National Center for Geographic Information and Analysis

A Hierarchical
Spatial Data Structure

for Global Geographic Information Systems

By:

Michael F. Goodchild & Yang Shiren

National Center for Geographic Information & Analysis
Geography Department, University of California
Santa Barbara, California 93106

Technical Paper 89-5

June, 1989

Simonett Center for Spatial Analysis
University of California
35 10 Phelps Hall
Santa Barbara, CA 931064060
Office (805) 8938224
Fax (805) 8938617
ncgia@ncgia.ucsb.edu

State University of New York
301 Wilkeson Quad, Box 610023
Buffalo NY 142610001
Office (716) 6452545
Fax (716) 6455957
ncgia@ubvms.cc.buffalo.edu

University of Maine
348 Boardman Hall
Orono ME 044695711
Office (207) 5812149
Fax (207) 5812206
ncgia@spatial.maine.edu

A HIERARCHICAL SPATIAL DATA STRUCTURE FOR GLOBAL GEOGRAPHIC INFORMATION SYSTEM

Michael F. Goodchild[1] Yang Shiren[1]

Abstract

Hierarchical spatial data structures offer distinct advantages of data compression and fast access, but are difficult to adapt to the globe. Following Dutton (1984, 1988a, 1988b), we propose to project the globe onto an octahedron, and then to recursively subdivide each of its eight triangular faces into four triangles. We provide procedures for addressing the hierarchy, and for computing addresses in the hierarchical structure from latitude and longitude, and vice versa. At any level in the hierarchy the finite elements are all triangles, but are only approximately equal in area and shape; we provide methods for computing area, and for finding the addresses of neighboring triangles.

Introduction

Hierarchical spatial data structures (HSDSs) such as the quadtree and octtree (see for example Sarnet 1984) have been adopted in numerous geographic information systems and spatial data bases. They offer advantages in data compression and sampling efficiency, since the depth of the tree, and thus the density of information, can be varied from one area to another in response to the variability of the phenomenon being represented. Numerous processes operate faster on HSDSs, particularly various forms of spatial search. The address of a cell in an HSDS embeds both of its spatial coordinates, and thus effectively compresses two dimensions to one (Mark and Goodchild, 1986). Cells lower in the tree have longer addresses, and the length of an address is therefore a direct measure of spatial resolution. This has led to the suggestion (Dutton, 1988b; Saalfeld, 1988) that HSDSs offer a powerful solution to the problems of accuracy in spatial databases, since the spatial resolution of a position can be determined directly from the length of its spatial address.

Three properties of quadtree and octtree implementations of HSDS are of particular interest in this paper: (1) at any level, the cells are equal in area; (2) at any level, cells are equal in shape; and (3) the data structure correctly encodes the adjacency relationships between cells. The value of an HSDS for analysis and modeling would clearly be reduced without these properties, particularly in modeling based on finite elements. Unfortunately it has proven difficult to find a method of hierarchically subdividing the earth's surface so that these properties are retained. Many global databases have been based on rectangular cells superimposed on simple cylindrical projections such as Mercator's or the cylindrical equidistant projection. However although these schemes may achieve one of our required properties (as a conformal projection, the Mercator projection achieves property (1)), we note that it is well known that no projection of the earth onto a plane can satisfy both of properties (1) and (2). Moreover any cylindrical projection must violate property (3) because of the interruption at the poles. A method based on a cylindrical equal area projection was proposed by Tobler and Chen (1986); cells at a given level have equal area, but unequal shape.

An HSDS called "*Triacon*" or "*Quaternary Triangular Mesh (QTM)*" was suggested by Dutton (1984, 1988b). In this paper we follow Dutton's approach in first projecting the earth onto an octahedron, and then recursively subdividing each of the eight triangular faces of the octahedron into four triangles. Each level of the hierarchy after the first thus contains four times as many triangular cells or elements as the previous level. We simplify Dutton's approach in our numbering of the triangles, in order to obtain an addressing system which provides easy transformation to and from latitude and longitude. Our scheme satisfies property (3), and although properties (1) and (2) are only approximately satisfied, each triangular cell has an area which can be computed from a simple expression.

The discussion is organized as follows. We first describe the coordinate systems used to develop the properties of the proposed global HSDS. Subsequent sections develop the transformations between coordinate systems, particularly between cell address and latitude/longitude. Section 4 discuss the calculation of cell area, and section 5 and section 6 present an algorithm for finding the neighbors of a cell and the average data file storage distance, which corresponds to the expected cost of transition from one data cell to its neighbors.

1. Coordinate Systems

In our proposed scheme, the entire earth is described by an octahedron. One quarter of each hemisphere is represented by an equilateral triangle and is then decomposed. In order that the hierarchy be symmetrical and isohedral, i.e., all cells are congruent and

[1] National Center for Geographic information and Analysis and Department of Geography, University of California at Santa Barbara, CA 93106

every cell can be mapped onto other cells through translation, reflection, rotation or a combination of these, the triangle is subdivided into four smaller equilateral triangles and each of them is further subdivided recursively until a required level is reached. When the four triangles are decomposed from their ancestor triangle, they are labeled 0, 1, 2 and 3. There are 24 possible distinct ways of labeling. Moreover while the initial subdivision occurs with the northern hemisphere triangles standing on their bases ("upward"), the southern hemisphere triangles stand on vertices ("downward"). In subsequent iterations, triangles in both upward and downward orientations must be subdivided in both hemispheres. If this orientation of triangles is considered, there are 48 possible labeling schemes. In order to limit the complexity of the addressing and conversion algorithms, we use the following labeling method in every recursive decomposition: (1) the center triangle is labeled cell 0; (2) the triangle vertically above (below) the central triangle is labeled cell 1; and (3) the triangles below (above) and left and right of triangle 0 are labeled cells 2 and 3 respectively. Note that the terms in the parentheses are used when the triangle being subdivided stands on a vertex("downward").

The initial representation of the globe as eight triangles is termed the level 0 subdivision; after *j* further subdivisions of each triangle we reach level j of the HSDS. Thus at level j, there are $8x4^j$ cells. For much of the discussion in this paper the level 0 subdivision will be ignored, and we will refer simply to the recursive subdivision of one quarter hemisphere. Figure la is a triangle decomposed to level 4, with each cell identified by its address which consists of four base4 digits, identifying the triangles selected at each level of subdivision. The full address including level 0 would require an initial base8 digit. Figure 1b shows the decimal address of cells in the triangular decomposition. Figure 1c shows the ordering of cells, and emphasizes the consistent choice of the left cell as cell 2 at every level, irrespective of whether the triangle is upward or downward.

In this study, we use the following coordinate systems:

(1). Positions on the globe are referenced by latitude ϕ and longitude λ.

(2). Each of the eight isosceles triangles of the level 0 octahedron contains one quarter hemisphere. Locations within each triangle are identified by Cartesian coordinates x and y, with respect to an origin in the lower left corner. The triangles are numbered 0 through 3 in the Northern hemisphere, and 4 through 7 in the Southern, in both cases in anticlockwise order when viewed from the North pole. The appropriate triangle can thus be found by dividing longitude by $\frac{\pi}{2}$ and truncating to an integer, and adding 4 for southern latitudes.

The triangle vertices are assumed to lie at $(0,0,)$ $(2^n,0)$ and $(2^{n-1}, 2^{n-1}\sqrt{3})$ in the (x, y) coordinate system, where n is the highest level of subdivision. We assume that x depends linearly on longitude for a given latitude, and that y depends linearly on latitude. Figure 2 shows the relationships between latitude and longitude and (x, y) schematically.

The left, right and bottom edges of the triangle in Figure 2 can be described by the following equations:

Left edge: $y = \sqrt{3}\,x$ *or* $\lambda = 0$ (1–1)

Right edge: $y = (2^n - x)\sqrt{3}$ *or* $\lambda = \frac{\pi}{2}$ (1–2)

Bottom edge: $y = 0$ *or* $\phi = 0$ (1–3)

From Figure 2 and expressions (1-1) to (1-3), we have the following expression for the relation between y and ϕ.

$$\phi = \frac{\pi}{2^n\sqrt{3}}\,y, \quad or \quad y = \frac{2^n\sqrt{3}}{\pi}\,\phi \tag{1–4}$$

From the point $p(x, y)$ in the Figure 2, we have

$$x = x_1 + x' \qquad where \qquad x_1 = \frac{1}{\sqrt{3}}\,y$$

and x' is the horizontal distance from $p(x, y)$ to the left edge of the triangle. Since the distance between x_2 and x_1 corresponds to the maximum longitude difference at latitude $\phi = \pi y/(2^n\sqrt{3})$ which is also defined as $\frac{\pi}{2}$, therefore, we have

$$\frac{x'}{x_2 - x_1} = \frac{2\lambda}{\pi}$$

From expression (1-1) and (1-2), we have

$$x_2 = 2^n - \frac{y}{\sqrt{3}}$$

Then

$$x' = \frac{2\lambda}{\pi}(x_2 - x_1) = \frac{2^{n+1}\lambda}{\pi}(1 - \frac{2^{1-n}}{\sqrt{3}}\,y)$$

and

$$x \quad = x_1 + x' = \frac{y}{\sqrt{3}} + \frac{2^{n+1}\lambda}{\pi}(1 - \frac{2^{1-n}}{\sqrt{3}}\,y) =$$

$$= \frac{2^n}{\pi}[\phi + 2\lambda(1 - \frac{2}{\pi}\phi)]$$

The expressions for transformation of longitude λ and latitude ϕ to x and y in the triangle are the following:

$$x = \frac{2^n}{\pi}[\phi + 2\lambda(1 - \frac{2}{\pi}\phi)] \qquad y = \frac{2^n\sqrt{3}}{\pi}\phi \tag{1–5}$$

or

$$\lambda = \frac{\pi}{2^{n+1}} \frac{\sqrt{3}x - y}{\sqrt{3} - 2^{1-n}y} \qquad \phi = \frac{\pi}{2^n\sqrt{3}}y \tag{1–6}$$

The parameters of the triangles at different levels are shown in Table 1. Table 2 shows the lengths of edges of the triangles at different levels of decomposition. At the 20-th level, the edges of triangles are less than $10m$, and 20 quaternary digits or 40 binary digits (approximately 12 decimal digits) are required for addressing.

2. Conversion of Triangle Address to Cartesian Coordinates

The addresses of vertices can be calculated by using the parameters listed in Table 1. For triangles decomposed to the k-th level, the *triangle address* is represented by k quaternary numbers

$$a_1, a_2, a_3, \ldots, a_k \tag{2-1}$$

where $1 \le k \le n$. In this section we consider the problem of determining the Cartesian coordinates of the centroid of a triangle with given triangle address. The coordinates of the triangle's vertices can be determined from the parameters listed in Table 1 and from knowledge of the triangle's orientation.

With the triangle cell ordering shown in Figure 1, there are the following relations:

(1). Let

$$NZ_k = \sum_{i=1}^{k-1}[a_i = 0] = \sum_{i=1}^{k-1}[\bar{a}_{i1} \cap \bar{a}_{i2}] \tag{2–2}$$

denote the number of zeros in a_1 to a_{k-1}. a_{i1} and a_{i2} are two binary digits representing each quaternary digit a_i. $\bar{a}_{i1}$, $\bar{a}_{i2}$ are logical negative of (or *NOT*) a_{i1}, a_{i2}. (a_{i1}, a_{i2}) = (0, 0), (0, 1), (1, 0), (1, 1) correspond to a_i = 0, 1, 2, 3 respectively.

If the level 0 triangle is upward, then the k -th level triangles with address

$$A = a_1, a_2, a_3, \ldots, a_k$$

are upward if NZ_k is even, and downward if NZ_k is odd. NZ_k is zero when $k = 1$.

We can readily generalize to include the base-8 digit representing the initial octahedral decomposition at level 0. If the base-8 digit is represented as three binary digits, then the level zero triangle is upward if the first digit is 0, downward if it is 1. The generalized definition of NZ_k is:

$$NZ_k = a_{00} + \sum_{i=1}^{k-1}[\bar{a}_{i1} \cap \bar{a}_{i2}]$$

where a_{00} is the first bit of the binary representation of the base-8 digit.

(2). For the triangles in the k-th level with the same ancestor triangle, i.e., having identical a_1 to a_{k-1}, then:

1. The triangle with $a_k = 1$ is a complex conjugate or a reflection of the triangle with $a_k = 0$.
2. The triangles with $a_k = 2$ and $a_k = 3$ are the left-down (or left-up) and right-down (or right-up) translation of the triangle with $a_k = 1$ respectively.

(3). The centroid coordinates of the 0 level (original) triangle are

$$(X_0, Y_0) = (2^{n-1}, \frac{2^{n-1}}{\sqrt{3}}) \tag{2-3}$$

(4). The relative distance of the triangle centroids at the k-th level $O_{k,0}$, $O_{k,1}$, $O_{k,2}$, *and* $O_{k,3}$ from the centroid of their ancestor at the (k-1) -th level $O_{k-1,0}$ are (see Table 3):

$$O_{k-1,0} - O_{k,0}: \quad (\Delta X_{k,0}, \Delta Y_{k,0}) = (0, 0) \tag{2–4a}$$

$$O_{k-1,0} - O_{k,1}: \quad (\Delta X_{k,1}, \Delta Y_{k,1}) = (0, \alpha \frac{2^{n-k}}{\sqrt{3}}) \tag{2–4b}$$

$$O_{k-1,0} - O_{k,2}: \quad (\Delta X_{k,2}, \Delta Y_{k,2}) = (-2^{n-k-1}, -\alpha \frac{2^{n-k-1}}{\sqrt{3}}) \tag{2–4c}$$

$$O_{k-1,0} - O_{k,3}: \quad (\Delta X_{k,3}, \Delta Y_{k,3}) = (2^{n-k-1}, -\alpha \frac{2^{n-k-1}}{\sqrt{3}}) \tag{2–4d}$$

where

$$\alpha = (-1)^{NZ_k}$$

The *Triangle Address* $A = a_1, a_2, \ldots, a_k$ can be converted to Cartesian coordinates by the following expressions:

$$X_k = 2^{n-1} + \sum_{i=1}^{k} [(-1)^{a_{i2}+1} a_{i1} 2^{n-i-1}] =$$

$$= [2^k + \sum_{i=1}^{k} [(-1)^{a_{i2}+1} a_{i1} 2^{k-i}]] 2^{n-k-1} \tag{2–5a}$$

$$Y_k = \frac{2^{n-1} + \sum_{i=1}^{k} [(-1)^{NZ_i + a_{i1}} [2(\bar{a}_{i1} \cap a_{i2}) + a_{i1}] 2^{n-i-1}]}{\sqrt{3}} =$$

$$= \frac{2^k + \sum_{i=1}^{k} [(-1)^{NZ_i + a_{i1}} [2(\bar{a}_{i1} \cap a_{i2}) + a_{i1}] 2^{k-i}]}{\sqrt{3}} 2^{n-k-1} \tag{2–5b}$$

Coordinates of the three vertices of the triangles can be calculated from:

Top vertex: $X_{k-Top} = X_k$ (2–6a)

$$Y_{k-Top} = Y_k + (-1)^{NZ_k} \frac{2^{n-k}}{\sqrt{3}} \tag{2–6b}$$

Left vertex: $X_{k-Left} = X_k - 2^{n-k-1}$ (2–7*a*)

$$Y_{k-Left} = Y_k + (-1)^{NZ_k+1}\frac{2^{n-k-1}}{\sqrt{3}} \quad (2\text{–}7b)$$

Right vertex: $X_{k-Right} = X_k + 2^{n-k-1}$ (2–8*a*)

$$Y_{k-Right} = Y_k + (-1)^{NZ_k+1}\frac{2^{n-k-1}}{\sqrt{3}} \quad (2\text{–}8b)$$

Example 1. Find the Cartesian coordinates of the centroid of triangle $A = 3023$.

$A = 3023 = 11\ 00\ 10\ 11$

or

$a_{11} = 1,\quad a_{12} = 1;\qquad a_{21} = 0,\quad a_{22} = 0;\qquad a_{31} = 1,\quad a_{32} = 0;\qquad a_{41} = 1,\quad a_{42} = 1.$

$$X = [2^4 + 1\times(-1)^{1+1}2^3 + 0\times(-1)^{0+1}2^2 + 1\times(-1)^{0+1}2^1 + 1\times(-1)^{1+1}2^0]\,2^{n-5} =$$

$$= [2^4 + 2^3 - 2^1 + 2^0]\,2^{n-5} = 23\times 2^{n-5}$$

$$Y = [2^4 + 1\times(-1)^{0+1}2^3 + 0\times(-1)^{0+0}2^2 + 1\times(-1)^{1+1}2^1 + 1\times(-1)^{1+1}2^0]\frac{2^{n-5}}{\sqrt{3}} =$$

$$= [2^4 - 2^3 + 2^1 + 2^0]\frac{2^{n-5}}{\sqrt{3}} = 11\times\frac{2^{n-5}}{\sqrt{3}}$$

Example 2. Find the Cartesian coordinates of the centroid of triangle $A = 1003$.

$A = 1003 = 01\ 00\ 00\ 11$

or

$a_{11} = 0,\quad a_{12} = 1;\qquad a_{21} = 0,\quad a_{22} = 0;\qquad a_{31} = 0,\quad a_{32} = 0;\qquad a_{41} = 1,\quad a_{42} = 1.$

$$X = [2^4 + 0\times(-1)^{1+1}2^3 + 0\times(-1)^{0+1}2^2 + 0\times(-1)^{1}2^{0+1} + 1\times(-1)^{1+1}2^0]\,2^{n-5} =$$

$$= [2^4 + 2^0]\,2^{n-5} = 17\times 2^{n-5}$$

$$Y = [2^4 + 2\times(-1)^{0+0}2^3 + 0\times(-1)^{0+0}2^2 + 0\times(-1)^{1+0}2^1 + 1\times(-1)^{2+1}2^0]\frac{2^{n-5}}{\sqrt{3}} =$$

$$= [2^5 - 2^0]\frac{2^{n-5}}{\sqrt{3}} = 31\times\frac{2^{n-5}}{\sqrt{3}}$$

3. Conversion of Cartesian Coordinates to Triangle Address

For implementation of the triangular tessellation data structure in a global geographic information system, it is necessary to convert the coordinates of the earth to the triangle address. Since longitude and latitude can be directly represented by Cartesian coordinates, the problem can be reduced to conversion of Cartesian coordinates into the triangle address. A recursive approximation algorithm for conversion of Cartesian coordinates to triangle address will be derived below.

Let an equilateral triangle be divided into four triangles and let the centroid of the parent triangle be denoted by $O_{j-1,0}$, the centroids of four son triangles are denoted as $O_{j,0}$, $O_{j,1}$, $O_{j,2}$, *and* $O_{j,3}$ as shown in Figure 3. We have $O_{j-1,0} = O_{j,0}$ and the distance between the centroid of the parent triangle and the centroids of the other three son triangles can be calculated by expressions (2-4a) to (2-4d).

If $p(x, y)$ is an arbitrary point with Cartesian coordinates (x,y) and is inside the parent triangle, the point $p(x, y)$ will be a point in the k-th son triangle if $p(x, y)$ is closest to the centroid of son triangle k ($k = 0, 1, 2,$ *or* 3). This can be seen by drawing three bisectors of the parent triangle and connecting $p(x, y)$ to the centroids of four son triangles as shown in Figure 3.

We start the recursive approximation procedures from level 0. To simplify the calculation, the distance squared is used instead of distance.

(1). Calculate the relative distance

$$\Delta x_0 = x - 2^{n-1}, \qquad \Delta y_0 = y - \frac{2^{n-1}}{\sqrt{3}} \tag{3–1}$$

between (x, y) and the centroid of the original triangle.

(2). Calculate

$$D(1,l) = (\Delta x_0 - \Delta X_{1,l})^2 + (\Delta y_0 - \Delta Y_{1,l})^2 \tag{3–2a}$$

for $l = 0, 1, 2, 3$, where $\Delta X_{1,l}$ and $\Delta Y_{1,l}$ are the relative distance from the centroids of the first level triangles to the centroid of the 0 level triangle as expressed in (2-4a) to (2-4d) and $\alpha = 1$ for $j = 1$ (level 1).

If

$$Min\,[D(1,0), D(1,1), D(1,2), D(1,3)] = D(1,k_1), \tag{3–2b}$$

then $p(x, y)$ is in the triangle k_1 of the first level triangles and we set

$$a_1 = k_1 \tag{3–2c}$$

and

$$\Delta x_1 = \Delta x_0 - \Delta X_{1,k_1}, \qquad \Delta y_1 = \Delta y_0 - \Delta Y_{1,k_1}. \tag{3–2d}$$

The recursive approximation algorithm for conversion of Cartesian coordinates to triangle address is as follows:

input $x,\ y,\ k$;

output $a_1, a_2, a_3, \ldots a_n$;

begin {main}

$\Delta x_0 = x - 2^n$; $\qquad \Delta y_0 = y - \dfrac{2^{n-1}}{\sqrt{3}}$

$j = 0$;

repeat

for $j = 1$ *to* k

for $l = 1$ *to* 3

$$NZ_j = a_{00} + \sum_{i=1}^{j-1} [a_j = 0]$$

$$D(j,l) = (\Delta x_{j-1} - \Delta X_{j,l})^2 + (\Delta y_{j-1} - \Delta Y_{j,l})^2$$

if

$$D(j, k_j) = \min [D(j,0), D(j,1), D(j,2), D(j,3)]$$

then

$$a_j = k_j$$

$$j = j + 1$$

continue

end

Example 3. Given the Cartesian triangle address $(x, y) = (23\times 2^{n-5}, \frac{11\times 2^{n-5}}{\sqrt{3}})$ find the triangular tiling address a_1, a_2, a_3, a_4.

(1). For $j = 0$ we have $\alpha = +1$

$$\Delta x_0 = (23-16)2^{n-5} = 7\times 2^{n-5}, \qquad \Delta y_0 = \frac{(11-16)2^{n-5}}{\sqrt{3}} = -5\times 2^{n-5}.$$

(2). For $j = 1$ we have $\alpha = +1$:

$$\Delta x_{1,0} = (7-0)\times 2^{n-5} = 7\times 2^{n-5}, \qquad \Delta y_{1,0} = \frac{-(5-0)\times 2^{n-5}}{\sqrt{3}} = \frac{-5\times 2^{n-5}}{\sqrt{3}}$$

$D(1,0) = 57.33333$

$$\Delta x_{1,1} = (7-0)\times 2^{n-5} = 7\times 2^{n-5}, \qquad \Delta y_{1,1} = \frac{-(5+16)\times 2^{n-5}}{\sqrt{3}} = \frac{-21\times 2^{n-5}}{\sqrt{3}}$$

$D(1,1) = 196$

$$\Delta x_{1,2} = (7+8)\times 2^{n-5} = 15\times 2^{n-5}, \qquad \Delta y_{1,2} = \frac{-(5-8)\times 2^{n-5}}{\sqrt{3}} = \frac{3\times 2^{n-5}}{\sqrt{3}}$$

$D(1,2) = 228$

$$\Delta x_{1,3} = (7-8)\times 2^{n-5} = -2^{n-5}, \qquad \Delta y_{1,3} = \frac{-(5-8)\times 2^{n-5}}{\sqrt{3}} = \frac{3\times 2^{n-5}}{\sqrt{3}}$$

$D(1,3) = 4$

$Min\ D(1,l) = D(1,3) = 4$

We have

$a_1 = 3 \quad \Delta x_1 = -2^{n-5} \qquad \Delta y_1 = \frac{3\times 2^{n-5}}{\sqrt{3}}.$

(3). For $j = 2$: We have $\alpha = +1$ since $a_1 \neq 0$,

$$\Delta x_{2,0} = (-1-0)\times 2^{n-5} = -2^{n-5}, \qquad \Delta y_{2,0} = \frac{(3-0)\times 2^{n-5}}{\sqrt{3}} = \frac{3\times 2^{n-5}}{\sqrt{3}}.$$

$D(2,0) = 4$

$$\Delta x_{2,1} = (-1-0)\times 2^{n-5} = -2^{n-5}, \qquad \Delta y_{2,1} = \frac{(3-8)\times 2^{n-5}}{\sqrt{3}} = \frac{-5\times 2^{n-5}}{\sqrt{3}}$$

$D(2,1) = 9.333333$

$$\Delta x_{2,2} = (-1+4)\times 2^{n-5} = 3\times 2^{n-5}, \qquad \Delta y_{2,2} = \frac{(3+4)\times 2^{n-5}}{\sqrt{3}} = \frac{7\times 2^{n-5}}{\sqrt{3}}$$

$D(2,2) = 25.333333$

$$\Delta x_{2,3} = (-1-4)\times 2^{n-5} = -2^{n-5}, \qquad \Delta y_{2,3} = \frac{(3+4)\times 2^{n-5}}{\sqrt{3}} = \frac{7\times 2^{n-5}}{\sqrt{3}}$$

$D(2,3) = 41.333333$

$Min\ D(2,l) = D(2,0) = 4$

We have

$a_2 = 0 \quad \Delta x_2 = -2^{n-5} \qquad \Delta y_2 = \frac{3\times 2^{n-5}}{\sqrt{3}}.$

(4). For $j = 3$: We have $\alpha = -1$ since $\sum_{i=1}^{2}(a_i = 0) = 1$.

$$\Delta x_{3,0} = (-1-0)\times 2^{n-5} = -2^{n-5}, \qquad \Delta y_{3,0} = \frac{(3-0)\times 2^{n-5}}{\sqrt{3}} = 3\times 2^{n-5}$$

$D(3,0) = 4$

$$\Delta x_{3,1} = (-1-0)\times 2^{n-5} = -2^{n-5}, \qquad \Delta y_{3,1} = \frac{(3+4)\times 2^{n-5}}{\sqrt{3}} = \frac{7\times 2^{n-5}}{\sqrt{3}}$$

$D(3,1) = 17.333333$

$$\Delta x_{3,2} = (-1+2)\times 2^{n-5} = 2^{n-5}, \qquad \Delta y_{3,2} = \frac{(3-2)\times 2^{n-5}}{\sqrt{3}} = \frac{2^{n-5}}{\sqrt{3}}$$

$D(3,2) = 1.333333$

$$\Delta x_{3,3} = (-1-2)\times 2^{n-5} = -3^{n-5}, \qquad \Delta y_{3,3} = \frac{(3-2)\times 2^{n-5}}{\sqrt{3}} = \frac{2^{n-5}}{\sqrt{3}}$$

$D(3,3) = 4.333333$

$Min\ D(3,l) = D(3,2) = 1.333333$

We have

$a_3 = 2 \quad \Delta x_3 = 2^{n-5} \qquad \Delta y_3 = \frac{2^{n-5}}{\sqrt{3}}.$

(5). For $j = 4$: We have $\alpha = -1$ since $\sum_{i=1}^{3}(a_i = 0) = 1.$

$$\Delta x_{4,0} = (1-0)\times 2^{n-5} = 2^{n-5}, \qquad \Delta y_{4,0} = \frac{(1-0)\times 2^{n-5}}{\sqrt{3}} = \frac{2^{n-5}}{\sqrt{3}}.$$

$D(4,0) = 1.333333$

$$\Delta x_{4,1} = (1-0)\times 2^{n-5} = 2^{n-5}, \qquad \Delta y_{4,1} = \frac{(1+2)\times 2^{n-5}}{\sqrt{3}} = \frac{3\times 2^{n-5}}{\sqrt{3}}$$

$D(4,1) = 4$

$$\Delta x_{4,2} = (1+1)\times 2^{n-5} = 2\times 2^{n-5}, \qquad \Delta y_{4,2} = \frac{(1-1)\times 2^{n-5}}{\sqrt{3}} = 0$$

$D(4,2) = 4$

$$\Delta x_{4,3} = (1-1)\times 2^{n-5} = 0, \qquad \Delta y_{4,3} = \frac{(1-1)\times 2^{n-5}}{\sqrt{3}} = 0$$

$D(4,3) = 0$

$Min\ D(4,l) = D(4,3) = 0$

We have

$a_4 = 3 \quad \Delta x_4 = 0 \qquad \Delta y_4 = 0.$

Therefore, the triangular tiling address is

$A = a_1\ a_2\ a_3\ a_4 = 3023.$

5. The Area of Decomposed Triangles

The process we have described for creating an HSDS for the globe does not satisfy properties (1) and (2) precisely; triangles at level $k > 0$ are not equal in area, and have varying shapes, although we believe that our scheme represents close to an optimum compromise between these conflicting objectives. In this section we examine the areas of triangles explicitly. We assume that the earth is spherical, although the results should generalize easily to the more accurate ellipsoid of revolution.

The earth surface area A between latitude ϕ_1 and ϕ_2 covered by a level 0 triangle is

$$A = \frac{\pi R^2}{2}(\sin\phi_2 - \sin\phi_1) \tag{4-1}$$

At level n, the total number of triangles in the belt between ϕ and $\phi + \frac{\pi}{2^{n+1}}$ is

$$N_\phi = 2\times\frac{2^{n+1}}{\pi}(\frac{\pi}{2}-\phi)-1 = \frac{2^{n+2}}{\pi}(\frac{\pi}{2}-\phi)-1$$

and the earth surface area of a triangle at level n is

$$\Delta A_\phi = \frac{A}{N_\phi} = \frac{\pi^2 R^2}{2^{n+3}} \frac{\sin(\phi+2^{-n-1}\pi)-\sin\phi}{\frac{\pi}{2}-\phi-2^{-n-2}\pi} \tag{4–2}$$

When $2^{n+1} \gg 1$, we have

$$\sin(\phi+2^{-n-1}\pi)-\sin\phi \approx \frac{\pi}{2^{n+1}}\cos\phi$$

and

$$\frac{\pi}{2}-\phi-2^{-n-2}\pi \approx \frac{\pi}{2}-\phi$$

The expression (4-2) can be written as

$$\Delta A_\phi \approx k\frac{\cos\phi}{\frac{\pi}{2}-\phi} = k\frac{\sin\chi}{\chi} = k\,sinc\,\chi \tag{4–3}$$

where

$$k = \frac{\pi^3 R^2}{2^{2n+4}}, \qquad \chi = \frac{\pi}{2}-\phi.$$

From expression (4-3), it is interesting to note that the area covered by high level decomposed triangle varies with the *sinc* function. In the range from $\chi = 0$ to $\chi = \frac{\pi}{2}$, $sinc\,\chi$ is a monotonically decreasing function of χ, or ΔA_ϕ is an increasing function of ϕ for $\phi = 0$ to $\phi = \frac{\pi}{2}$. For $\phi = 0$ and $\phi = \frac{\pi}{2}$, we have

$$\Delta A_0 = \frac{2k}{\pi}, \qquad \Delta A_{\frac{\pi}{2}} = k$$

and

$$\frac{\Delta A_{\frac{\pi}{2}}}{\Delta A_0} = \frac{\pi}{2} \tag{4–4}$$

That is, for high level decomposed triangles, the corresponding area increases $\frac{\pi}{2} = 1.5708$ times when latitude changes from 0 to $\frac{\pi}{2}$. Only the triangles along a given latitude have the same area and the area changes with latitude according to the *sinc* function as shown in Figure 4.

5. Algorithm to Find Neighbors of Triangles

It is often necessary to find the *three directly connected neighbors* of a given triangle with address

$$A = a_1, a_2, a_3, \ldots, a_k.$$

We denote the three neighbors as *Top, Left* and *Right* neighbors. The direction of the neighbors depends on whether the triangle A is upward or downward as follows:

		Top	Left	Right
Upward	(NZ even)	S	NW	NE
Downward	(NZ odd)	N	SW	SE

We use the codes

$$T = t_1, t_2, t_3, \ldots, t_k,$$
$$L = l_1, l_2, l_3, \ldots, l_k,$$
$$R = r_1, r_2, r_3, \ldots, r_k,$$

to represent the addresses of the Top, Left and Right neighbors respectively. Both the triangle and its neighbor are inside a triangle of the j -th level if $a_1, a_2, \ldots, a_{j-1}$ does not change. The problem is to determine the level of triangle within which a neighbor of a given triangle is contained, and to change the code of $a_j, a_{j+1}, \ldots, a_k$ for the top, left and right neighbors of a given triangle separately.

Recall that the triangles are ordered as follows:

(1). the center, top, left and right triangles within a triangle are ordered 0, 1, 2, and 3 respectively,

(2). the triangle with $a_j = 1$ is a reflection of the triangle with $a_j = 0$,

(3). the triangles with $a_j = 2$ and $a_j = 3$ are the left-down (or left-up) and right-down (or right-up) translation of the triangle with $a_j = 1$ respectively.

The neighbor addresses can be searched using the the following conversion table:

a_i	Top	Left	Right
0	1 #	2 #	3 #
1	0 #	3	2
2	2	1	0 #
3	3	0 #	1

for a_i (i *from* k *to* j), where # is the search terminate symbol. It can be implemented as follows.

(1). To find the top neighbor, starting from $i = k$, a_i changes to t_i:

$t_i = 2$ if $a_i = 2$
$t_i = 3$ if $a_i = 3$

and process a_{i-1} for t_{i-1}. On the other hand if $a_i \neq 2$ and $a_i \neq 3$, then

$t_i = 1$ if $a_i = 0$
$t_i = 0$ if $a_i = 1$

and the search finishes.

We define this i as j and set

$$t_k = a_k$$

for $i = 1$ to $i = j-1$

(2). For the left neighbor, starting from $i = k$, a_i changes to l_i:

$l_i = 3$ if $a_i = 1$

$l_i = 1$ if $a_i = 2$

and process a_{i-1} for l_{i-1}. On the other hand if $a_i \neq 1$ and $a_i \neq 3$, then

$l_i = 2$ if $a_i = 0$

$l_i = 0$ if $a_i = 3$

and the search finishes. We define this i as j and set

$$l_k = a_k$$

for $i = 1$ to $i = j-1$

(3). For the right neighbor, starting from $i = k$. a_i changes to r_i:

$r_i = 2$ if $a_i = 1$

$r_i = 1$ if $a_i = 3$

and process a_{i-1} for r_{i-1}. On the other hand if $a_i \neq 1$ and $a_i \neq 3$, then

$r_i = 3$ if $a_i = 0$

$r_i = 0$ if $a_i = 2$

and the search finishes. We define this i as j and set

$$r_k = a_k$$

for $i = 1$ to $i = j-1$

The algorithm for finding neighbor addresses described above is easy to implement as only $k-j+1$ quaternary digits have to be determined by simple criteria and the other $j-1$ digits are only a copy of the corresponding digits in the given triangle. The average number of quaternary digits which need to be changed to find a neighbor can be determined as follows.

(1). The probability of changing only the last (k-th) digit is $\frac{1}{2}$;

(2). The probability of changing j quaternary digits is 2^{-j}

Therefore, the *average number of steps of calculation* is

$$S_{av} = \sum_{j=1}^{k} \frac{j}{2^j} + \frac{k}{2^k} = 2 - \frac{2}{2^k}$$

S_{av} is less than 2 quaternary or 4 binary digits.

The neighbor finding algorithm can be used for searching hexagons with a given triangle included or to find the twelve neighbors of a given triangle (Figure 5). Let a given triangle be denoted by T_0 and the other five triangles in a hexagon are T_1, T_2, T_3, T_4 and T_5 respectively. They can be found as follows whether the given triangle is upward or downward.

For a hexagon with the given triangle as a top triangle:

$$T_1 = R(T_0), \quad T_2 = T(T_1), \quad T_3 = L(T_2), \quad T_4 = L(T_3)$$

$$T_5 = T(T_4)$$

For a hexagon with the given triangle as a left triangle:

$$T_1 = L(T_0), \quad L_2 = L(T_1), \quad T_3 = T(T_2), \quad T_4 = R(T_3)$$

$$T_5 = R(T_4)$$

For a hexagon with the given triangle as a right triangle:

$$T_1 = T(T_0), \qquad T_2 = R(T_1), \qquad T_3 = R(T_2), \qquad T_4 = T(T_3)$$

$$T_5 = L(T_4)$$

where $T_j = T(T_i)$, $T_j = L(T_i)$, and $T_j = R(T_i)$, implies that the triangle T_j is the top, left and right neighbor of triangle T_i respectively. The first four triangles in the three hexagons above are the twelve neighbors of the given triangle as shown in Figure 5.

Another application of the neighbor finding algorithm is that a chain code with a series of codes T (*top*), L (*left*) and R (*right*) can be used to describe lines or borders of areas.

6. Average Data File Storage Distance

One of the important indices in data file structures for large geographical information systems is the *Average Data File Storage Distance* (Goodchild and Grandfield, 1983; Mark and Goodchild, 1986). This is defined as the average absolute difference between the addresses of neighboring cells or tiles; in our case, each triangle is assumed to have three neighbors. Goodchild and Grandfield (1983) used the index in a study of the data compression achieved by different ordering of a lattice, whereas Goodchild (1989) argued its usefulness in predicting the time required to access data base partitions in very large spatial archives. In this analysis we are concerned only with subdivisions of the level 0 triangles, and ignore the differences which occur across edges of the octahedron.

The average data file storage distance is the sum of absolute differences between adjacent triangular cells D_{total} divided by the number of edges OC_{total}, or

$$D_{av} = \frac{D_{total}}{OC_{total}}$$

The total differences for triangular cells can be represented as

$$D_j = 4\times D_{j+1} + \Delta D_j$$

and

$$OC_j = 4\times OC_{j+1} + \Delta OC_j$$

where $4\times D_{j+1}$ and $4\times OC_{j+1}$ are the total differences and number of edges of triangular cells for level $j+1$ triangles and ΔD_j and ΔOC_j are the distances and edges added at the j -th level (See Figure 1b).

$$\Delta D_j = 6\times 2^{k-j}\times 4^{k-j} = 6\times 2^{3(k-j)}$$

is the sum of edge cell values of triangles 1, 2 and 3 minus the sum of edge cell values of triangle 0 at the j -th level, and

$$\Delta OC_j = 3\times 2^{k-j}$$

is the total number of edges added at the j -th level. Therefore

$$D_j = 6\times \sum_{i=2(k-j)}^{3(k-j)} 2^i = 6\times 2^{2(k-j)}\times \sum_{i=0}^{k-j} 2^i$$

and

$$OC_j = 3\times\sum_{i=k-1}^{2(k-1)} 2^i = 3\times 2^{k-1}\times\sum_{i=0}^{k-j^i}$$

The average distance at j -th level is

$$D_{av,j} = \frac{D_j}{OC_j} = 2^{k-j+1}$$

We have $j = 1$ when the 0 level triangle is decomposed to k -th level. In this case

$$D_{av} = 2^k$$

7. Conclusions

The hierarchical data structure which we have described in this paper satisfies one of our original requirements in full, by preserving the relationships between neighboring cells. The distortions of area inherent in the structure, and described by the sinc function, range up to a factor of 1.57 at the poles. Triangles become increasingly equilateral toward the center of each level 0 triangle at higher levels of subdivision, but the triangles adjacent to each level 0 vertex always contain one right angle. Our requirements of equal area and equal shape are thus satisfied only approximately.

In this structure every object on the earth's surface can be indexed by the address of the smallest enclosing triangle. The length of the address is then a direct index of the object's size. To find the smallest enclosing triangle of a polygon, we simply determine the triangle address of one of its vertices to some arbitrary but high level k, and then identify a largest value $j \leq k$ such that all other vertices share the same quaternary digits 1 through j. For example, the US, which spans two level 0 triangles, has a null address, while the block formed by 3rd and 4th Streets, Broadway and Fulton in the City of Troy, New York has the address 0223022113013 (level 13). The approximate edge length of a level 16 triangle is 150m, or the rough dimensions of a city block, according to Table 2. However while the Broadway and Fulton faces of the block are both wholly within level 16 triangles, the smallest triangle enclosing the entire block is at level 13.

Length of address can also be used as a measure of uncertainty of position, by identifying the smallest triangle which encloses the union of the object's possible positions. For example, the accuracy currently provided by the Global Positioning System (GPS or NavStar) is about 20m. The corresponding length of address for any point on the earth's surface is 19 quaternary digits or 38 bits, any further precision being spurious. For comparison, to achieve 20m precision in latitude/longitude coordinates, it is necessary to specify location to the nearest second, which requires 7 decimal digits plus sign for longitude and 6 digits plus sign for latitude.

The ideal workstation for global systems modeling would allow the user to browse freely through data distributed over the surface of the globe. With datasets based on rectangular subdivision of a cylindrical projection it is relatively easy to browse in the equatorial region, but difficult near the poles because of high levels of distortion and interruption at the pole itself. Similar problems occur using rectangular subdivision of any other standard projection. For example, the orthographic projection gives a view of the globe as it would appear from space. However it would be time consuming to re-compute and redisplay the projection for every change of viewpoint.

Recent developments in 3D graphics display technology may make browsing on the globe much more practical. Instead of projecting to a plane, a solid is represented digitally by a polyhedron with triangular faces, and displayed in perspective directly from a display list of triangles. The graphical rendering (color or texture) of each triangle can be controlled directly from its attributes. Workstations which can display polyhedra of 10,000 triangles in 1 second are currently available for less than $20,000, and we can expect orders of magnitude improvement in these specifications in the near future. Thus we are able with current technology to create a browse of a global dataset at level 6 (approximately I degree resolution). For spatial variables such as land/water, subdivision can be much higher in some areas because of the relative homogeneity of continents and oceans. Thus the developing technology of 3D display based on polyhedra with triangular faces gives a powerful argument for trianglebased tessellation over more conventional methods.

The results presented in this paper suggest several potentially fruitful areas for further work. We have thus far ignored the non-spherical nature of the earth in calculating triangle areas. We also intent to pursue the development of algorithms, particularly to build the triangle data structure from vector data, such as the world's coastlines. In the longer term, we plan to develop a prototype workstation for global data based on the triangular structure and triangle display lists.

Reference

Dutton, G., 1984. Geodesic modelling of planetary relief, *Cartographica*, 21: 188 207.

Dutton, G., 1988a. Computational aspects of quaternary triangular meshes. Unpublished

Dutton, G., 1988b. Modelling locational uncertainty via hierarchical tessellation, Paper presented at NCGIA Specialist Meeting on Accuracy of Spatial Databases, Montecito, CA, December.

Goodchild, M.F., 1989. Optimal tiling for large cartographic databases. *Proceedings, AutoCarto 9*. ASPRS/ACSM, Falls Church, VA, 4445 1.

Mark, D.M. and M.F. Goodchild, 1986. On the ordering of twodimensional space: introduction and relation to tesseral principles. In B. Diaz and S. Bell, editors, *Spatial Processing Using Tesseral Methods*. National Environmental Research Council, Reading, UK, 179192.

Saalfeld, A., 1988. Census Bureau research concerns for accuracy of spatial data. Paper presented at NCGIA Specialist Meeting on Accuracy of Spatial Databases, Montecito, CA, December.

Samet, H., 1984. The quadtree and related hierarchical data structures. Computing Surveys 16:187260.

Tobler, W. R. and Zitan Chen, 1986. A quadtree for global information storage, *Geographical Analysis* 18:36071

Level	edge length	height	center to left side	center to bottom	center to top
0	2^n	$2^{n-1}\sqrt{3}$	2^{n-1}	$\frac{2^{n-1}}{\sqrt{3}}$	$\frac{2^n}{\sqrt{3}}$
1	2^{n-1}	$2^{n-2}\sqrt{3}$	2^{n-2}	$\frac{2^{n-2}}{\sqrt{3}}$	$\frac{2^{n-1}}{\sqrt{3}}$
2	2^{n-2}	$2^{n-3}\sqrt{3}$	2^{n-3}	$\frac{2^{n-3}}{\sqrt{3}}$	$\frac{2^{n-2}}{\sqrt{3}}$
j	2^{n-j}	$2^{n-j-1}\sqrt{3}$	2^{n-j-1}	$\frac{2^{n-j-1}}{\sqrt{3}}$	$\frac{2^{n-j}}{\sqrt{3}}$

Table 1. Basic geometric parameters of decomposed triangles

level	deg min sec	along equator (longitude) km	along meridian (latitude) km
0	90°	10018.5380	9983.8912
2	22°30′	2504.6345	2495.9953
4	5°37′30″	626.1586	623.9988
6	1°24′22.5″	156.5397	155.9997
8	21′5.625″	39.1349	38.9999
10	5′16.40625″	9.7837	9.7500
12	1′19.1015″	2.4459	2.4375
14	19.77539″	611.5m	609.4m
16	4.9438″	153.9m	152.3m
18	1.23596″	38.2m	38.1m
20	0.3089904″	9.55m	9.525m

Table 2. Length of triangle edges at increasing levels of subdivision of a spheroid with radius 6378*km* along Equator and 6356*km* along meridian

level	$O_{0,j} - O_{j-1}$ $(\Delta X,\ \Delta Y)$	$O_{1,j} - O_{j-1}$ $(\Delta X,\ \Delta Y)$	$O_{2,j} - O_{j-1}$ $(\Delta X,\ \Delta Y)$	$O_{3,j} - O_{j-1}$ $(\Delta X,\ \Delta Y)$
1	$(0, 0)$	$(0, \frac{2^{n-1}}{\sqrt{3}})$	$(-2^{n-2}, -\frac{2^{n-2}}{\sqrt{3}})$	$(2^{n-2}, -\frac{2^{n-2}}{\sqrt{3}})$
2	$(0, 0)$	$(0, \alpha \frac{2^{n-2}}{\sqrt{3}})$	$(-2^{n-3}, -\alpha \frac{2^{n-3}}{\sqrt{3}})$	$(2^{n-3}, -\alpha \frac{2^{n-3}}{\sqrt{3}})$
.	. .	. .	. .	. .
j	$(0, 0)$	$(0, \alpha \frac{2^{n-j}}{\sqrt{3}})$	$(-2^{n-j-1}, -\alpha \frac{2^{n-j-1}}{\sqrt{3}})$	$(2^{n-j-1}, -\alpha \frac{2^{n-j-1}}{\sqrt{3}})$

Table 3. Relative distances of successor triangles' centroids to the centroid of their ancestor triangle.

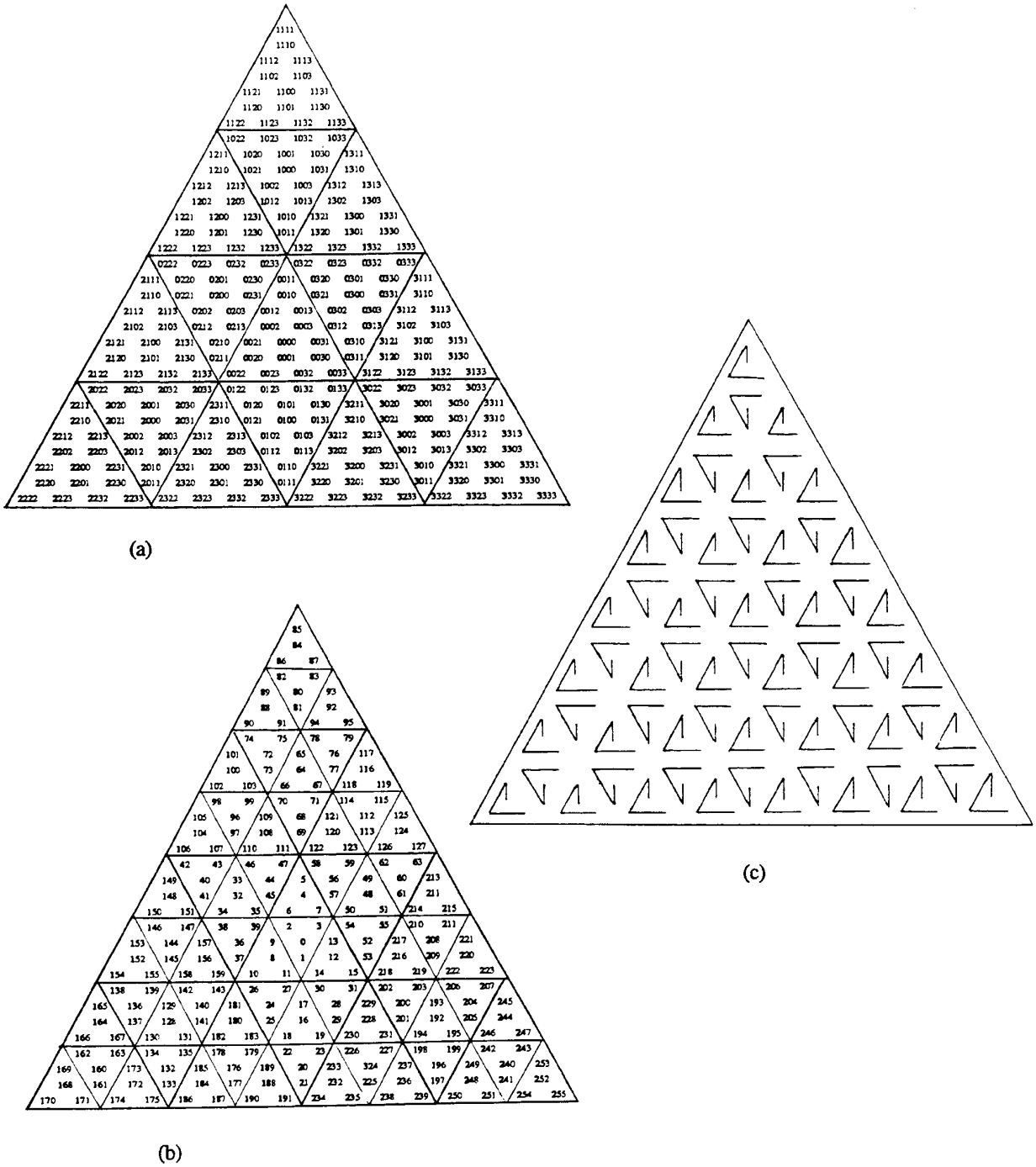

Figure 1 The triangular data structure applied to a quarter hemisphere, showing (a) quaternary addressing, (b)decimal addressing, and (c) the ordering of level 4 triangles

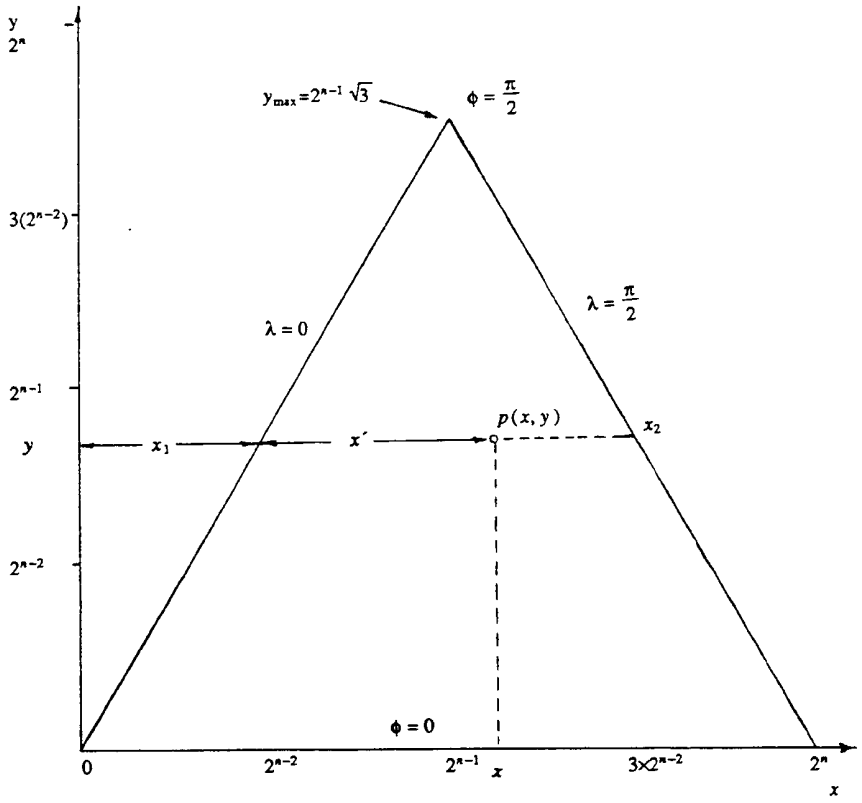

Figure 2. Relationship between the coordinate systems

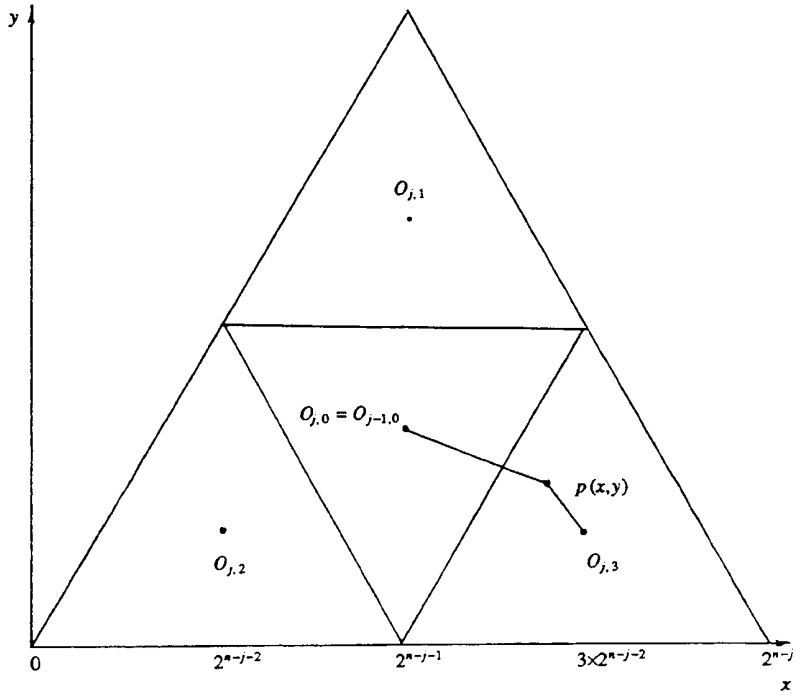

Figure 3. Centoids of triangular cells and distance to these centroids

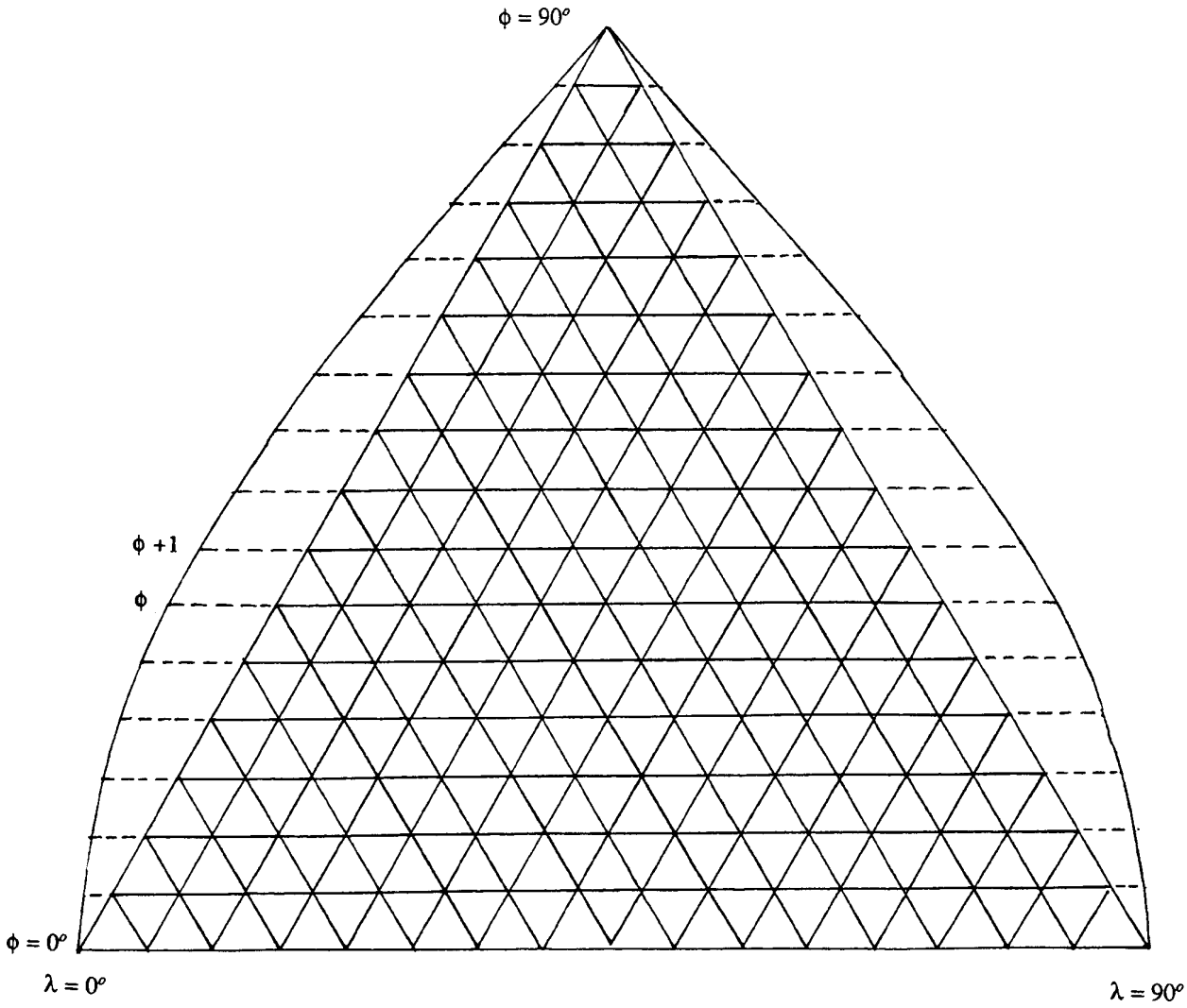

Figure 4. Distribution of the area of triangular cells at different latitudes (sinc function)

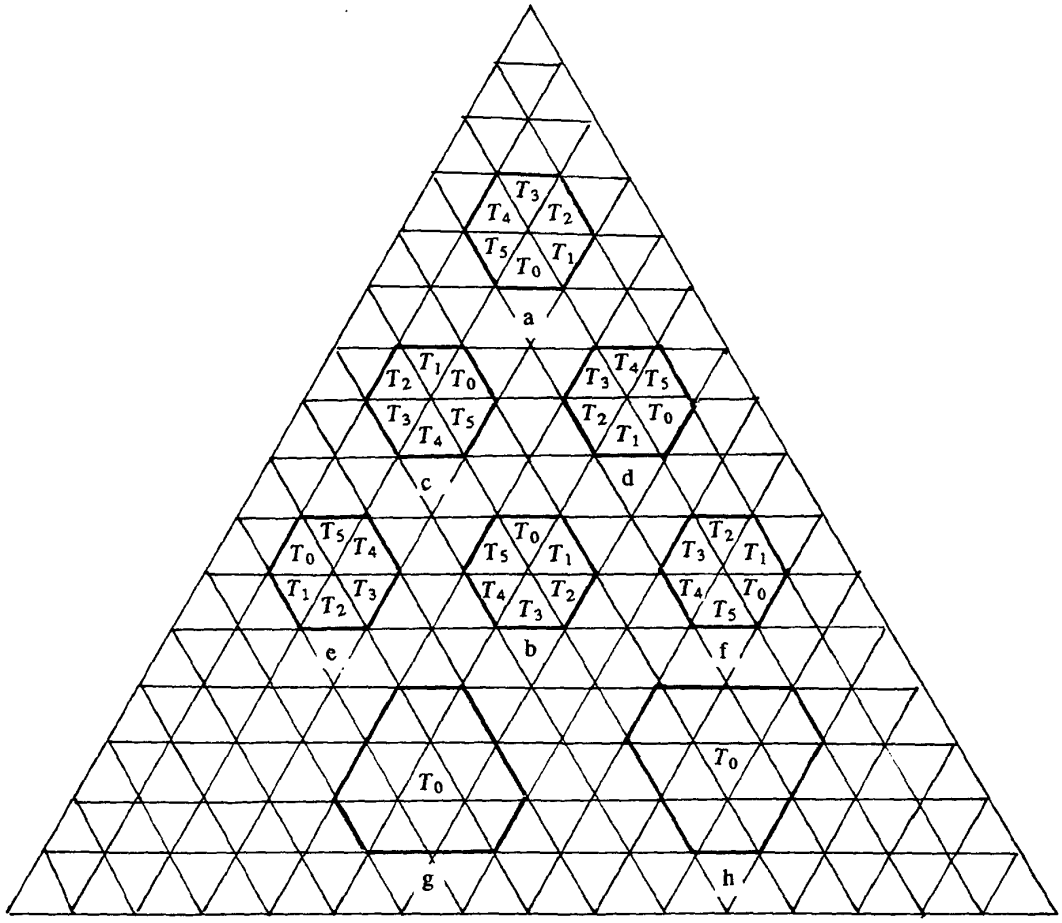

Figure 5. Hexagon including a given triangle, and twelve neighbors of a given triangle

a. Top hexagon with given triangle upward

b. Top hexagon with given triangle downward

c. Left hexagon with given triangle upward

d. Left hexagon with given triangle downward

e. Right hexagon with given triangle upward

f. Right hexagon with given triangle downward

g. Twelve neighbors with given triangle upward

h. Twelve neighbors with given triangle downward