T', n) = 2rc_5^{-1}(T')^2 \log_2 n + d)^{d+3}\delta$. Let c_6 be a large enough constant such that $(T' \log_2 n)^{c_6^{1/(d-1)}} \geqq 2^d(4\psi(T', n) + 6)^d$. The constants c_1 and c_2 in Lemma 2 are given by $c_1 = c_5/2$ and $c_2 = c_6/c_4$.

For each key $k \in K$, let

$$\tau(k) = \{b: b \in (\beta_J \cap \beta(n^{d-1})), k \in (b \cap K)\},$$

and let

$$\hat{V}(k) = \{v: k \in data(v), \exists b \in \tau(k) \quad \text{s.t.} \quad Q(b) \text{ satisfies } cond(v)\}.$$

We note that for each box $b \in \tau(k)$ there exists a vertex $v \in \hat{V}(k)$ such that $Q(b)$ satisfies $cond(v)$. For each $k \in J \cap K$, $|\tau(k)| \geqq c_4(\log_2 n)^{d-1}$, and $\hat{V}(k) \subseteq V(k)$, where $V(k)$ is the set of boxes defined in Lemma 2 in § 2.2.

Let K' be the largest subset of $J \cap K$ such that

$$\forall k \in K' \quad |\hat{V}(k)| \geqq \frac{c_4(\log_2 n)^{d-1}}{c_6(\log_2 T' + \log_2 \log_2 n)^{d-1}}.$$

If $|K'| \geqq c_5 n/2$ then the conditions in Lemma 2 are satisfied. So assume that $|(J \cap K) - K'| \geqq c_5 n/2$. With each key k in $(J \cap K) - K'$ we can associate a distinguished vertex $\mu(k)$ and a distinguished set of canonical boxes $A(k)$ satisfying the following conditions:

(1) $|A(k)| \geqq c_6(\log_2 T' + \log_2 \log_2 n)^{d-1}$.

(2) $A(k) \subseteq \tau(k)$.

(3) $\mu(k) \in \hat{V}(k)$.

(4) For all $b \in A(k)$, $Q(b)$ satisfies $cond(\mu(k))$.

The query time for a box in $\beta(n^{d-1})$ cannot exceed $T' \log_2 n$, and hence for all $k \in ((J \cap K) - K')$, $|data(\mu(k))| \leqq T' \log_2 n$. Then $|\{\mu(k): k \in ((J \cap K) - K')\}| \geqq c_5 n/2T' \log_2 n$. Let η be the set of all vertices ν such that there are at least δ vertices in the set $\{\mu(k): k \in ((J \cap K) - K')\}$ that are also present in the subtree rooted at v. As the degree of each vertex in the tree is at most r (r a fixed constant), $|\eta| \geqq c_5 n/2r\delta T' \log_2 n$.

For a vertex u, let $num(u)$ be the number of canonical boxes b such that $b \in \beta(n^{d-1})$ and $Q(b)$ satisfies $cond(u)$. Suppose we can show that for each vertex u in η, $num(u) \geqq \psi(T', n)$. From the lower bound on the number of vertices in η it would follow that

$$\sum_{u \in \eta} num(u) \geqq |\eta| \psi(T', n) > n(\log_2 n + d)^{d+1} T'.$$

Since $|\beta(n^{d-1})| \leqq n(\log_2 n + d)^{d-1}$, we could then conclude that there is a $b \in \beta(n^{d-1})$ such that $Q(b)$ satisfies $cond(u)$ for at least $T' \log_2 n + 1$ vertices u in η, and that would be a contradiction.

Let u be an arbitrary vertex in η other than the root. We have to show that $num(u) \geqq \psi(T', n)$. With u one can associate distinct keys $k_1, k_2, \cdots, k_\delta$ in $(J \cap K) - K'$, such that for each box b in $\bigcup_{i=1}^{\delta} A(k_i)$, $Q(b)$ satisfies $cond(u)$. Let $b(k_i) = \bigcap_{b \in A(k_i)} b$. Then for $1 \leqq i \leqq \delta$, $b(k_i) \in \beta_J$, and $b(k_i)$ contains key k_i. By Lemma 5, each of $b(k_i)$ has volume at most $n^{d-1}/(4\psi(T', n) + 6)^d$.

Among $b(k_1), b(k_2), \cdots, b(k_\delta)$ we can find $2^d(T' \log_2 n + 1)^d (\log_2 n + 1)$ distinct boxes of identical dimensions, say $b_1, b_2, \cdots, b_l$. This is possible since the number of possibilities for the dimensions of a canonical box is at most $(\log_2 n + 1)^d$, and by Theorem 4 each of $b(k_i)$ can contain at most $\log_2 n$ keys in K. Corresponding to $b_1, b_2, \cdots, b_l$, for $1 \leqq m \leqq d$, let

$$L_m = \{[\pi_{m1}(b_j), \pi_{m2}(b_j)): 1 \leqq j \leqq l\}.$$

All the intervals in L_m are of the same length, and any two intervals in L_m do not overlap. Let $t = T' \log_2 n$. Suppose that for some m, $1 \leqq m \leqq d$, there exist $t+1$ intervals $[\pi_{m1}(b_{j_1}), \pi_{m2}(b_{j_1})), [\pi_{m1}(b_{j_2}), \pi_{m2}(b_{j_2})), \cdots, [\pi_{m1}(b_{j_{t+1}}), \pi_{m2}(b_{j_{t+1}}))$, in L_m such that for $1 \leqq q \leqq t$, $\pi_{m1}(b_{j_{q+1}}) - \pi_{m2}(b_{j_q}) \geqq \psi(T', n)(\pi_{m2}(b_{j_1}) - \pi_{m1}(b_{j_1}))$. To each of these $t+1$ intervals $[\pi_{m1}(b_{j_q}), \pi_{m2}(b_{j_q}))$ there corresponds a distinct box $\hat{b}_q$ such that $[\pi_{m1}(\hat{b}_q), \pi_{m2}(\hat{b}_q)) = [\pi_{m1}(b_{j_q}), \pi_{m2}(b_{j_q}))$, $\hat{b}_q \in (\beta_J \cap \beta(n^{d-1}))$, and $Q(\hat{b}_q)$ satisfies $cond(u)$. The pathlength from the root to u does not exceed $t-1$, and so we may apply Lemma 8 to these $t+1$ boxes and conclude that there are at least $\psi(T', n)$ boxes b in $\beta(n^{d-1})$ such that $Q(b)$ satisfies $cond(u)$.

We shall now show that there exists a required collection of $T' \log_2 n+1$ intervals in some set L_m. Assume that such a collection of intervals does not exist. Then for each m, $1 \leqq m \leqq d$, all the intervals in the set L_m can be covered by $T' \log_2 n$ intervals, each of length at most $(2\psi(T', n)+3)(\pi_{m2}(b_1)-\pi_{m1}(b_1))$. Each of the $dT' \log_2 n$ covering intervals is closed on the left and open on the right, and has integer end points in the range 1 to $n+1$. It then follows that the boxes $b_1, b_2, \cdots, b_l$, are themselves contained in the union of $(T' \log_2 n)^d$ boxes, each of volume at most $(2\psi(T', n)+3)^d vol(b_1)$. As each of $b_1, b_2, \cdots, b_l$ contains a distinct key, and $l \geqq 2^d(T' \log_2 n+1)^d(\log_2 n+1)$, and $vol(b_1) \leqq (4\psi(T', n)+6)^{-d} n^{d-1}$, it follows that there is a box of volume at most n^{d-1} that contains at least $2^d(\log_2 n+1)$ keys in K. This box of volume at most n^{d-1} can be covered by 2^d canonical boxes of volume n^{d-1}. Thus there must be a canonical box of volume n^{d-1} that contains at least $\log_2 n+1$ keys in K, and since K is a set of keys satisfying the conditions in Theorem 4 this is not possible. $\square$

5. Conclusion. We have obtained space-time trade-offs for orthogonal range query in two models, the arithmetic model and the tree model. Most data structures used in practice are rooted trees and so it may be worth studying more problems in the context of Model B. We conclude by raising questions related to the tree model.

(1) Drawing an analogy with decision trees, what happens when the conditions associated with tree edges are allowed to be comparisons involving linear or higher order polynomial functions of the input? Do the bounds weaken in such a situation?

(2) What kind of bounds can one obtain for queries other than orthogonal range query, say circular range query or polyhedral query?

(3) Can the lower bounds in the tree model be extended to data structures that are directed acyclic graphs?

REFERENCES

[1] J. L. BENTLEY AND H. A. MAURER, *Efficient worst-case data structures for range searching*, Acta Inform., 13 (1980), pp. 155–168.

[2] B. CHAZELLE, *Filtering search: A new approach to query answering*, in Proc. 24th Annual IEEE Symposium on Foundations of Computer Science, 1983, pp. 122–132.

[3] ———, *Lower bounds on the complexity of multidimensional searching*, in Proc. 27th Annual IEEE Symposium on Foundations of Computer Science, 1986, pp. 87–96.

[4] R. COLE AND C. K. YAP, *Geometric retrieval problems*, in Proc. 24th Annual IEEE Symposium on Foundations of Computer Science, 1983, pp. 112–121.

[5] M. L. FREDMAN, *The inherent complexity of dynamic data structures which accommodate range queries*, in Proc. 21st Annual IEEE Symposium on Foundations of Computer Science, 1980, pp. 191–200.

[6] ———, *A lower bound on the complexity of orthogonal range queries*, J. Assoc. Comput. Mach., 28 (1981), pp. 696–705.

[7] ———, *Lower bounds on the complexity of some optimal data structures*, SIAM J. Comput., 10 (1981), pp. 1–10.

[8] H. GABOW, J. BENTLEY, AND R. TARJAN, *Scaling and related techniques for geometric problems*, in Proc. 16th Annual ACM Symposium on Theory of Computing, 1984, pp. 135–143.

[9] G. S. LUEKER, *A data structure for orthogonal range queries*, in Proc. 19th Annual IEEE Symposium on Foundations of Computer Science, 1978, pp. 28–34.

[10] P. M. VAIDYA, *Space-time tradeoffs for orthogonal range queries*, in Proc. 17th Annual ACM Symposium on Theory of Computing, 1985, pp. 169–174.

[11] D. E. WILLARD, *New data structures for orthogonal range queries*, Tech. Report TR-22-78, Aiken Computer Laboratory, Harvard University, Cambridge, MA, 1978.

[12] ———, *Polygon retrieval*, SIAM J. Comput., 11 (1982), pp. 149–165.

[13] A. C. YAO, *Space-time tradeoff for answering range queries*, in Proc. 14th Annual ACM Symposium on Theory of Computing, 1982, pp. 128–136.

[14] ———, *On the complexity of maintaining partial sums*, SIAM J. Comput., 14 (1985), pp. 277–288.