Analysis of the n-Dimensional Quadtree Decomposition for Arbitrary Hyperrectangles

Christos Faloutsos, *Member, IEEE,* H.V. Jagadish, *Member, IEEE,* and Yannis Manolopoulos, *Member, IEEE Computer Society*

Abstract—We give a closed-form expression for the average number of *n*-dimensional quadtree nodes ("pieces" or "blocks") required by an *n*-dimensional hyperrectangle aligned with the axes. Our formula includes as special cases the formulae of previous efforts for two-dimensional spaces [8]. It also agrees with theoretical and empirical results that the number of blocks depends on the hypersurface of the hyperrectangle and not on its hypervolume. The practical use of the derived formula is that it allows the estimation of the space requirements of the *n*-dimensional quadtree decomposition. Quadtrees are used extensively in two-dimensional spaces (geographic information systems and spatial databases in general), as well in higher dimensionality spaces (as oct-trees for three-dimensional spaces, e.g., in graphics, robotics, and three-dimensional medical images [2]). Our formula permits the estimation of the space requirements for data hyperrectangles when stored in an index structure like a (*n*-dimensional) quadtree, as well as the estimation of the search time for query hyperrectangles, for the so-called linear quadtrees [17]. A theoretical contribution of the paper is the observation that the number of blocks is a piece-wise linear function of the sides of the hyperrectangle.

Index Terms—Regular decomposition, geometric data, quadtrees, oct-trees, GIS, robotics.

1 INTRODUCTION

HIERARCHICAL decomposition of space plays an important role in every application that involves geometric data. The idea is that the space is decomposed recursively into smaller and smaller pieces, until the content of each such piece is homogeneous. The problem solved in this paper is the analytical estimation of the number of pieces that an *n*-dimensional rectangle (hyperrectangular region) is decomposed into.

Consider a **two-dimensional image** represented as a $2^k \times 2^k$ array of 1×1 squares. Each such square is called a **pixel**. The length $K = 2^k$ of the side of the image is called the **granularity** of the image. A geometric object within such an image is represented by turning the appropriate pixels to black, while the background is considered white. More than one geometric object may exist in an image. A **block** is a $2^m \times 2^m$ square ($0 \le m \le k$) obtained as the result of recursive decomposition of the image into quadrants and sub-quadrants. We focus on representing one object only. An object within an image is decomposed into blocks as in Fig. 1. For example, in this figure the square $[0, 2] \times [2, 4]$ is a block, while the square $[1, 3] \times [2, 4]$ is not.

For a two-dimensional object, the result of such a decomposition is termed as a *region quadtree*. Such a hierarchical decomposition approach has been used in several areas, including:

- In graphics and robotics (three-dimensional space) [3].
- In geographic information systems and spatial databases. The TIGER project at the U.S. Bureau of Census uses a linear quadtree representation to store all the points of interest in the map of U.S.A. [22]. A similar approach has also been used by Shaffer in the QUILT system for geographic and spatial databases [21], as well as by Orenstein in the extensible data base management system PROBE [18].
- In traditional databases, where records with *n* attributes correspond to points in an *n*-dimensional space. Many methods have been suggested to store such a collection of data, utilizing the hierarchical decomposition approach (e.g., k-d trees [4], quadtrees and their variations [11]).
- In spatiotemporal and scientific databases, where time introduces one more axis [16].
- In image databases, e.g., [2], where three-dimensional brain scans have to be stored. Regions in these brain scans can be encoded using oct-trees, to save space and to achieve faster response on range queries.
- In Grand-Challenge databases [5] (e.g., with meteorological, environmental, sensor data, etc.). In general, these databases contain large multidimensional arrays, (e.g., tuples of the form (x, y, z, t, *temperature*)), which can be stored in some multiresolution, hierarchical fashion, clustering related (i.e., nearby) points together.
- Whenever a transformation is used (e.g., a two-dimensional rectangle corresponds to a four-dimensional point [9], [12]; a polyhedron is mapped to a high-dimensionality point [15]).

We focus on rectilinear hyperrectangles, that is, *n*-d rectangles with sides aligned with the axis. The problem we examine here is the following:

- *C. Faloutsos is with the Department of Computer Science, University of Maryland, College Park, MD 20742-3255. E-mail: christos@cs.umd.edu.*
- *H.V. Jagadish is with AT&T Bell Laboratories, Murray Hill, NJ 07974.*
- *Y. Manolopoulos is with the Department of Informatics, Aristotle University, Thessaloniki, Greece 54006.*

Manuscript received 20 Dec. 1994.
For information on obtaining reprints of this article, please send e-mail to: transkde@computer.org, and reference IEEECS Log Number 104392.0.

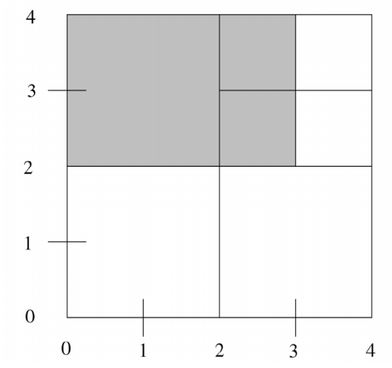


Fig. 1. The shaded rectangle is decomposed into three blocks.

Given a rectilinear hyperrectangle of size $s_1 \times s_2 \times \ldots s_n$,
Find the number of blocks that it will span on the average.

Previous attempts have been restricted to two-dimensional rectangles: Dyer in [6] presented an analysis for the best, worst and average case of a square of size $2^n \times 2^n$, giving an approximate formula for the average case. Shaffer in [20] gives a closed formula for the exact number of blocks that such a square requires when anchored at a given position (x, y); he also gives the formula for the average number of blocks for such squares (averaged over all the possible positions). In a previous paper [8], we generalized some of these formulae for arbitrary (two-dimensional) rectangles. Analysis of the closely related Peano and Hilbert space filling curves for two-dimensional spaces was presented in [14] and [19].

In this paper, we generalize the formulae for n-dimensional rectangles. The derived formulae are useful whenever a hierarchical decomposition is used for higher-dimensionality spaces, either for data hyperrectangles, or for query hyperrectangles. In all these cases, the number of pieces that a hyperrectangle decomposes into clearly affects the space overhead and the search time. Therefore, it is essential for query optimization in spatial/temporal databases [1].

The proposed methodology is as follows:

1) Find the formulae when the sides of the hyperrectangles are of the form $2^{m_i} - 1$, for every dimension i = 1, 2, ..., n. Let's call these hyperrectangles **magic**. One important observation is the fact that the solution for magic rectangles is simple.
2) Prove that the formula for a nonmagic hyperrectangle can be derived by a linear interpolation from the surrounding magic hyperrectangles.

The paper is organized as follows. Section 2 gives some preliminary definitions and examples. Section 3 gives the solution (closed-form formulae) for the magic hyperrectangles. Section 4 establishes a theorem that the solution for nonmagic hyperrectangles can be derived by using linear interpolation. Section 5 gives closed formulae for the average number of blocks in the case of two-dimensional rectangles and three-dimensional parallelepipeds. Section 6 makes some observations and suggests future research directions.

2 PRELIMINARIES

A hyperrectangle is represented as $(x_1, s_1, x_2, s_2, \ldots, x_n, s_n)$ where x_i (i = 1, ..., n) is the ith coordinate of the **anchor** (i.e., the corner with the smallest coordinate values or the 'lower left' corner; this is the corner closest to the origin, since all the coordinates are nonnegative) and s_i is the size of the hyperrectangle on the ith dimension. Table 1 shows the symbols and their definitions.

TABLE 1
DEFINITION OF SYMBOLS

Symbol	Definition
n	Number of dimensions
$x_1, \ldots, x_n$	Coordinates of the lowest corner of the hyperrectangle (i.e., the one closest to the origin)
s_i	Length of the hyperrectangle in ith dimension
$b(x_1, s_1, \ldots, x_n, s_n)$	Number of blocks to cover a specific hyperrectangle
$\bar{b}(s_1, s_2, \ldots, s_n)$	Average number of blocks to cover the hyperrectangle of the query size
$K = 2^k$	Granularity = side of the 'universe' in hyperpixels

DEFINITION 1. *The average number of blocks for a rectangle of sides* $(s_1, s_2, \ldots, s_n)$ *is given by:*

$$\bar{b}(s_1, s_2, \ldots, s_n) = \frac{1}{K^n} \sum_{x_1=0}^{K-1} \cdots \sum_{x_n=0}^{K-1} b(x_1, s_1, x_2, s_2, \ldots, x_n, s_n) \tag{1}$$

where $K = 2^k$ is the granularity. Intuitively, we let the hyperrectangle go to each and every possible position, and we average the number of blocks that the hyperrectangle decomposes into, at each position. Notice that:

- K should be large enough so that the $K \times K \ldots \times K$ hypercube completely encloses the hyperrectangle under examination. In other words: $s_i \le K$ for i = 1, ..., n.
- The hyperrectangle wraps around the edges. This assumption has been used in all the previous analyses of quadtrees [6], [8].

Some important observations, that allow recursive decomposition of the problem:

OBSERVATION 1 (Slicing). If the starting coordinate x_i on the ith axes of a hyperrectangle is odd, then we can "slice off" a hyperplane of width one, that is perpendicular to the ith dimension and starts at x_i. In such a case, the number of blocks of the two pieces added together is the same as the number of blocks of the whole hyperrectangle, in this given position. Without loss of generality, assume the hyperrectangle starts at an odd point in the first dimension. Then:

$$\begin{aligned} &b(2x_1 + 1, s_1, x_2, s_2, \ldots, x_n, s_n) = \\ &b(2x_1 + 1, 1, x_2, s_2, \ldots, x_n, s_n) \\ &+ b(2x_1 + 2, s_1 - 1, x_2, s_2, \ldots, x_n, s_n) \end{aligned}$$

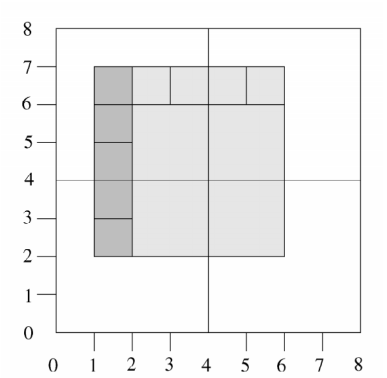


Fig. 2. Slicing from the left, when the rectangle starts at an odd point (the left slice is more heavily shaded).

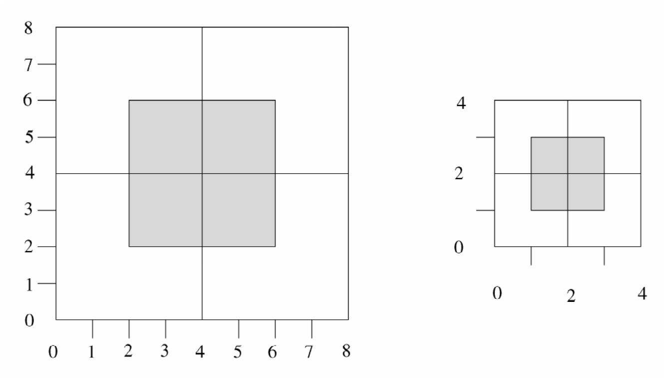


Fig. 3. Halving the granularity.

JUSTIFICATION. No block of size $2^k(k > 1)$ starts at an odd coordinate. Thus, the blocks of the sliced-off hyperplane can not be combined with the rest of the blocks of the hyperrectangle, to form larger blocks.

Clearly, the same principle can be used if the hyperrectangle **ends** at an odd point. Fig. 2 illustrates the slicing principle for a two-dimensional space.

OBSERVATION 2 (Unit). If any one dimension of a hyperrectangle is of unit size, then it can be covered only with unit size blocks. Thus, the number of blocks required to cover it is equal to its volume and is obtained as the product of the sides, independent of position. That is:

$$b(x_1, s_1, x_2, s_2, \ldots, x_m, 1, \ldots, x_n, s_n) = \prod_{i=1}^{n} s_i$$

JUSTIFICATION. The blocks will either start or end at an odd coordinate; thus, as in Observation 1, they cannot be combined to form larger blocks.

OBSERVATION 3 (Shrinking). If a hyperrectangle starts and ends at even numbers in all dimensions, then we can make the granularity coarser, maintaining the same number of blocks:

$$b(2x_1, 2s_1, 2x_2, 2s_2, \ldots, 2x_n, 2s_n) = b(x_1, s_1, x_2, s_2, \ldots, x_n, s_n)$$

JUSTIFICATION. Every block of size 2^k ($k \geq 2$) in the original address space corresponds 1-to-1 to a 2^{k-1} block of the "shrinked" space.

Fig. 3 gives a two-dimensional example of the idea.

The above observations, for $n = 2$ dimensional address space, have been used in [21] and [8].

3 SOLUTION FOR MAGIC HYPERRECTANGLES

DEFINITION 2. *A rectangle is called* ***magic*** *iff each side s_i is of the form $2^{m_i} - 1$.*

LEMMA 1 (*Magic hyperrectangles*). *If a rectangle is magic, then the number of blocks it decomposes to is* ***independent*** *of the position of the anchor:*

$$b(x_1, 2^{m_1} - 1, x_2, 2^{m_2} - 1, \ldots, x_n, 2^{m_n} - 1) = \text{constant} \quad \forall (x_1, x_2, \ldots, x_n)$$

PROOF. Without loss of generality, let s_1 be the smallest side of the hyperrectangle. For every dimension i, we can apply the Slicing Observation exactly once, because every side s_i is odd. After that, all the sides are even, and the anchor points are even as well. So we can apply the Shrinking Observation; the resulting rectangle will still be magic: for every dimension i, after slicing and shrinking we will have a side of size: $(s_i - 1)/2 = (2^{m_i} - 1 - 1)/2 = 2^{m_i - 1} - 1$. Applying this step inductively, and using the Unit Observation as the base case, we have the required lemma. □

COROLLARY 1. *For magic hyperrectangles, we have*:

$$\bar{b}(s_1, s_2, s_n) = b(x_1, s_1, x_2, s_2, \ldots, x_n, s_n) \quad \forall (x_1, x_2, \ldots, x_n)$$

Based on this corollary, we can quickly derive formulae for magic rectangles, bypassing (1).

3.1 Solution for Magic Hypercubes

Consider first a magic hyperrectangle with all its sides the same size, that is, a hypercube. Let this size be $2^m - 1$.

LEMMA 2. *For a magic hypercube the number of blocks is*:

$$\bar{b}(2^m - 1, \ldots, 2^m - 1) = (2^m - 1)^n - (2^n - 1) \sum_{t=1}^{m-1} (2^t - 1)^n$$

PROOF. Independent of the position of the anchor, we "slice off" one slice in each dimension and then shrink. Thus:

$$b(x_1, 2^m - 1, \ldots, x_n, 2^m - 1) = (2^m - 1)^n - (2^m - 2)^n + \bar{b}(2^{m-1} - 1, \ldots, 2^{m-1} - 1) \quad (2)$$

where the first two terms give the number of blocks contained in the slices, and the last term calculates the number of internal blocks. Solving this recursive relation (2) we have:

$$\bar{b}\left(2^m-1,\ \ldots,\ 2^m-1\right)=\sum_{t=1}^{m}\left(\left(2^t-1\right)^n-\left(2^t-2\right)^n\right) \quad (3)$$

or

$$\bar{b}\left(2^m-1,\ \ldots,\ 2^m-1\right)=\left(2^m-1\right)^n-\left(2^n-1\right)\sum_{t=1}^{m-1}\left(2^t-1\right)^n$$

This completed the proof. □

Next, we try to find an approximation for large values of m. Specifically, we try to relate it to the hypersurface S of the cube. Clearly, for a magic n-d hypercube of side $(2^m - 1)$, we have:

$$\begin{aligned} S &= 2n\left(2^m-1\right)^{(n-1)} \\ &\approx 2n2^{m(n-1)} \end{aligned} \quad (4)$$

COROLLARY 2. *For a magic hypercube with a large side* ($m \gg 1$), *the number of blocks is approximated by half of the hypersurface S, times a constant that depends on the dimensionality n*:

$$\bar{b}\left(2^m-1,\ \ldots,\ 2^m-1\right)\approx\frac{S}{2}\frac{2^{n-1}}{2^{n-1}-1}$$

PROOF. The tth term of (3) can be approximated by keeping the first two terms of each binomial expansion:

$$\left(2^t-1\right)^n-\left(2^t-2\right)^n\approx n2^{t(n-1)}$$

Adding the above terms, (3) gives

$$\begin{aligned}\bar{b}\left(2^m-1,\ \ldots,\ 2^m-1\right)&\approx\sum_{t=1}^{m}n2^{t(n-1)}\\ &\approx n\frac{2^{(n-1)(m+1)}}{2^{n-1}-1}\\ &\approx\frac{2^{n-1}}{2^{n-1}-1}n2^{(n-1)m}\end{aligned}$$

From (4), we have that, for large m:

$$\bar{b}\left(2^m-1,\ \ldots,\ 2^m-1\right)\approx\frac{2^{n-1}}{2^{n-1}-1}\frac{S}{2}$$

which completes the proof. □

We can examine some interesting cases:

- For $n \gg 1$, the factor vanishes to one, and the average number of blocks is approximately $S/2$.
- For 2d space, which is of much interest, the factor is 2; thus the average number of blocks is approximately the perimeter of the rectangle. More accurately, we obtain, from (3)

$$\begin{aligned}\bar{b}\left(2^m-1,\ 2^m-1\right)&=4\left(2^m-1\right)-3m\\ &=S-3m\\ &\approx S\end{aligned}$$

  This agrees with the result of Hunter and Steiglitz [13], stating that the number of quadtree nodes for a polygon is proportional to its perimeter.
- Similarly, for $n = 3$, the factor is $4/3$; working out the details from (3), we have

$$\bar{b}\left(2^m-1,\ 2^m-1,\ 2^m-1\right)=4*4^m-18*2^m+7m+14$$

  which leads to the approximation:

$$\begin{aligned}\bar{b}\left(2^m-1,\ 2^m-1,\ 2^m-1\right)&\approx 2/3\,S\\ &\approx S/2*4/3\end{aligned}$$

  That is, for a magic cube, the average number of blocks is $\approx 2/3$ of its surface.

3.2 Extension to Any Magic Hyperrectangle

For a magic hyperrectangle, without loss of generality, let $s_1 = 2^m - 1$ be its smallest side. Also, let $s_i = 2^{m+d_i} - 1$ where $d_i \geq 0$. In other words, we assume that: $d_1 = 0$.

LEMMA 3. *For any magic hyperrectangle the number of blocks is*:

$$\bar{b}\left(2^m-1,\ 2^{m+d_2}-1,\ \ldots,\ 2^{m+d_n}-1\right)=$$
$$\prod_{i=1}^{n}\left(2^{m+d_i}-1\right)-\left(2^n-1\right)\sum_{j=1}^{m-1}\prod_{i=1}^{n}\left(2^{m-j+d_i}-1\right)$$

PROOF. Using the Slicing and Shrinking Observations as we did for the magic hypercubes, we have:

$$\begin{aligned}&\bar{b}\left(2^m-1,\ 2^{m+d_2}-1,\ \ldots,\ 2^{m+d_n}-1\right)=\\ &\left(2^m-1\right)\left(2^{m+d_2}-1\right)\ldots\left(2^{m+d_n}-1\right)\\ &-\left(2^m-2\right)\left(2^{m+d_2}-2\right)\ldots\left(2^{m+d_n}-2\right)\\ &+\bar{b}\left(2^{m-1}-1,\ 2^{m-1+d_2}-1,\ \ldots,\ 2^{m-1+d_n}-1\right)\end{aligned}$$

Solving the recursion (it bottoms after m steps), we have:

$$\bar{b}\left(2^m-1,\ 2^{m+d_2}-1,\ \ldots,\ 2^{m+d_n}-1\right)=$$
$$\sum_{t=1}^{m}\left(\prod_{i=1}^{n}\left(2^{t+d_i}-1\right)-\prod_{i=1}^{n}\left(2^{t+d_i}-2\right)\right) \quad (5)$$

or

$$\bar{b}\left(2^m-1,\ 2^{m+d_2}-1,\ \ldots,\ 2^{m+d_n}-1\right)=$$
$$\prod_{i=1}^{n}\left(2^{m+d_i}-1\right)-\left(2^n-1\right)\sum_{j=1}^{m-1}\prod_{i=1}^{n}\left(2^{m-j+d_i}-1\right)$$ □

Again, we try to find an approximation for large m.

COROLLARY 3. *For large m*, (5) *can be approximated by*:

$$\bar{b}\left(2^m-1,\ 2^{m+d_2}-1,\ \ldots,\ 2^{m+d_n}-1\right)\approx\frac{S}{2}*\frac{2^{n-1}}{2^{n-1}-1}$$

PROOF. By using a reasoning similar to that of the case of square rectangles and by using the following expression for the hypersurface:

$$\begin{aligned}S&=2*\left(s_1^{-1}+\ldots+s_n^{-1}\right)*s_1*s_2*\ldots*s_n\\ &\approx 2*2^{-m}*\left(2^{-d_1}+\ldots+2^{-d_n}\right)*2^{nm}*2^{\sum_i d_i}\end{aligned}$$

The tth term of (5) can be approximated using only the first two terms of the expansion of each product. Thus:

$$\prod_{i=1}^{n}\left(2^{t+d_i}-1\right)-\prod_{i=1}^{n}\left(2^{t+d_i}-2\right)\approx$$
$$2^{(n-1)*t}*2^{\sum_i d_i}*\left(2^{-d_1}+\ldots+2^{-d_n}\right)$$
$$\approx 2^{(n-1)*t}*\frac{S}{2}*2^{-(n-1)*m}$$

Adding all the above terms from $t = 1$ to $t = m$ completes the proof. □

4 PROOF OF LINEARITY

In the previous section, we solved the problem for magic hyperrectangles. Here, we show how to solve the problem for arbitrary rectangles using *linear interpolation*.

LEMMA 4. *If $x_1 + s_1$ is odd, then*:

$$b(x_1,\ s_1,\ x_2,\ s_2,\ \ldots,\ x_n,\ s_n)=$$
$$b(x_1,\ s_1-1,\ x_2,\ s_2,\ \ldots,\ x_n,\ s_n)+C_1$$

where C_1 is a constant independent of the specific values of x_1 and s_1.

PROOF. The hypercubes to cover the incremental volume (shaded part, in Fig. 4a) are forced to be no more than one unit in the first dimension, and therefore one unit in each dimension. The number of hypercubes required is simply $s_2 \times s_3 \times \ldots \times s_n$, by following the Unit Observation. Define C_1 to be $\prod_{i=2}^{n} s_i$ to complete the proof. □

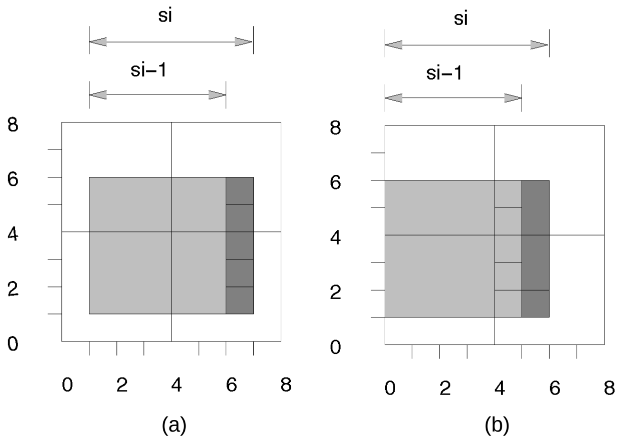


Fig. 4. Illustration for Lemmas 4-5: The incremental volume (with darker shade) results in a different number of blocks; however, the difference does not depend on s_i (= s_1, in this case). The rectangle ends at (a) an odd x_1 coordinate; (b) an even x_1 coordinate, which is not a multiple of 4.

LEMMA 5. *If $x_1 + s_1$ is even, but not divisible by four, then*:

$$b(x_1,\ s_1,\ x_2,\ s_2,\ \ldots,\ x_n,\ s_n)=$$
$$b(x_1,\ s_1-1,\ x_2,\ s_2,\ \ldots,\ x_n,\ s_n)+C_2$$

where C_2 is a constant independent of the specific values of x_1 and s_1.

PROOF. Now, some of the hypercubes already used to cover the hyperrectangle may be merged with the new layer added into larger blocks, two units on the side, on the even boundaries. The number of such mergers possible is determined solely by the size and position in dimensions 2, ..., n and is independent of x_1 and s_1. Call the number of additional blocks required C_2. □

Fig. 4b illustrates the situation: The larger rectangle will need two blocks of dimensions 2×2 and two blocks of dimensions 1×1, while the smaller rectangle will need five 1×1 blocks; however, the difference does not depend on the length s_1.

LEMMA 6. *If $x_1 + s_1$ is divisible by 2^{j-1} but not by 2^j, and $s_1 \geq 2^{j-1}$ then*:

$$b(x_1,\ s_1,\ x_2,\ s_2,\ \ldots,\ x_n,\ s_n)=$$
$$b(x_1,\ s_1-1,\ x_2,\ s_2,\ \ldots,\ x_n,\ s_n)+C_j$$

where C_j is a constant independent of the specific values of x_1 and s_1.

PROOF. Similar to Lemma 5. The additional condition imposing a minimum limit on s_1 is required since clearly no more mergers are possible beyond the length of the side s_1. Yet, the construction in the lemma could require mergers into blocks up to 2^{j-1} on the side. □

LEMMA 7. *If $x_1 + s_1$ is divisible by 2^j and $2^{m-1} \leq s_1 < 2^m \leq 2^j$, then*:

$$b(x_1,\ s_1,\ x_2,\ s_2,\ \ldots,\ x_n,\ s_n)=$$
$$b(x_1,\ s_1-1,\ x_2,\ s_2,\ \ldots,\ x_n,\ s_n)+C_m$$

where C_m is a constant independent of the specific values of x_1 and s_1.

PROOF. Similar to Lemma 6. Since s_1 is too small, the merger of blocks cannot continue until a side of 2^j is reached. Instead, it stops at an earlier point, and this point is determined by the magic points between which s_1 lies but is otherwise independent of s_1 and x_1. □

Now we are in the position to state the main theorems.

THEOREM 1. *For an arbitrary hyperrectangle with sides $(s_1, s_2, \ldots, s_n)$, where $2^{m-1} \leq s_1 < 2^m - 1$, we have*:

$$\bar{b}(s_1,\ s_2,\ \ldots,\ s_n)-\bar{b}(s_1-1,\ s_2,\ \ldots,\ s_n)=$$
$$\bar{b}(s_1+1,\ s_2,\ \ldots,\ s_n)-\bar{b}(s_1,\ s_2,\ \ldots,\ s_n)$$

PROOF. Consider the expected number of hypercube blocks to cover a hyperrectangle $\bar{b}(s_1-1,\ s_2,\ \ldots,\ s_n)$. If $s_1 - 1$ is increased to s_1, then, following the lemmas above, the increase in the value $\bar{b}(\ )$ is independent of the specific value of s_1, as long as a magic threshold is not crossed. Since the value of x_1 is arbitrary, independent of the specific value of s_1 we have that $x_1 + s_1$ is divisible by two with probability 1/2, by four with probability 1/4, and so on. Therefore, the number of additional blocks required is C_1 with probability $1/2$, C_2 with probability $1/2^2$, and so C_j with probability $1/2^j$, until C_m with probability $1/2^m$ and C_{m+1} with probability $1/2^m$. Thus, all cases are taken in consideration and their respective probabilities sum to unity. Note, also, that divisibility by higher powers of 2 does not alter the constant, and hence we can sum all these terms into a single term. Call this summation C:

$$C = C_1/2 + C_2/4 + \ldots + C_m/2^m + C_{m+1}/2^m \quad (6)$$

Exactly the same summation C is obtained if s_1 is now increased to $s_1 + 1$. Thus, the theorem is established. □

In other words, the function $b(s_1, s_2, \ldots, s_n)$ is *piece-wise linear* on its arguments, with "break points" whenever a value s_i is a magic number. Theorem 1 can be used to do linear interpolation, as follows: Let s_i be the side of the rectangle on the ith dimension, and let m_i and M_i be the magic numbers that surround s_i, that is

$$m_i = 2^k - 1 \le s_i < 2^{k+1} - 1 = M_i \quad (7)$$

Then, we have:

$$\begin{aligned}\bar{b}(s_1, s_2, \ldots, s_i, \ldots, s_n) = {} & \bar{b}(s_1, s_2, \ldots, m_i, \ldots, s_n) * \\ & (M_i - s_i)/(M_i - m_i) + \\ & \bar{b}(s_1, s_2, \ldots, M_i, \ldots, s_n) * \\ & (s_i - m_i)/(M_i - m_i)\end{aligned} \quad (8)$$

Based on that, we can compute the value of $\bar{b}(\ )$ at any point. The next theorem gives the details.

THEOREM 2. *Let $R = s_1 \times s_2 \ldots \times s_n$ be a hyperrectangle; let m_1 and M_1 be the magic values that contain s_1 (i.e., $m_1 = 2^j - 1 \le s_1 < 2^{j+1} - 1 = M_1$), with similar definitions for m_i and M_i. There are 2^n magic rectangles that we can generate (for each dimension i, we have two choices: m_i and M_i, for a total of 2^n choices). The average number of blocks for R is determined by a linear interpolation among the values of the above 2^n magic rectangles.*

PROOF. Consider each dimension in turn and increase the size from m_i to M_i in steps of 1. Each step increases the average number of blocks by the same amount, on account of Theorem 1. While Theorem 1 was established for the first dimension, by arguments of symmetry it holds for all other dimensions as well. Therefore, the increase from m_i to s_i is a linear interpolation of the increase from m_i to M_i. The order in which the dimensions are considered is immaterial. □

Table 2 shows the values for $\bar{b}(\ )$ for the two-dimensional case, with boldface numbers for the magic rectangles. Notice that the rest of the numbers can be derived by linear interpolation among the four magic rectangles nearest to the point of interest (e.g., for the $\bar{b}(\ )(5, 2)$, the corresponding magic rectangles are (3, 1), (3, 3), (7, 1), (7, 3)). In the next section, we illustrate Theorem 2, deriving the formulae for $\bar{b}(\ )$ for two-dimensional and three-dimensional spaces. We also give some examples of how to do the interpolation.

5 INTERESTING SPECIAL CASES: TWO- AND THREE-DIMENSIONAL RECTANGLES

In this section, we illustrate the steps of the lemmas and theorems of the previous section by deriving closed-form exact formulae for the average number of blocks a two-dimensional and a three-dimensional rectangle. Following the steps of the previous section, we first calculate the number of blocks for any magic rectangular object, and then we give exact formulae for any (nonmagic) rectangular object.

5.1 Two-Dimensional Rectangles

This case has been analyzed in [8]. Here, we show how those results can be derived as special cases of the Theorems and Lemmas of the previous section.

LEMMA 8. *The average number of blocks $\bar{b}(\ )$ that a magic rectangle in two-dimensional space decomposes into is:*

$$\bar{b}\left(2^m - 1,\ 2^{m+d_2} - 1\right) = 2\left(2^m - 1\right)\left(2^{d_2} + 1\right) - 3m \quad (9)$$

PROOF. From (5) we have:

$$\bar{b}\left(2^m - 1,\ 2^{m+d_2} - 1\right) = \sum_{t=1}^{m}\left(\prod_{i=1}^{2}\left(2^{t+d_i} - 1\right) - \prod_{i=1}^{2}\left(2^{t+d_i} - 2\right)\right) \quad (10)$$

It is sufficient to prove that the right hand parts of relations (9) and (10) are equal. The proof follows by induction on m. For $m = 1$ both sides of the equation are equal to: $2^{d_2+1} - 1$. For $m = 2$ both sides are equal to: $3 * 2^{d_2+1}$. We assume that the above relation holds for $m = k$:

$$\sum_{t=1}^{k}\left(\prod_{i=1}^{2}\left(2^{t+d_i} - 1\right) - \prod_{i=1}^{2}\left(2^{t+d_i} - 2\right)\right) = 2\left(2^k - 1\right)\left(2^{d_2} + 1\right) - 3k$$

We will prove that it holds for $m = k + 1$:

$$\sum_{t=1}^{k+1}\left(\prod_{i=1}^{2}\left(2^{t+d_i} - 1\right) - \prod_{i=1}^{2}\left(2^{t+d_i} - 2\right)\right) = 2\left(2^{k+1} - 1\right)\left(2^{d_2} + 1\right) - 3(k+1)$$

It is sufficient to prove that the left-hand part of the above equation is:

$$\begin{aligned}& 2\left(2^{k+1} - 1\right)\left(2^{d_2} + 1\right) - 3(k+1) = \\ & 2\left(2^k - 1\right)\left(2^{d_2} + 1\right) - 3k + \left(2^{k+1} - 1\right)\left(2^{k+1+d_2} - 1\right) - \\ & \left(2^{k+1} - 2\right)\left(2^{k+1+d_2} - 2\right)\end{aligned}$$

After some simple algebra, we derive that the above lemma holds. □

TABLE 2
NUMBER OF BLOCKS FOR TWO-DIMENSIONAL RECTANGLES

s_1 \ s_2	1	2	3	4	5	6	7	8
1	**1**	2	**3**	4	5	6	**7**	8
2	2	3.25	4.5	5.75	7	8.25	9.5	10.75
3	**3**	4.5	**6**	<u>7.5</u>	9	10.5	**12**	13.5
4	4	5.75	7.5	<u><u>9.0625</u></u>	10.625	12.1875	13.75	15.3125
5	5	7	9	10.625	12.25	13.875	15.5	17.125
6	6	8.25	10.5	12.1875	13.875	15.5625	17.25	18.9375
7	**7**	9.5	**12**	<u>13.75</u>	15.5	17.25	**19**	20.75
8	8	10.75	13.5	15.3125	17.125	18.9375	20.75	22.515625

Notes: Magic rectangles are in boldface. The underlined entries are examined in the examples.

Table 2 gives the average number of blocks a rectangle is decomposed into, when its sides s_1 and s_2 are smaller than nine. The entries were calculated by exhaustive enumeration, using the definition of (1). Entries corresponding to magic rectangles are in boldface. The remaining entries can be derived by a linear interpolation among the appropriate magic rectangles. Next, we illustrate how the linear interpolation is done:

EXAMPLE 1**.** The entry for $s_1 = 3$, $s_2 = 4$ is computed as follows: s_1 is already a magic number; for s_2, the enclosing magic numbers are $3(=2^2-1)$ and $7(=2^3-1)$. Thus, we need to interpolate only on the second axis:

$$\begin{aligned}\bar{b}(3,\ 4) &= \bar{b}(3,\ 3)*(7-4)/(7-3)+\bar{b}(3,\ 7)*(4-3)/(7-3)\\ &= \frac{1}{7-3}(6*3+12*1)\\ &= 7.5\end{aligned}$$

EXAMPLE 2**.** Consider the entry $\bar{b}(7,\ 4)$ (underlined in Table 2). Again, $s_1 = 7$ is a magic number; the s_2 number is surrounded by the magic numbers 3 and 7. Thus, we need to interpolate only on the second axis:

$$\begin{aligned}\bar{b}(7,4) &= \bar{b}(7,3)*(7-4)/(7-3)+\bar{b}(7,7)*(4-3)/(7-3)\\ &= \frac{1}{7-3}(12*3+19*1)\\ &= 13.75\end{aligned}$$

EXAMPLE 3**.** Consider the entry $\bar{b}(4,\ 4)$ (doubly underlined, in Table 2). The enclosing magic numbers for both axis are "3" and "7." The interpolation would give

$$\begin{aligned}\bar{b}(4,\ 4) &= \bar{b}(3,\ 3)*\frac{7-4}{7-3}*\frac{7-4}{7-3}+\bar{b}(3,\ 7)*\frac{7-4}{7-3}*\frac{4-3}{7-3}\\ &\quad+\bar{b}(7,\ 3)*\frac{4-3}{7-3}*\frac{7-4}{7-3}+\bar{b}(7,\ 7)*\frac{4-3}{7-3}*\frac{4-3}{7-3}\\ &= \frac{1}{(7-3)*(7-3)}(6*3*3+12*3*1+\\ &\qquad 12*1*3+19*1*1)\\ &= \frac{1}{16}(54+36+36+19)\\ &= 9.0625\end{aligned}$$

Equivalently, we could have done a linear interpolation among the values $\bar{b}(3,4)$ and $\bar{b}(7,4)$, that we have already computed in Examples 1 and 2:

$$\begin{aligned}\bar{b}(4,\ 4) &= \bar{b}(3,\ 4)*(7-4)/(7-3)+\bar{b}(7,\ 4)*(4-3)/(7-3)\\ &= 9.0625\end{aligned}$$

Next, we trace the steps of the proof of Theorem 1, giving a closed formula for the constant C.

LEMMA 9**.** *Given that the rectangle with sides $(s_1,\ s_2)$ is magic, then the average number of blocks for a rectangle with sides $(s_1 + 1,\ s_2)$ is:*

$$\bar{b}(s_1+1, s_2) = \bar{b}(s_1, s_2) + 2^{m+d_2-2max} - 3*2^{-max} + 2$$

where $max = \lfloor \log(\min(s_1 + 1,\ s_2)) \rfloor$, log is the base-2 logarithm and $s_i = 2^{m+d_i} - 1$.

PROOF**.** See Appendix A. □

It is evident that in a two-dimensional space the constant C of Theorem 1 is given by:

$$C = 2^{m+d_2-2max} - 3*2^{-max} + 2$$

We can rewrite this expression as:

$$(s_2+1)*2^{-2max} - 3*2^{-max} + 2 \tag{11}$$

from which we can see that this constant C is independent of x_1, s_1. We use the symbol $C(s_i)$ to emphasize the dependency on s_i: Thus:

$$C(s_i) = (s_i+1)*2^{-2max} - 3*2^{-max} + 2 \tag{12}$$

5.2 Three-Dimensional Rectangles

In this subsection, we examine the case of a parallelepiped and we derive a formula for the constant C of Theorem 2.

LEMMA 10**.** *The average number of blocks that a magic parallelepiped decomposes into is:*

$$\bar{b}\left(2^m - 1,\ 2^{m+d_2} - 1,\ 2^{m+d_3} - 1\right) =$$
$$\sum_{t=1}^{m}\left(\prod_{i=1}^{3}\left(2^{t+d_i} - 1\right) - \prod_{i=1}^{3}\left(2^{t+d_i} - 2\right)\right)$$
$$= \frac{4}{3}\left(2^{2m} - 1\right)\left(2^{d_2} + 2^{d_3} + 2^{d_2+d_3}\right)$$
$$-6\left(2^m - 1\right)\left(1 + 2^{d_2} + 2^{d_3}\right) + 7m$$

PROOF. By induction on m. □

LEMMA 11. *Given that three-dimensional parallelepiped with sides* (s_1, s_2, s_3) *is magic, then the average number of blocks for a parallelepiped with sides* ($s_1 + 1$, s_2, s_3) *is*:

$$\bar{b}\left(s_1 + 1,\ s_2,\ s_3\right) = \bar{b}\left(s_1,\ s_2,\ s_3\right) + 2^{2m+d_2+d_3-3max} +$$
$$8 - \frac{7}{3}\left(2^{max+1} + 2^{-max}\right) - 2^m\left(2^{d_2} + 2^{d_3}\right)\left(\frac{7}{9}2^{-2max} - \frac{7}{6}max\right) + \frac{2}{9}$$

where $max = \lfloor \log(\min(s_1 + 1,\ s_2,\ s_3)) \rfloor$ *and, as before,* $s_i = 2^{m+d_i} - 1 (i = 1,\ 2,\ 3)$.

PROOF. See Appendix B. □

From Lemma 11, we understand why the constant C of Theorem 1 is a quantity independent of s_1. However, we observe that it depends on the other two sides s_2 and s_3. This is the reason why for the case of a three-dimensional space we have to denote this quantity as $C(s_i, s_j)$, where:

$$C\left(s_i,\ s_j\right) = \left(s_i + 1\right) * \left(s_j + 1\right) * 2^{-3max} + \left(s_i + s_j + 2\right) *$$
$$\left(\frac{7}{9}2^{-2max} - \frac{7}{6}max + \frac{2}{9}\right) - \frac{7}{3}\left(2^{max+1} + 2^{-max}\right) + 8 \qquad (13)$$

Table 3 gives the average number of blocks a parallelepiped is composed of, when its sides are smaller than 6. Entries in boldface correspond to magic parallelepipeds. All the entries have been computed using exhaustive enumeration, from the definition of (1).

6 DISCUSSION AND CONCLUSIONS

We have examined the problem of determining the number of quadtree blocks that an n-dimensional rectangle will be decomposed into on the average. There are two interesting observations:

- Our approach (Theorem 2 and (5)) generalizes all the older approaches on two-dimensional rectangles [6], [8], [20]. For $n = 2$ dimensions, our formula reduces to the corresponding formula of [8], which was shown to include the formulas in [6], [20] for the average number of blocks.
- It generalizes the observation of Hunter and Steiglitz [13] that the expected number of quadtree blocks is proportional to the perimeter of the polygon. Our formula shows that, for two-dimensional rectangles, the expected number of quadtree blocks is approximately the perimeter of the rectangle, while for higher dimensionalities $n \gg 1$, it is roughly half of the hypersurface.

TABLE 3
NUMBER OF BLOCKS
FOR THREE-DIMENSIONAL PARALLELEPIPEDS

s_1 \ s_2	1	2	3	4	5	s_3
1	**1**	2	**3**	4	5	1
	2	4	6	8	10	2
	3	6	**9**	12	15	3
	4	8	12	16	20	4
	5	10	15	20	25	5
2	2	4	6	8	10	1
	4	7.125	10.25	13.375	16.5	2
	6	10.25	14.5	18.75	23	3
	8	13.375	18.75	24.125	29.5	4
	4	16.5	23	29.5	36	5
3	**3**	6	**9**	12	15	1
	6	10.25	14.5	18.75	23	2
	9	14.5	**20**	25.5	31	3
	12	18.75	25.5	32.25	39	4
	15	23	31	39	47	5
4	4	8	12	16	20	1
	8	13.375	18.75	24.125	29.5	2
	12	18.75	25.5	32.25	39	3
	16	24.125	32.25	40.265625	48.28125	4
	20	29.5	39	48.28125	57.5625	5
5	5	10	15	20	25	1
	10	16.5	23	29.5	36	2
	15	23	31	39	47	3
	20	29.5	39	48.28125	57.5625	4
	25	36	47	57.5625	68.125	5

Magic parallelepipeds are in boldface.

The contributions of this paper are both practical and theoretical. From the practical point of view, the number of quadtree blocks of a decomposition is important, because it determines the number of nodes that a main-memory-based quadtree will require; the number of entries in a linear quadtree that will be required; also, the number of pieces that a range query will be decomposed into (which will be proportional to the response time for this query).

From the theoretical point of view, it proposes a methodology which we believe will be useful in the analysis of other quadtree-related methods (e.g., methods using space-filling curves, such as the z-ordering [17], Gray codes [7], or the Hilbert curve [10]). The methodology consists of two steps:

Step 1 Solve the problem for the “magic” rectangles (which is easy)

Step 2 Show that the formula for an arbitrary rectangle can be derived by linear interpolation from suitable “magic” rectangles.

Future work includes the extension of this method for the analysis of rectilinear polygons (including concave ones), as well as the analysis for space filling curves for two-dimensional and n-dimensional spaces.

APPENDIX A
LEMMA FOR THE TWO-DIMENSIONAL CASE

LEMMA 9. *Given that the rectangle with sides* (s_1, s_2) *is magic, then the average number of blocks for a rectangle with sides* ($s_1 + 1$, s_2) *is*:

$$\bar{b}\left(s_1 + 1,\ s_2\right) = \bar{b}\left(s_1,\ s_2\right) + 2^{m+d_2-2max} - 3 * 2^{-max} + 2$$

where $max = \lfloor \log(\min(s_1 + 1, s_2)) \rfloor$.

PROOF. First, let's assume that the rectangle does not wrap around the edges ($x_1 + s_1, x_2 + s_2 \le K$). With probability 1/2 we have: $(x_1 + s_1 + 1) \bmod 2 \ne 0$ (the end point in the first dimension is an odd number). Then, according to the Slicing and Unit Observations, the new number of blocks is:

$$b(x_1, s_1 + 1, x_2, s_2) = \bar{b}(s_1, s_2) + s_2$$

With probability equal to 1/4 we have: $(x_1 + s_1 + 1) \bmod 2 = 0$ but $(x_1 + s_1 + 1) \bmod 4 \ne 0$. Then:

$$b(x_1, s_1 + 1, x_2, s_2) = \bar{b}(s_1, s_2) + s_2 - \left(\left\lfloor \frac{x_2 + s_2}{2} \right\rfloor - \left\lceil \frac{x_2}{2} \right\rceil\right)\left(2^1 + 2^1 - 1\right) \quad (14)$$

The product in the previous relation stands for the number of blocks we have to subtract because mergings have been performed. The first two terms in the second parenthesis respectively stand for the number of pixels of the original magic rectangle (2^1) and for the number of the pixels of the additional slice (2^1) that merge in one 2×2 block. Thus, the third term in the parenthesis (i.e., –1) stands for the greater formed block we have to take into account. The first parenthesis of the product gives the number of greater blocks that may be formed.

Since s_2 is an odd integer (of the form $2^{m+d_2} - 1$), it is easily verifiable that:

$$\left\lfloor \frac{x_2 + s_2}{2} \right\rfloor - \left\lceil \frac{x_2}{2} \right\rceil = \left\lfloor \frac{s_2}{2} \right\rfloor$$

Thus, relation (14) becomes:

$$b(x_1, s_1 + 1, x_2, s_2) = \bar{b}(s_1, s_2) + s_2 - \left\lfloor \frac{s_2}{2} \right\rfloor\left(2^1 + 2^1 - 1\right) \quad (15)$$

With probability equal to 1/8 we have: $(x_1 + s_1 + 1) \bmod 4 = 0$ but $(x_1 + s_1 + 1) \bmod 8 \ne 0$. Then:

$$\bar{b}(x_1, s_1 + 1, x_2, s_2) = \bar{b}(s_1, s_2) + s_2 - \left(\left\lfloor \frac{x_2 + s_2}{4} \right\rfloor - \left\lceil \frac{x_2}{4} \right\rceil\right)\left(2^1 + 2^2 + 2^2 - 1\right) - \left\lfloor \frac{s_2 - 4\left(\left\lfloor \frac{x_2 + s_2}{4} \right\rfloor - \left\lceil \frac{x_2}{4} \right\rceil\right)}{2} \right\rfloor\left(2^1 + 2^1 - 1\right) \Rightarrow$$

$$b(x_1, s_1 + 1, x_2, s_2) = \bar{b}(s_1, s_2) + s_2 - \left\lfloor \frac{s_2}{4} \right\rfloor\left(2^1 + 2^2 + 2^2 - 1\right) - \left\lfloor \frac{s_2 - 4\left(\left\lfloor \frac{s_2}{4} \right\rfloor\right)}{2} \right\rfloor\left(2^1 + 2^1 - 1\right)$$

Since: $\sum_{j=1}^{i} 2^j = 2\left(2^i - 1\right)$, the above relation becomes:

$$b(x_1, s_1 + 1, x_2, s_2) = \bar{b}(s_1, s_2) + s_2 - \left\lfloor \frac{s_2}{4} \right\rfloor 3\left(2^2 - 1\right)\left\lfloor \frac{s_2 \bmod 4}{2} \right\rfloor 3\left(2^1 - 1\right)$$

Suppose that: $8 \le \min(s_1 + 1, s_2) < 16$. Then, with probability equal to 1/8, we have: $(x_1 + s_1 + 1) \bmod 4 = 0$ and $(x_1 + s_1 + 1) \bmod 8 = 0$. Thus:

$$b(x_1, s_1 + 1, x_2, s_2) = \bar{b}(s_1, s_2) + s_2 - \left\lfloor \frac{s_2}{8} \right\rfloor 3\left(2^3 - 1\right) - \left\lfloor \frac{s_2 \bmod 8}{4} \right\rfloor 3\left(2^2 - 1\right) - \left\lfloor \frac{s_2 \bmod 4}{2} \right\rfloor 3\left(2^1 - 1\right)$$

Following this reasoning, similar expressions can be derived for large values of s_1, s_2 and such that $K/2 < \min(s_1 + 1, s_2) \le K$, $\forall K = 2^k$.

Secondly, suppose that the rectangle wraps around in one dimension only (i.e., $x_2 + s_2 > K$). Then, expression (14) should be rewritten as:

$$b(x_1, s_1 + 1, x_2, s_2) = \bar{b}(s_1, s_2) + s_2 - \left(\left\lfloor \frac{x_2 + s_2 - K}{2} \right\rfloor + \left\lfloor \frac{K - x_2}{2} \right\rfloor\right)\left(2^1 + 2^1 - 1\right)$$

However, the latter expression may be reduced to (15). This way, the set of equations derived by assuming that the rectangle wraps around only one edge reduces to the set of equations produced to describe the no-wrapping rectangle. The same result holds even if the rectangle wraps around both edges.

Thus, by considering all the positions possibly taken by the end point in the 1st dimension, we conclude to the following expression:

$$\bar{b}(s_1 + 1, s_2) = \bar{b}(s_1, s_2) + s_2 - \sum_{i=1}^{max-1} \frac{1}{2^{i+1}}\left(\left\lfloor \frac{s_2}{2^i} \right\rfloor 3\left(2^i - 1\right) + \sum_{j=2}^{i} \left\lfloor \frac{s_2 \bmod 2^j}{2^{j-1}} \right\rfloor 3\left(2^{j-1} - 1\right)\right) - \frac{1}{2^{max}}\left(\left\lfloor \frac{s_2}{2^{max}} \right\rfloor 3\left(2^{max} - 1\right) + \sum_{j=2}^{max} \left\lfloor \frac{s_2 \bmod 2^j}{2^{j-1}} \right\rfloor 3\left(2^{j-1} - 1\right)\right)$$

which is averaged and independent of the anchor point (x_1, x_2). Since: $s_2 = 2^{m+a_2} - 1$, the floor functions are simplified to unity and after some algebra on geometric series the lemma is proved. Notice, also, that if $d_2 > 0$ then $max = \log(s_1 + 1) = m$, whereas if $d_2 = 0$ then $max = \log(s_2) = m - 1$. □

APPENDIX B
LEMMA FOR THE THREE-DIMENSIONAL CASE

LEMMA 11. *Given that three-dimensional parallelepiped with sides* (s_1, s_2, s_3) *is magic, then the average number of blocks for a parallelepiped with sides* $(s_1 + 1, s_2, s_3)$ *is:*

$$\bar{b}(s_1+1,\ s_2,\ s_3) = \bar{b}(s_1,\ s_2,\ s_3) + 2^{2m+d_2+d_3-3max} +$$
$$8 - \frac{7}{3}\left(2^{max+1} + 2^{-max}\right) -$$
$$2^m\left(2^{d_2} + 2^{d_3}\right)\left(\frac{7}{9}2^{-2max} - \frac{7}{6}max + \frac{2}{9}\right)$$

where $max = \lfloor \log(\min(s_1+1,\ s_2,\ s_3)) \rfloor$ *and, as before,* $s_i = 2^{m+d_i} - 1 (i = 1, 2, 3)$.

PROOF. We follow the same reasoning as for the case of Lemma 9. If $(x_1 + s_1 + 1) \bmod 2 \neq 0$ (which may happen with probability 1/2), then according to the Slicing and Unit Observations we calculate the new number of blocks to be:

$$b(x_1,\ s_1+1,\ x_2,\ s_2,\ x_3,\ s_3)$$
$$= \bar{b}(s_1,\ s_2,\ s_3) + s_2 * s_3$$

If $(x_1 + s_1 + 1) \bmod 2 = 0$ but $(x_1 + s_1 + 1) \bmod 4 \neq 0$, then with probability equal to 1/4 we have:

$$b(x_1,\ s_1+1,\ x_2,\ s_2,\ x_3,\ s_3) = \bar{b}(s_1,\ s_2,\ s_3) + s_2 * s_3 -$$
$$\left\lfloor \frac{s_2}{2} \right\rfloor \left\lfloor \frac{s_3}{2} \right\rfloor \left(\left(2^1\right)^2 + \left(2^1\right)^2 - 1 \right)$$

In an analogous manner, with probability equal to 1/8 (for the case $(x_1 + s_1 + 1) \bmod 4 = 0$ but $(x_1 + s_1 + 1) \bmod 8 \neq 0$), we have:

$$b(x_1,\ s_1+1,\ x_2,\ s_2,\ x_3,\ s_3) = \bar{b}(s_1,\ s_2,\ s_3) + s_2 * s_3 -$$
$$\left\lfloor \frac{s_2}{4} \right\rfloor \left\lfloor \frac{s_3}{4} \right\rfloor \left(\left(2^1\right)^2 + \left(2^2\right)^2 + \left(2^2\right)^2 - 1 \right) - \left(\left(2^1\right)^2 + \left(2^1\right)^2 - 1 \right)$$
$$= \bar{b}(s_1,\ s_2,\ s_3) + s_2 * s_3 - \left\lfloor \frac{s_2}{4} \right\rfloor \left\lfloor \frac{s_3}{4} \right\rfloor \left(\frac{7}{3}\left(4^2 - 1\right) \right) \left(\frac{7}{3}\left(4^1 - 1\right) \right)$$

Thus, by generalizing and considering all the positions possibly taken by the end point in the first dimension, we conclude to the following expression:

$$\bar{b}(s_1+1,\ s_2,\ s_3) = \bar{b}(s_1,\ s_2,\ s_3) + s_2 * s_3 -$$
$$\sum_{i=1}^{max-1} \frac{1}{2^{i+1}} \left(\left\lfloor \frac{s_2}{2^i} \right\rfloor \left\lfloor \frac{s_3}{2^i} \right\rfloor \frac{7}{3}\left(4^i - 1\right) + \sum_{j=2}^{i} \frac{7}{3}\left(4^{j-1} - 1\right) \right) -$$
$$\frac{1}{2^{max}} \left(\left\lfloor \frac{s_2}{2^{max}} \right\rfloor \left\lfloor \frac{s_3}{2^{max}} \right\rfloor \frac{7}{3}\left(4^{max} - 1\right) + \sum_{j=2}^{max} \frac{7}{3}\left(4^{j-1} - 1\right) \right)$$

which is averaged and independent of the anchor point (x_1, x_2, x_3). After some algebra the expression of the lemma follows. □

ACKNOWLEDGMENTS

The authors would like to thank the anonymous referees, and especially referees #4 and #7, for their extremely careful review of the paper and their perceptive feedback. The work of Christos Faloutsos was partially supported by the National Science Foundation (IRI-8958546 and IRI-9205273), with matching funds from EMPRESS Software Inc. and Thinking Machines Inc. The work of Yannis Manolopoulos was performed with the Dept. of Computer Science and the Inst. for Systems Research (ISR) at the University of Maryland, College Park, MD.

REFERENCES

[1] W.G. Aref and H. Samet, "Optimization Strategies for Spatial Query Processing," *Proc. VLDB (Very Large Date Bases)*, pp. 81-90, Sept. 1991.

[2] M. Arya, W. Cody, C. Faloutsos, J. Richardson, and A. Toga, "Qbism: Extending a dbms to Support 3d Medical Images," *Proc. 10th Int'l Conf. Data Eng. (ICDE)*, pp. 314-325, Feb. 1994.

[3] D. Ballard and C. Brown, *Computer Vision*. Prentice Hall, 1982.

[4] J.L. Bentley, "Multidimensional Binary Search Trees Used for Associative Searching," *Comm. ACM*, vol. 18, no. 9, pp. 509-517, Sept. 1975.

[5] Mathematical Committee on Physical and NSF Engineering Sciences, *Grand Challenges: High Performance Computing and Communications*, National Science Foundation, 1992. The FY 1992 U.S. Research and Development Program.

[6] C.R. Dyer, "The Space Efficiency of Quadtrees," *Computer Graphics and Image Processing*, vol. 19, no. 4, pp. 335-348, Aug. 1982.

[7] C. Faloutsos, "Gray Codes for Partial Match and Range Queries," *IEEE Trans. Software Eng.*, vol. 14, no. 10, pp. 1,381-1,392, Oct. 1988. Early version available as UMIACS-TR-87-4, also CS-TR-1796.

[8] C. Faloutsos, "Analytical Results on the Quadtree Decomposition of Arbitrary Rectangles," *Pattern Recognition Letters*, vol. 13, no. 1, pp. 31-40, Jan. 1992.

[9] C. Faloutsos and Y. Rong, "Dot: A Spatial Access Method Using Fractals," *Proc. IEEE Conf. Data Eng.*, pp. 152-159, Kobe, Japan, Apr. 1991. Early version available as UMIACS-TR-89-31, CS-TR-2214.

[10] C. Faloutsos and S. Roseman, "Fractals for Secondary Key Retrieval, *Proc. Eighth ACM SIGACT-SIGMOD-SIGART Symp. Principles of Database Systems (PODS)*, pp. 247-252, Mar. 1989. Also available as UMIACS-TR-89-47 and CS-TR-2242.

[11] R.A. Finkel and J.L. Bentley, "Quadtrees: A Data Structure for Retrieval on Composite Keys," *ACTA Information*, vol. 4, no. 1, pp. 1-9, 1974.

[12] K. Hinrichs and J. Nievergelt, "The Grid File: A Data Structure to Support Proximity Queries on Spatial Objects," *Proc. WG'83 (Int'l Workshop Graph Theoretic Concepts in Computer Science)*, pp. 110-113, 1983.

[13] G.M. Hunter and K. Steiglitz, "Operations on Images Using Quad Trees," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 1, no. 2, pp. 145-153, Apr. 1979.

[14] H.V. Jagadish, "Linear Clustering of Objects with Multiple Attributes," *Proc. ACM SIGMOD Conf.*, pp. 332-342, May 1990.

[15] H.V. Jagadish, "Spatial Search with Polyhedra," *Proc. Sixth IEEE Int'l Conf. Data Eng.*, Feb. 1990.

[16] C.P. Kolovson and M. Stonebraker, "Segment Indexes: Dynamic Indexing Techniques for Multi-Diminsional Interval Data," *Proc. ACM SIGMOD*, pp. 138-147, May, 1991.

[17] J. Orenstein, "Spatial Query Processing in an Object-Oriented Database System," *Proc. ACM SIGMOD*, pp. 326-336, May 1986.

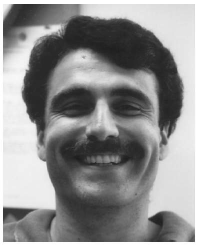
[18] J.A. Orenstein and F.A. Manola, "Probe Spatial Data Modeling and Query Processing in an Image Database Application," *IEEE Trans. Software Eng.*, vol. 14, no. 5, pp. 611-629, May 1988.

[19] Y. Rong and C. Faloutsos, "Analysis of the Clustering Property of Peano Curves," Technical Report CS-TR-2792, UMIACS-TR-91-151, Univ. of Maryland, Dec. 1991.

[20] C.A. Shaffer, "A Formula for Computing the Number of Quadtree Node Fragments Created by a Shift," *Pattern Recognition Letters*, vol. 7, no. 1, pp. 45-49, Jan. 1988.

[21] C.A. Shaffer, H. Samet, and R.C. Nelson, "Quilt: A Geographic Information System Based on Quadtrees," Technical Report CS-TR-1885.1, Univ. of Maryland, Dept. of Computer Science, July 1987.

[22] M. White, "N-Trees: Large Ordered Indexes for Multi-Dimensional Space," Application Mathematics Research Staff, Statistical Research Div., U.S. Bureau of the Census, Dec. 1981.



Christos Faloutsos received the BSc degree in electrical engineering (1981) from the National Technical University of Athens, Greece, and the MSc and PhD degrees in computer science from the University of Toronto, Canada. He is an associate professor in the Department of Computer Science at the University of Maryland, College Park. In 1989, he received the Presidential Young Investigator Award from the National Science Foundation. His research interests include physical database design, searching methods for text, geographic information systems, and indexing methods for medical and multimedia databases. He is a member of the IEEE.

H.V. Jagadish received his PhD from Stanford University in 1985, and since then, has been with the Computing Systems Research laboratory at AT&T Bell Laboratories in Murray Hill, New Jersey. His research interests include the management of multimedia information and the use of database technology for communications networks. He is a member of the IEEE.

Yannis Manolopoulos received his five-year diploma in electrical engineering and his PhD degree in computer engineering from the Aristotle University in Thessaloniki, Greece, in 1981 and 1986, respectively. He has been on the academic staff at Aristotle University since 1987 and is currently an associate professor with the Department of Informatics. He spent two sabbatical years at the University of Toronto, Canada, and the University of Maryland at College Park. He is co-author of more than 50 articles in refereed journals and conference proceedings. He is also the author of two textbooks (in Greek) on data structures and file structures which are recommended in the vast majority of Computer Science/Engineering departments in Greece. His research interests include multimedia, spatial, temporal, and text databases. He is a member of the ACM, IEEE Computer Society, Greek Computer Society, and Technical Chamber of Greece.