U.S. Department of Commerce
National Oceanic and Atmospheric Administration
National Weather Service
National Centers for Environmental Prediction
5200 Auth Road
Camp Springs, MD 20746-4304

Office Note 460

CONSTRUCTION OF A HILBERT CURVE ON THE SPHERE WITH AN ISOMETRIC PARAMETRIZATION OF AREA

R. James Purser*, Manuel de Pondeca,
Science Applications International Corp., Beltsville, Maryland

Sei-Young Park
Korean Meteorological Administration/FPD

June 29, 2009

* email: jim.purser@noaa.gov

Abstract

Of the various space-filling curves invented since their introduction by Peano the Hilbert curve is probably the simplest to define and to work with. Its value as a tool for optimizing both adaptive and static domain decompositions is now well established in the practice of massively parallel computing. Other important applications of interest to atmospheric scientists involve the thinning of data from remote sensing instruments in regions where these data are distributed with excessive density relative to the assimilation grid resolution, or the aggregation of clusters of such dense data into more manageable numbers of approximately equivalent surrogates known as 'super-observations'. Hilbert curves also enable observational data to be partitioned into disjoint subsets suitable for applying cross-validation methods to the problem of estimating statistical parameters for an assimilation scheme. In these applications involving randomly distributed observations it is desirable that equal intervals in the parameter of the space-filling curve map to equal areas of the globe. This brief note shows how this can be most conveniently done for the sphere in a way that does not unduly distort the pattern traced by the curve. The extensions of these methods to curves filling spherical domains that include the vertical, and even the temporal dimensions, are also discussed briefly.

1. Introduction

Space-filling curves were first introduced by Peano (1890) and quickly refined by Hilbert (1891) and many others (for example, see Sagan 1994). In the atmospheric sciences Behrens and Zimmermann (2000) have shown how the ability to parameterize all the points in space in a continuous and serial manner enables efficient algorithms to solve the problem of optimizing the decomposition of domains for parallel computation in a dynamically adaptive way. Even for the static geometry of subdomains for a massively-parallel climate model, Dennis (2003) has shown that the Hilbert-curve technique provides a superior domain decomposition compared to other standard algorithms. A very different exploitation of the serial parameterization of space was shown by de Pondeca et al. (2006) to lead to natural methods for grouping excessively dense data into super-observations ('super-obs') of a more manageable quantity, or simply to thin such data in a way guaranteed not to leave inadvertent voids. The present note briefly reviews the construction of the classical Hilbert curve in a plane square, before tackling the main result, which is to show how the 'isometric' parameterization of the square, in which equal curve parameters map to equal areas (or volumes) of the space filled by the curve, can be extended to a reasonably simple version of the Hilbert curve covering the sphere and preserving the same isometric property.

2. The Hilbert curve on the square

The construction of a Hilbert curve is a recursive procedure which begins with a 'tour' through one or several squares. The rules of construction can be stated very simply.

(i) Each square is visited once during the tour.

(ii) Entry is made through a definite edge and in the neighborhood of one of its corners.

TABLE 1. A BINARY CARTESIAN COORDINATE TABULATION OF ONE STEP OF THE REFINEMENT OF A GENERIC TRANSIT OF THE HILBERT CURVE IN A UNIT SQUARE REGION, $0 \leq x < 1$, $0 \leq y < 1$. BISECTING THE x AND THE y RANGES SO AS TO CUT THIS SQUARE INTO QUARTERS, AND LABELING EACH SUB-SQUARE BY ITS CORNER CLOSEST TO THE ORIGIN, WE SEE THAT THE FOUR PAIRS OF BINARY COORDINATES IN THE LEFT PAIR OF COLUMNS DESCRIBES A 'TOUR' THROUGH THESE SUB-SQUARES. THE ENTRY AND EXIT SUB-SQUARES HAVE PRECISELY ONE EDGE, $[(0,0),(0,1.)]$, OF THE ORIGINAL SQUARE IN COMMON, ARE THUS SUFFICIENT TO INDICATE THAT, UPON REFINEMENT, THE ENTRY AND EXIT POINTS REMAIN THESE SAME TWO CORNERS OF THE UNIT SQUARE. THE RIGHT PAIR OF COLUMNS SHOW THE TOUR AT A RESOLUTION ENHANCED BY A FACTOR OF TWO WHERE EACH SUCCESSIVE PAIR OF ROWS INDICATES THE EDGE OF THE RESPECTIVE SUB-SQUARE THROUGH WHICH ENTRY AND EXIT OF THIS SEGMENT OF THE REFINED CURVE INTO THE SUB-SQUARE IS MADE.

(.0	.0)	(.00	.00)
		(.01	.00)
(.1	.0)	(.10	.00)
		(.10	.01)
(.1	.1)	(.10	.10)
		(.10	.11)
(.0	.1)	(.01	.11)
		(.00	.11)

(iii) Exit is made through another definite edge and in the neighborhood of another corner not diagonally-opposite to the first one.

(iv) Upon refinement by quartering an existing square, the rules of transit are maintained through each of the four daughter squares in such a way as to preserve the original entry and exit edges and neighborhoods.

The allowable ways (within proper rotations by a right-angle) of transitting a square are shown schematically in Fig. 1, where it is evident that transits labeled D, E, and F are simply the mirror images of those marked A, B, and C, respectively. Transit D is also the reversal of A, while B and C are the reverses of each other, as are E and F.

Given the sequence of original abutting squares visited, and the entry-edge of the first square of the sequence, the rules (i)—(ii) restrict the configuration of tours at this first stage to just two – one that enters the first edge near the right corner and one that enters it near the left corner of that edge, since, after resolving this initial choice, the rest of the tour is then determined uniquely. Moreover, the quartering rule (iv) of refinement, together with the other rules applied in the daughter squares uniquely define the tour at this higher resolution. The refinement rules

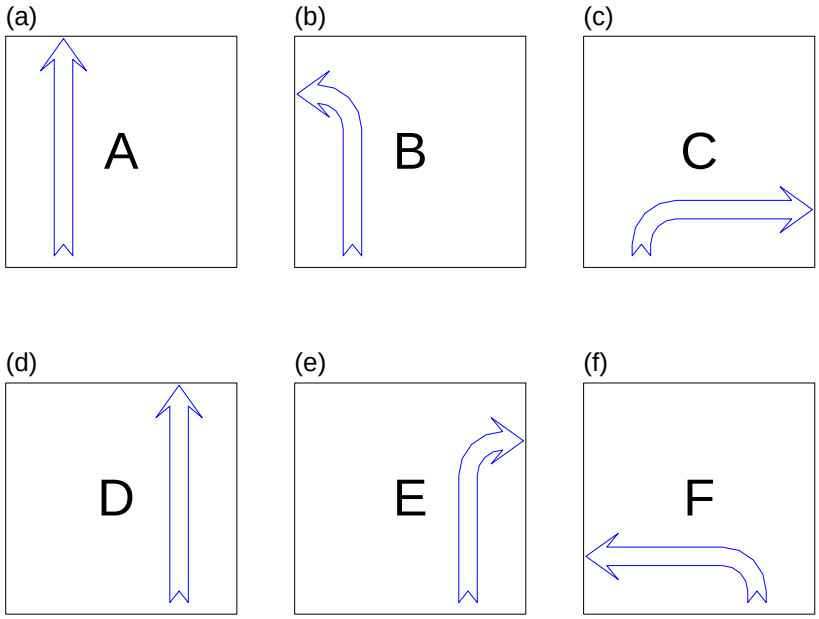

Figure 1. The different modes of transit of a square in an elementary segment of the tour at a finite stage of the construction of the Hilbert curve. The final three cases, D, E, F, are just the mirror-images of the first three, A, B, C, respectively. The transit enters the square through a given edge near one corner and departs through one of the three other edges, and alway close to a corner which, with the first-mentioned corner, forms an edge of the square .

for transits A, B and C are shown schematically in Fig. 2.

We can list the refinement rules for each transit mode as follows, where each segment of a tour is read from left to right:

$$\begin{aligned} A &\rightarrow CFBE, \\ B &\rightarrow CFBD, \\ C &\rightarrow ACEB, \\ D &\rightarrow FCEB, \\ E &\rightarrow FCEA, \\ F &\rightarrow DFBE. \end{aligned}$$

Through iterations of the refinement rules, all geometrically possible couplings of the transit types can appear in a consecutive pair except: AB, CD, DE, FA.

The process of successive refinement is seen more clearly when we join the centers of the smallest sub-squares of each stage, in the order in which they are visited, by a polygonal curve. Fig. 3 shows the first few steps in the refinement for the generic square with a transit consistent with types A and B of Fig. 2.

An equivalent symbolic representation of the refinement process is obtained by encoding the steps of the polygonal curve as binary expansions. In Table 1 the left column shows the binary expansion (requiring only a single digit beyond the binary point) of successive half-width sub-squares visited in the coarsest representation of the tour. The coordinates give the lower

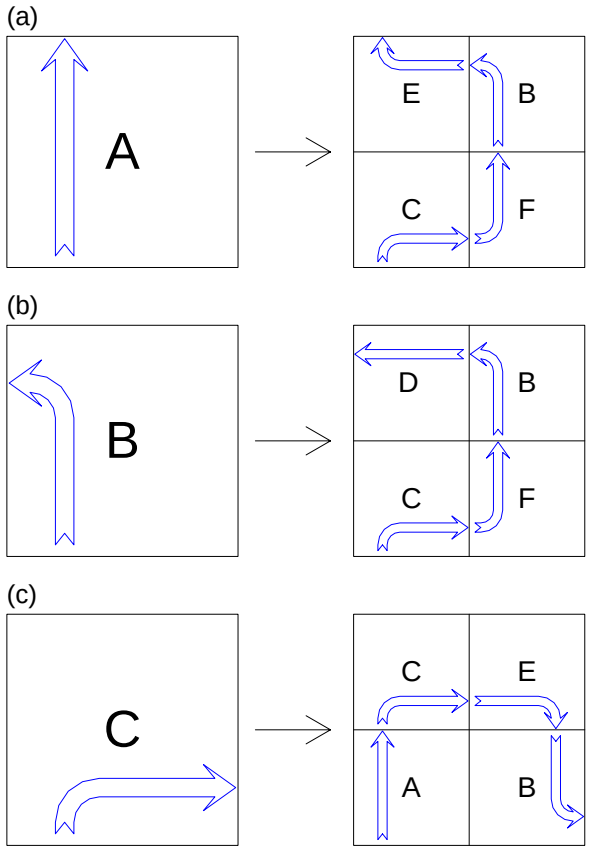

Figure 2. Illustration of the refinement rules for coarse transits, A, B and C, when the old square is quartered. The construction rules preclude any other pattern in each case.

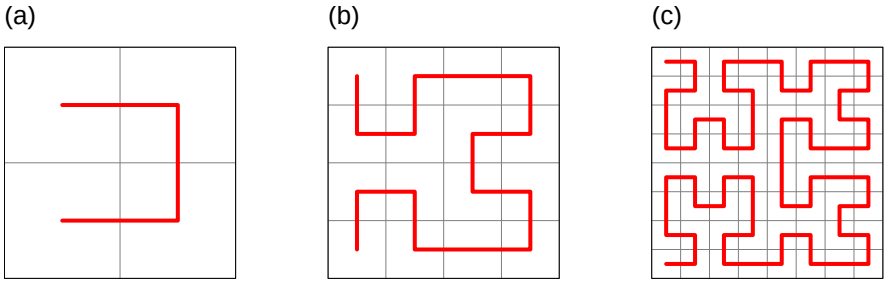

Figure 3. Graphical depiction of the first few refinements of the Hilbert curve that fills a square. The thick lines show the polygonal path connecting the ceneters of the subsquares at each level of refinement. The entry and exit points are associated with the left edge of the domain square in this case.

left corner of each sub-square. The first, (.0,.0), and the last, (.0,.1), of these can be taken to infer that the edge of the domain square at whose end points the entry and exit points lie is the left edge, $[(0.,0.),(0.,1.)]$. Similarly, the four successive pairs of points listed in the right column can be taken to indicate the directed edges of the corresponding squares in the left column of the table. This information is what is needed by the algorithm that generates the next stage of the refinement.

The initial state of the iterated curve can be parameterized by a segment of the real line that assigns equal measure to each square visited. Then, at each stage of refinement, the daughter squares of a parent square can equally share the parameters of that parent square by simply dividing this portion of the parameter range by four and making the assignment in the proper sequence to preserve continuity. The equal-measure property is then inherited throughout the process of sequential refinements. This is very naturally done by employing base-4 parameters so that each iteration of refinement adds a new 'digit' to the numerical representation of the parameter. The base-4 representation also makes the process of locating the image of a given parameter quite straight-forward once the configuration of the initial unrefined tour is specified. In a similar way, we can employ base-8 digits for the parameters of the successive stages of the refinement of the 3-dimensional space-filling curve in a cube, and base-16 for a curve filling a hyper-cubic region of three space and one time dimensions, as will be discussed in section 4.

The true Hilbert curve is the limiting case of the refinement procedure but the 'isometric' property of the image area being proportional to the parameter segment length still holds, and is a property which is desirable to preserve when we generalize the construction to the surface of the sphere. The next section describes a continuous mapping between the sphere and the plane squares forming the surface of a cube, that possesses the equal-area mapping property. When this property holds, the desired isometry is assured between Hilbert curve parameter segments and the spherical areas they map to.

3. Isometric Hilbert curve exploiting an equal-area mapping between the faces of a cube and the sphere

The Cartesian map coordinates, longitude and sine(latitude), map the sphere to a rectangle in an equal-area way and, by linearly stretching the map to give it a $2:1$ aspect ratio, it may be tiled by two squares and filled with a Hilbert curve. However, this simple solution involves an infinite degree of map-deformation at the poles and makes the Hilbert curve unsuitable as a tool for constructing super-obs at high latitudes.

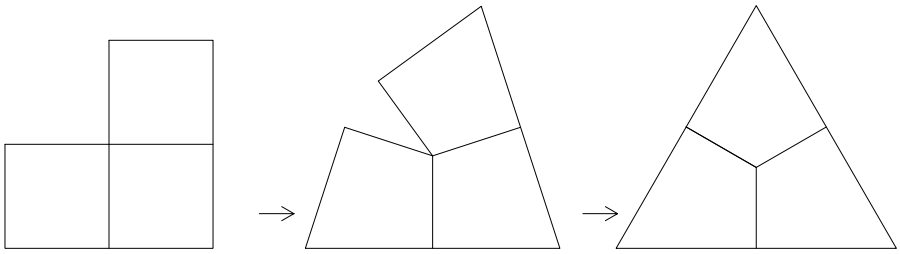

Figure 4. Cartoon showing how a triangular region may be covered by three quadrilaterals. However, the Hilbert curve's entry and exit points are not both located at the ends of an edge of the triangle; one end of the Hilbert curve can be put at a corner of the triangle and the other end at the center of the opposite side, as is evident from Fig. 5 below.

We can seek a less distorting mapping between the sphere and the six plane squares forming the faces of cube — the 'cubed sphere' (whose use in providing gridded domains for numerical weather prediction has been discussed in, for examples, Ronchi et al. 1996, Rančić et al. 1996, McGregor 1996). The simplest of these is the gnomonic cube, obtained by central projection

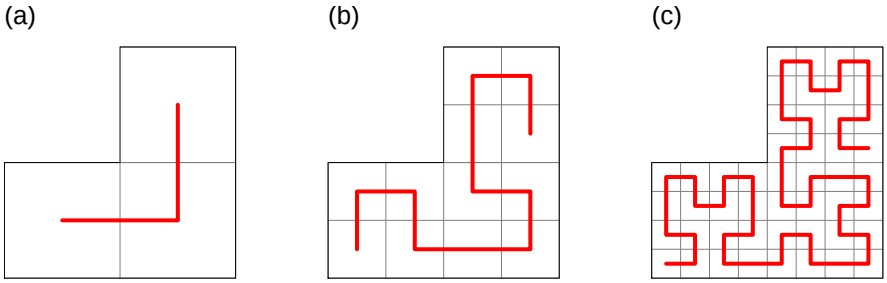

Figure 5. Like Fig. 3 but now a depiction of the first few refinements of the Hilbert curve that fills an **L**-shaped region of three main sub-squares which, as shown in Fig. 4, can be folded into a triangular region (such as the octant of a sphere). In this case, the end points of the refined curve cannot both be at the ends of the same edge of the triangle; instead we arrange the tour to put one end at a corner of the triangle and the other end at the midpoint of the opposite edge.

between the cube and the sphere inscribed within it. The faces of the cube are squares which can be linked by a Hilbert curve filling each one in turn. However, this would mean leaving and reentering each hemisphere several times before completely covering the global domain. A more elegant tour that *does* complete one hemisphere before entering the other is possible. We first subdivide each face of the cube into four smaller squares (24 in all), and cluster the trios of the sub-squares that meet at each of the eight original corners into an 'octant' of the sphere. Effectively, each octant of three sub-squares projects to a triangular shape and the tour can be constructed so that each successive eighth of the Hilbert parameter range maps to the interior of such an octant. Figure 4 shows how an **L**-shaped region of three sub-squares can be folded into a triangle. But Fig. 5 shows that the Hilbert curve refinement can no longer place both the entry and exit points at the ends of one edge of the triangle; instead, the most convenient configuration has one end of the parameter segment mapping to a corner of the triangle and the other end mapping to the midpoint of the opposite edge. Table 2 shows the binary-expansion representation of the coarse-scale polygonal tour, in the same style as Table 1.

TABLE 2. A BINARY CARTESIAN COORDINATE TABULATION OF ONE STEP OF THE REFINEMENT OF A GENERIC TRANSIT OF THE HILBERT CURVE IN AN **L**-SHAPED REGION IN TWO DIMENSIONS.

(.0 .0)	(.00 .00) (.01 .00)
(.1 .0)	(.10 .00) (.10 .01)
(.1 .1)	(.10 .10) (.11 .10)

Completing the tour through eight octants and two hemispheres in turn is accomplished as illustrated schematically in Fig. 6. The relatively high degree of symmetry should keep the distortion of the mapping to the cube fairly small. However, the most direct (gnomonic) mapping is not of the equal-area type. Fig. 7a shows the image of a random uniformly distributed set of sample points from the Hilbert curve mapped, first, to the randomly-oriented

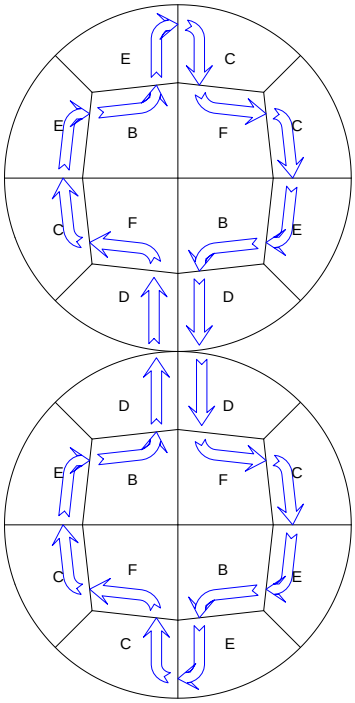

Figure 6. A cyclic tour satisfying the rules of construction for an arrangement of quadrilaterals covering the north and south hemispheres. Each hemisphere is visited and completed in turn, as is each octant (quarter hemisphere).

gnomonic cube, and then to the sphere by central projection. Following this step is an equal-area projection from the sphere to a plane disk. This second mapping is accomplished by a rotationally-symmetric mapping from the sphere to a map radius r on the disk, where

$$r = 2\sin(\theta/2), \tag{3.1}$$

for a co-latitude (angle to the north pole) of θ radians. The geometrical construction of this equal-area mapping between the unit sphere and the disk of radius two units is shown in schematic cross-section in Fig. 8. But it is quite evident from Fig. 7a that the random orientation of the cube is betrayed by the marked pattern of the density inhomogeneity of the random scatter produced by gnomonic mapping between the sphere and the cube.

The remedy is a slightly more complicated continuous transformation which is constructed as follows. Let us consider Earth-centered Cartesians, (X, Y, Z), with X and Y spanning the equatorial plane, and Z pointing along the north polar axis. Take one representative face of the cube to be the one containing the north pole, and span this face with gnomonic coordinates,

$$x = \frac{X}{Z}, \tag{3.2a}$$

$$y = \frac{Y}{Z}, \tag{3.2b}$$

or

$$X = Zx, \tag{3.3a}$$

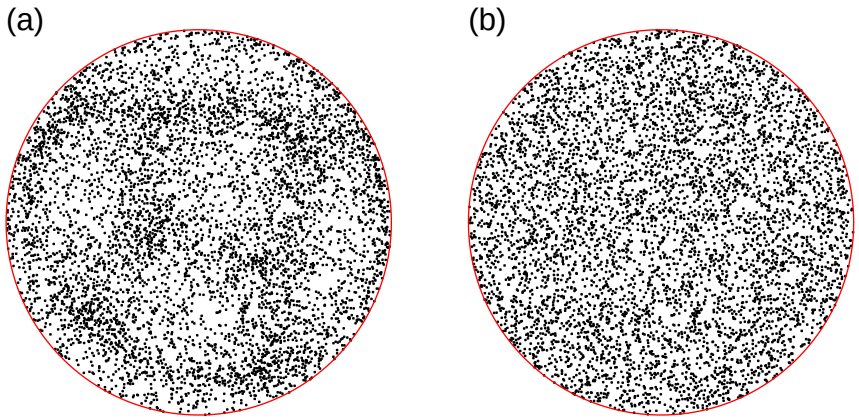

Figure 7. Each panel shows a scatter of points chosen randomly and with uniform distribution over the length of the parameterized Hilbert curve that covers a square dihedron. In both cases, a continuous mapping is defined between the cube and the sphere, and thence to the disk shown, where this last step is via the area-preserving mapping defined as illustrated by Fig. 8. However, in (a) we see the results obtained when the step mapping between the cube and the sphere is not itself an equal-area mapping, and the systematical irregularities of the density in this randomly oriented view are clear to the eye. In (b) the equal-area mapping between cube and sphere guarantees an essentailly uniform scatter plot that does not betray the orientation of the cube. The Hilbert curve is thus said to possess an 'isometric' parameterization over the sphere.

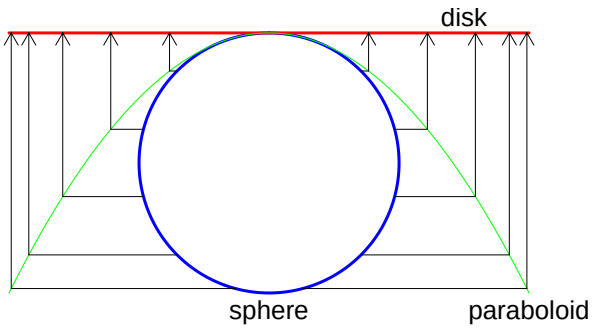

Figure 8. A schematic depiction of the geometry of the equal-area projection between the surface of the unit sphere and the disk of radius 2 units. The projection from the sphere to the mediating paraboloid is radially outward from the polar axis.

$$Y = Zy, \tag{3.3b}$$
$$Z = (1 + x^2 + y^2)^{-1/2}. \tag{3.3c}$$

As defined, the map coordinates of this entire square face range within $-1 \leq x \leq 1$ and $-1 \leq y \leq 1$. If A denotes the area on the sphere and a the corresponding map area, it is not hard to see from the geometry of the situation that

$$\frac{dA}{da} = Z^3. \tag{3.4}$$

At the map corners, where $Z = \sqrt{3}/3$ the Jacobian dA/da drops to less than a fifth, which explains the irregular density in the scatter plot of Fig. 7a.

In order to remedy this defect we first find the spherical area associated to each concentric square of the map so that another centered square of sides given by the square-root of this

quantity can form the basis of an alternative cubed-sphere mapping. A second step is then to adjust the positions tangentially to the sides of each concentric square to ensure all mapped areas become true. The first step could be done with the help of the Gauss-Bonnet theorem by inspecting the angle at the corner of the image on the sphere of each concentric map-square. Alternatively, as we demonstrate below, the map Jacobian (3.4) can be integrated in y, and then x, to obtain the same result together with additional information which we will need. Owing to the eight-fold symmetry of the square face under consideration, we need only focus on the representative octant, $x \geq y \geq 0$.

Let $I(x, y)$ be the quantity defined by the integral,

$$\begin{aligned} I(x,y) &= \int_0^y Z^3(x,y')dy', \\ &= \frac{1}{(1+x^2)} \int_0^{\hat{y}} (1+y'^2)^{-3/2}\, dy', \end{aligned} \tag{3.5}$$

where

$$\hat{y} = \frac{y}{(1+x^2)^{1/2}}. \tag{3.6}$$

Thus,

$$I(x,y) = \frac{y}{(1+x^2)(1+x^2+y^2)^{1/2}}. \tag{3.7}$$

Integrating $J(x) = I(x, x)$ with respect to x will give the area within the 45° triangular wedge out to this x. At $x = 1$ this will give us $1/48$ of the area of the sphere. Since we wish the square root of the area integral to be proportional to a new coordinate for the map, then for this new map also to extend to unity at the edge of the cube, we define the new coordinate p in the following way:

$$\begin{aligned} \frac{\pi p^2(x)}{12} &= \int_0^x J(x')\, dx', \\ &= \int_0^x \frac{x'}{(1+x'^2)(1+2x'^2)^{1/2}} dx' \\ &= \frac{1}{2} \arcsin\left(\frac{x^2}{1+x^2}\right). \end{aligned} \tag{3.8}$$

Thus,

$$p(x) = \left[\frac{6}{\pi} \arcsin\left(\frac{x^2}{1+x^2}\right)\right]^{1/2}. \tag{3.9}$$

For a second new coordinate, q, to form, with p, a pair with a constant Jacobian with respect to the sphere, we need:

$$\begin{aligned} T = \frac{q(x,y)}{p(x)} &= \frac{I(x,y)}{J(x)}, \\ &= \frac{y}{x}\left(\frac{1+2x^2}{1+x^2+y^2}\right)^{1/2}. \end{aligned} \tag{3.10}$$

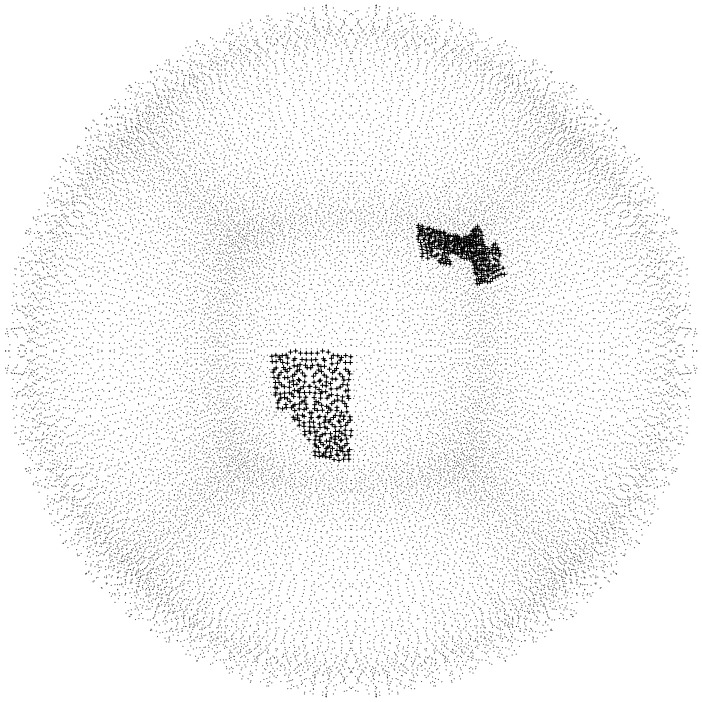

Figure 9. A scatter-plot showing the location of points at equal spacing along the gnomonic cubic surface Hilbert curve projected first to the sphere, then via an equal-area mapping, to a disk. The two patches of the more prominent special symbols occupy the same range of the Hilbert curve parameter but, because they are located at the corner and near the center of the cube's square face, they project to unequal geographical areas.

Corresponding transformations apply in the other sectors of each square panel.

The new mapping itself is continuous everywhere but its derivatives are now discontinuous across the 45° diagonals of each square face in addition to being discontinuous (as in the original gnomonic mapping) across cube edges. However, the map distortions remain relatively small and the mapping has the advantage of being easily invertible. The image on the gnomonic square of the equal-area map point, (p, q) when $p \geq q > 0$ is obtained by first finding x inverting:

$$S = \sin\left(\frac{\pi p^2}{6}\right) = \frac{x^2}{1+x^2} \tag{3.11}$$

that is,

$$x = \left(\frac{S}{1-S}\right)^{1/2}. \tag{3.12}$$

We then obtain:

$$y = Tx\left(\frac{1+x^2}{1+2x^2-T^2x^2}\right)^{1/2}. \tag{3.13}$$

Fig. 7b shows the result of generating random points uniformly scattered over the (p, q) map panels, transforming to the sphere (via the gnomonic map) and thence back to the equal-area disk projection. Here again, we are able to recover a perfectly uniform random distribution.

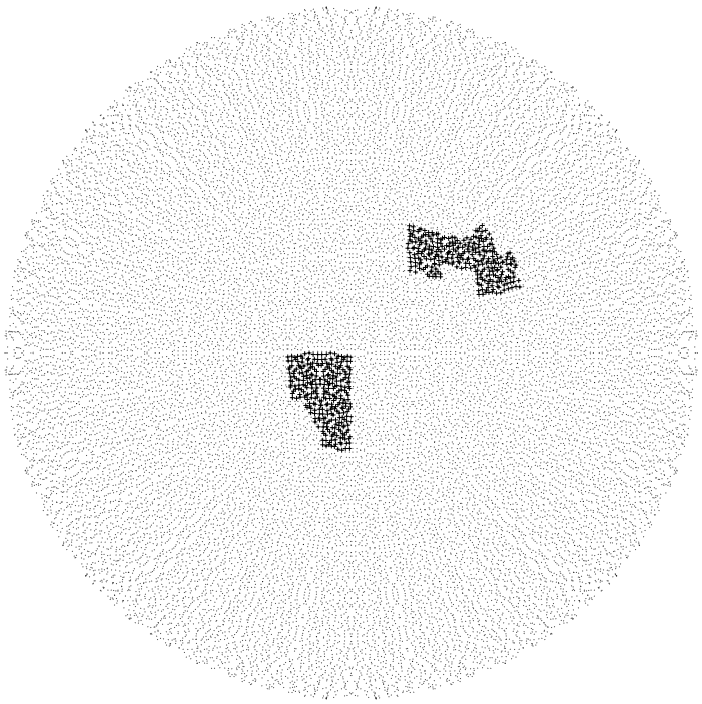

Figure 10. The same as in Fig. 9 except that an intermediate tranformation replacing the gnomonic cube with a cubic mapping that has the equal area mapping property. The two patches of special symbols, whose ranges of the Hilbert curve parameter are equal, now project also to equal geographical areas – obeying the desired isometric property.

One application of the Hilbert curve is to facilitate the construction of super-obs. The isometric property is especially valuable in this context as it means that, when the observations belonging to a segment of the Hilbert curve of a given parameter length are grouped into a single superob, the geograhical area of the region covered by the parameter segment is immediately known. Fig. 9 shows a pattern formed by a very large number of 'obs' placed at a perfectly uniform spacing along the Hilbert curve that uses the mapping of Fig. 7a, except without the random rotation of map orientation. Two small segments of equal length in the parameter space are selected for consideration for consolidation as 'super-obs', one near the center of one face of the mapping cube, and one near a corner. Here again, we see the systematic modulation of the density of the original scatter, but in this case, we also see that the two 'superobs' have areas that are obviously very different. This undesirable situation is corrected when we switch to the isometric Hilbert curve, with results shown in Fig. 10, that provide super-obs belonging to identical geographical areas.

The shapes of the two super-ob patches shown in Fig. 10 are not exactly the shapes they have on the original spherical surface owing to the slight distortion they incur as a consequence of the mapping to the disk. While we would not expect the footprint of a super-ob to be circular, we would at least hope that some suitable measure of its aspect ratio did not depart too far from unity. The ratio of the square-roots of the principal components of the centered second moments of the patch would provide one crude measure. In the case of super-ob footprints

TABLE 3. Like Table 1 but now showing the pattern by which the coarsely resolved path of a segment of the Hilbert curve through a generic cubic region may be refined. In this case, the entry and exit points at each refinement stage are as close as the resolution will allow to $(0, 0, 0)$ and $(0, 0, 1.)$, which also bracket one unambiguous edge of the original cube.

(.0 .0 .0)	(.00 .00 .00) (.01 .00 .00)
(.1 .0 .0)	(.10 .00 .00) (.10 .01 .00)
(.1 .1 .0)	(.10 .10 .00) (.10 .11 .00)
(.0 .1 .0)	(.01 .11 .00) (.01 .11 .01)
(.0 .1 .1)	(.01 .11 .10) (.01 .11 .11)
(.1 .1 .1)	(.10 .11 .11) (.10 .10 .11)
(.1 .0 .1)	(.10 .01 .11) (.10 .00 .11)
(.0 .0 .1)	(.01 .00 .11) (.00 .00 .11)

obtained by taking a continuous segment of the Hilbert curve in a plane square domain, we can see from the central portion of Fig. 3c that this aspect ratio estimate could attain a maximum value of about four (where four consecutive filled sub-squares lie in a straight line), which remains acceptable for practical purposes. We would like to be assured that this ratio will never be unduly amplified by the compounding effect of the distortion of the equal-area mapping between the cube and the sphere. Fig. 11 shows an array of small circles on the surface of the sphere mapped back to the interior of one octant of one face of the cube through the equal-area mapping we have described. The maximum distortion occurs near the outer edge, $p = 1$, but is clearly not very large, even here.

4. Incorporating dimensions of height and time

There is no intrinsic restriction to the dimensionality of the Hilbert curve. We can add the third dimension (height) and refine a circuitous tour of a cube. The tour is now too complicated (even at coarse resolution) to show clearly by graphical means, and the power and versatility of the binary tabulations become evident. The tour through a generic cube moves through its eight main sub-cubes as indicated in Table 3, where the entry-exit edge (for the unit-sided cube) in the orientation of this example has coordinates, $[(0, 0, 0), (0, 0, 1)]$. By analogy to the two-dimension case of Table 1, the eight successive pairs of points listed to two binary places of precision in the right column of the table indicate not only the sub-cubes visited (information

TABLE 4. Like Table 3 but now showing the pattern by which the coarsely resolved path of a segment of the Hilbert curve through a prism with **L**-shaped cross-section cubic region may be refined. As in Table 3, the entry and exit points at each refinement stage are as close as the resolution will allow to $(0, 0, 0)$ and $(1., 0, 0)$. The **L**-shape can be closed up into a triangle, thus making the shape dealt with here apply to a triangular prism whose entry and exit points are at the ends of an edge belonging to a triangular face.

(.0 .0 .0)	(.00 .00 .00) (.00 .00 .01)
(.0 .0 .1)	(.00 .00 .10) (.01 .00 .10)
(.1 .0 .1)	(.10 .00 .10) (.10 .01 .10)
(.1 .1 .0)	(.10 .10 .10) (.11 .10 .10)
(.1 .1 .0)	(.11 .10 .01) (.11 .10 .00)
(.1 .0 .0)	(.11 .01 .00) (.11 .00 .00)

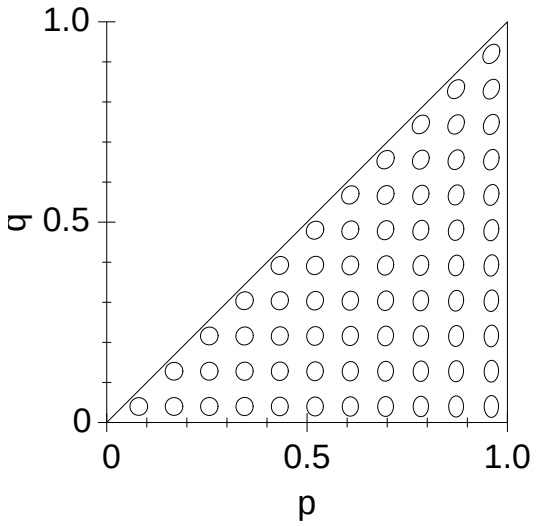

Figure 11. A mapping to one octant of one face of the equal-area cube of an array of identical small circles on the sphere, showing that even the worst distortion incurred by this mapping is of moderate magnitude.

requiring only the first digit of the expansion of each component, as is given in the left column), but, as in the two-dimensional example of the square, the right column data infer the respective edges of each of these eight sub-cubes to which the entry and exit points of the associated segments adhere. The sequence of 3-bit binaries given in the left column are identical to the so-called "3-bit Gray code" sequence (Gray 1953, Gilbert 1958) originally invented to stabilize

TABLE 5. Like Table 4 but now showing the pattern by which the coarsely resolved path of a segment of the Hilbert curve through a prism with **L**-shaped cross-section when the entry and exit points at each refinement stage are as close as the resolution will allow to $(0, 0, 0)$ and $(0, 0, 1.)$.

(.0 .0 .0)	(.00 .00 .00) (.01 .00 .00)
(.1 .0 .0)	(.10 .00 .00) (.10 .01 .00)
(.1 .1 .0)	(.10 .10 .00) (.10 .10 .01)
(.1 .1 .1)	(.10 .10 .11) (.10 .10 .11)
(.1 .0 .1)	(.10 .01 .11) (.10 .00 .11)
(.0 .0 .1)	(.01 .00 .11) (.00 .00 .11)

angular measurements given by optical shaft encoders in mechanical engineering applications (since only one bit changes at a time).

In refining the cubic Hilbert curve we encounter an ambiguity concerning the way a transit associated with a given edge of a sub-cube is interpreted as a tour through that cube's own eight sub-cubes. However, the two possibilities are mirror-images of each other, so the ambiguity can be resolved by simply always choosing this eight-cube tour to be the one chirally identical (i.e., via a *proper* rotation) to the tour defined in Table 3.

TABLE 6. Like Table 1 but now showing the pattern by which the coarsely resolved path of a segment of the Hilbert curve through a generic hypercubic region may be refined. In this case, the entry and exit points at each refinement stage are as close as the resolution will allow to $(0, 0, 0)$ and $(0, 0, 1.)$, which also bracket one unambiguous edge of the original cube.

(.0 .0 .0 .0)	(.00 .00 .00 .00) (.01 .00 .00 .00)
(.1 .0 .0 .0)	(.10 .00 .00 .00) (.10 .01 .00 .00)
(.1 .1 .0 .0)	(.10 .10 .00 .00) (.10 .11 .00 .00)
(.0 .1 .0 .0)	(.01 .11 .00 .00) (.01 .11 .01 .00)
(.0 .1 .1 .0)	(.01 .11 .10 .00) (.01 .11 .11 .00)
(.1 .1 .1 .0)	(.10 .11 .11 .00) (.10 .10 .11 .00)
(.1 .0 .1 .0)	(.10 .01 .11 .00) (.10 .00 .11 .00)
(.0 .0 .1 .0)	(.01 .00 .11 .00) (.01 .00 .11 .01)
(.0 .0 .1 .1)	(.01 .00 .11 .10) (.01 .00 .11 .11)
(.1 .0 .1 .1)	(.10 .00 .11 .11) (.10 .01 .11 .11)
(.1 .1 .1 .1)	(.10 .10 .11 .11) (.10 .11 .11 .11)
(.0 .1 .1 .1)	(.01 .11 .11 .11) (.01 .11 .10 .11)
(.0 .1 .0 .1)	(.01 .11 .01 .11) (.01 .11 .00 .11)
(.1 .1 .0 .1)	(.10 .11 .00 .11) (.10 .10 .00 .11)
(.1 .0 .0 .1)	(.10 .01 .00 .11) (.10 .00 .00 .11)
(.0 .0 .0 .1)	(.01 .00 .00 .11) (.01 .00 .00 .11)

In the case of a tour through a 'thick' spherical shell, using the preferred spherical-cube isometric mapping described in the previous section, then, if upon each of the 24 fundamental sub-squares of this spherical mapping are stacked at least one pair of cubes of this footprint size, we will need to consider a Hilbert tour through triangular-section prisms that unfold, in a manner analogous to Fig. 4, into **L**-section prisms now composed of six cubes (two wide, and two high). If the depth of the domain *only* amounts to this double stack, we need only consider tours through the prism that enter and exit at the same trianglar face and, in this case, unlike the two-dimensional situation, we *can* now enter and exit at the ends of a common edge. This allows the transfer between one octant and the next in the three-dimensional tour analogous to that shown in Fig. 6 to occur always at the equator, allowing all octants to be treated

identically. In the style of Table 3 (and its predecessors) the appropriate tour through the **L**, or its reverse, is as tabulated in Table 4. However, if the domain through which the Hilbert curve is even deeper than this, we shall require *more* than the depth of one of these six-cube basic prisms to be stacked up and, at some of the transits, we shall need entry and exit points of such a prism to be at different heights. Table 5 treats this case, with the edge defined by the entry and exit points now being vertical, or perpendicular to the triangular (or **L**-shaped) base. Careful inspection of the binary digits in this case reveals that this tour exhibits mirror symmetry in the horizontal midplane (something that is not possible to achieve in the case of the tour of Table 4, when entry and exit occur at the same altitude).

The preceding discussion does not address the question of how one should decide whether the domain counts as 'thick' in the vertical. In other words, we are in need of an effective *metric* by which we can gauge and compare the relative displacements in the horizontal and vertical directions in a way that is meaningful in the context of the applications. In applications to data assimilation, which we have implicitly assumed here, the effective metric in the two directions might come from the horizontal and vertical scales of the covariances of background error. These in turn will usually tend to relate closely to the ratio of scales given by classical quasi-geostrophic theory. Locally, horizontal to vertical scale ratios tend to be not far from the ratio of the Brunt-Väisälä frequency, N, to the effective rotational frequency, Coriolis, f, or Earth-rotational, Ω. By integrating a characteristic representative value for this ratio up through the atmosphere, we can decide whether the domain considered belongs to the 'thick' or 'thin' category, and plan the space-filling tour accordingly. As an example, let us suppose that the ratio of scales in the lowest 10 km, where the Brunt-Väisälä frequency tends to be relatively low, is 100:1, while in the next 20 km of altitude let this ratio be 200:1. Then, by this crude measure, the horizontal distance equivalent to the entire 30 km depth we have considered is $D = 10 \times 100 + 20 \times 200 = 5000$km, which corresponds to half of one side of the spherical octant. Thus, by this reckoning, we would not need to invoke the 'thick' atmosphere assumption for an assimilation domain that extends only up to 30 km.

A similar consideration needs to be applied if the domain from which data are gathered and compared is three-dimensional in the sense of horizontal plus time, or fully four-dimensional, with horizontal, vertical and temporal dimensions all being substantially present. In the latter case, not only must we determine the relative vertical 'thickness' (in the appropriate metric comparison to the Earth's quarter-circumference) but also the magnitude of the 'duration' of the period under consideration, transformed to Earth units. In this case the crucial quantity relating the effective equivalence between horizontal distances and time durations has the units of a speed. If we take this speed to be a moderately large, but not atypical advection speed, say 50 km/hour, then the time window of any typical assimilation, say six hours, translates into a horizontal distance (in this case, 300 km) very much smaller than the quarter-circumference (10000 km) we might use as the criterion determining whether a domain that is 'thick' in the time dimension. Thus, the temporal dimension of any four-dimensional Hilbert curve we might wish to use over a global domain would not become activated until several generations of horizontal and vertical refinement of the largest 'cubes' have already occurred in the symbolic expansion of the Hilbert curve parameter. In other words, a 'mixed-radix' expansion is appropriate, with base-8 used for the first few digits (corresponding to horizontal and vertical degrees of freedom only) followed by the full base-16 continuation at the later digits (corresponding to

the resolution of excursions also in the time dimension at small distance scales). Thus, it is necessary to be able to set rules for the construction of the Hilbert curve within hyper-cubes, but not necessarily within the generalizations of the prisms ('hyper-prisms'?) we needed to consider for the largest global scales in three dimensions. To complete this section, we therefore provide a tabulation of the pattern of refinement of a generic tour through a hyper-cube in Table 6. As before, we recognize the Gray code binary sequence working down the left column of this table.

5. Remarks

The two equal-area mappings from appropriate sets of adjoining square to the sphere can be trivially adapted to a mapping from half of these tiles to a hemisphere, and hence to a plane disk, again, with minimal distortion. But, by shrinking the disk and transforming back to the sphere with the inverse of this last mapping step, we also get a way of mapping a set of square tiles to any circular spherical cap in an equal-area fashion, and can thus cover such a region with a Hilbert curve isometrically.

Although the techniques described here have focused mainly on the case of two-dimensional maps, we have shown that the procedure for constructing a three-dimensional Hilbert curve through a finite depth of the atmosphere is very similar in principle. A metric is required to specify the horizontal equivalent distance for each vertical infinitesimal increment. If, by this measure, the atmosphere is a shallow one, then several generations of horizontal refinement will precede the first vertical refinement, at which point the refined transits thread cubes instead of squares with each refinement requiring the creation of eight new segments instead of four. If the atmosphere is very deep, it may be required to stack one or several 'cubes' on the footprints made by the spherical decomposition into squares at the coarsest level, and proceed from the outset to refine each cube in all three of its edge directions. A further generalization allows the Hilbert curve to be extended, in addition to the space dimensions, into the time domain. Here again it becomes necessary to decide on the formal equivalence of a unit of time with the units of horizontal and vertical space but, as we have indicated, the construction of the space-time-filling curve presents no fundamental difficulties. Observations that are localized in space and time can then be serially ordered by the parameter of the Hilbert curve that threads them together, and any subsequent super-obbing, data thinning or data subsets partitioned for cross-validation become significantly easier in the new, effectively one-dimensional rendering of the data.

Acknowledgments

We wish to thank Drs. Dave Parrish and Todd Spindler for their helpful reviews of this work.

References

Behrens, J., and J. Zimmermann 2000 Parallelizing an unstructured grid generator with a space-filling curve approach. *Lecture Notes in Computer Science* Springer, Berlin, pp. 815–823.

Dennis, J. M.	2003	Partitioning with space-filling curves on the cubed-sphere. In: *Proceedings of the International Parallel and Distributed Processing Symposium., April 2003,* IEEE.
De Pondeca, M., S.-Y. Park, R. J. Purser, and G. DiMego	2006	Applications of Hilbert curves to the selection of subsets of spatially inhomogeneous observational data for cross-validation and to the construction of super-observations. *Eos Trans. AGU 87*,(52) Fall Meeting. Suppl., Abstract A31A-0868.
Gilbert, E. N.	1958	Gray codes and paths on the n-cube. *Bell System Tech. J.*, **37,** 815–826.
Gray, F.	1953	Pulse code communication. United States Patent Number 2632058.
Hilbert, D.	1891	Über die stetige Abbildung einer Linie auf ein Flächenstück. *Math. Ann.,* **38,** 459–460.
McGregor, J. L.	1996	Semi-Lagrangian advection on conformal-cubic grids. *Mon. Wea. Rev.,* **124,** 1311–1322.
Peano, G.,	1890	Sur une courbe, qui remplit toute une aire plane. *Math. Ann.,* **36,** 157–160.
Rančić, M., R. J. Purser, and F. Mesinger	1996	A global shallow-water model using an expanded spherical cube: Gnomonic versus conformal coordinates. *Quart. J. Roy. Meteor. Soc.,* **122,** 959–982.
Ronchi, C., Iacono, R. and Paolucci, P. S.	1996	The 'cubed sphere': A new method for the solution of partial differential equations in spherical geometry. *J. Comput. Phys.,* **124,** 93–114.
Sagan, H.	1994	*Space-Filling Curves.* Springer-Verlag, New York.